Moon, Hyungsik Roger; Shum, Matthew; Weidner, Martin

Working Paper

Estimation of random coefficients logit demand models with interactive fixed effects

cemmap working paper, No. CWP12/17

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Moon, Hyungsik Roger; Shum, Matthew; Weidner, Martin (2017) : Estimation of random coefficients logit demand models with interactive fixed effects, cemmap working paper, No. CWP12/17, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2017.1217

This Version is available at:
https://hdl.handle.net/10419/189704

cemmap
centre for microdata methods and practice

Estimation of random coefficients logit demand models with interactive fixed effects

Hyungsik Roger Moon
Matthew Shum
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP12/17

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Estimation of Random Coefficients Logit Demand Models with Interactive Fixed Effects*

Hyungsik Roger Moon[‡§] Matthew Shum[¶] Martin Weidner[‖]

First version: October 2009; This version: February 15, 2017

Abstract

We extend the Berry, Levinsohn and Pakes (BLP, 1995) random coefficients discrete-choice demand model, which underlies much recent empirical work in IO. We add interactive fixed effects in the form of a factor structure on the unobserved product characteristics. The interactive fixed effects can be arbitrarily correlated with the observed product characteristics (including price), which accommodates endogeneity and, at the same time, captures strong persistence in market shares across products and markets. We propose a two-step least squares-minimum distance (LS-MD) procedure to calculate the estimator. Our estimator is easy to compute, and Monte Carlo simulations show that it performs well. We consider an empirical illustration to US automobile demand.

Keywords: discrete-choice demand model, interactive fixed effects, factor analysis, panel data, random utility model.

JEL codes: C23, C25.

*We thank participants in presentations at Georgetown, Johns Hopkins, Ohio State, Penn State, Rice, Texas A&M, UC Davis, UC Irvine, UCLA, Chicago Booth, Michigan, UPenn, Wisconsin, Southampton, the 2009 California Econometrics Conference and the 2010 Econometric Society World Congress for helpful comments. Chris Hansen, Han Hong, Sung Jae Jun, Jinyong Hahn, and Rosa Matzkin provided very helpful discussions. Moon acknowledges the NSF for financial support via SES 0920903. Weidner acknowledges support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001, and from the European Research Council grant ERC-2014-CoG-646917-ROMIA.

‡Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253. Email: moonr@usc.edu

§Department of Economics, Yonsei University, Seoul, Korea.

¶Division of Humanities and Social Sciences, California Institute of Technology, MC 228-77, Pasadena, CA 91125. Email: mshum@caltech.edu.

‖Department of Economics, University College London, Gower Street, London WC1E 6BT, U.K., and CeMMAP. Email: m.weidner@ucl.ac.uk.

1 Introduction

The Berry, Levinsohn and Pakes (1995) (hereafter BLP) demand model, based on the random coefficients logit multinomial choice model, has become the workhorse of demand modeling in empirical industrial organization and antitrust analysis. An important virtue of this model is that it parsimoniously and flexibly captures substitution possibilities between the products in a market. At the same time, the nested simulated GMM procedure proposed by BLP accommodates possible endogeneity of the observed product-specific regressors, notably price. This model and estimation approach has proven very popular (e.g. Nevo (2001), Petrin (2002); surveyed in Ackerberg et. al. (2007)).

Taking a cue from recent developments in panel data econometrics (e.g. Bai and Ng (2006), Bai (2009), and Moon and Weidner (2015a; 2015b)), we extend the standard BLP demand model by adding interactive fixed effects to the unobserved product characteristic, which is the main "structural error" in the BLP model. This interactive fixed effect specification combines market (or time) specific fixed effects with product specific fixed effects in a multiplicative form, which is often referred to as a factor structure.

Our factor-based approach extends the baseline BLP model in two ways. First, we offer an alternative to the usual moment-based GMM approach. The interactive fixed effects "soak up" some important channels of endogeneity, which may obviate the need for instrumental variables of endogenous regressors such as price. This is important as such instruments may not be easy to identify in practice. Moreover, our analysis of the BLP model with interactive fixed effects illustrates that the problem of finding instruments for price (which arises in any typical demand model) is distinct from the problem of underidentification of some model parameters (such as the variance parameters for the random components), which arises from the specific nonlinearities in the BLP random coefficients demand model. In our setting, the fixed effects may obviate the need for instruments to control for price endogeneity but, as we will point out, we still need to impose additional moment conditions in order to identify these nonlinear parameters. Second, even if endogeneity persists in the presence of the interactive fixed effects, the instruments only need to be exogenous with respect to the residual part of the unobserved product characteristics, which is not explained by the interactive fixed effect. This may expand the set of variables which may be used as instruments.

To our knowledge, the current paper presents the first application of some recent developments in the econometrics of long panels (with product and market fixed effects) to the workhorse demand model in empirical IO. Relative to the existing panel factor literature (for instance, Bai (2009), and Moon and Weidner (2015a; 2015b)) that assume a linear regression with exogenous regressors, the nonlinear model that we consider here

poses both identification and estimation challenges. Namely, the usual principal components approach for linear factor models with exogenous regressors is inadequate due to the nonlinearity of the model and the potentially endogenous regressors. At the same time, the conventional GMM approach of BLP cannot be used for identification and estimation due to the presence of the interactive fixed effects.

We propose an alternative identification and estimation scheme which we call the *Least Squares-Minimum Distance (LS-MD)* method. It consists of two steps. The first step is a least squares regression of the mean utility on the included product-market specific regressors, factors, and the instrumental variables. The second step minimizes the norm of the least squares coefficient of the instrumental variables in the first step. This estimation approach is similar to the two stage estimation method for a class of instrumental quantile regressions in Chernozhukov and Hansen (2006). We show that under regularity conditions that are comparable to the standard GMM problem, the parameter of interest is point identified and its estimator is consistent. We also derive the limit distribution under asymptotic sequences where both the number of products and the number of markets converge to infinity. In practice, the estimator is simple and straightforward to compute. Monte Carlo simulations demonstrate its good small-sample properties.

Our work complements some recent papers in which alternative estimation approaches and extensions of the standard random coefficients logit model have been proposed, including Villas-Boas and Winer (1999), Knittel and Metaxoglou (2014), Dube, Fox and Su (2012), Harding and Hausman (2007), Bajari, Fox, Kim and Ryan (2011), and Gandhi, Kim and Petrin (2010).

We illustrate our estimator on a dataset of market shares for automobiles, inspired by the exercise in BLP. This application illustrates that our estimator is easy to compute in practice. Significantly, we find that, once factors are included in the specification, the estimation results under the assumption of exogenous and endogenous price are quite similar, suggesting that the factors are indeed capturing much of the unobservable product and time effects leading to price endogeneity.

The paper is organized as follows. Section 2 introduces the model. In Section 3 we discuss how to identify the model when valid instruments are available. In Section 4 we introduce the LS-MD estimation method. Consistency and asymptotic normality are discussed in Section 5. Section 6 contains Monte Carlo simulation results, and Section 7 discusses the empirical example. Section 8 concludes. In the appendix we list the assumptions for the asymptotic analysis and provide technical derivations and proofs of results in the main text.

Notation

We write A' for the transpose of a matrix or vector A. For column vectors v the Euclidean norm is defined by $\|v\| = \sqrt{v'v}$. For the n-th largest eigenvalues (counting multiple eigenvalues multiple times) of a symmetric matrix B we write $\mu_n(B)$. For an $m \times n$ matrix A the Frobenius norm is $\|A\|_F = \sqrt{\text{Tr}(AA')}$, and the spectral norm is $\|A\| = \max_{0 \neq v \in \mathbb{R}^n} \frac{\|Av\|}{\|v\|}$, or equivalently $\|A\| = \sqrt{\mu_1(A'A)}$. Furthermore, we use $P_A = A(A'A)^\dagger A'$ and $M_A = \mathbb{1}_m - A(A'A)^\dagger A'$, where $\mathbb{1}_m$ is the $m \times m$ identity matrix, and $(A'A)^\dagger$ denotes a generalized inverse, since A may not have full column rank. The vectorization of an $m \times n$ matrix A is denoted $\text{vec}(A)$, which is the $mn \times 1$ vector obtained by stacking the columns of A. For square matrices B, C, we use $B > C$ (or $B \geq C$) to indicate that $B - C$ is positive (semi) definite. We use ∇ for the gradient of a function, *i.e.* $\nabla f(x)$ is the vector of partial derivatives of f with respect to each component of x. We use "wpa1" for "with probability approaching one".

2 Model

The random coefficients logit demand model is an aggregate market-level model, formulated at the individual consumer-level. Consumer i's utility of product j in market[1] t is given by

$$u_{ijt} = \delta_{jt}^0 + \epsilon_{ijt} + X_{jt}' \, v_i \, , \tag{2.1}$$

where ϵ_{ijt} is an idiosyncratic product-specific preference shock, and $v_i = (v_{i1}, \ldots, v_{iK})'$ is an idiosyncratic characteristic preference. The mean utility is defined as

$$\delta_{jt}^0 = X_{jt}' \beta^0 + \xi_{jt}^0 \, , \tag{2.2}$$

where $X_{jt} = (X_{1,jt}, \ldots, X_{K,jt})'$ is a vector of K observed product characteristics (including price), and $\beta^0 = \left(\beta_1^0, \ldots, \beta_K^0\right)'$ is the corresponding vector of coefficients. Following BLP, ξ_{jt}^0 denotes unobserved product characteristics of product j, which can vary across markets t. This is a "structural error", in that it is observed by all consumers when they make their decisions, but is unobserved by the econometrician. In this paper, we focus on the case where these unobserved product characteristics vary across products and markets according to a factor structure:

$$\xi_{jt}^0 = \lambda_j^{0\prime} \, f_t^0 + e_{jt} \, , \tag{2.3}$$

[1]The t subscript can also denote different time periods.

where $\lambda_j^0 = \left(\lambda_{1j}^0, \ldots, \lambda_{Rj}^0\right)'$ is a vector of factor loadings corresponding to the R factors[2] $f_t^0 = \left(f_{1t}^0, \ldots, f_{Rt}^0\right)'$, and e_{jt} is a product and market specific error term. Here $\lambda_j^{0\prime} f_t^0$ represent interactive fixed effects, in that both the factors f_t^0 and factor loadings λ_j^0 are unobserved to the econometrician, and can be correlated arbitrarily with the observed product characteristics X_{jt}. We assume that the number of factors R is known.[3] The superscript zero indicates the true parameters, and objects evaluated at the true parameters. Let $\lambda^0 = (\lambda_{jr}^0)$ and $f^0 = (\lambda_{tr}^0)$ be $J \times R$ and $T \times R$ matrices, respectively.

The factor structure in equation (2.3) approximates reasonably some unobserved product and market characteristics of interest in an interactive form. For example, television advertising is well-known to be composed of a product-specific component as well as an annual cyclical component (peaking during the winter and summer months).[4] The factors and factor loadings can also explain strong correlation of the observed market shares over both products and markets, which is a stylized fact in many industries that has motivated some recent dynamic oligopoly models of industry evolution (e.g. Besanko and Doraszelski (2004)). The standard BLP estimation approach, based on moment conditions, allows for weak correlation across markets and products, but does not admit strong correlation due to shocks that affect all products and markets simultaneously, which we model via the factor structure.

To begin with, we assume that the regressors X_{jt} are exogenous with respect to the errors e_{jt}, that is, X_{jt} and e_{jt} are uncorrelated for given (j, t). This assumption, however, is only made for ease of exposition, and in both Section 4.1 below and in the empirical illustration, we consider the more general case where regressors (such as price) may be endogenous. Notwithstanding, regressors which are strictly exogenous with respect to e_{jt} can still be endogenous with respect to the ξ_{jt}^0, due to correlation of the regressors with the factors and factor loadings. Thus, including the interactive fixed effects may "eliminate" endogeneity problems, so that instruments for endogeneity may no longer be needed. This possibility of estimating a demand model without searching for instruments may be of great practical use in antitrust analysis.

[2]Depending on the specific application one has in mind one may have different interpretations for λ_j and f_t. For example, in the case of national brands sold in different markets it seems more natural to interpret λ_j as the underlying factor (a vector product qualities) and f_t as the corresponding loadings (market specific tastes for these qualities). For convenience, we refer to f_t as factors and λ_j as factor loadings throughout the whole paper, which is the typical naming convention in applications where t refers to time.

[3]Known R is also assumed in Bai (2009) and Moon and Weidner (2015a) for the linear regression model with interactive fixed effects. Allowing for R to be unknown presents a substantial technical challenge even for the linear model, and therefore goes beyond the scope of the present paper.

[4]cf. *TV Dimensions* (1997).

Moreover, when endogeneity persists even given the interactive fixed effects, then our approach may allow for a larger set of IV's. For instance, one criticism of the so-called "Hausman" instruments (cf. Hausman (1997)) – that is, using the price of product j in market t' as an instrument for the price of product j in market t – is that they may not be independent of "nationwide" demand shocks – that is, product-specific shocks which are correlated across markets. Our interactive fixed effect $\lambda_j' f_t$ can be interpreted as one type of nationwide demand shock, where the λ_j factor loadings capture common (nationwide) components in the shocks across different markets t and t'. Since the instruments in our model can be arbitrarily correlated with λ_j and f_t, the use of Hausman instruments in our model may be (at least partially) immune to the aforementioned criticism.

Next, we introduce the key equations for market shares in the random-coefficient logit demand model. Following Berry, Levinsohn, and Pakes (1995), the probability that agent i chooses product j in market t takes the multinomial logit form:

$$\pi_{jt}(\delta_t, X_t, v_i) = \frac{\exp\left(\delta_{jt} + X_{jt}' v_i\right)}{1 + \sum_{l=1}^{J} \exp\left(\delta_{lt} + X_{lt}' v_i\right)} \,. \tag{2.4}$$

We do not observe individual specific choices, but market shares of the J products in the T markets. The market share of product j in market t is given by

$$s_{jt}(\alpha^0, \delta_t, X_t) = \int \pi_{jt}(\delta_t, X_t, v)\, dG_{\alpha^0}(v) \,, \tag{2.5}$$

where $G_{\alpha^0}(v)$ is the known distribution of consumer taste v_i over the product characteristic, and α^0 is an $L \times 1$ vector of parameters of this distribution.[5]

Underlying these derivations are assumptions that (i) the distributions of $\epsilon = (\epsilon_{ijt})$ and $v = (v_i)$ are mutually independent, and are also independent of $X = (X_{jt})$ and $\xi^0 = (\xi_{jt}^0)$; (ii) ϵ_{ijt} follows a marginal type I extreme value distribution iid across i and j (but not necessarily independent across t).[6] Moreover, the most often used specification for the distribution of random coefficients in the literature is to assume that they have a multivariate normal distribution, that is, $v \sim \mathcal{N}(0, \Sigma^0)$, where Σ^0 is a $K \times K$ matrix of parameters, which can be subject to constraints (e.g. only one or a few regressors may have random coefficients, in which case the components of Σ^0 are only non-zero for these regressors), and α^0 consists of the independent parameters in Σ^0.[7]

[5]The dependence of $\pi_{jt}(\delta_t, X_t, v_i)$ and $s_{jt}(\alpha^0, \delta_t, X_t)$ on t stems from the arguments δ_t and X_t.

[6]When the index t refers to time (or otherwise possesses some natural ordering), then sequential exogeneity is allowed throughout the whole paper, that is, X_{jt} can be correlated with past values of the errors, $e_{js}, s < t$. The errors e_{jt} are assumed to be independent across j and t, but heteroscedasticity is allowed.

[7]We focus in this paper on the case where the functional form of the distribution function G_α is known by the researcher. Recent papers have addressed estimation when this is not known; e.g. Bajari, Fox, Kim and Ryan (2011), (2012).

The observables in this model are the market shares s_{jt} and the regressors X_{jt}.[8] In addition, we need M instruments $Z_{jt} = (Z_{1,jt}, \ldots, Z_{M,jt})'$ to construct extra (unconditional) moment conditions, in addition to the unconditional moment conditions constructed by X_{jt}, in order to estimate the parameters α, with $M \geq L$. These additional instruments are also needed in the usual BLP estimation procedure, even in the absence of the factor structure. Suppose that X_{jt} is exogenous with respect to $\xi^0_{j,t}$. From this, we construct unconditional moment conditions $\mathbb{E}(X_{jt}\xi^0_{j,t}) = 0$. Then, extra moment conditions are still required to identify the covariance parameters in the random coefficients distribution. Notice that those Z's may be non-linear functions of the exogeneous X's, so we do not necessarily need to observe additional exogenous variables.[9]

Let $s = (s_{jt})$, $X_k = (X_{k,jt})$, $Z_m = (Z_{m,jt})$ and $e = (e_{jt})$ be $J \times T$ matrices, and also define the tensors $X = (X_{k,jt})$ and $Z = (Z_{m,jt})$, which contain all observed product characteristics and instruments. In the presence of the unobserved factor structure, it is difficult to identify regression parameters of regressors X_k that have a factor structure themselves, which includes product invariant and time invariant regressors. Our assumptions below rule out all those X_k and Z_m that have a low rank when considered as a $J \times T$ matrix.[10] The unknown parameters are α^0, β^0, λ^0, and f^0.

The existing literature on demand estimation usually considers asymptotic sequences with either J growing large and T fixed, or T growing large and J fixed. Under these standard asymptotic sequences, the estimation of the nuisance parameters λ^0 and f^0 creates a Neyman and Scott (1948) incidental parameter problem: because the number of nuisance parameters grows with the sample size, the estimators for the parameters of interest become inconsistent. Following some recent panel data literature, for example, Hahn and Kuersteiner (2002; 2004) and Hahn and Newey (2004), we handle this problem by considering asymptotic sequences where both J and T become large. Under this alternative asymptotic, the incidental parameter problem is transformed into the issue of

[8] In the present paper we assume that the true market shares $s_{jt} = s_{jt}(\delta^0_t)$ are observed. Berry, Linton and Pakes (2004) explicitly consider sampling error in the observed market shares in their asymptotic theory. Here, we abstract away from this additional complication and focus on the econometric issues introduced by the factor structure in ξ^0.

[9] If one is willing to impose the conditional moment condition $\mathbb{E}(e_{jt}|X_{jt}) = 0$, then valid Z_{jt} can be constructed as non-linear transformations of X_{jt}.

[10] This is exactly analogous to the usual short panel case, in which the presence of fixed effects for each cross-sectional unit precludes identification of the coefficients on time-invariant regressors. If the number of factors R is known accurately, then the coefficients of these low-rank regressors can be identified, but the necessary regularity conditions are relatively cumbersome. For ease of exposition we will therefore rule out both low-rank regressors and low-rank instruments by our assumptions below, and we refer to Bai (2009) and Moon and Weidner (2015a) for a further discussion of this topic.

asymptotic bias in the limiting distribution of the estimators of the parameters of interest. This asymptotic bias can be characterized and corrected for. Our Monte Carlo simulations suggest that the alternative asymptotic provides a good approximation of the properties of our estimator at finite sample sizes, as long as J and T are moderately large.

3 Identification

Given the non-linearity of the model, questions regarding the identification of the model parameters of interest are naturally raised. In the following we provide conditions under which the parameters α and β as well as the product $\lambda f'$ are identified. We do not consider how to identify λ and f separately, because they only enter into the model jointly as $\lambda f'$.[11]

Following standard identification arguments (e.g. Matzkin (2013)), our proof demonstrates identification by showing the existence of an injective mapping from the model parameters $(\alpha, \beta, \lambda f')$ and the distribution of the random elements of the model (e, X, Z) to the distribution of the observed data (s, X, Z), where the random elements of the model are comprised of unobserved error terms, product characteristics, and instruments and the observed data are the market shares, product characteristics, and instruments.[12]

Our identification result utilizes a population distribution of a full $J \times T$ panel of observables (s, X, Z), conditional on parameters α, β and $\lambda f'$. The fact that we have nuisance parameters λ_j and f_t in both panel dimensions makes the distribution of the full $J \times T$ panel of observables a natural starting point for the identification discussion (where J and T are are finite constants in this section). Normally in the large N, T panel data literature (e.g. in Hahn and Newey (2004), Bai (2009), etc.) there is no explicit identification discussion, but consistency as $N, T \to \infty$ (or in our case $J, T \to \infty$) is shown directly. The reason is that there is no fixed population distribution that corresponds to the sample as both panel dimensions become large. Thus, when going from identification to estimation there will not be a simple analog principle that allows to treat the sample as multiple draws from the population. This is a general conceptual issue, independent of our paper. The inference results below therefore do not follow immediately from the identification

[11] The transformation $\lambda \to \lambda S$ and $f \to fS^{-1}$ gives observationally equivalent parameters for any non-degenerate $R \times R$ matrix S. Once the product $\lambda f'$ is identified, one can impose further normalization restrictions to identify λ and f separately, if desired.

[12] Injectivity implies that the mapping is one-to-one – and hence invertible – along the relevant range. The range of this mapping excludes some distributions of (s, X, Z); for instance, distributions in which some of the market shares take zero values with non-zero probability cannot be generated by our model, due to the multinomial logit structure. See Gandhi, Lu, and Shi (2013) for additional discussion of estimating discrete-choice demand models when some of the products are observed to have zero market shares.

result presented in this section; in particular, the incidental parameter problem (Neyman and Scott (1948)) related to inference of λ_j and f_t needs to be properly addressed.

For our identification result we assume, as in BLP, that there exists a one-to-one relationship between market shares and mean utilities, as summarized by the following assumption. Let $\mathcal{B}_\alpha \subset \mathbb{R}^L$ be a given parameter set for α.

Assumption INV (Invertibility Assumptions). *We assume that equation (2.5) is invertible, that is, for each market t the mean utilities $\delta_t = (\delta_{1t}, \ldots, \delta_{Jt})$ are unique functions of $\alpha \in \mathcal{B}_\alpha$, the market shares $s_t = (s_{1t}, \ldots, s_{Jt})$, and the regressors $X_t = (X_{1t}, \ldots, X_{Jt})$. We denote these functions by $\delta_{jt}(\alpha, s_t, X_t)$.*[13]

Berry, Gandhi, and Haile (2013) provide general conditions under which this invertibility assumption is satisfied, and Berry and Haile (2014) and Chiappori and Komunjer (2009) utilize this inverse mapping in their nonparametric identification results.

Using Assumption INV and the specifications (2.2) and (2.3) we have

$$\delta_{jt}^0 = \delta_{jt}(\alpha^0, s_t, X_t) = \sum_{k=1}^{K} \beta_k^0 X_{k,jt} + \sum_{r=1}^{R} \lambda_{jr}^0 f_{tr}^0 + e_{jt} \,. \tag{3.1}$$

In JT-vector notation this equation can be written as $\delta^{\text{vec}}(\alpha^0) = x\beta^0 + \sum_{r=1}^{R} f_{\cdot r}^0 \otimes \lambda_{\cdot r}^0 + e^{\text{vec}}$, where $\delta^{\text{vec}}(\alpha) = \text{vec}[\delta(\alpha, s, X)]$ and $e^{\text{vec}} = \text{vec}(e)$ are JT-vectors, and x is a $JT \times K$ matrix with columns $x_{\cdot,k} = \text{vec}\,(X_k)$. For simplicity we suppress the dependence of $\delta^{\text{vec}}(\alpha)$ on s and X. It is furthermore convenient to define the $JT \times M$ matrix z with columns $z_{\cdot,m} = \text{vec}\,(Z_m)$, the mean utility difference $d(\alpha) = \delta^{\text{vec}}(\alpha) - \delta^{\text{vec}}(\alpha^0)$, and the unobserved utility difference $\Delta\xi_{\alpha,\beta} = d(\alpha) - x(\beta - \beta^0)$. Both $d(\alpha)$ and $\Delta\xi_{\alpha,\beta}$ are JT vectors. Note that $\Delta\xi_{\alpha,\beta}$ is simply the vectorized difference of the residual unobserved product characteristic at (α, β) and (α^0, β^0). In the following the indices j and t run from 1 to J and 1 to T, respectively.

Assumption ID (Assumptions for Identification).

(i) The second moments of $\delta_{jt}(\alpha)$, X_{jt} and Z_{jt} exist for all α, and all j, t.

(ii) $\mathbb{E}(e_{jt}) = 0$.

(iii) $\mathbb{E}(X_{jt} e_{jt}) = 0$, $\mathbb{E}(Z_{jt} e_{jt}) = 0$, *for all j, t.*[14]

(iv) $\mathbb{E}[(x, z)'(\mathbb{1}_T \otimes M_{(\lambda,\lambda^0)})(x, z)] \geq b\, \mathbb{1}_{K+M}$, *for some $b > 0$ and all $\lambda \in \mathbb{R}^{J \times R}$.*[15]

[13] Note that the dependence of $\delta_{jt}(\alpha, s_t, X_t)$ on t stems from the arguments s_t and X_t.

[14] The exogeneity assumption $\mathbb{E}(X_{jt} e_{jt}) = 0$ can be relaxed. Regression coefficients β^{end} on endogenous regressors need to be included in the parameter vector α, i.e. α is replaced by $(\alpha, \beta^{\text{end}})$, see Section 4.1 below.

[15] Here, $P_{(\lambda,\lambda^0)} = (\lambda, \lambda^0)[(\lambda, \lambda^0)'(\lambda, \lambda^0)]^\dagger(\lambda, \lambda^0)'$, where $\dagger$ refers to a generalized inverse, and $M_{(\lambda,\lambda^0)} = \mathbb{1}_J - P_{(\lambda,\lambda^0)}$ are the $J \times J$ matrices that project onto and orthogonal to the span of (λ, λ^0).

(v) For all $(\alpha,\beta)\neq(\alpha^0,\beta^0)$, and all $\lambda\in\mathbb{R}^{J\times R}$ we assume that[15]
$\mathbb{E}\big[\Delta\xi'_{\alpha,\beta}(x,z)\big]\mathbb{E}\big[(x,z)'(x,z)\big]^{-1}\mathbb{E}\big[(x,z)'\Delta\xi_{\alpha,\beta}\big] > \mathbb{E}\left[\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{(\lambda,\lambda^0)}\right)\Delta\xi_{\alpha,\beta}\right]$.

In this assumption, and also for the remainder of the paper, we treat the fixed effects λ^0 and f^0 as non-random parameters, that is, all expectations in Assumption ID are implicitly conditional on λ^0 and f^0. The assumptions are discussed in Section 3.1 below.

To formulate our identification result we need to introduce some additional notation. We denote the set of joint distributions of e, X, Z by $\mathcal{F}_{e,X,Z}$, and the set of joint distributions of s, X, Z (the observables) by $\mathcal{F}_{s,X,Z}$. The model described in Section 2 gives unique market shares s for any given e, X, Z and parameters α, β, $\lambda f'$. The model therefore also uniquely describes the distribution of observables for a given distribution $F_{e,X,Z}\in\mathcal{F}_{e,X,Z}$ and parameters α, β, $\lambda f'$, and we denote this distribution of observables given by the model as $\Gamma(\alpha,\beta,\lambda f',F_{e,X,Z})\in\mathcal{F}_{s,X,Z}$. We say that two distributions $F_1,F_2\in\mathcal{F}_{s,X,Z}$ are equal if the corresponding joint cdf's are the same, and we write $F_1=F_2$ in that case. Analogously, we define equality on $\mathcal{F}_{e,X,Z}$.

Theorem 3.1 (Identification). *Let Assumption INV be satisfied. Let $F^0_{e,X,Z}\in\mathcal{F}_{e,X,Z}$ be such that it satisfies Assumption ID. Let $F_{e,X,Z}\in\mathcal{F}_{e,X,Z}$ and consider two sets of parameters $(\alpha,\beta,\lambda f')$ and $(\alpha^0,\beta^0,\lambda^0 f^{0\prime})$. Then, $\Gamma(\alpha,\beta,\lambda f',F_{e,X,Z})=\Gamma(\alpha^0,\beta^0,\lambda^0 f^{0\prime},F^0_{e,X,Z})$ implies that $\alpha=\alpha^0$, $\beta=\beta^0$, $\lambda f'=\lambda^0 f^{0\prime}$ and $F_{e,X,Z}=F^0_{e,X,Z}$.*

The theorem states that if the distribution of observables $F^0_{s,X,Z}=\Gamma(\alpha^0,\beta^0,\lambda^0 f^{0\prime},F^0_{e,X,Z})$ is generated from the parameters $(\alpha^0,\beta^0,\lambda^0 f^{0\prime})$ and $F^0_{e,X,Z}$, satisfying Assumption ID, then any other $(\alpha,\beta,\lambda f')$ and $F_{e,X,Z}$ that generate the same distribution of observables $F^0_{s,X,Z}=\Gamma(\alpha,\beta,\lambda f',F_{e,X,Z})$ must be equal to the original $(\alpha^0,\beta^0,\lambda^0 f^{0\prime})$ and $F^0_{e,X,Z}$. In other words, we can uniquely recover the model parameters from the distribution of observables. Two observationally equivalent model structures $(\alpha^0,\beta^0,\lambda^0 f^{0\prime},F^0_{e,X,Z})$ and $(\alpha,\beta,\lambda f',F_{e,X,Z})$ need to be identical.

The key tool for the proof of Theorem 3.1 is the the expected least squares objective function

$$Q\left(\alpha,\beta,\gamma,\lambda,f;F^0_{s,X,Z}\right)=\mathbb{E}_0\left\{\sum_{j=1}^{J}\sum_{t=1}^{T}\left[\delta_{jt}(\alpha)-X'_{jt}\beta-Z'_{jt}\gamma-\lambda'_j f_t\right]^2\right\},$$

where $\gamma\in\mathbb{R}^L$ is an auxiliary parameter, and $\mathbb{E}_0$ refers to the expectation under the distribution of observables $F^0_{s,X,Z}$,[16] which is assumed to be generated from the model, i.e. $F^0_{s,X,Z}=\Gamma(\alpha^0,\beta^0,\lambda^0 f^{0\prime},F^0_{e,X,Z})$, with $F^0_{e,X,Z}$ satisfying Assumption ID.

[16]Normally, we refer to $\mathbb{E}_0$ simply as $\mathbb{E}$. We only use different notation here to stress at which point the argument $F^0_{s,X,Z}$ enters into $Q\left(\alpha,\beta,\gamma,\lambda,f;F^0_{s,X,Z}\right)$.

The true value of the auxiliary parameter γ is zero, because of the exclusion restriction on Z_{jt}. In the proof of Theorem 3.1 we show that under our assumptions the minimizer of $Q\left(\alpha, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$ over $(\beta, \lambda, f, \gamma)$, for fixed α, only satisfies $\gamma = 0$ if and only if $\alpha = \alpha^0$. Thus, by using the expected least squares objective function as a tool we can uniquely identify α^0 from the distribution of obervables $F^0_{s,X,Z}$. Having identified α^0 we can identify β^0 and $\lambda^0 f^{0\prime}$ simply as the unique minimizers of $Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$. These findings immediately preclude observational equivalence, *viz* two sets of distinct parameters $(\alpha^0, \beta^0, \lambda^0 f^{0\prime}) \neq (\alpha^1, \beta^1, \lambda^1 f^{1\prime})$ which are both consistent with the observed distribution $F^0_{s,X,Z}$. For complete details we refer to the proof in the appendix. Furthermore, our identification argument is constructive, as it leads naturally to the LS-MD estimator which we introduce in subsequent sections.

3.1 Discussion of the Identification Conditions

In this section we discuss the conditions of the identification theorem. First, we note that when no factors are present ($R = 0$), then our identification Assumptions ID below essentially require that the unconditional moment conditions $\mathbb{E}(X_{jt}e_{jt}) = 0$ and $\mathbb{E}(Z_{jt}e_{jt}) = 0$ uniquely identify the model parameters α and β, thus following the original identification strategy in BLP (1995).[17]

Assumption (i) demands existence of second moments, assumption (ii) requires the error process to have zero mean, and assumption (iii) imposes exogeneity of the product characteristics X_{jt} and the instruments Z_{jt} with respect to the error e_{jt} (endogenous regressors are discussed in Section 4.1). Apart from the term $M_{(\lambda,\lambda^0)}$, Assumption ID(iv) is a standard non-collinearity condition on the product characteristics and the instruments – which jointly appear as regressors in the first step of (4.1). The generalized condition $\mathbb{E}[(x,z)'(\mathbb{1}_T \otimes M_{(\lambda,\lambda^0)})(x,z)] \geq b > 0$ requires non-collinearity of the regressors even after projecting out all directions proportional to the true factor loading λ^0 and to any other possible factor loadings λ. A sufficient condition for this assumption is the rank condition $\text{rank}[\mathbb{E}(\Xi\,\Xi')] > 2R$ for any non-zero linear combination $\Xi = \beta \cdot X + \gamma \cdot Z$. This rank condition, for example, rules out product-invariant regressors and instruments, as already mentioned above.

Those parts of the conditions (i) to (iv) that do not contain Z_{jt} are used to identify β^0 and $\lambda^0 f^{0\prime}$ when α^0 is already identified. These conditions are typical regularity con-

[17] As such, our identification results do not add to the literature on non-parametric identification of the BLP model (as in Berry and Haile (2014), Chiappori and Komunjer (2009), Bajari, Fox, Kim and Ryan (2011)); our concern is, rather, to show that the logit demand model with parametrically-distributed random coefficients can still be identified after the introduction of the interactive fixed effects.

ditions for identification of a linear regression model with a modification only required in condition (iv) to accommodate the interactive fixed effects. (See also Moon and Weidner (2015b).)

The key additional assumption that we need for identification of α^0 is Assumption ID(v). Note that $\Delta\xi_{\alpha^0,\beta^0} = 0$, that is, both the left and right side of the inequality in assumption (v) are zero for $(\alpha, \beta) = (\alpha^0, \beta^0)$, which is why this case is explicitly ruled out in the assumption. The left hand side of the inequality in assumption (v) is the sum of squares of that part of $\Delta\xi_{\alpha,\beta}$ that is explained by the regressors x and the instruments z. The right hand side is the sum of squares of that part of $\Delta\xi_{\alpha,\beta}$ that is explained by the true factor loading λ^0 and an arbitrary other factor loading λ. Thus, the condition is a relevance condition on the instruments, which requires that the explanatory power of the regressors and the instruments needs to be larger than the explanatory power of λ and λ^0 for $\Delta\xi_{\alpha,\beta}$.

A more concrete intuition for Assumption ID(v) can be obtained in the case without factors. Without factors, the identification condition simplifies to $\forall(\alpha, \beta) \neq (\alpha^0, \beta^0)$:

$$\mathbb{E}\left[\Delta\xi'_{\alpha,\beta}(x,z)\right]\mathbb{E}\left[(x,z)'(x,z)\right]^{-1}\mathbb{E}\left[(x,z)'\Delta\xi_{\alpha,\beta}\right] > 0. \tag{3.2}$$

This can be shown to be equivalent to the statement $\forall\alpha \neq \alpha^0$:

$$\mathbb{E}\left[d(\alpha)'(x,z)\right]\mathbb{E}\left[(x,z)'(x,z)\right]^{-1}\mathbb{E}\left[(x,z)'d(\alpha)\right] > \mathbb{E}\left[d(\alpha)'x\right]\mathbb{E}\left(x'x\right)^{-1}\mathbb{E}\left[x'd(\alpha)\right]. \tag{3.3}$$

We see that this condition is nothing more than the usual instrument relevance condition (for z in this case) underlying the typical GMM approach in estimating BLP models. It can also be shown to be equivalent to the condition that for all $\alpha \neq \alpha^0$ the matrix $\mathbb{E}[(d(\alpha), x)'(x, z)]$ has full rank (equal to $K + 1$).

The matrix valued function $\delta(\alpha) = \delta(\alpha, s, X)$ was introduced as the inverse of equation (2.5) for the market shares $s_{jt}(\delta_t)$. Thus, once a functional form for $s_{jt}(\delta_t)$ is chosen and some distributional assumptions on the data generating process are made, it is in principle possible to analyze Assumption ID(v) further and to discuss validity and optimality of the instruments. Unfortunately, too little is known about the properties of $\delta(\alpha)$ to enable a general analysis.[18] For this reason, in our Monte Carlo simulations in section 6 below, we provide both analytical and and numerical verifications for Assumption ID(v) for the specific setup there.

The final remark is that Assumption ID(v) also restricts the family of the distribution of the random coefficient. As a very simple example, suppose that we would specify the

[18]This is a problem not only with our approach, but also with the estimators in BLP, and for Berry, Linton and Pakes (2004).

distribution G_α for the random vector v as $v \sim \mathcal{N}(\alpha_1, \alpha_2)$, where $\alpha = (\alpha_1, \alpha_2)$, and we would also include a constant in the vector of regressors X_{jt}. Then, the regression coefficient on the constant and α_1 cannot be jointly identified (because they both shift mean utility by a constant, but have no other effect), and Assumption ID(v) will indeed be violated in this case.

4 LS-MD Estimator

If δ_{jt}^0 is known, then the above model reduces to the linear panel regression model with interactive fixed effects. Estimation of this model was discussed under fixed T asymptotics in, for example, Holtz-Eakin, Newey and Rosen (1988), and Ahn, Lee, Schmidt (2001), and for $J, T \to \infty$ asymptotics in Bai (2009), and Moon and Weidner (2015a; 2015b).

The computational challenge in estimating the model (3.1) lies in accommodating both the model parameters (α, β), which in the existing literature has mainly been done in a GMM framework, as well as the nuisance elements λ_j, f_t, which in the existing literature have been treated using a principal components decomposition in a least-squares context (e.g., Bai (2009), and Moon and Weidner (2015a; 2015b)). Our estimation procedure – which mimics the identification proof discussed previously – combines both the GMM approach to demand estimation and the least squares approach to the interactive fixed effect model.

Definition: the *least squares-minimum distance (LS-MD)* estimators for α and β are defined by

Step 1 (least squares): for given α let

$$\delta(\alpha) = \delta(\alpha, s, X) ,$$

$$\left(\tilde{\beta}_\alpha , \tilde{\gamma}_\alpha , \tilde{\lambda}_\alpha , \tilde{f}_\alpha\right) = \underset{\{\beta,\gamma,\lambda,f\}}{\operatorname{argmin}} \sum_{j=1}^{J}\sum_{t=1}^{T} \left[\delta_{jt}(\alpha) - X_{jt}'\beta - Z_{jt}'\gamma - \lambda_j' f_t\right]^2 ,$$

Step 2 (minimum distance):

$$\widehat{\alpha} = \underset{\alpha\in\mathcal{B}_\alpha}{\operatorname{argmin}} \, \tilde{\gamma}_\alpha' \, W_{JT} \, \tilde{\gamma}_\alpha ,$$

Step 3 (least squares):

$$\delta(\widehat{\alpha}) = \delta(\widehat{\alpha}, s, X) ,$$

$$\left(\widehat{\beta} , \widehat{\lambda} , \widehat{f}\right) = \underset{\{\beta,\lambda,f\}}{\operatorname{argmin}} \sum_{j=1}^{J}\sum_{t=1}^{T} \left[\delta_{jt}(\widehat{\alpha}) - X_{jt}'\beta - \lambda_j' f_t\right]^2 . \tag{4.1}$$

Here, $\beta \in \mathbb{R}^K$, $\delta(\alpha, s, X)$, X_k and Z_m are $J \times T$ matrices, λ is $J \times R$, f is $T \times R$, W_{JT} is a positive definite $M \times M$ weight matrix, $\mathcal{B}_\alpha \subset \mathbb{R}^L$ is an appropriate parameter set for α.

Steps 1 and 2 are nested, because $\tilde{\gamma}_\alpha$ defined by step 1 needs to be calculated multiple times while performing the numerical optimization in step 2, but step 3 only needs to be performed once after the calculation of $\widehat{\alpha}$ in step 2 is finished. Step 1 resembles the linear least-squares estimators with interactive fixed effects considered in Bai (2009) and Moon and Weidner (2015a), but because our model also includes the nonlinear parameter α, this step is nested within step 2, which involves iteration over different candidate values for α.

In step 1, we include the IV's Z_m as auxiliary regressors, with coefficients $\gamma \in \mathbb{R}^M$. Step 2 is based on imposing the exclusion restriction on the IV's, which requires that $\gamma = 0$, at the true value of α. Thus, we first estimate β, λ, f, and the instrument coefficients γ by least squares for fixed α, and subsequently we estimate α by minimizing the norm of $\tilde{\gamma}_\alpha$ with respect to α.

Step 3 in (4.1), which defines $\widehat{\beta}$, is just a repetition of step 1, but with $\alpha = \widehat{\alpha}$ and $\gamma = 0$. One could also use the step 1 estimator $\tilde{\beta}_{\widehat{\alpha}}$ to estimate β. Under the assumptions for consistency of $(\widehat{\alpha}, \widehat{\beta})$ presented below, this alternative estimator is also consistent for β^0. However, in general $\tilde{\beta}_{\widehat{\alpha}}$ has a larger variance than $\widehat{\beta}$, since irrelevant regressors are included in the estimation of $\tilde{\beta}_{\widehat{\alpha}}$.

For given α, β and γ the optimal factors and factor loadings in the least squares problems in step 1 (and step 3) of (4.1) turn out to be the principal components estimators for λ and f. These incidental parameters can therefore be concentrated out easily, and the remaining objective function for β and γ turns out to be given by an eigenvalue problem (see e.g. Moon and Weidner (2015a; 2015b) for details), namely

$$\left(\tilde{\beta}_\alpha\,,\,\tilde{\gamma}_\alpha\right) = \underset{\{\beta,\gamma\}}{\operatorname{argmin}} \sum_{r=R+1}^{T} \mu_r\left[\left(\delta(\alpha) - \beta\cdot X - \gamma\cdot Z\right)'\left(\delta(\alpha) - \beta\cdot X - \gamma\cdot Z\right)\right]\,, \qquad (4.2)$$

where $\beta\cdot X = \sum_{k=1}^{K}\beta_k X_k$, $\gamma\cdot Z = \sum_{m=1}^{M}\gamma_m Z_m$, and $\mu_r(.)$ refers to the r'th largest eigenvalue of the argument matrix. This formulation greatly simplifies the numerical calculation of the estimator, since eigenvalues are easy and fast to compute, and we only need to perform numerical optimization over β and γ, not over λ and f.

The step 1 optimization problem in (4.1) has the same structure as the interactive fixed effect regression model. Thus, for $\alpha = \alpha^0$ it is known from Bai (2009), and Moon and Weidner (2015a; 2015b) that (under their assumptions) $\tilde{\beta}_{\alpha^0}$ is $\sqrt{JT}$-consistent for β^0 and asymptotically normal as $J, T \to \infty$ with $J/T \to \kappa^2$, $0 < \kappa < \infty$.

Step 1 also involves solving for the vector of δ's which solves the market share equations (2.5), at a given value for α. This computational problem is well-studied in the BLP literature.[19]

[19] We solve it using nonlinear equation solvers, which is a relatively standard procedure from the existing BLP

The LS-MD estimator we propose above is distinctive, because of the inclusion of the instruments Z as regressors in the first-step. This can be understood as a generalization of an estimation approach for a linear regression model with endogenous regressors. Consider a simple structural equation $y_1 = Y_2\alpha + e$, where the endogenous regressors Y_2 have the reduced form specification $Y_2 = Z\delta + V$, and e and V are correlated. The two stage least squares estimator of α is $\widehat{\alpha}_{2\text{SLS}} = (Y_2'P_ZY_2)^{-1}\, Y_2'P_Zy_1$, where $P_Z = Z\,(Z'Z)^{-1}\,Z'$. In this set up, it is possible to show that $\widehat{\alpha}_{2\text{SLS}}$ is also an LS-MD estimator with a suitable choice of the weight matrix. Namely, in the first step the OLS regression of $(y_1 - Y_2\alpha)$ on Z yields the OLS estimator $\tilde{\gamma}_\alpha = (Z'Z)^{-1}\,Z'\,(y_1 - Y_2\alpha)$. Then, in the second step minimizing the distance $\tilde{\gamma}_\alpha' W \tilde{\gamma}_\alpha$ with respect to α gives $\widehat{\alpha}(W) = [Y_2'Z(Z'Z)^{-1}W(Z'Z)^{-1}Z'Y_2]^{-1}$ $[Y_2'Z(Z'Z)^{-1}W(Z'Z)^{-1}Z'y_1]$. Choosing $W = Z'Z$ thus results in $\widehat{\alpha} = \widehat{\alpha}\,(Z'Z) = \widehat{\alpha}_{2\text{SLS}}$. Obviously, for our nonlinear model, strict 2SLS is not applicable; however, our estimation approach can be considered a generalization of this alternative iterative estimator, in which the exogenous instruments Z are included as "extra" regressors in the initial least-squares step.[20]

The two-step procedure in the LS-MD estimation is similar to the two stage estimation method in Chernozhukov and Hansen (2006) that investigated endogenous quantile regressions.

4.1 Extension: regressor endogeneity with respect to e_{jt}

So far, we have assumed that the regressors X could be endogenous only through the factors $\lambda_j' f_t$, and they are exogenous wrt e. However, this could be restrictive in some applications, for example, when price p_{jt} is determined by ξ_{jt} contemporaneously. Hence, we consider here the possibility that the regressors X could also be correlated with e. This is readily accommodated within our framework. Let $X^{\text{end}} \subset X$ denote the endogenous regressors, with $\dim(X^{\text{end}}) = K_2$. (Hence, the number of exogenous regressors equals $K - K_2$.) Similarly, let β^{end} denote the coefficients on these regressors, while β continues to denote the coefficients on the exogenous regressors. Correspondingly, we assume that M, the number of instruments, exceeds $L + K_2$.

literature. Its validity is ensured by results (in Berry, Levinsohn, Pakes (1995)) showing that, for fixed α, these equations constitute a contraction mapping, and the nonlinear equation solver recovers the (unique) fixed point.

[20]Moreover, the presence of the factors makes it inappropriate to use the moment condition-based GMM approach proposed by BLP, see Appendix B. Moment based approaches to factor model estimation like Holtz-Eakin, Newey and Rosen (1988) and Ahn, Lee, Schmidt (2001; 2013) would also have to be non-trivially extended to handle the random coefficient parameter α in the presence of two dimensional incidental parameters in a nonlinear framework, but we have not explored this possibility.

Definition: the *least-squares minimum distance (LS-MD)* estimators for α and β with endogenous regressors X^{end} is defined by:

step 1: for given $\alpha^{\text{end}} = (\alpha, \beta^{\text{end}})$ let

$$\delta(\alpha) \;=\; \delta(\alpha,\, s,\, X)\;,$$

$$\left(\tilde{\beta}_{\alpha^{\text{end}}}\,,\; \tilde{\gamma}_{\alpha^{\text{end}}}\,,\; \tilde{\lambda}_{\alpha^{\text{end}}}\,,\; \tilde{f}_{\alpha^{\text{end}}}\right) \;=\; \underset{\{\beta,\,\gamma,\,\lambda,\,f\}}{\text{argmin}} \sum_{j=1}^{J}\sum_{t=1}^{T}\left[\delta_{jt}(\alpha) - X_{jt}^{\text{end}\prime}\beta^{\text{end}} - X_{jt}'\beta - Z_{jt}'\gamma - \lambda_j' f_t\right]^2,$$

step 2:

$$\widehat{\alpha}^{\text{end}} = (\widehat{\alpha}, \widehat{\beta}^{\text{end}}) \;=\; \underset{\alpha^{\text{end}}\in\mathcal{B}_\alpha\times\mathcal{B}_\beta^{\text{end}}}{\text{argmin}}\; \tilde{\gamma}_{\alpha^{\text{end}}}'\, W_{JT}\, \tilde{\gamma}_{\alpha^{\text{end}}}\;,$$

step 3:

$$\delta(\widehat{\alpha}) \;=\; \delta(\widehat{\alpha},\, s,\, X)\;,$$

$$\left(\widehat{\beta}\,,\; \widehat{\lambda}\,,\; \widehat{f}\right) \;=\; \underset{\{\beta\in\mathbb{R}^K\,,\,\lambda,\,f\}}{\text{argmin}} \sum_{j=1}^{J}\sum_{t=1}^{T}\left[\delta_{jt}(\widehat{\alpha}) - X_{jt}^{\text{end}\prime}\beta^{\text{end}} - X_{jt}'\beta - \lambda_j' f_t\right]^2, \tag{4.3}$$

where $\mathcal{B}_\alpha$ and $\mathcal{B}_\beta^{\text{end}}$ are parameter sets for α and β^{end}.

The difference between this estimator, and the previous one for which all the regressors were assumed exogenous, is that the estimation of β^{end}, the coefficients on the endogenous regressors $\tilde{X}$, has been moved to the second step. The estimation procedure in (4.3) can me mapped into our original LS-MD procedure in (4.1), if we make the following formal replacements:

$$\alpha^{\text{end}} = (\alpha, \beta^{\text{end}}) \mapsto \alpha, \qquad\qquad \delta(\alpha) - \beta^{\text{end}} \cdot X^{\text{end}} \mapsto \delta(\alpha). \tag{4.4}$$

Thus, by changing the meaning of α and $\delta(\alpha)$ accordingly, the identification result above is still valid, and all results below on the consistency, asymptotic distribution and bias correction of the LS-MD estimator (4.1) with only (sequentially) exogenous regressors directly generalize to the estimator (4.3) with more general endogenous regressors. Given this discussion, we see that the original BLP (1995) model can be considered a special case of our model in which factors are absent (i.e. $R = 0$).

5 Consistency and Asymptotic Distribution

In this section we present our results on the properties of the LS-MD estimator $\widehat{\alpha}$ and $\widehat{\beta}$ defined in (4.1) under the asymptotics $J, T \to \infty$.

Assumption 1 (Assumptions for Consistency).

(i) $\sup_{\alpha\in\mathcal{B}_\alpha\setminus\alpha^0} \frac{\|\delta(\alpha)-\delta(\alpha^0)\|_F}{\|\alpha-\alpha^0\|} = \mathcal{O}_p(\sqrt{JT})$,

$\|X_k\|_F = \mathcal{O}_p(\sqrt{JT})$, $\|Z_m\|_F = \mathcal{O}_p(\sqrt{JT})$, *for* $k = 1,\dots,K$ *and* $m = 1,\dots,M$.

(ii) $\|e\| = \mathcal{O}_p(\sqrt{\max(J,T)})$.

(iii) $\frac{1}{JT}\operatorname{Tr}(X_k e') = o_p(1)$, $\frac{1}{JT}\operatorname{Tr}(Z_m e') = o_p(1)$, *for* $k = 1,\dots,K$ *and* $m = 1,\dots,M$.[21]

(iv) $\min_{\lambda\in\mathbb{R}^{J\times R}} \left\{\mu_{K+M}\left[\frac{1}{JT}(x,z)'(\mathbb{1}_T\otimes M_{(\lambda,\lambda^0)})(x,z)\right]\right\} \geq b$, *wpa1, for some* $b>0$.

(v) There exists $b>0$ *such that wpa1 for all* $\alpha\in\mathcal{B}_\alpha$ *and* $\beta\in\mathbb{R}^K$

$$\left[\tfrac{1}{JT}\Delta\xi'_{\alpha,\beta}(x,z)\right]\left[\tfrac{1}{JT}(x,z)'(x,z)\right]^{-1}\left[\tfrac{1}{JT}(x,z)'\Delta\xi_{\alpha,\beta}\right] - \max_{\lambda\in\mathbb{R}^{J\times R}}\left[\tfrac{1}{JT}\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{(\lambda,\lambda^0)}\right)\Delta\xi_{\alpha,\beta}\right] \geq b\|\alpha-\alpha^0\|^2 + b\|\beta-\beta^0\|^2.$$

(vi) $W_{JT} \to_p W > 0$.

Theorem 5.1 (Consistency). *Let Assumption 1 hold, and let* $\alpha^0\in\mathcal{B}_\alpha$. *In the limit* $J,T\to\infty$ *we then have* $\widehat{\alpha} = \alpha^0 + o_p(1)$, *and* $\widehat{\beta} = \beta^0 + o_p(1)$.

The proof of Theorem 5.1 is given in the appendix. The similarity between Assumption 1 and Assumption ID is obvious, so that for the most part we can refer to Section 3.1 for the interpretation of these assumptions, and in the following we focus on discussing the differences between the consistency and identification assumptions. The one additional assumption is the last one, which requires existence of a positive definite probability limit of the weight matrix W_{JT}.

Apart from a rescaling with appropriate powers of JT, the Assumptions $1(i)$, (iii), (iv), and (v) are almost exact sample analogs of their identification counterparts in Assumption ID. The two main differences are that assumption (i) also imposes a Lipschitz-like continuity condition on $\delta(\alpha)$ around α^0, and that the right hand-side of the inequality in assumption (v) is not just zero, but a quadratic form in $(\alpha-\alpha^0)$ and $(\beta-\beta^0)$ — the latter is needed, because expressions which are exactly zero in the identification proof are now only converging to zero asymptotically.

Assumption $1(ii)$ imposes a bound on the the spectral norm of e, which is satisfied as long as e_{jt} has mean zero, has a uniformly bounded fourth moment (across j,t,J,T) and is weakly correlated across j and t.[22] The assumption is therefore the analog of Assumption ID(ii).

[21] We can relax the exogeneity assumption $\frac{1}{JT}\operatorname{Tr}(X_k e') = o_p(1)$. For all endogenous regressor the corresponding regression coefficients β^{end} need to be included in the parameter vector α, see the replacement (4.4) above.

[22] Such a statement on the spectral norm of a random matrix is a typical result in random matrix theory. The difficulty – and the reason why we prefer such a high-level assumption on the spectral norm of e – is to

At finite J, T, a sufficient condition for existence of $b > 0$ such that the inequality in Assumption 1(iv) is satisfied, is $\text{rank}(\Xi) > 2R$ for any non-zero linear combination Ξ of X_k and Z_m. This rank condition rules out product-invariant and market-invariant product characteristics X_k and instruments Z_m, since those have rank 1 and can be absorbed into the factor structure.[23] There are many reformulations of this rank condition, but in one formulation or another this rank condition can be found in any of the above cited papers on linear factor regressions, and we refer to Bai (2009), and Moon and Weidner (2015a) for a further discussion.

Next, we present results on the limiting distribution of $\widehat{\alpha}$ and $\widehat{\beta}$. Some further regularity condition are necessary to derive the limiting distribution of our LS-MD estimator, and those are summarized in Assumption 2 to 4 in the appendix. These assumptions are straightforward generalization of the assumptions imposed by Moon and Weidner (2015a; 2015b) for the linear model, except for part (i) of Assumption 4, which demands that $\delta(\alpha)$ can be linearly approximated around α^0 such that the Frobenius norm of the remainder term of the expansion is of order $o_p(\sqrt{JT}\|\alpha - \alpha^0\|)$ in any $\sqrt{J}$ shrinking neighborhood of α^0. Notice also that Assumption 4(iv) implies $\mathbb{E}(e_{jt}|X_{jt}, Z_{jt}) = 0$, while so far we only required e_{jt} to be uncorrelated with X_{jt} and Z_{jt}.

Theorem 5.2. *Let Assumptions 1, 2, 3 and 4 be satisfied, and let α^0 be an interior point of $\mathcal{B}_\alpha$. In the limit $J,T \to \infty$ with $J/T \to \kappa^2$, $0 < \kappa < \infty$, we then have*

$$\sqrt{JT}\begin{pmatrix} \widehat{\alpha} - \alpha^0 \\ \widehat{\beta} - \beta^0 \end{pmatrix} \underset{d}{\to} \mathcal{N}\left(\kappa B_0 + \kappa^{-1} B_1 + \kappa B_2,\ \left(G\mathcal{W}G'\right)^{-1} G\mathcal{W}\Omega\mathcal{W}G' \left(G\mathcal{W}G'\right)^{-1}\right),$$

with the formulas for G, $\mathcal{W}$, Ω, B_0, B_1 and B_2 given in the appendix C.

The proof of Theorem 5.2 is provided in the appendix. Analogous to the least squares estimator in the linear model with interactive fixed effects, there are three bias terms in the limiting distribution of the LS-MD estimator. The bias term κB_0 is only present if regressors or instruments are pre-determined, that is, if X_{jt} or Z_{jt} are correlated with $e_{j\tau}$ for $t > \tau$ (but not for $t = \tau$, since this would violate weak exogeneity). A reasonable

specify the meaning of "weakly correlated across j and t". The extreme case is obviously independence across j and t, but weaker assumptions are possible. We refer to the discussion in Moon and Weidner (2015a) for other examples.

[23] Inclusion of product-invariant and market-invariant characteristics ("low-rank regressors") does not hamper the identification and estimation of the regression coefficients on the other ("high-rank") regressors. This is because including low-rank regressors is equivalent to increasing the number of factors R, and then imposing restrictions on the factors and factors loadings of these new factors. Conditions under which the coefficients of low-rank regressors can be estimated consistently are discussed in Moon and Weidner (2015a).

interpretation of this bias terms thus requires that the index t refers to time, or has some other well-defined ordering. The other two bias terms $\kappa^{-1}B_1$ and κB_2 are due to heteroscedasticity of the idiosyncratic error e_{jt} across firms j and markets t, respectively. The first and last bias terms are proportional to κ, and thus are large when T is small compared to J, while the second bias terms is proportional to κ^{-1}, and thus is large when T is large compared to J. Note that no asymptotic bias is present if regressors and instruments are strictly exogenous and errors e_{jt} are homoscedastic. There is also no asymptotic bias when $R = 0$, since then there are no incidental parameters. For a more detailed discussion of the asymptotic bias, we again refer to Bai (2009) and Moon and Weidner (2015a).

While the structure of the asymptotic bias terms is analogous to the bias encountered in linear models with interactive fixed effects, we find that the structure of the asymptotic variance matrix for $\widehat{\alpha}$ and $\widehat{\beta}$ is analogous to the GMM variance matrix. The LS-MD estimator can be shown to be equivalent to the GMM estimator if no factors are present. In that case the weight matrix $\mathcal{W}$ that appears in Theorem 5.2 can be shown to be the probability limit of the GMM weight matrix that is implicit in our LS-MD approach and, thus, our asymptotic variance matrix exactly coincides with the one for GMM (see also Appendix B). If factors are present, there is no GMM analog of our estimator, but the only change in the structure of the asymptotic variance matrix is the appearance of the projectors M_{f^0} and M_{λ^0} in the formulas for G, Ω and $\mathcal{W}$. The presence of these projectors implies that those components of X_k and Z_m which are proportional to f^0 and λ^0 do not contribute to the asymptotic variance, that is, do not help in the estimation of $\widehat{\alpha}$ and $\widehat{\beta}$. This is again analogous the standard fixed effect setup in panel data, where time-invariant components do not contribute to the identification of the regression coefficients.

Using the explicit expressions for the asymptotic bias and variance of the LS-MD estimator, one can provide estimators for this asymptotic bias and variance. By replacing the true parameter values (α^0, β^0, λ^0, f^0) by the estimated parameters ($\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\lambda}$, $\widehat{f}$), the error term (e) by the residuals ($\widehat{e}$), and population values by sample values it is easy to define estimators $\widehat{B}_0$, $\widehat{B}_1$, $\widehat{B}_2$, $\widehat{G}$, $\widehat{\Omega}$ and $\widehat{\mathcal{W}}$ for B_0, B_1, B_2, G, Ω and $\mathcal{W}$. This is done explicitly in appendix C.4.

Theorem 5.3. *Let the assumptions of Theorem 5.2 and Assumption 5 be satisfied. In the limit $J,T \to \infty$ with $J/T \to \kappa^2$, $0 < \kappa < \infty$ we then have $\widehat{B}_1 = B_1 + o_p(1)$, $\widehat{B}_2 = B_2 + o_p(1)$, $\widehat{G} = G + o_p(1)$, $\widehat{\Omega} = \Omega + o_p(1)$ and $\widehat{\mathcal{W}} = \mathcal{W} + o_p(1)$. If in addition the bandwidth parameter h, which enters in the definition of $\widehat{B}_0$, satisfies $h \to \infty$ and $h^5/T \to 0$, then we also have $\widehat{B}_0 = B_0 + o_p(1)$.*

The proof is again given in the appendix. Theorem 5.3 motivates the introduction of

the bias corrected estimator

$$\begin{pmatrix} \widehat{\alpha}^* \\ \widehat{\beta}^* \end{pmatrix} = \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} - \frac{1}{T}\widehat{B}_0 - \frac{1}{J}\widehat{B}_1 - \frac{1}{T}\widehat{B}_2 \, . \tag{5.1}$$

Under the assumptions of Theorem 5.3 the bias corrected estimator is asymptotically unbiased, normally distributed, and has asymptotic variance $(G\mathcal{W}G')^{-1} G\mathcal{W}\Omega\mathcal{W}G' (G\mathcal{W}G')^{-1}$, which is consistently estimated by $\left(\widehat{G}\widehat{\mathcal{W}}\widehat{G}'\right)^{-1} \widehat{G}\widehat{\mathcal{W}}\widehat{\Omega}\widehat{\mathcal{W}}\widehat{G}' \left(\widehat{G}\widehat{\mathcal{W}}\widehat{G}'\right)^{-1}$. These results allow inference on α^0 and β^0.

From the standard GMM analysis it is know that the $(K+M)\times(K+M)$ weight matrix $\mathcal{W}$ which minimizes the asymptotic variance is given by $\mathcal{W} = c\,\Omega^{-1}$, where c is an arbitrary scalar. If the errors e_{jt} are homoscedastic with variance σ_e^2 we have $\Omega = \sigma_e^2 \operatorname{plim}_{J,T\to\infty} \frac{1}{JT}\left(x^{\lambda f}, z^{\lambda f}\right)'\left(x^{\lambda f}, z^{\lambda f}\right)$, with $x^{\lambda f}$ and $z^{\lambda f}$ defined in Appendix C. In this case it is straightforward to show that the optimal $\mathcal{W} = \sigma_e^2\,\Omega^{-1}$ is attained by choosing

$$W_{JT} = \frac{1}{JT} z' M_{x^{\lambda f}} z \, . \tag{5.2}$$

Under homoscedasticity this choice of weight matrix is optimal in the sense that it minimizes the asymptotic variance of our LS-MD estimator, but nothing is known about the efficiency bound in the presence of interactive fixed effects, that is, a different alternative estimator could theoretically have even lower asymptotic variance.

The unobserved factor loading λ^0 and factor f^0 enter into the definition of $x^{\lambda f}$ and thus also into the optimal W_{JT} in (5.2). A consistent estimator for the optimal W_{JT} can be obtained by estimating λ^0 and f^0 in a first stage LS-MD estimation, using an arbitrary positive definite weight matrix.

Under heteroscedasticity of e_{jt} there are in general not enough degrees of freedom in W_{JT} to attain the optimal $\mathcal{W}$. The reason for this is that we have chosen the first stage of our estimation procedure to be an ordinary least squares step, which is optimal under homoscedasticity but not under heteroscedasticity. By generalizing the first stage optimization to weighted least squares one would obtain the additional degrees of freedom to attain the optimal $\mathcal{W}$ also under heteroscedasticity, but in the present paper we will not consider this possibility further.

6 Monte Carlo Simulations

We consider a model with only one regressors $X_{jt} = p_{jt}$, which we refer to as price. The data generating process for mean utility and price is given by

$$\begin{aligned} \delta_{jt} &= \beta^0 \, p_{jt} \, + \, \lambda_j^0 \, f_t^0 \, + \, e_{jt}, \\ p_{jt} &= \max\left(0.2, \, 1 + \tilde{p}_{jt} + \lambda_j^0 \, f_t^0\right), \end{aligned} \tag{6.1}$$

where λ_j^0, f_t^0, e_{jt} and $\tilde{p}_{jt}$ are mutually independent and are all independent and identically distributed across j and t as $\mathcal{N}(0,1)$. In the data generating process the number of factors is $R = 1$. For the number of factors used in the estimation procedure, R_{EST}, we consider the correctly specified case $R_{\text{EST}} = R = 1$, the misspecified case $R_{\text{EST}} = 0$, and the case where the number of factors is overestimated $R_{\text{EST}} = 2$. We have truncated the data generating process for price so that p_{jt} takes no values smaller than 0.2.

The market shares are computed from the mean utilities according to equation (2.4) and (2.5), where we assume a normally distributed random coefficient on price p_{jt}, i.e. $v \sim \mathcal{N}(0, \alpha^2)$. We chose the parameters of the model to be $\beta^0 = -3$ and $\alpha^0 = 1$. These parameters corresponds to a distribution of consumer tastes where more than 99% of consumers prefer low prices.

Although the regressors are strictly exogenous with respect to e_{jt}, we still need an instrument to identify α. We choose $Z_{jt} = p_{jt}^2$, the squared price. Thus, the number of instruments is $M = 1$. We justify the choice of squared price as an instrument in subsection 6.1 by verifying the instrument relevance Assumption 1(v) is satisfied for our simulation design.

Simulation results for three different samples sizes $J = T = 20$, 50 and 80, and three different choices for the number of factors in estimation $R_{\text{EST}} = 0$, 1, and 2 are presented in Table 1. We find that the estimators for $\widehat{\alpha}$ and $\widehat{\beta}$ to be significantly biased when $R_{\text{EST}} = 0$ factors are chosen in the estimation. This is because the factor and factor loading enter into the distribution of the regressor p_{jt} and the instrument Z_{jt}, which makes them endogenous with respect to the total unobserved error $\xi_{jt}^0 = \lambda_j^0 \, f_t^0 \, + \, e_{jt}$, and results in the estimated model with $R_{\text{EST}} = 0$ to be misspecified. The standard errors of the estimators are also much larger for $R_{\text{EST}} = 0$ than for $R_{\text{EST}} > 0$, since the variation of the total unobserved error ξ_{jt}^0 is larger than the variation of e_{jt}, which is the residual error after accounting for the factor structure.

For the correctly specified case $R_{\text{EST}} = R = 1$ we find the biases of the estimators $\widehat{\alpha}$ and $\widehat{\beta}$ to be negligible relative to the standard errors. For $J = T = 20$ the absolute value of the biases is about one tenth the standard errors, and the ratio is even smaller for the larger sample sizes. As the sample size increases from $J = T = 20$ to $J = T = 50$ and

		$R_{\text{EST}}=0$		$R_{\text{EST}}=1$		$R_{\text{EST}}=2$	
J,T	statistics	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\alpha}$	$\widehat{\beta}$
20,20	bias	0.4255	-0.3314	0.0067	-0.0099	0.0024	-0.0050
	std	0.1644	0.1977	0.0756	0.0979	0.0815	0.1086
	rmse	0.4562	0.3858	0.0759	0.0983	0.0815	0.1086
50,50	bias	0.4305	-0.3178	0.0005	-0.0012	0.0022	-0.0024
	std	0.0899	0.0984	0.0282	0.0361	0.0293	0.0369
	rmse	0.4398	0.3326	0.0282	0.0361	0.0293	0.0369
80,80	bias	0.4334	-0.3170	-0.0009	0.0010	0.0003	-0.0003
	std	0.0686	0.0731	0.0175	0.0222	0.0176	0.0223
	rmse	0.4388	0.3253	0.0175	0.0222	0.0176	0.0223

Table 1: Simulation results for the data generating process (6.1), using 1000 repetitions. We report the bias, standard errors (std), and square roots of the mean square errors (rmse) of the LS-MD estimator $(\widehat{\alpha}, \widehat{\beta})$. The true number of factors in the process is $R = 1$, but we use $R_{\text{EST}} = 0, 1$, and 2 in the estimation.

$J = T = 80$ one finds the standard error of the estimators to decrease at the rate $1/\sqrt{JT}$, consistent with our asymptotic theory.

The result for the case $R_{\text{EST}} = 2$ are very similar to those for $R_{\text{EST}} = 1$, that is, overestimating the number of factors does not affect the estimation quality much in our simulation, at least as long as R_{EST} is small relative to the sample size J, T.[24] The biases for the estimators found for $R_{\text{EST}} = 2$ are still negligible and the standard errors are about 10% larger for $R_{\text{EST}} = 2$ than for $R_{\text{EST}} = 1$ at $J = T = 20$, and even less than 10% larger for the larger sample sizes. The result that choosing $R_{\text{EST}} > R$ has only a small effect on the estimator is not covered by the asymptotic theory in this paper, where we assume $R_{\text{EST}} = R$, but is consistent with the analytical results found in Moon and Weidner (2015b) for the linear model with interactive fixed effects.

We have chosen a data generating process for our simulation where regressors and instruments are strictly exogenous (as opposed to pre-determined) with respect to e_{jt}, and where the error distribution e_{jt} is homoscedastic. According to our asymptotic theory there is therefore no asymptotic bias in the estimators $\widehat{\alpha}$ and $\widehat{\beta}$, which is consistent with the results in Table 1. The simulation results for the bias corrected estimators $\widehat{\alpha}^*$ and $\widehat{\beta}^*$

[24] In pure factor models consistent inference procedures on the number of factors are known, e.g. Bai and Ng (2002), Harding (2007), Onatski (2010), and Ahn and Horenstein (2013). In our model the number of factor can be estimated by applying those pure factor model techniques to the residuals $\widehat{\xi} = \delta(\widehat{\alpha}) - \widehat{\beta} \cdot X$, where $\widehat{\alpha}$ and $\widehat{\beta}$ are LS-MD estimator obtained with $R_{\text{EST}} \geq R$. Showing consistency of this procedure, however, goes beyond the scope of the current paper.

are reported in Table 3 in the appendix, but there is virtually no effect from bias correction here, that is, the results in Table 1 and Table 3 are almost identical.

Table 3 also reports the average estimated standard error based on our asymptotic variance estimator, as well as the empirical size of a nominal 5% t-test for the hypothesis that the respective parameter equals its true value. Those are not particularly interesting for $R_{\text{EST}} < R$, where the model is badly misspecified. For $R_{\text{EST}} \geq R$ we find that for small sample sizes ($J = T = 20$) our standard errors underestimate the dispersion of the estimator distributions by around 20%, and the t-test is oversized accordingly. For larger sample sizes ($J = T = 80$) our standard errors are still a bit too small, but only by around 5% or less, thus resulting in empirical sizes quite close to the nominal size.

6.1 Remarks: Instrument Choice

For the special case where there is only one normally distributed random coefficient attached to the regressor p_{jt}, one can write equation (2.5) as

$$s_{jt}(\alpha, \delta_t, X_t) = \frac{1}{\sqrt{2\pi}\alpha} \int \frac{\exp(\delta_{jt} + p_{jt}v)}{1 + \sum_{l=1}^{J} \exp(\delta_{lt} + p_{lt}v)} \exp\left(-\frac{v^2}{2\alpha^2}\right) dv. \tag{6.2}$$

For $x \geq 0$ we have the general inequalities $1 \geq (1+x)^{-1} \geq 1 - x$. Applying this to (6.2) with $x = \sum_{l=1}^{J} \exp(\delta_{lt} + p_{lt}v)$ one obtains $s_{jt}^{\text{up}}(\alpha, \delta_t, X_t) \geq s_{jt}(\alpha, \delta_t, X_t) \geq s_{jt}^{\text{low}}(\alpha, \delta_t, X_t)$, where

$$\begin{aligned}
s_{jt}^{\text{up}}(\alpha, \delta_t, X_t) &= \frac{1}{\sqrt{2\pi}\alpha} \int \exp(\delta_{jt} + p_{jt}v) \exp\left(-\frac{v^2}{2\alpha^2}\right) dv \\
&= \exp\left(\delta_{jt} + \alpha^2 p_{jt}^2/2\right), \\
s_{jt}^{\text{low}}(\alpha, \delta_t, X_t) &= \frac{1}{\sqrt{2\pi}\alpha} \int \exp(\delta_{jt} + p_{jt}v) \left[1 - \sum_{l=1}^{J} \exp(\delta_{lt} + p_{lt}v)\right] \exp\left(-\frac{v^2}{2\alpha^2}\right) dv \\
&= s_{jt}^{\text{up}}(\alpha, \delta_t, X_t)\Bigg[1 - \underbrace{\sum_{l=1}^{J} \exp\left(\delta_{lt} + \alpha^2 p_{lt}^2/2 + \alpha^2 p_{jt} p_{lt}\right)}_{=\nu_{jt}(\alpha, \delta_t)}\Bigg].
\end{aligned} \tag{6.3}$$

Here, the integrals over v that appear in the upper and lower bound are solvable analytically, so that we obtain convenient expressions for $s_{jt}^{\text{up}}(\alpha, \delta_t, X_t)$ and $s_{jt}^{\text{low}}(\alpha, \delta_t, X_t)$.

Consider the specification (6.1) for β negative and large (in absolute value) relative to α^2. Then δ_{jt} is also negative and large in absolute value, which implies that the $\nu_{jt} = \nu_{jt}(\alpha, \delta_t)$ defined in (6.3) is small. For $\nu_{jt} \ll 1$, as here, the lower and upper bounds are almost identical, which implies $s_{jt}(\alpha, \delta_t, X_t) \approx \exp\left(\delta_{jt} + \alpha^2 p_{jt}^2/2\right)$, where $\approx$ means almost equal under that approximation. Solving for the mean utility yields $\delta_{jt}(\alpha, s_t, X_t) \approx$

$\log s_{jt}(\alpha, \delta_t, X_t) - \alpha^2 p_{jt}^2/2$. The difference between $\delta_{jt}(\alpha, s_t, X_t)$ and $\delta_{jt}^0 = \delta_{jt}(\alpha^0, s_t, X_t)$ can then be approximated by

$$\delta_{jt}(\alpha, s_t, X_t) - \delta_{jt}^0 \;\approx\; -\frac{p_{jt}^2}{2}\left[\alpha^2 - (\alpha^0)^2\right]. \tag{6.4}$$

This shows that whenever the approximation $\nu_{jt} \ll 1$ is justified, then the squared price p_{jt}^2 is a valid instrument to identify α. More precisely, equation (6.4) implies that the LS-MD estimator with instrument p_{jt}^2 is approximately equivalent to the least squares estimator for the linear model with outcome variable $Y_{jt} = \beta p_{jt} + \alpha^2 p_{jt}^2 + \lambda_j' f_t + e_{jt}$. Consistency of this least squared estimator for β and α^2 in the presence of the parameters λ_j and f_t is discussed in Bai (2009) and Moon and Weidner (2015a).

We have thus shown that $\nu_{jt} \ll 1$ is a sufficient condition for validity of the instrument p_{jt}^2. However, for the data-generating process with parameters $\alpha^0 = 1$ and $\beta^0 = -3$ used in the Monte Carlo simulation this is not a good approximation — when calculating ν_{jt} in that setup one typically finds values much larger than one. Therefore, we next confirm by numerical methods that p_{jt}^2 is also a valid instrument when $\nu_{jt} \ll 1$ does not hold.

The Instrument Relevance Condition: Some Numerical Evidence

We want to verify the instrument relevance Assumption 1(v) for the data generating process (6.1) in the Monte Carlo Simulations with parameters $\beta^0 = -3$, and $\alpha^0 = 1$. For this purpose we define

$$\begin{aligned}
\rho_{\mathrm{IV}}(\alpha, \beta) &= \frac{\left[\frac{1}{JT}\Delta\xi_{\alpha,\beta}'(x,z)\right]\left[\frac{1}{JT}(x,z)'(x,z)\right]^{-1}\left[\frac{1}{JT}(x,z)'\,\Delta\xi_{\alpha,\beta}\right]}{\frac{1}{JT}\Delta\xi_{\alpha,\beta}'\Delta\xi_{\alpha,\beta}}, \\
\rho_{\mathrm{F}}(\alpha, \beta) &= \frac{\max_{\lambda\in\mathbb{R}^{J\times R}}\left[\frac{1}{JT}\Delta\xi_{\alpha,\beta}'\left(\mathbb{1}_T \otimes P_{(\lambda,\lambda^0)}\right)\Delta\xi_{\alpha,\beta}\right]}{\frac{1}{JT}\Delta\xi_{\alpha,\beta}'\Delta\xi_{\alpha,\beta}}, \\
\Delta\rho(\alpha, \beta) &= \rho_{\mathrm{IV}}(\alpha, \beta) - \rho_{\mathrm{F}}(\alpha, \beta).
\end{aligned} \tag{6.5}$$

$\rho_{\mathrm{IV}}(\alpha, \beta)$ is the amount of $\Delta\xi_{\alpha,\beta}$ explained by the instruments and regressors relative to the total variation of $\Delta\xi_{\alpha,\beta}$, i.e. the relative explanatory power of the instruments. $\rho_{\mathrm{F}}(\alpha, \beta)$ is the maximum amount of $\Delta\xi_{\alpha,\beta}$ explained by R factor loadings relative the total variation of $\Delta\xi_{\alpha,\beta}$, i.e. the relative explanatory power of the factors. $\rho_{\mathrm{IV}}(\alpha, \beta)$ and $\rho_{\mathrm{F}}(\alpha, \beta)$ take values betweens 0 and 1.

The difference between the explanatory power of the instruments and regressors and the explanatory power of the factors is given by $\Delta\rho(\alpha, \beta)$. Assumption 1($v$) requires that $\Delta\rho(\alpha, \beta) > 0$ for all $\alpha \in \mathcal{B}_\alpha$ and $\beta \in \mathbb{R}^K$.

Figure 1 contains plots of $\rho_{\mathrm{IV}}(\alpha, \beta)$, $\rho_{\mathrm{F}}(\alpha, \beta)$ and $\Delta\rho(\alpha, \beta)$ as a function of α and β for one particular draw of the data generating process with $J = T = 80$. The sample size is

sufficiently large that for different draws the plots in Figure 1 look essentially identical.[25] Although the data generating process only contains one factor, we used $R = 2$ factors in the calculation of $\rho_{\mathrm{F}}(\alpha, \beta)$ and $\Delta\rho(\alpha, \beta)$ in Figure 1, in order to verify Assumption 1(v) also for the case where the number of factors is overestimated (denoted R_{EST}=2 above) — since $\rho_{\mathrm{F}}(\alpha, \beta)$ is an increasing function of R, we thus also verify the conditions of $R = 1$.

For the given draw and within the examined parameter range one finds that $\rho_{\mathrm{IV}}(\alpha, \beta)$ varies between 0.69 and 1.00, $\rho_{\mathrm{F}}(\alpha, \beta)$ varies between 0.34 and 0.87, and $\Delta\rho(\alpha, \beta)$ varies between 0.03 and 0.49, in particular $\Delta\rho(\alpha, \beta) > 0$, which is what we wanted to verify.

The variation in $\Delta\rho(\alpha, \beta)$ in this example is mostly driven by the variation in $\rho_{\mathrm{F}}(\alpha, \beta)$, since $\rho_{\mathrm{IV}}(\alpha, \beta)$ for the most part is quite close to one, that is, the explanatory power of the instruments and regressors is very large. The analytical approximation above showed that for $\nu_{jt} \ll 1$ the regressor p_{jt} and the instrument p_{jt}^2 perfectly predict $\Delta\xi_{\alpha,\beta}$, that is, we have $\rho_{\mathrm{IV}}(\alpha, \beta) \approx 1$ under that approximation. Our numerical result now shows that p_{jt}^2 can be a sufficiently powerful instrument also outside the validity range of this approximation.

7 Empirical illustration: estimation of demand for new automobiles

As an illustration of our procedure, we estimate an aggregate random coefficients logit model of demand for new automobiles, modeled after the analysis in BLP (1995). We compare specifications with and without factors, and with and without price endogeneity. Throughout, we allow for one normally-distributed random coefficient, attached to price.[26]

For this empirical illustration, we use the same data as was used in BLP (1995), which are new automobile sales from 1971-1990.[27] However, our estimation procedure requires a balanced panel for the principal components step. Since there is substantial entry and exit of individual car models, we aggregate up to manufacturer-size level, and assume that consumers choose between aggregate composites of cars.[28] Furthermore, we also reduce

[25] The appendix contains additional details on the numerical calculation of $\rho_F(\alpha, \beta)$.

[26] In principle, multiple random coefficients could be accommodated in a straightforward manner; as this application is primarily illustrative, we do not consider this here.

[27] In such a setting, where we have a single national market evolving over time, we can interpret λ_j as (unobserved) national advertising for brand j, which may be roughly constant across time, and f_t represents the effectiveness or "success" of the advertising, which varies over time. Indeed, for the automobile sector (which is the topic of our empirical example), the dollar amount of national brand-level advertising does not vary much across years, but the success of the ad campaign does vary.

[28] This resembles the treatment in Esteban and Shum's (2007) empirical study of the new and used car markets,

our sample window to the sixteen years 1973-1988. In Table 5, we list the 23 car aggregates employed in our analysis, along with the across-year averages of the variables.

Except from the aggregation, our variables are the same as in BLP. Market share is given by total sales divided by the number of households in that year. Price is measured in \$1000 of 1983/84 dollars. Our unit for "horse power over weight" (hp/weight) is 100 times horse power over pound. "Miles per dollar" (mpd) is obtained from miles per gallons divided by real price per gallon, and measured in miles over 1983/84 dollars. Size is given by length times width, and measured in $10^{-4}\,\text{inch}^2$.

We construct instruments using the idea of Berry (1994). The instruments for a particular aggregated model and year are given by the averages of hp/weight, mpd and size, over all cars produced by different manufactures in the same year. As the weight matrix in the second step of the LS-MD procedure we use $W_{JT} = \frac{1}{JT} z' M_x z$, which is the optimal weight matrix under homoscedasticity of e_{jt} and for $R = 0$.[29]

Results. Table 2 contains estimation results from four specifications of the model. In specification A, prices are considered exogenous (wrt e_{jt}), but one factor is present, which captures some degree of price endogeneity (wrt. ξ_{jt}). Specification B also contains one factor, but treats prices as endogenous, even conditional on the factor. Specification C corresponds to the BLP (1995) model, where prices are endogenous, but no factor is present. Finally, in specification D, we treat prices as exogenous, and do not allow for a factor. This final specification is clearly unrealistic, but is included for comparison with the other specifications. In table 2 we report the bias corrected LS-MD estimator (this only makes a difference for specification A and B), which accounts for bias due to heteroscedasticity in the error terms, and due to pre-determined regressors (we choose bandwidth $h = 2$ in the construction of $\widehat{B}_0$). The estimation results without bias correction are reported in table 4. It turns out, that it makes not much difference, whether the LS-MD estimator, or its bias corrected version are used. The t-values of the bias corrected estimators are somewhat larger, but apart from the constant, which is insignificant anyways, the bias correction changes neither the sign of the coefficients nor the conclusion whether the coefficients are

which likewise required a balanced panel.

[29] We do not change the weight matrix when estimating specifications with $R = 1$, because we do not want differences in the results for different values of R to be attributed to the change in W_{JT}.

We include a constant regressor in the model, although this is a "low-rank" regressor, which is ruled out by our identification and consistency assumptions. However, as discussed in a footnote above the inclusion of a low-rank regressor does not hamper the identification and estimation of the regression coefficients of the other ("high-rank") regressors. One certainly wants to include a constant regressor when estimating the model with no factors ($R = 0$), so to make results easily comparable we include it in all our model specifications.

	Specifications:							
	A: $R = 1$ exogenous p		B: $R = 1$ endogenous p		C: $R = 0$ endogenous p		D: $R = 0$ exogenous p	
price	-4.109	(-3.568)	-3.842	(-4.023)	-1.518	(-0.935)	-0.308	(-1.299)
hp/weight	0.368	(1.812)	0.283	(1.360)	-0.481	(-0.314)	0.510	(1.981)
mpd	0.088	(2.847)	0.117	(3.577)	0.157	(0.870)	0.030	(1.323)
size	5.448	(3.644)	5.404	(3.786)	0.446	(0.324)	1.154	(2.471)
α	2.092	(3.472)	2.089	(3.837)	0.894	(0.923)	0.171	(1.613)
const	3.758	(1.267)	0.217	(0.117)	-3.244	(-0.575)	-7.827	(-8.984)

Table 2: Parameter estimates (and t-values) for four different model specifications (no factor $R = 0$ vs. one factor $R = 1$; exogenous price vs. endogenous price). α is the standard deviation of the random coefficient distribution (only price has a random coefficient), and the regressors are p (price), hp/weight (horse power per weight), mpd (miles per dollar), size (car length times car width), and a constant.

significant at 5% level.

In Specification A, most of the coefficients are precisely estimated. The price coefficient is -4.109, and the characteristics coefficients take the expected signs. The α parameter, corresponding to the standard deviation of the random coefficient on price, is estimated to be 2.092. These point estimates imply that, roughly 97% of the time, the random price coefficient is negative, which is as we should expect.

Compared to this baseline, Specification B allows price to be endogenous (even conditional on the factor). The point estimates for this specifications are virtually unchanged from those in Specification A, except for the constant term. Overall, the estimation results for the specifications A and B are very similar, and show that once factors are taken into account it does not make much difference whether price is treated as exogenous or endogenous. This suggests that the factors indeed capture most of the price endogeneity in this application.

In contrast, the estimation results for specifications C and D, which are the two specifications without any factors, are very different qualitatively. The t-values for specification C are rather small (i.e. standard errors are large), so that the difference in the coefficient estimates in these two specifications are not actually statistically significant. However, the differences in the t-values themselves shows that it makes a substantial difference for the no-factor estimation results whether price is treated as exogenous or endogenous.

Specifically, in Specification C, the key price coefficient and α are substantially smaller in magnitude; furthermore, the standard errors are large, so that none of the estimates are significant at usual significance levels. Moreover, the coefficient on *hp/weight* is negative,

which is puzzling. In Specification D, which corresponds to a BLP model, but without price endogeneity, we see that the price coefficient is reduced dramatically relative to the other specifications, down to -0.308.

Elasticities. The sizeable differences in the magnitudes of the price coefficients across the specification with and without factors suggest that these models may imply economically meaningful differences in price elasticities. For this reason, we compute the matrices of own- and cross-price elasticities for Specifications B (see Table (6)) and C (see Table (7)). The elasticities were computed using the data in 1988, the final year of our sample. Comparing these two sets of elasticities, the most obvious difference is that the elasticities – both own- and cross-price – for Specification C, corresponding to the standard BLP model without factors, are substantially smaller (about one-half in magnitude) than the Specification B elasticities. For instance, reading down the first column of Table (6), we see that a one-percent increase in the price of a small Chevrolet car would result in a 28% reduction in its market share, but increase the market share for large Chevrolet cars by 1.5%. For the results in Table (7), however, this same one-percent price increase would reduce the market share for small Chevrolet cars by only 13%, and increase the market share for large Chevrolet cars by less than half a percent.

On the whole, then, this empirical illustration shows that our estimation procedure is feasible even for moderate-sized datasets like the one used here. Including interactive fixed effects delivers results which are strikingly different than those obtained from specifications without these fixed effects.

8 Conclusion

In this paper, we consider an extension of the popular BLP random coefficients discrete-choice demand model, which underlies much recent empirical work in IO. We add interactive fixed effects in the form of a factor structure on the unobserved product characteristics. The interactive fixed effects can be arbitrarily correlated with the observed product characteristics (including price), which accommodate endogeneity and, at the same time, captures strong persistence in market shares across products and markets. We propose a two-step least squares-minimum distance (LS-MD) procedure to calculate the estimator. Our estimator is easy to compute, and Monte Carlo simulations show that it performs well.

The model in this paper is, to our knowledge, the first application of factor-modeling to a nonlinear setting with endogenous regressors. Since many other models used in

applied settings (such as duration models in labor economics, and parametric auction models in IO) have these features, we believe that factor-modeling may prove an effective way of controlling for unobserved heterogeneity in these models. We are exploring these applications in ongoing work.

A Additional Tables and Figures

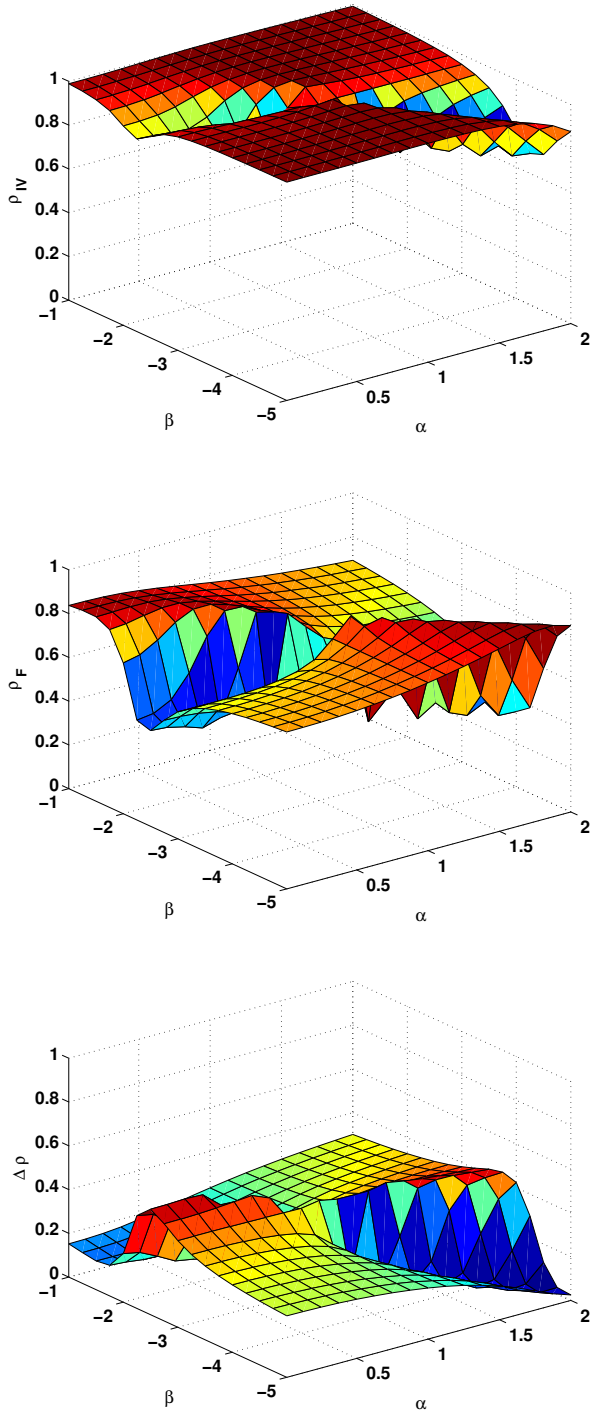

Figure 1: For one draw of the data generating process used in the Monte Carlo design with $J = T = 80$ we plot $\rho_{\mathrm{IV}}(\alpha, \beta)$, $\rho_{\mathrm{F}}(\alpha, \beta)$ and $\Delta\rho(\alpha, \beta)$ defined in (6.5) as a function of α and β. The number of factors used in the calculation of $\rho_{\mathrm{F}}(\alpha, \beta)$ is $R = 2$, although only one factor is present in the data generating process.

J,T	statistics	$R_{\text{EST}}=0$ $\widehat{\alpha}^*$	$\widehat{\beta}^*$	$R_{\text{EST}}=1$ $\widehat{\alpha}^*$	$\widehat{\beta}^*$	$R_{\text{EST}}=2$ $\widehat{\alpha}^*$	$\widehat{\beta}^*$
20,20	bias	0.4255	-0.3314	0.0042	-0.0068	0.0001	-0.0023
	std	0.1644	0.1977	0.0759	0.0981	0.0818	0.1085
	rmse	0.4562	0.3858	0.0760	0.0983	0.0817	0.1084
	mean(SE)	0.0938	0.1300	0.0660	0.0870	0.0632	0.0833
	emp. size	0.96	0.65	0.09	0.07	0.15	0.12
50,50	bias	0.4305	-0.3178	0.0000	-0.0006	0.0017	-0.0018
	std	0.0899	0.0984	0.0283	0.0362	0.0293	0.0368
	rmse	0.4398	0.3326	0.0282	0.0361	0.0293	0.0369
	mean(SE)	0.0418	0.0551	0.0270	0.0344	0.0265	0.0338
	emp. size	1.00	0.99	0.07	0.06	0.10	0.10
80,80	bias	0.4334	-0.3170	-0.0012	0.0012	0.0001	0.0000
	std	0.0686	0.0731	0.0175	0.0222	0.0176	0.0223
	rmse	0.4388	0.3253	0.0175	0.0222	0.0176	0.0223
	mean(SE)	0.0272	0.0354	0.0171	0.0215	0.0169	0.0213
	emp. size	1.00	1.00	0.07	0.07	0.06	0.06

Table 3: Simulation results for the data generating process (6.1), using 1000 repetitions. We report the bias, standard errors (std), square roots of the mean square errors (rmse), and the average of the estimated standard error (mean SE) of the bias corrected LS-MD estimator $(\widehat{\alpha}^*, \widehat{\beta}^*)$. In addition, we report the empirical size of a nominal size 5% t-test based on $\widehat{\alpha}^*$ and $\widehat{\beta}^*$ for the hypothesis that the parameter equals its true value. The true number of factors in the process is $R=1$, but we use $R_{\text{EST}}=0,1$, and 2 in the estimation.

	Specifications:			
	A: $R=1$ exogenous p		B: $R=1$ endogenous p	
price	-3.112	(-2.703)	-2.943	(-3.082)
hp/weight	0.340	(1.671)	0.248	(1.190)
mpd	0.102	(3.308)	0.119	(3.658)
size	4.568	(3.055)	4.505	(3.156)
α	1.613	(2.678)	1.633	(3.000)
const	-0.690	(-0.232)	-2.984	(-1.615)

Table 4: Parameter estimates (and t-values) for model specification A and B. Here we report the LS-MD estimators without bias correction, while in table 2 we report the bias corrected LS-MD estimators.

Product#	Make	Size Class	Manuf.	Mkt Share % (avg)	Price (avg)	hp/weight (avg)	mpd (avg)	size (avg)
1	CV (Chevrolet)	small	GM	1.39	6.8004	3.4812	20.8172	1.2560
2	CV	large	GM	0.49	8.4843	3.5816	15.9629	1.5841
3	OD (Oldsmobile)	small	GM	0.25	7.6786	3.4789	19.1946	1.3334
4	OD	large	GM	0.69	9.7551	3.6610	15.7762	1.5932
5	PT (Pontiac)	small	GM	0.46	7.2211	3.4751	19.3714	1.3219
6	PT	large	GM	0.31	8.6504	3.5806	16.6192	1.5686
7	BK (Buick)	all	GM	0.84	9.2023	3.6234	16.9960	1.5049
8	CD (Cadillac)	all	GM	0.29	18.4098	3.8196	13.6894	1.5911
9	FD (Ford)	small	Ford	1.05	6.3448	3.4894	21.7885	1.2270
10	FD	large	Ford	0.63	8.9530	3.4779	15.7585	1.6040
11	MC (Mercury)	small	Ford	0.19	6.5581	3.6141	22.2242	1.2599
12	MC	large	Ford	0.32	9.2583	3.4610	15.9818	1.6053
13	LC (Lincoln)	all	Ford	0.16	18.8322	3.7309	13.6460	1.7390
14	PL (Plymouth)	small	Chry	0.31	6.2209	3.5620	22.7818	1.1981
15	PL	large	Chry	0.17	7.7203	3.2334	15.4870	1.5743
16	DG (Dodge)	small	Chry	0.35	6.5219	3.6047	23.2592	1.2031
17	DG	large	Chry	0.17	7.8581	3.2509	15.4847	1.5681
18	TY (Toyota)	all	Other	0.54	7.1355	3.7103	24.3294	1.0826
19	VW (Volkswagen)	all	Other	0.17	8.2388	3.5340	24.0027	1.0645
20	DT/NI (Datsen/Nissan)	all	Other	0.41	7.8120	4.0226	24.5849	1.0778
21	HD (Honda)	all	Other	0.41	6.7534	3.5442	26.8501	1.0012
22	SB (Subaru)	all	Other	0.10	5.9568	3.4718	25.9784	1.0155
23	REST	all	Other	1.02	10.4572	3.6148	19.8136	1.2830

Table 5: Summary statistics for the 23 product-aggregates used in estimation.

	CV s	CV l	OD s	OD l	PT s	PT l	BK	CD	FD s	FD l	MC s	MC l	LC	PL s	PL l	DG s	DG l	TY	VW	DT/NI	HD	SB	REST
CV s	-28.07	0.82	0.70	1.70	0.96	0.31	2.77	0.14	1.32	2.38	0.41	1.45	0.03	0.32	0.22	0.44	0.31	1.57	0.57	1.74	0.91	0.15	6.58
CV l	1.50	-34.54	0.72	2.02	0.79	0.21	3.27	0.73	0.97	3.54	0.37	2.15	0.16	0.21	0.21	0.30	0.30	1.21	0.40	1.62	0.71	0.10	10.17
OD s	1.29	0.72	-35.78	2.08	0.72	0.18	3.36	1.15	0.84	3.90	0.35	2.37	0.25	0.17	0.20	0.25	0.28	1.06	0.34	1.53	0.63	0.08	11.35
OD l	0.98	0.64	0.65	-35.80	0.59	0.13	3.37	2.09	0.64	4.34	0.30	2.63	0.45	0.12	0.17	0.18	0.25	0.84	0.25	1.36	0.51	0.06	12.86
PT s	1.76	0.80	0.72	1.90	-32.51	0.26	3.09	0.38	1.14	3.02	0.39	1.84	0.08	0.26	0.22	0.37	0.31	1.39	0.48	1.70	0.81	0.12	8.56
PT l	2.17	0.81	0.68	1.55	0.98	-26.85	2.53	0.06	1.40	1.97	0.41	1.21	0.01	0.35	0.22	0.48	0.31	1.65	0.61	1.72	0.94	0.16	5.37
BK	0.99	0.64	0.66	2.09	0.60	0.13	-34.47	2.04	0.65	4.33	0.30	2.62	0.44	0.12	0.18	0.18	0.25	0.84	0.25	1.36	0.51	0.06	12.81
CD	0.00	0.01	0.01	0.08	0.00	0.00	0.12	-6.97	0.00	0.36	0.00	0.21	3.67	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	1.19
FD s	2.03	0.82	0.71	1.71	0.95	0.31	2.79	0.15	-28.99	2.41	0.41	1.47	0.03	0.32	0.22	0.44	0.31	1.56	0.57	1.74	0.90	0.15	6.67
FD l	0.61	0.50	0.55	1.95	0.42	0.07	3.13	4.23	0.40	-34.69	0.23	2.80	0.90	0.06	0.14	0.10	0.20	0.56	0.15	1.07	0.34	0.04	14.05
MC s	1.57	0.77	0.72	1.99	0.81	0.22	3.24	0.63	1.02	3.41	-34.49	2.07	0.14	0.23	0.21	0.32	0.30	1.26	0.42	1.64	0.74	0.11	9.77
MC l	0.62	0.50	0.55	1.95	0.43	0.07	3.14	4.15	0.41	4.64	0.23	-36.50	0.88	0.06	0.14	0.11	0.20	0.56	0.15	1.08	0.35	0.04	14.03
LC	0.00	0.01	0.02	0.09	0.01	0.00	0.15	20.15	0.00	0.41	0.00	0.24	-23.81	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	1.39
PL s	2.21	0.79	0.64	1.40	0.98	0.34	2.29	0.03	1.42	1.65	0.40	1.01	0.01	-23.54	0.21	0.49	0.30	1.66	0.63	1.67	0.95	0.16	4.42
PL l	1.47	0.75	0.72	2.03	0.78	0.21	3.29	0.78	0.96	3.59	0.37	2.18	0.17	0.21	-35.26	0.30	0.29	1.19	0.39	1.61	0.70	0.10	10.33
DG s	2.17	0.81	0.68	1.55	0.98	0.33	2.54	0.06	1.40	1.99	0.41	1.22	0.01	0.35	0.22	-26.80	0.31	1.64	0.61	1.72	0.94	0.16	5.41
DG l	1.47	0.75	0.72	2.03	0.78	0.21	3.29	0.78	0.96	3.59	0.37	2.18	0.17	0.21	0.20	0.30	-35.18	1.19	0.39	1.61	0.70	0.10	10.33
TY	1.94	0.81	0.72	1.79	0.93	0.29	2.91	0.22	1.25	2.65	0.41	1.62	0.05	0.30	0.22	0.41	0.31	-30.16	0.54	1.73	0.87	0.14	7.41
VW	2.13	0.82	0.69	1.61	0.97	0.32	2.63	0.09	1.37	2.13	0.41	1.31	0.02	0.34	0.22	0.47	0.31	1.62	-27.86	1.73	0.93	0.15	5.85
DT/NI	1.49	0.76	0.72	2.02	0.79	0.21	3.28	0.75	0.97	3.55	0.37	2.16	0.16	0.21	0.21	0.30	0.29	1.20	0.40	-33.74	0.71	0.10	10.22
HD	1.88	0.81	0.72	1.83	0.91	0.28	2.97	0.26	1.22	2.77	0.40	1.69	0.06	0.29	0.22	0.40	0.31	1.47	0.52	1.72	-31.39	0.13	7.77
SB	2.16	0.82	0.68	1.58	0.98	0.33	2.57	0.07	1.39	2.04	0.41	1.25	0.02	0.35	0.22	0.47	0.31	1.64	0.61	1.73	0.94	-27.60	5.58
REST	0.56	0.47	0.53	1.91	0.40	0.07	3.07	4.71	0.37	4.65	0.22	2.80	1.00	0.06	0.13	0.09	0.19	0.51	0.13	1.02	0.32	0.03	-25.42

Table 6: Estimated price elasticities for specification B in $t = 1988$. Rows (i) correspond to market shares (s_{jt}), and columns (j) correspond to prices (p_{jt}) with respect to which elasticities are calculated.

	CV s	CV l	OD s	OD l	PT s	PT l	BK	CD	FD s	FD l	MC s	MC l	LC	PL s	PL l	DG s	DG l	TY	VW	DT/NI	HD	SB	REST
CV s	-12.95	0.46	0.46	0.48	0.46	0.47	0.48	1.45	0.46	0.51	0.46	0.51	1.41	0.48	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.47	0.51
CV l	0.43	-15.20	0.49	0.53	0.45	0.41	0.53	2.46	0.43	0.60	0.46	0.59	2.39	0.40	0.47	0.41	0.47	0.43	0.42	0.47	0.44	0.41	0.61
OD s	0.41	0.47	-15.79	0.53	0.44	0.39	0.53	2.83	0.41	0.61	0.46	0.61	2.73	0.37	0.47	0.39	0.47	0.42	0.40	0.47	0.42	0.39	0.63
OD l	0.38	0.45	0.48	-16.57	0.41	0.35	0.53	3.40	0.38	0.63	0.44	0.63	3.28	0.33	0.45	0.35	0.45	0.39	0.36	0.45	0.40	0.36	0.65
PT s	0.44	0.47	0.49	0.51	-14.32	0.44	0.51	2.01	0.45	0.56	0.47	0.56	1.95	0.44	0.47	0.44	0.47	0.45	0.44	0.47	0.45	0.44	0.58
PT l	0.46	0.44	0.43	0.44	0.44	-11.76	0.44	1.09	0.46	0.45	0.44	0.45	1.07	0.51	0.44	0.48	0.44	0.45	0.47	0.44	0.45	0.48	0.46
BK	0.38	0.45	0.48	0.53	0.42	0.35	-16.54	3.38	0.38	0.63	0.44	0.63	3.26	0.33	0.45	0.35	0.45	0.39	0.36	0.45	0.40	0.36	0.65
CD	0.03	0.06	0.07	0.10	0.05	0.03	0.10	-7.85	0.03	0.15	0.06	0.15	5.14	0.02	0.06	0.03	0.06	0.04	0.03	0.06	0.04	0.03	0.16
FD s	0.46	0.46	0.47	0.48	0.46	0.47	0.48	1.48	-13.03	0.51	0.46	0.51	1.44	0.48	0.46	0.47	0.46	0.46	0.46	0.46	0.46	0.47	0.52
FD l	0.33	0.42	0.45	0.52	0.37	0.30	0.51	4.22	0.33	-17.46	0.41	0.63	4.06	0.27	0.42	0.30	0.42	0.35	0.31	0.42	0.36	0.30	0.65
MC s	0.43	0.47	0.49	0.53	0.45	0.42	0.52	2.35	0.43	0.59	-14.99	0.59	2.28	0.41	0.47	0.42	0.47	0.44	0.42	0.47	0.44	0.42	0.60
MC l	0.33	0.42	0.45	0.52	0.38	0.30	0.52	4.20	0.33	0.63	0.41	-17.44	4.03	0.27	0.42	0.30	0.42	0.35	0.31	0.42	0.36	0.30	0.66
LC	0.04	0.07	0.08	0.11	0.05	0.03	0.10	5.75	0.04	0.16	0.06	0.16	-8.59	0.02	0.07	0.03	0.07	0.04	0.03	0.07	0.04	0.03	0.17
PL s	0.45	0.40	0.40	0.39	0.42	0.48	0.39	0.79	0.45	0.39	0.41	0.39	0.77	-10.42	0.40	0.48	0.40	0.43	0.47	0.40	0.43	0.47	0.39
PL l	0.42	0.47	0.49	0.53	0.45	0.41	0.53	2.51	0.42	0.60	0.46	0.60	2.43	0.40	-15.28	0.41	0.47	0.43	0.41	0.47	0.44	0.41	0.61
DG s	0.46	0.44	0.44	0.44	0.44	0.48	0.44	1.10	0.46	0.45	0.44	0.45	1.08	0.51	0.44	-11.80	0.44	0.45	0.47	0.44	0.45	0.48	0.46
DG l	0.42	0.47	0.49	0.53	0.45	0.41	0.53	2.51	0.42	0.60	0.46	0.60	2.43	0.40	0.47	0.41	-15.28	0.43	0.41	0.47	0.44	0.41	0.61
TY	0.46	0.47	0.48	0.49	0.46	0.46	0.49	1.69	0.46	0.53	0.46	0.53	1.64	0.46	0.47	0.46	0.47	-13.58	0.46	0.47	0.46	0.46	0.54
VW	0.46	0.45	0.45	0.46	0.45	0.48	0.45	1.24	0.46	0.48	0.45	0.47	1.21	0.50	0.45	0.48	0.45	0.46	-12.28	0.45	0.45	0.47	0.48
DT/NI	0.42	0.47	0.49	0.53	0.45	0.41	0.53	2.48	0.43	0.60	0.46	0.60	2.40	0.40	0.47	0.41	0.47	0.43	0.42	-15.22	0.44	0.41	0.61
HD	0.45	0.47	0.48	0.50	0.46	0.45	0.50	1.79	0.45	0.54	0.47	0.54	1.74	0.46	0.47	0.45	0.47	0.45	0.45	0.47	-13.83	0.45	0.55
SB	0.46	0.44	0.44	0.45	0.45	0.48	0.45	1.15	0.46	0.46	0.44	0.46	1.13	0.50	0.44	0.48	0.44	0.45	0.47	0.44	0.45	-12.00	0.47
REST	0.32	0.41	0.45	0.51	0.37	0.29	0.51	4.37	0.32	0.63	0.40	0.63	4.19	0.26	0.41	0.29	0.41	0.34	0.30	0.41	0.35	0.29	-17.59

Table 7: Estimated price elasticities for specification C (BLP case) in $t = 1988$. Rows (i) correspond to market shares (s_{jt}), and columns (j) correspond to prices (p_{jt}) with respect to which elasticities are calculated.

B Alternative GMM approach

In this section we show that in the presence of factors a moment based estimation approach along the lines originally proposed by BLP is inadequate. The moment conditions imposed by the model are

$$\begin{aligned} \mathbb{E}\left[e_{jt}\left(\alpha^0,\, \beta^0,\, \lambda^0 f^{0\prime}\right) X_{k,jt}\right] &= 0\,, \qquad & k &= 1,\ldots,K\,, \\ \mathbb{E}\left[e_{jt}\left(\alpha^0,\, \beta^0,\, \lambda^0 f^{0\prime}\right) Z_{m,jt}\right] &= 0\,, & m &= 1,\ldots,M\,, \end{aligned} \tag{B.1}$$

where $e_{jt}(\alpha,\, \beta,\, \lambda f') = \delta_{jt}(\alpha,\, s_t,\, X_t) - \sum_{k=1}^{K} \beta_k\, X_{k,jt} - \sum_{r=1}^{R} \lambda_{jr} f_{tr}$. Note that we write the residuals e_{jt} as a function of the $J \times T$ matrix $\lambda f'$ in order to avoid the ambiguity of the decomposition into λ and f. The corresponding sample moments read

$$\begin{aligned} m_k^X(\alpha,\, \beta,\, \lambda f') &= \frac{1}{JT} \operatorname{Tr}\left(e(\alpha,\, \beta,\, \lambda f')\, X_k'\right)\,, \\ m_m^Z(\alpha,\, \beta,\, \lambda f') &= \frac{1}{JT} \operatorname{Tr}\left(e(\alpha,\, \beta,\, \lambda f')\, Z_m'\right)\,. \end{aligned} \tag{B.2}$$

We also define the sample moment vectors $m^X(\alpha,\, \beta,\, \lambda f') = \left(m_1^X, \ldots, m_K^X\right)'$ and $m^Z(\alpha,\, \beta,\, \lambda f') = \left(m_1^Z, \ldots, m_M^Z\right)'$. An alternative estimator for α, β, λ and f is then given by[30]

$$\begin{aligned} \left(\hat{\lambda}_{\alpha,\beta}\,,\ \hat{f}_{\alpha,\beta}\right) &= \underset{\{\lambda,\, f\}}{\operatorname{argmin}} \sum_{j=1}^{J} \sum_{t=1}^{T} e_{jt}^2(\alpha,\, \beta,\, \lambda f')\,. \\ \left(\hat{\alpha}^{\mathrm{GMM}},\, \hat{\beta}^{\mathrm{GMM}}\right) &= \underset{\{\alpha \in \mathcal{B}_\alpha,\, \beta\}}{\operatorname{argmin}} \begin{pmatrix} m^X(\alpha,\, \beta,\, \hat{\lambda}_{\alpha,\beta}\hat{f}'_{\alpha,\beta}) \\ m^Z(\alpha,\, \beta,\, \hat{\lambda}_{\alpha,\beta}\hat{f}'_{\alpha,\beta}) \end{pmatrix}' \mathcal{W}_{JT} \begin{pmatrix} m^X(\alpha,\, \beta,\, \hat{\lambda}_{\alpha,\beta}\hat{f}'_{\alpha,\beta}) \\ m^Z(\alpha,\, \beta,\, \hat{\lambda}_{\alpha,\beta}\hat{f}'_{\alpha,\beta}) \end{pmatrix}\,, \end{aligned} \tag{B.3}$$

where $\mathcal{W}_{JT}$ is a positive definite $(K+M) \times (K+M)$ weight matrix. The main difference between this alternative estimator and our estimator (4.1) is that the least-squares step is used solely to recover estimates of the factors and factor loadings (principal components estimator), while the structural parameters (α, β) are estimated in the GMM second step. The relation between $\hat{\alpha}$ and $\hat{\beta}$ defined in (4.1) and $\hat{\alpha}^{\mathrm{GMM}}$ and $\hat{\beta}^{\mathrm{GMM}}$ defined in (B.3) is as follows

[30] The minimizing $\hat{\lambda}_{\alpha,\beta}$ and $\hat{f}_{\alpha,\beta}$ are the least squares estimators, or equivalently, the principal components estimators, e.g. $\hat{\lambda}_{\alpha,\beta}$ consists of the eigenvectors corresponding to the R largest eigenvalues of the $J \times J$ matrix

$$\left(\delta(\alpha,\, s,\, X) - \sum_{k=1}^{K} \beta_k\, X_k\right)\left(\delta(\alpha,\, s,\, X) - \sum_{k=1}^{K} \beta_k\, X_k\right)'\,.$$

(i) Let $R = 0$ (no factors) and set

$$\mathcal{W}_{JT} = \begin{pmatrix} \left(\frac{1}{JT}x'x\right)^{-1} & 0_{K\times M} \\ 0_{M\times K} & 0_{M\times M} \end{pmatrix} + \begin{pmatrix} -(x'x)^{-1}x'\,z \\ \mathbb{1}_M \end{pmatrix} \left(\frac{1}{JT}z'M_xz\right)^{-1} W_{JT} \left(\frac{1}{JT}z'M_xz\right)^{-1} \begin{pmatrix} -(x'x)^{-1}x'\,z \\ \mathbb{1}_M \end{pmatrix}', \tag{B.4}$$

where x is a $JT \times K$ matrix and z is a $JT \times M$ matrix, given by $x_{\cdot,k} = \mathrm{vec}\,(X_k)$, $k = 1, \ldots, K$, and $z_{\cdot,m} = \mathrm{vec}\,(Z_m)$, $m = 1, \ldots, M$. Then $\hat{\alpha}$ and $\hat{\beta}$ solve (4.1) with weight matrix W_{JT} if and only if they solve (B.3) with this weight matrix $\mathcal{W}_{JT}$,[31] i.e. in this case we have $(\hat{\alpha}, \hat{\beta}) = (\hat{\alpha}^{\mathrm{GMM}}, \hat{\beta}^{\mathrm{GMM}})$.

(ii) Let $R > 0$ and $M = L$ (exactly identified case). Then a solution of (4.1) also is a solution of (B.3), but not every solution of (B.3) needs to be a solution of (4.1).

(iii) For $M > L$ and $R > 0$ there is no straightforward characterization of the relationship between the estimators in (4.1) and (B.3).

We want to discuss the exactly identified case $M = L$ a bit further. The reason why in this case every solution of (4.1) also solves (B.3) is that the first order conditions (FOC's) wrt to β and γ of the first stage optimization in (4.1) read $m^X(\hat{\alpha},\, \hat{\beta},\, \hat{\lambda}_{\hat{\alpha},\hat{\beta}}\hat{f}'_{\hat{\alpha},\hat{\beta}}) = 0$ and $m^Z(\hat{\alpha},\, \hat{\beta},\, \hat{\lambda}_{\hat{\alpha},\hat{\beta}}\hat{f}'_{\hat{\alpha},\hat{\beta}}) = 0$, which implies that the GMM objective function of (B.3) is zero, i.e. minimized. The reverse statement is not true, because for $R > 0$ the first stage objective function in (4.1) is not a quadratic function of β and γ anymore once one concentrates out λ and f, and it can have multiple local minima that satisfy the FOC. Therefore, $\hat{\alpha}^{\mathrm{GMM}}$ and $\hat{\beta}^{\mathrm{GMM}}$ can be inconsistent, while $\hat{\alpha}$ and $\hat{\beta}$ are consistent, which is the main reason to consider the latter in this paper.

To illustrate this important difference between $\hat{\alpha}^{\mathrm{GMM}}$, $\hat{\beta}^{\mathrm{GMM}}$ and $\hat{\alpha}$, $\hat{\beta}$, we want to give a simple example for a linear model in which the least squares objective function has multiple local minima. Consider a DGP where $Y_{jt} = \beta^0 X_{jt} + \lambda_j^0 f_t^0 + e_{jt}$, with $X_{jt} = 1 + 0.5\tilde{X}_{jt} + \lambda_j^0 f_t^0$, and $\tilde{X}_{jt}$, e_{jt}, λ_j^0 and f_t^0 are all identically distributed as $\mathcal{N}(0,1)$,

[31] With this weight matrix $\mathcal{W}_{JT}$ the second stage objective function in (B.3) becomes

$$\begin{aligned} &(d(\alpha) - x\beta)'\; x\,(x'x)^{-1}\,x'\;(d(\alpha) - x\beta)\,/JT + d'(\alpha)\,M_x\,z\,(z'M_xz)^{-1}W_{JT}(z'M_xz)^{-1}\,z'\,M_x\,d(\alpha) \\ &= (d(\alpha) - x\beta)'\,P_x\,(d(\alpha) - x\beta)\,/JT + \tilde{\gamma}'_\alpha\,W_{JT}\,\tilde{\gamma}_\alpha\;, \end{aligned}$$

where $d(\alpha) = \mathrm{vec}(\delta(\alpha,\, s,\, X) - \delta(\alpha^0,\, s,\, X))$. Here, β only appears in the first term, and by choosing $\beta = \hat{\beta} = (x'x)^{-1}x'd(\alpha)$ this term becomes zero. Thus, we are left with the second term, which is exactly the second stage objective function in (4.1) in this case, since for $R = 0$ by the Frisch-Waugh theorem we have $\tilde{\gamma}_\alpha = (z'M_xz)^{-1}\,z'\,M_x\,d(\alpha)$.

mutually independent, and independent across j and t. Here, the number of factors $R = 1$, and we assume that Y_{jt} and X_{jt} are observed and that $\beta^0 = 0$. The profiled least squares objective function in this model, which corresponds to our inner loop, is given by $L(\beta) = \sum_{r=2}^{T} \mu_r \left[(Y - \beta X)'(Y - \beta X)\right]$. For $J = T = 100$ and a concrete draw of Y and X, this objective function is plotted in figure 2. The shape of this objective function is qualitatively unchanged for other draws of Y and X, or larger values of J and T. As predicted by our consistency result, the global minimum of $L(\beta)$ is close to $\beta^0 = 0$, but another local minimum is present, which does neither vanish nor converge to $\beta^0 = 0$ when J and T grow to infinity. Thus, the global minimum of $L(\beta)$ gives a consistent estimator, but the solution to the FOC $\partial L(\beta)/\partial\beta = 0$ gives not. In this example, the principal components estimator of $\lambda(\beta)$ and $f(\beta)$, which are derived from $Y - \beta X$, become very bad approximations for λ^0 and f^0 for $\beta \gtrsim 0.5$. Thus, for $\beta \gtrsim 0.5$, the fixed effects are essentially not controlled for anymore in the objective function, and the local minimum around $\beta \approx 0.8$ reflects the resulting endogeneity problem.

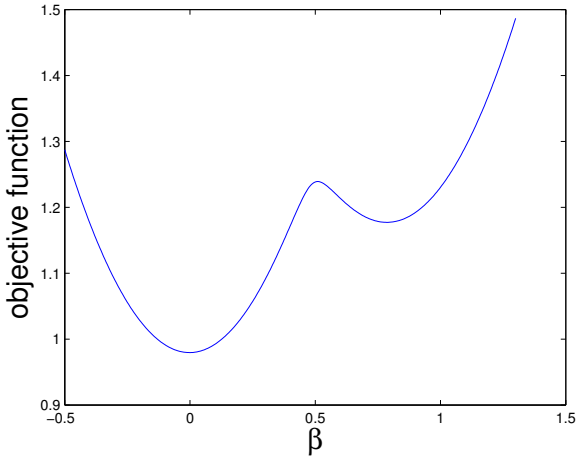

Figure 2: Example for multiple local minima in the least squares objective function $L(\beta)$. The global minimum can be found close to the true value $\beta^0 = 0$, but another local minimum exists around $\beta \approx 0.8$, which renders the FOC inappropriate for defining the estimator $\hat{\beta}$.

C Details for Theorems 5.2 and 5.3

C.1 Formulas for Asymptotic Variance Terms

We define the $JT \times K$ matrix $x^{\lambda f}$, the $JT \times M$ matrix $z^{\lambda f}$, and the $JT \times L$ matrix g by

$$x^{\lambda f}_{.,k} = \text{vec}\left(M_{\lambda^0} X_k M_{f^0}\right) \ , \quad z^{\lambda f}_{.,m} = \text{vec}\left(M_{\lambda^0} Z_m M_{f^0}\right) \ , \quad g_{.,l} = -\text{vec}\left(\nabla_l \, \delta(\alpha^0)\right) \ , \qquad \text{(C.1)}$$

where $k = 1, \ldots, K$, $m = 1, \ldots, M$, and $l = 1, \ldots, L$. Note that $x^{\lambda f} = (\mathbb{1}_T \otimes M_{\lambda^0}) x^f$, $z^{\lambda f} = (\mathbb{1}_T \otimes M_{\lambda^0}) z^f$, and g is the vectorization of the gradient of $\delta(\alpha)$, evaluated at the true parameter. We introduce the $(L+K)\times(L+K)$ matrix G and the $(K+M)\times(K+M)$ matrix Ω as follows

$$G = \operatorname*{plim}_{J,T\to\infty} \frac{1}{JT} \begin{pmatrix} g' x^{\lambda f} & g' z^{\lambda f} \\ x^{\lambda f\prime} x^{\lambda f} & x^{\lambda f\prime} z^{\lambda f} \end{pmatrix}, \quad \Omega = \operatorname*{plim}_{J,T\to\infty} \frac{1}{JT} \left(x^{\lambda f}, z^{\lambda f} \right)' \operatorname{diag}(\Sigma_e^{\text{vec}}) \left(x^{\lambda f}, z^{\lambda f} \right), \tag{C.2}$$

where $\Sigma_e^{\text{vec}} = \text{vec} \left\{ \left[\mathbb{E}\left(e_{jt}^2 \right) \right]_{\substack{\text{j}=1,\ldots,\text{J} \\ \text{t}=1,\ldots,\text{T}}} \right\}$ is the JT-vector of vectorized variances of e_{jt}. Finally, we define the $(K+M)\times(K+M)$ weight matrix $\mathcal{W}$ by

$$\begin{aligned} \mathcal{W} = \operatorname*{plim}_{J,T\to\infty} \Bigg[& \begin{pmatrix} \left(\frac{1}{JT} x^{\lambda f\prime} x^{\lambda f}\right)^{-1} & 0_{K\times M} \\ 0_{M\times K} & 0_{M\times M} \end{pmatrix} + \begin{pmatrix} -(x^{\lambda f\prime} x^{\lambda f})^{-1} x^{\lambda f\prime} z^{\lambda f} \\ \mathbb{1}_M \end{pmatrix} \\ & \times \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \begin{pmatrix} -(x^{\lambda f\prime} x^{\lambda f})^{-1} x^{\lambda f\prime} z^{\lambda f} \\ \mathbb{1}_M \end{pmatrix}' \Bigg]. \end{aligned} \tag{C.3}$$

Existence of these probability limits is imposed by Assumption 3 below.

C.2 Formulas for Asymptotic Bias Terms

Here we provide the formulas for the asymptotic bias terms B_0, B_1 and B_2 that enter into Theorem 5.2. Let the $J \times 1$ vector $\Sigma_e^{(1)}$, the $T \times 1$ vector $\Sigma_e^{(2)}$, and the $T \times T$ matrices $\Sigma_k^{X,e}$, $k = 1, \ldots, K$, and $\Sigma_m^{Z,e}$, $m = 1, \ldots, M$, be defined by

$$\begin{aligned} \Sigma_{e,j}^{(1)} &= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left(e_{jt}^2\right) , & \Sigma_{e,t}^{(2)} &= \frac{1}{J} \sum_{j=1}^{J} \mathbb{E}\left(e_{jt}^2\right) , \\ \Sigma_{k,t\tau}^{X,e} &= \frac{1}{J} \sum_{j=1}^{J} \mathbb{E}\left(X_{k,jt}\, e_{j\tau}\right) , & \Sigma_{m,t\tau}^{Z,e} &= \frac{1}{J} \sum_{j=1}^{J} \mathbb{E}\left(Z_{m,jt}\, e_{j\tau}\right) , \end{aligned} \tag{C.4}$$

where $j = 1, \ldots, J$ and $t, \tau = 1, \ldots, T$. Furthermore, let

$$
\begin{aligned}
b_k^{(x,0)} &= \operatorname*{plim}_{J,T\to\infty} \operatorname{Tr}\left(P_{f^0}\,\Sigma_k^{X,e}\right) , \\
b_k^{(x,1)} &= \operatorname*{plim}_{J,T\to\infty} \operatorname{Tr}\left[\operatorname{diag}\left(\Sigma_e^{(1)}\right) M_{\lambda^0}\, X_k\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right] , \\
b_k^{(x,2)} &= \operatorname*{plim}_{J,T\to\infty} \operatorname{Tr}\left[\operatorname{diag}\left(\Sigma_e^{(2)}\right) M_{f^0}\, X_k'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\right] , \\
b_m^{(z,0)} &= \operatorname*{plim}_{J,T\to\infty} \operatorname{Tr}\left(P_{f^0}\,\Sigma_m^{Z,e}\right) , \\
b_m^{(z,1)} &= \operatorname*{plim}_{J,T\to\infty} \operatorname{Tr}\left[\operatorname{diag}\left(\Sigma_e^{(1)}\right) M_{\lambda^0}\, Z_m\, f^0\,(f^{0\prime}f^0)^{-1}\,(\lambda^{0\prime}\lambda^0)^{-1}\,\lambda^{0\prime}\right] , \\
b_m^{(z,2)} &= \operatorname*{plim}_{J,T\to\infty} \operatorname{Tr}\left[\operatorname{diag}\left(\Sigma_e^{(2)}\right) M_{f^0}\, Z_m'\,\lambda^0\,(\lambda^{0\prime}\lambda^0)^{-1}\,(f^{0\prime}f^0)^{-1}\,f^{0\prime}\right] ,
\end{aligned}
\tag{C.5}
$$

and we set $b^{(x,i)} = \left(b_1^{(x,i)}, \ldots, b_K^{(x,i)}\right)'$ and $b^{(z,i)} = \left(b_1^{(z,i)}, \ldots, b_M^{(z,i)}\right)'$, for $i = 0, 1, 2$. With these definitions we can now give the expression for the asymptotic bias terms which appear in Theorem 5.2, namely

$$
B_i = -\left(G\mathcal{W}G'\right)^{-1} G\mathcal{W} \begin{pmatrix} b^{(x,i)} \\ b^{(z,i)} \end{pmatrix}, \tag{C.6}
$$

where $i = 0, 1, 2$.

C.3 Additional Assumptions for Asymptotic Distribution and Bias Correction

In addition to Assumption 1, which guarantees consistency of the LS-MD estimator, we also require the Assumptions 2, 3 and 4 to derive the limiting distribution of the estimator in Theorem 5.2, and Assumption 5 to provide consistent estimators for the asymptotic bias and asymptotic covariance matrix in Theorem 5.3. These additional assumptions are presented below.

Assumption 2. *We assume that the limits of $\lambda^{0\prime}\lambda^0/J$ and $f^{0\prime}f^0/T$ are finite and have full rank, i.e. (a) $\lim_{J,T\to\infty}\left(\lambda^{0\prime}\lambda^0/J\right) > 0$, (b) $\lim_{J,T\to\infty}\left(f^{0\prime}f^0/T\right) > 0$.*

Assumption 2 guarantees that $\|\lambda^0\|$ and $\|f^0\|$ grow at a rate of $\sqrt{J}$ and $\sqrt{T}$, respectively. This is a so called "strong factor" assumption that makes sure that the influence of the factors is sufficiently large, so that the principal components estimators $\widehat{\lambda}$ and $\widehat{f}$ can pick up the correct factor loadings and factors.

Assumption 3. *We assume existence of the probability limits G, Ω, $\mathcal{W}$, $b^{(x,i)}$ and $b^{(z,i)}$, $i = 0, 1, 2$. In addition, we assume $G\mathcal{W}G' > 0$ and $G\mathcal{W}\Omega\mathcal{W}G' > 0$.*

Assumption 4.

(i) There exist $J \times T$ matrices $r^{\Delta}(\alpha)$ and $\nabla_l \delta(\alpha^0)$, $l = 1, \dots, L$, such that

$$\delta(\alpha) - \delta(\alpha^0) = \sum_{l=1}^{L} (\alpha_l - \alpha_l^0)\, \nabla_l \delta(\alpha^0) + r^{\Delta}(\alpha) \;,$$

and

$$\frac{1}{\sqrt{JT}} \|\nabla_l \delta(\alpha^0)\|_F = \mathcal{O}_p(1) \;, \qquad \text{for } l = 1, \dots, L \;,$$

$$\sup_{\{\alpha:\, \sqrt{J}\|\alpha - \alpha^0\| < c,\, \alpha \neq \alpha^0\}} \frac{\frac{1}{\sqrt{JT}} \|r^{\Delta}(\alpha)\|_F}{\|\alpha - \alpha^0\|} = o_p(1) \;, \qquad \text{for all } c > 0 \;.$$

(ii) $\|\lambda_j^0\|$ and $\|f_t^0\|$ are uniformly bounded across j, t, J and T.

(iii) The errors e_{jt} are independent across j and t, they satisfy $\mathbb{E} e_{jt} = 0$, and $\mathbb{E}(e_{jt})^{8+\epsilon}$ is bounded uniformly across j, t and J, T, for some $\epsilon > 0$.

(iv) The regressors X_k, $k = 1, \dots, K$, (both high- and low rank regressors) and the instruments Z_m, $m = 1, \dots, M$, can be decomposed as $X_k = X_k^{\text{str}} + X_k^{\text{weak}}$ and $Z_m = Z_m^{\text{str}} + Z_m^{\text{weak}}$. The components X_k^{str} and Z_m^{str} are strictly exogenous, i.e. $X_{k,jt}^{\text{str}}$ and $Z_{m,jt}^{\text{str}}$ are independent of $e_{j\tau}$ for all j, i, t, τ. The components X_k^{weak} and Z_m^{weak} are weakly exogenous, and we assume

$$X_{k,jt}^{\text{weak}} = \sum_{\tau=1}^{t-1} c_{k,j\tau}\, e_{j,t-\tau} \;, \qquad Z_{m,jt}^{\text{weak}} = \sum_{\tau=1}^{t-1} d_{m,j\tau}\, e_{j,t-\tau} \;,$$

for some coefficients $c_{k,j\tau}$ and $d_{m,j\tau}$ that satisfy

$$|c_{k,j\tau}| < \alpha^{\tau} \;, \qquad |d_{k,j\tau}| < \alpha^{\tau} \;,$$

where $\alpha \in (0,1)$ is a constant that is independent of $\tau = 1, \dots, T-1$, $j = 1 \dots J$, $k = 1, \dots, K$ and $m = 1, \dots, M$. We also assume that $\mathbb{E}[(X_{k,jt}^{\text{str}})^{8+\epsilon}]$ and $\mathbb{E}[(Z_{m,jt}^{\text{str}})^{8+\epsilon}]$ are bounded uniformly over j, t and J, T, for some $\epsilon > 0$.

Assumption 1(*ii*) and (*iii*) are implied by Assumption 4, so it would not be necessary to impose those explicitly in Theorem 5.2. Part (*ii*), (*iii*) and (*iv*) of Assumption 4 are identical to Assumption 5 in Moon and Weidner (2015a; 2015b), except for the appearance of the instruments Z_m here, which need to be included since they appear as additional regressors in the first step of our estimation procedure. Part (*i*) of Assumption 4 can for example be justified by assuming that within any $\sqrt{J}$-shrinking neighborhood of α^0 we have wpa1 that $\delta_{jt}(\alpha)$ is differentiable, that $|\nabla_l \delta_{jt}(\alpha)|$ is uniformly bounded across j, t, J and T, and that $\nabla_l \delta_{jt}(\alpha)$ is Lipschitz continuous with a Lipschitz constant that is

uniformly bounded across j, t, J and T, for all $l = 1, \dots L$. But since the assumption is only on the Frobenius norm of the gradient and remainder term, one can also conceive weaker sufficient conditions for Assumption 4(i).

Assumption 5. *For all $c > 0$ and $l = 1, \dots, L$ we have*

$$\sup_{\{\alpha:\, \sqrt{JT}\|\alpha-\alpha^0\|<c\}} \|\nabla_l \delta(\alpha) - \nabla_l \delta(\alpha^0)\|_F = o_p(\sqrt{JT}).$$

This last assumption is needed to guarantee consistency of the bias and variance estimators that are presented in the following.

C.4 Bias and Variance Estimators

Here we present consistent estimators for the matrices G, Ω, and $\mathcal{W}$, which enter into the asymptotic variance of the LS-MD estimator, and for the vectors B_0, B_1 and B_2, which enter into the asymptotic bias of the estimator. Consistency of these estimators is stated in Theorem 5.3.

Given the LS-MD estimators $\widehat{\alpha}$ and $\widehat{\beta}$, we can define the residuals

$$\widehat{e} = \delta(\widehat{\alpha},\, s,\, X) - \sum_{k=1}^{K} \widehat{\beta}_k\, X_k - \widehat{\lambda}\widehat{f}' \,. \tag{C.7}$$

We also define the $JT \times K$ matrix $\widehat{x}^{\lambda f}$, the $JT \times M$ matrix $\widehat{z}^{\lambda f}$, and the $JT \times L$ matrix $\widehat{g}$ by

$$\widehat{x}^{\lambda f}_{\cdot,k} = \text{vec}\left(M_{\widehat{\lambda}} X_k M_{\widehat{f}}\right) \,, \quad \widehat{z}^{\lambda f}_{\cdot,m} = \text{vec}\left(M_{\widehat{\lambda}} Z_m M_{\widehat{f}}\right) \,, \quad \widehat{g}_{\cdot,l} = -\text{vec}\left(\nabla_l\, \delta(\widehat{\alpha})\right) \,, \tag{C.8}$$

where $k = 1, \dots, K$, $m = 1, \dots, M$, and $l = 1, \dots, L$. The definition of $\widehat{\Sigma}_e^{\text{vec}}$, $\widehat{\Sigma}_e^{(1)}$ and $\widehat{\Sigma}_e^{(2)}$ is analogous to that of Σ_e^{vec}, $\Sigma_e^{(1)}$ and $\Sigma_e^{(2)}$, but with $\mathbb{E}(e_{jt}^2)$ replaced by $\widehat{e}_{jt}^2$. The $T \times T$ matrices $\widehat{\Sigma}_k^{X,e}$, $k = 1, \dots, K$, and $\widehat{\Sigma}_m^{Z,e}$, $m = 1, \dots, M$, are defined by

$$\begin{aligned}
\widehat{\Sigma}^{X,e}_{k,t\tau} &= \begin{cases} \frac{1}{J} \sum_{j=1}^{J} X_{k,jt}\, \widehat{e}_{j\tau} & \text{for } 0 < t-\tau \leq h \\ 0 & \text{otherwise} \end{cases} \\
\widehat{\Sigma}^{Z,e}_{m,t\tau} &= \begin{cases} \frac{1}{J} \sum_{j=1}^{J} Z_{m,jt}\, \widehat{e}_{j\tau} & \text{for } 0 < t-\tau \leq h \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{C.9}$$

where $t, \tau = 1, \ldots, T$, and $h \in \mathbb{N}$ is a bandwidth parameter. Using these objects we define

$$\begin{aligned}
\widehat{G} &= \frac{1}{JT} \begin{pmatrix} \widehat{g}' \, \widehat{x}^{\lambda f} & \widehat{g}' \, \widehat{z}^{\lambda f} \\ \widehat{x}^{\lambda f \prime} \, \widehat{x}^{\lambda f} & \widehat{x}^{\lambda f \prime} \, \widehat{z}^{\lambda f} \end{pmatrix} , \\
\widehat{\Omega} &= \frac{1}{JT} \left(\widehat{x}^{\lambda f}, \widehat{z}^{\lambda f} \right)' \operatorname{diag}(\widehat{\Sigma}_e^{\text{vec}}) \left(\widehat{x}^{\lambda f}, \widehat{z}^{\lambda f} \right) , \\
\widehat{b}_k^{(x,0)} &= \operatorname{Tr} \left(P_{\widehat{f}} \, \widehat{\Sigma}_k^{X,e} \right) , \\
\widehat{b}_k^{(x,1)} &= \operatorname{Tr} \left[\operatorname{diag} \left(\widehat{\Sigma}_e^{(1)} \right) M_{\widehat{\lambda}} \, X_k \, \widehat{f} \, (\widehat{f}' \widehat{f})^{-1} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, \widehat{\lambda}' \right] , \\
\widehat{b}_k^{(x,2)} &= \operatorname{Tr} \left[\operatorname{diag} \left(\widehat{\Sigma}_e^{(2)} \right) M_{\widehat{f}} \, X_k' \, \widehat{\lambda} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, (\widehat{f}' \widehat{f})^{-1} \, \widehat{f}' \right] , \\
\widehat{b}_m^{(z,0)} &= \operatorname{Tr} \left(P_{\widehat{f}} \, \widehat{\Sigma}_m^{Z,e} \right) , \\
\widehat{b}_m^{(z,1)} &= \operatorname{Tr} \left[\operatorname{diag} \left(\widehat{\Sigma}_e^{(1)} \right) M_{\widehat{\lambda}} \, Z_m \, \widehat{f} \, (\widehat{f}' \widehat{f})^{-1} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, \widehat{\lambda}' \right] , \\
\widehat{b}_m^{(z,2)} &= \operatorname{Tr} \left[\operatorname{diag} \left(\widehat{\Sigma}_e^{(2)} \right) M_{\widehat{f}} \, Z_m' \, \widehat{\lambda} \, (\widehat{\lambda}' \widehat{\lambda})^{-1} \, (\widehat{f}' \widehat{f})^{-1} \, \widehat{f}' \right] ,
\end{aligned} \tag{C.10}$$

for $k = 1, \ldots, K$ and $m = 1, \ldots, M$. We set $\widehat{b}^{(x,i)} = \left(\widehat{b}_1^{(x,i)}, \ldots, \widehat{b}_K^{(x,i)} \right)'$ and $\widehat{b}^{(z,i)} = \left(\widehat{b}_1^{(z,i)}, \ldots, \widehat{b}_M^{(z,i)} \right)'$, for $i = 0, 1, 2$. The estimator of $\mathcal{W}$ is given by

$$\begin{aligned}
\widehat{\mathcal{W}} = &\begin{pmatrix} \left(\frac{1}{JT} \widehat{x}^{\lambda f \prime} \widehat{x}^{\lambda f} \right)^{-1} & 0_{K \times M} \\ 0_{M \times K} & 0_{M \times M} \end{pmatrix} + \begin{pmatrix} -(\widehat{x}^{\lambda f \prime} \widehat{x}^{\lambda f})^{-1} \widehat{x}^{\lambda f \prime} \, \widehat{z}^{\lambda f} \\ \mathbb{1}_M \end{pmatrix} \left(\frac{1}{JT} \widehat{z}^{\lambda f \prime} M_{\widehat{x}^{\lambda f}} \widehat{z}^{\lambda f} \right)^{-1} \\
&W_{JT} \left(\frac{1}{JT} \widehat{z}^{\lambda f \prime} M_{\widehat{x}^{\lambda f}} \widehat{z}^{\lambda f} \right)^{-1} \begin{pmatrix} -(\widehat{x}^{\lambda f \prime} \widehat{x}^{\lambda f})^{-1} \widehat{x}^{\lambda f \prime} \, \widehat{z}^{\lambda f} \\ \mathbb{1}_M \end{pmatrix}' .
\end{aligned} \tag{C.11}$$

Finally, for $i = 0, 1, 2$, we have

$$\widehat{B}_i = - \left(\widehat{G} \widehat{\mathcal{W}} \widehat{G}' \right)^{-1} \widehat{G} \widehat{\mathcal{W}} \begin{pmatrix} \widehat{b}^{(x,i)} \\ \widehat{b}^{(z,i)} \end{pmatrix} . \tag{C.12}$$

The only subtlety here lies in the definition of $\widehat{\Sigma}_k^{X,e}$ and $\widehat{\Sigma}_m^{Z,e}$, where we explicitly impose the constraint that $\widehat{\Sigma}_{k,t\tau}^{X,e} = \widehat{\Sigma}_{m,t\tau}^{Z,e} = 0$ for $t - \tau \leq 0$ and for $t - \tau > h$, where $h \in \mathbb{N}$ is a bandwidth parameter. On the one side ($t - \tau \leq 0$) this constraint stems from the assumption that X_k and Z_m are only correlated with past values of the errors e, not with present and future values, on the other side ($t - \tau > h$) we need the bandwidth cutoff to guarantee that the variance of our estimator for B_0 converges to zero. Without imposing this constraint and introducing the bandwidth parameter, our estimator for B_0 would be inconsistent.

C.5 Correlation of e_{jt} across j

Our assumptions impose that the error term e_{jt} is independent both across products j and over markets/time t. We allow for the regressors X_{jt} to be pre-determined, for example,

lagged dependent variables are allowed, and would therefore run into identification problems if we also allowed e_{jt} to be to be correlated over time. However, it would not cause any conceptual problem to allow weak correlation of e_{jt} across products j. The above formulas for the asymptotic variance and bias of the LS-MD estimator would need to be modified as follows:

- The diagonal matrix $\text{diag}(\Sigma_e^{\text{vec}})$ that appears in the definition of Ω in equation (C.2) needs to be replaced by the potentially non-diagonal $JT \times JT$ variance-covariance matrix of the JT-vector of error terms $\text{vec}(e)$. Otherwise, the formula for the asymptotic variance-covariance matrix $(G\mathcal{W}G')^{-1} G\mathcal{W}\Omega\mathcal{W}G' (G\mathcal{W}G')^{-1}$ of the LS-MD estimator is unchanged.
- The diagonal matrix $\text{diag}\left(\Sigma_e^{(1)}\right)$ that enters into the definition of $b_k^{(x,1)}$ and $b_m^{(z,1)}$ in equation (C.5) needs to be replaced with the potentially non-diagonal $J \times J$ matrix $\frac{1}{T}\sum_{t=1}^{T} \Sigma_t^{(e)}$, where $\Sigma_t^{(e)}$ is the variance-covariance matrix of the J-vector $(e_{jt} : j = 1, \ldots, J)$. After this change and the change of Ω already described above, the asymptotic bias terms B_0, B_1 and B_2 are still given by equation (C.6).

Those two modifications to the asymptotic variance and bias are very much in line with the results in Bai (2009), who allows for cross-sectional dependence in the error term in a linear model with interactive fixed effects. We leave the question of bias correction and inference for the case of cross-sectional dependence in e_{jt} for future work.

D Proofs

In addition to the vectorizations x, $x^{\lambda f}$, z, $z^{\lambda f}$, g, and $d(\alpha)$, which were already defined above, we also introduce the $JT \times K$ matrix x^f, the $JT \times M$ matrix z^f, and the $JT \times 1$ vector ε by

$$x^f_{\cdot,k} = \text{vec}\left(X_k M_{f^0}\right), \qquad z^f_{\cdot,m} = \text{vec}\left(Z_m M_{f^0}\right), \qquad \varepsilon = \text{vec}\left(e\right),$$

where $k = 1, \ldots, K$ and $m = 1, \ldots, M$.

D.1 Proof of Identification

Proof of Theorem 3.1. To show that any two different *parameters* cannot be observational equivalent, we introduce the following functional

$$Q\left(\alpha, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right) = \mathbb{E}_0 \left\| \delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda f' \right\|_F^2,$$

where $\mathbb{E}_0$ refers to the expectation under the distribution of observables $F^0_{s,X,Z}$, which is implied by the model, i.e. $F^0_{s,X,Z} = \Gamma(\alpha^0, \beta^0, \lambda^0 f^{0\prime}, F^0_{e,X,Z})$.

First, we show that under Assumption ID(i)-(iv), the minima of the function $Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$ over $(\beta,\ \gamma,\ \lambda,\ f)$ satisfies $\beta = \beta^0$, $\gamma = 0$, and $\lambda f' = \lambda^0 f^{0\prime}$. Using model (3.1) and Assumption ID(ii) and (iii) we find

$$
\begin{aligned}
&Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right) \\
&= \mathbb{E}_0 \operatorname{Tr}\left\{[\delta(\alpha^0) - \beta \cdot X - \gamma \cdot Z - \lambda f']'[\delta(\alpha^0) - \beta \cdot X - \gamma \cdot Z - \lambda f']\right\} \\
&= \mathbb{E}_0 \operatorname{Tr}\left\{[(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0\prime} - \lambda f' + e]'[(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0\prime} - \lambda f' + e]\right\} \\
&= \mathbb{E}_0 \operatorname{Tr}(e'e) + \underbrace{\mathbb{E}_0 \operatorname{Tr}\left\{[(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0\prime} - \lambda f']'[(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0\prime} - \lambda f']\right\}}_{=Q^*(\beta,\gamma,\lambda,f;F^0_{s,X,Z})}.
\end{aligned}
\tag{D.1}
$$

Note that $Q^*(\beta, \gamma, \lambda, f; F^0_{s,X,Z}) \geq 0$ and that $Q^*(\beta^0, 0, \lambda^0, f^0; F^0_{s,X,Z}) = 0$. Thus, the minimum value of $Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$ equals $\mathbb{E}_0 \operatorname{Tr}(e'e)$ and all parameters that minimize $Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$ must satisfy $Q^*(\beta, \gamma, \lambda, f; F^0_{s,X,Z}) = 0$. We have for any λ and f

$$
\begin{aligned}
Q^*(\beta, \gamma, \lambda, f; F^0_{s,X,Z}) &\geq \mathbb{E}_0 \operatorname{Tr}\{[(\beta^0 - \beta) \cdot X - \gamma \cdot Z]' M_{(\lambda,\lambda^0)}[(\beta^0 - \beta) \cdot X - \gamma \cdot Z]\} \\
&= [(\beta^0 - \beta)', \gamma']\mathbb{E}_0[(x, z)'(\mathbb{1}_T \otimes M_{(\lambda,\lambda^0)})(x, z)][(\beta^0 - \beta)', \gamma']' \\
&\geq b\left(\|\beta - \beta^0\|^2 + \|\gamma\|^2\right)^2,
\end{aligned}
\tag{D.2}
$$

where the last line holds by Assumption ID(iv). This shows that $\beta = \beta^0$ and $\gamma = 0$ is necessary to minimize $Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$. Since $\operatorname{Tr}(AA') = 0$ for a matrix A implies $A = 0$, we find that $Q^*(\beta^0, 0, \lambda, f; F^0_{s,X,Z}) = 0$ implies $\beta = \beta^0$, $\gamma = 0$ and $\lambda^0 f^{0\prime} - \lambda f' = 0$. We have thus shown that $Q\left(\alpha^0, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right)$ is minimized if and only if $\beta = \beta^0$, $\gamma = 0$ and $\lambda f' = \lambda^0 f^{0\prime}$.

For the second part, we introduce a second functional; for a given α we define:

$$
\gamma(\alpha; F^0_{s,X,Z}) \in \operatorname{argmin}_\gamma \min_{\beta,\lambda,f} Q\left(\alpha, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right). \tag{D.3}
$$

We show that under Assumption ID(i)-(v), $\gamma(\alpha; F^0_{s,X,Z}) = 0$ implies $\alpha = \alpha^0$. From part (i) we already know that $\gamma(\alpha^0; F^0_{s,X,Z}) = 0$. The proof proceeds by contradiction. Assume that $\gamma(\alpha; F^0_{s,X,Z}) = 0$ for $\alpha \neq \alpha^0$. By definition of $\gamma(\cdot)$ in Eq. (D.3), this implies that there exists $\tilde{\beta}$, $\tilde{\lambda}$ and $\tilde{f}$ such that

$$
Q\left(\alpha, \tilde{\beta}, 0, \tilde{\lambda}, \tilde{f}; F^0_{s,X,Z}\right) \leq \min_{\beta,\gamma,\lambda,f} Q\left(\alpha, \beta, \gamma, \lambda, f; F^0_{s,X,Z}\right). \tag{D.4}
$$

Using model (3.1) and our assumptions we obtain the following lower bound for the lhs of inequal-

ity (D.4)

$$
\begin{aligned}
Q\left(\alpha,\tilde{\beta},0,\tilde{\lambda},\tilde{f};F^0_{s,X,Z}\right) &= \mathbb{E}_0\mathrm{Tr}\left[\left(\delta(\alpha)-\tilde{\beta}\cdot X-\tilde{\lambda}\tilde{f}'\right)'\left(\delta(\alpha)-\tilde{\beta}\cdot X-\tilde{\lambda}\tilde{f}'\right)\right] \\
&= \mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)-(\tilde{\beta}-\beta^0)\cdot X+\lambda^0 f^0-\tilde{\lambda}\tilde{f}'+e\right)' \\
&\qquad\qquad \left(\delta(\alpha)-\delta(\alpha^0)-(\tilde{\beta}-\beta^0)\cdot X+\lambda^0 f^0-\tilde{\lambda}\tilde{f}'+e\right)\Big] \\
&= 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big]+\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)-(\tilde{\beta}-\beta^0)\cdot X+\lambda^0 f^0-\tilde{\lambda}\tilde{f}'\right)' \\
&\qquad\qquad \left(\delta(\alpha)-\delta(\alpha^0)-(\tilde{\beta}-\beta^0)\cdot X+\lambda^0 f^0-\tilde{\lambda}\tilde{f}'\right)\Big] \\
&\geq 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big]+\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)-(\tilde{\beta}-\beta^0)\cdot X\right)'M_{(\tilde{\lambda},\lambda^0)} \\
&\qquad\qquad \left(\delta(\alpha)-\delta(\alpha^0)-(\tilde{\beta}-\beta^0)\cdot X\right)\Big] \\
&= 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big]+\mathbb{E}_0\left[\Delta\xi'_{\alpha,\tilde{\beta}}\left(\mathbb{1}_T\otimes M_{(\tilde{\lambda},\lambda^0)}\right)\Delta\xi_{\alpha,\tilde{\beta}}\right] \\
&= 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big]+\mathbb{E}_0\left[\Delta\xi'_{\alpha,\tilde{\beta}}\Delta\xi_{\alpha,\tilde{\beta}}\right]-\mathbb{E}_0\left[\Delta\xi'_{\alpha,\tilde{\beta}}\left(\mathbb{1}_T\otimes P_{(\tilde{\lambda},\lambda^0)}\right)\Delta\xi_{\alpha,\tilde{\beta}}\right].
\end{aligned}
\tag{D.5}
$$

Similarly, we obtain the following upper bound for the rhs of the above inequality (D.4)

$$
\begin{aligned}
\min_{\beta,\gamma,\lambda,f} Q\left(\alpha,\beta,\gamma,\lambda,f;F^0_{s,X,Z}\right) &\leq \min_{\beta,\gamma} Q\left(\alpha,\beta,\gamma,\lambda^0,f^0;F^0_{s,X,Z}\right) \\
&= \min_{\beta,\gamma}\mathbb{E}_0\mathrm{Tr}\left[\left(\delta(\alpha)-\beta\cdot X-\gamma\cdot Z-\lambda^0 f^{0\prime}\right)'\left(\delta(\alpha)-\beta\cdot X-\gamma\cdot Z-\lambda^0 f^{0\prime}\right)\right] \\
&= \min_{\beta,\gamma}\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)-(\beta-\beta^0)\cdot X-\gamma\cdot Z+e\right)' \\
&\qquad\qquad \left(\delta(\alpha)-\delta(\alpha^0)-(\beta-\beta^0)\cdot X-\gamma\cdot Z+e\right)\Big] \\
&= 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big]+\min_{\beta,\gamma}\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)-(\beta-\beta^0)\cdot X-\gamma\cdot Z\right)' \\
&\qquad\qquad \left(\delta(\alpha)-\delta(\alpha^0)-(\beta-\beta^0)\cdot X-\gamma\cdot Z\right)\Big] \\
&= 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big] \\
&\qquad +\min_{\beta,\gamma}\mathbb{E}_0\left[\left(\Delta\xi_{\alpha,\tilde{\beta}}-x(\beta-\tilde{\beta})-z\gamma\right)'\left(\Delta\xi_{\alpha,\tilde{\beta}}-x(\beta-\tilde{\beta})-z\gamma\right)\right] \\
&= 2\mathbb{E}_0\mathrm{Tr}\Big[\left(\delta(\alpha)-\delta(\alpha^0)+\tfrac{1}{2}e\right)'e\Big]+\mathbb{E}_0\left[\Delta\xi'_{\alpha,\tilde{\beta}}\Delta\xi_{\alpha,\tilde{\beta}}\right] \\
&\qquad -\mathbb{E}_0\big[\Delta\xi'_{\alpha,\tilde{\beta}}(x,z)\big]\mathbb{E}_0\big[(x,z)'(x,z)\big]^{-1}\mathbb{E}_0\big[(x,z)'\Delta\xi_{\alpha,\tilde{\beta}}\big].
\end{aligned}
\tag{D.6}
$$

Plugging these bounds in the original inequality we obtain

$$
\mathbb{E}_0\big[\Delta\xi'_{\alpha,\tilde{\beta}}(x,z)\big]\mathbb{E}_0\big[(x,z)'(x,z)\big]^{-1}\mathbb{E}_0\big[(x,z)'\Delta\xi_{\alpha,\tilde{\beta}}\big] \leq \mathbb{E}_0\left[\Delta\xi'_{\alpha,\tilde{\beta}}\left(\mathbb{1}_T\otimes P_{(\tilde{\lambda},\lambda^0)}\right)\Delta\xi_{\alpha,\tilde{\beta}}\right], \tag{D.7}
$$

which is a contradiction to Assumption ID(v).

We have thus shown that $\gamma(\alpha;F^0_{s,X,Z})=0$ implies $\alpha=\alpha^0$, which shows that α^0 is uniquely identified from $F^0_{s,X,Z}$. Using that α^0 is identified, we can now use the first part of the proof, and uniquely identify β^0 and $\lambda^0 f^{0\prime}$ from $F^0_{s,X,Z}$ as the unique minimizers of $Q(\alpha^0,\beta,\gamma,\lambda,f;F^0_{s,X,Z})$. Note that these findings immediately preclude observational equivalence, *viz* two sets of distinct parameters $(\alpha^0,\beta^0,\lambda^0,f^0)\neq(\alpha^1,\beta^1,\lambda^1,f^1)$ which are both consistent with the observed distribution $F^0_{s,X,Z}$.

Assumption INV guarantees that for given α^0, β^0 and $\lambda^0 f^{0\prime}$ the map $F^0_{s,X,Z} = \Gamma(\alpha^0, \beta^0, \lambda^0 f^{0\prime}, F^0_{e,X,Z})$ from $F^0_{e,X,Z}$ to $F^0_{s,X,Z}$ is invertible, i.e. we can uniquely identify $F^0_{e,X,Z}$ from $F^0_{s,X,Z}$. ■

D.2 Proof of Consistency

Proof of Theorem 5.1. # Part 1: We show that for any consistent estimator $\widehat{\alpha}$ (not necessarily the LS-MD estimator) we have $\tilde{\beta}_{\widehat{\alpha}} = \beta^0 + o_p(1)$ and $\tilde{\gamma}_{\widehat{\alpha}} = o_p(1)$. Thus, for this part of the proof assume that $\widehat{\alpha} = \alpha^0 + o_p(1)$. This part of the proof is a direct extension of the consistency proof in Moon and Weidner (2015b). We denote the least square objective function by $Q_{JT}(\alpha, \beta, \gamma, \lambda, f) = \frac{1}{JT} \left\| \delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda f' \right\|_F^2$. We first establish a lower bound on $Q_{JT}(\widehat{\alpha}, \beta, \gamma, \lambda, f)$. We have for all λ, f:

$$\begin{aligned}
Q_{JT}(\widehat{\alpha}, \beta, \gamma, \lambda, f) &= \frac{1}{JT} \mathrm{Tr}\left[\left(\delta(\widehat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f'\right)' \left(\delta(\widehat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f'\right) \right] \\
&\geq \frac{1}{JT} \mathrm{Tr}\left[\left(\delta(\widehat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f'\right)' M_{(\lambda, \lambda^0)} \left(\delta(\widehat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f'\right) \right] \\
&= \frac{1}{JT} \mathrm{Tr}\Big[\left(\left(\delta(\widehat{\alpha}) - \delta(\alpha^0)\right) + e - (\beta - \beta^0) \cdot X - \gamma \cdot Z \right)' M_{(\lambda, \lambda^0)} \\
&\qquad\qquad \left(\left(\delta(\widehat{\alpha}) - \delta(\alpha^0)\right) + e - (\beta - \beta^0) \cdot X - \gamma \cdot Z \right) \Big] \\
&\geq b \left\| \beta - \beta^0 \right\|^2 + b \left\| \gamma \right\|^2 + o_p\left(\left\| \beta - \beta^0 \right\| + \left\| \gamma - \gamma^0 \right\| \right) + \frac{1}{JT} \mathrm{Tr}\left(ee'\right) + o_p(1).
\end{aligned} \tag{D.8}$$

where in the last line we used Assumption 1(i), (ii), (iii), and (iv). Here are some representative examples of how the bounds in this last step are obtained from these assumptions:

$$\begin{aligned}
&\frac{1}{JT} \mathrm{Tr}\left[\left((\beta - \beta^0) \cdot X - \gamma \cdot Z \right)' M_{(\lambda, \lambda^0)} \left((\beta - \beta^0) \cdot X - \gamma \cdot Z \right) \right] \\
&\qquad = (\beta', \gamma') [\tfrac{1}{JT} (x, z)' (\mathbb{1}_T \otimes M_{(\lambda, \lambda^0)}) (x, z)] (\beta', \gamma')' \\
&\qquad \geq b(\beta', \gamma')(\beta', \gamma')' = b \left\| \beta - \beta^0 \right\|^2 + b \left\| \gamma \right\|^2, \\
&\left| \frac{1}{JT} \mathrm{Tr}\left[(\delta(\widehat{\alpha}) - \delta(\alpha^0))' M_{(\lambda, \lambda^0)} \left((\beta - \beta^0) \cdot X \right) \right] \right| \\
&\qquad \leq \frac{1}{JT} \left\| \delta(\widehat{\alpha}) - \delta(\alpha^0) \right\|_F \left\| M_{(\lambda, \lambda^0)} \left((\beta - \beta^0) \cdot X \right) \right\|_F \\
&\qquad \leq \frac{1}{JT} \left\| \delta(\widehat{\alpha}) - \delta(\alpha^0) \right\|_F \left\| (\beta - \beta^0) \cdot X \right\|_F \\
&\qquad = \mathcal{O}_p(1) \left\| \widehat{\alpha} - \alpha^0 \right\| \left\| \beta - \beta^0 \right\| = o_p(\left\| \beta - \beta^0 \right\|), \\
&\left| \frac{1}{JT} \mathrm{Tr}\left[e' M_{(\lambda, \lambda^0)} \left((\beta - \beta^0) \cdot X \right) \right] \right| \\
&\qquad = \left| \frac{1}{JT} \mathrm{Tr}\left[e' \left((\beta - \beta^0) \cdot X \right) \right] \right| + \left| \frac{1}{JT} \mathrm{Tr}\left[e' P_{(\lambda, \lambda^0)} \left((\beta - \beta^0) \cdot X \right) \right] \right| \\
&\qquad \leq o_p(1) \|\beta - \beta^0\| + \frac{R}{JT} \|e\| \left\| (\beta - \beta^0) \cdot X \right\| \\
&\qquad \leq o_p(1) \|\beta - \beta^0\| + \frac{R}{JT} \|e\| \left\| (\beta - \beta^0) \cdot X \right\|_F = o_p(\left\| \beta - \beta^0 \right\|).
\end{aligned} \tag{D.9}$$

See the supplementary material in Moon and Weidner (2015a) for further details regarding the algebra here. Applying the same methods, we also obtain

$$Q_{JT}(\widehat{\alpha}, \beta^0, 0, \lambda^0, f^0) = \frac{1}{JT} \operatorname{Tr}(ee') + o_p(1). \tag{D.10}$$

Since we could choose $\beta = \beta^0$, $\gamma = 0$, $\lambda = \lambda^0$ and $f = f^0$ in the first step minimization of the LS-MD estimator, the optimal LS-MD first stage parameters at $\widehat{\alpha}$ need to satisfy $Q_{JT}(\widehat{\alpha}, \tilde{\beta}_{\widehat{\alpha}}, \tilde{\gamma}_{\widehat{\alpha}}, \tilde{\lambda}_{\widehat{\alpha}}, \tilde{f}_{\widehat{\alpha}}) \leq Q_{JT}(\widehat{\alpha}, \beta^0, 0, \lambda^0, f^0)$. Using the above results thus gives

$$b \, \|\tilde{\beta}_{\widehat{\alpha}} - \beta^0\|^2 + b \, \|\tilde{\gamma}_{\widehat{\alpha}}\|^2 + o_p\left(\|\tilde{\beta}_{\widehat{\alpha}} - \beta^0\| + \|\tilde{\gamma}_{\widehat{\alpha}} - \gamma^0\|\right) + o_p(1) \, \leq \, 0 \, . \tag{D.11}$$

It follows that $\|\tilde{\beta}_{\widehat{\alpha}} - \beta^0\| = o_p(1)$ and $\tilde{\gamma}_{\widehat{\alpha}} = o_p(1)$.

Part 2: Now, let $\widehat{\alpha}$ be the LS-MD estimator. We want to show that $\widehat{\alpha} - \alpha^0 = o_p(1)$. From part 1 of the proof we already know that $\tilde{\gamma}_{\alpha^0} = o_p(1)$. In the second step of the LS-MD estimator the optimal choice $\widehat{\alpha}$ minimizes $\tilde{\gamma}'_{\widehat{\alpha}} \, W_{JT} \, \tilde{\gamma}_{\widehat{\alpha}}$, which implies that

$$\tilde{\gamma}'_{\widehat{\alpha}} \, W_{JT} \, \tilde{\gamma}_{\widehat{\alpha}} \, \leq \, \tilde{\gamma}'_{\alpha^0} \, W_{JT} \, \tilde{\gamma}_{\alpha^0} \, = \, o_p(1) \, , \tag{D.12}$$

and therefore $\tilde{\gamma}_{\widehat{\alpha}} = o_p(1)$. Here we used that W_{JT} converges to a positive definite matrix in probability. Analogous to the identification proof we are now going to find an upper and a lower bound for $Q_{JT}\left(\widehat{\alpha}, \tilde{\beta}_{\widehat{\alpha}}, \tilde{\gamma}_{\widehat{\alpha}}, \tilde{\lambda}_{\widehat{\alpha}}, \tilde{f}_{\widehat{\alpha}}\right)$. In the rest of this proof we drop the subscript $\widehat{\alpha}$ on $\tilde{\beta}$, $\tilde{\gamma}$, $\tilde{\lambda}$ and $\tilde{f}$. Using model (3.1) and our assumptions we obtain the following lower bound

$$\begin{aligned}
Q_{JT}\left(\widehat{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\lambda}, \tilde{f}\right) &= \tfrac{1}{JT} \operatorname{Tr}\left[\left(\delta(\widehat{\alpha}) - \tilde{\beta} \cdot X - \tilde{\gamma} \cdot Z - \tilde{\lambda}\tilde{f}'\right)' \left(\delta(\widehat{\alpha}) - \tilde{\beta} \cdot X - \tilde{\gamma} \cdot Z - \tilde{\lambda}\tilde{f}'\right)\right] \\
&= \tfrac{1}{JT} \operatorname{Tr}\Big[\left(\delta(\widehat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + \lambda^0 f^0 - \tilde{\lambda}\tilde{f}' + e\right)' \\
&\qquad \left(\delta(\widehat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + \lambda^0 f^0 - \tilde{\lambda}\tilde{f}' + e\right)\Big] \\
&\geq \tfrac{1}{JT} \operatorname{Tr}\Big[\left(\delta(\widehat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + e\right)' M_{(\tilde{\lambda}, \lambda^0)} \\
&\qquad \left(\delta(\widehat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + e\right)\Big] \\
&= \tfrac{1}{JT} \operatorname{Tr}\Big[\left(\delta(\widehat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X\right)' M_{(\tilde{\lambda}, \lambda^0)} \left(\delta(\widehat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X\right)\Big] \\
&\quad + \tfrac{2}{JT} \operatorname{Tr}\Big[\left(\delta(\widehat{\alpha}) - \delta(\alpha^0) + \tfrac{1}{2}e\right)' e\Big] + o_p(\|\widehat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1) \\
&= \tfrac{1}{JT} \left[\Delta\xi'_{\widehat{\alpha}, \tilde{\beta}} \left(\mathbb{1}_T \otimes M_{(\tilde{\lambda}, \lambda^0)}\right) \Delta\xi_{\widehat{\alpha}, \tilde{\beta}}\right] \\
&\quad + \tfrac{2}{JT} \operatorname{Tr}\Big[\left(\delta(\widehat{\alpha}) - \delta(\alpha^0) + \tfrac{1}{2}e\right)' e\Big] + o_p(\|\widehat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1) \\
&= \tfrac{1}{JT} \left[\Delta\xi'_{\widehat{\alpha}, \tilde{\beta}} \Delta\xi_{\widehat{\alpha}, \tilde{\beta}}\right] - \tfrac{1}{JT} \left[\Delta\xi'_{\widehat{\alpha}, \tilde{\beta}} \left(\mathbb{1}_T \otimes P_{(\tilde{\lambda}, \lambda^0)}\right) \Delta\xi_{\widehat{\alpha}, \tilde{\beta}}\right] \\
&\quad + \tfrac{2}{JT} \operatorname{Tr}\Big[\left(\delta(\widehat{\alpha}) - \delta(\alpha^0) + \tfrac{1}{2}e\right)' e\Big] + o_p(\|\widehat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1).
\end{aligned} \tag{D.13}$$

The bounds used here are analogous to those in (D.9), and we again refer to the supplementary material in Moon and Weidner (2015a).

Similarly, we obtain the following upper bound

$$
\begin{aligned}
&Q_{JT}\left(\widehat{\alpha},\tilde{\beta}_{\widehat{\alpha}},\tilde{\gamma}_{\widehat{\alpha}},\tilde{\lambda}_{\widehat{\alpha}},\tilde{f}_{\widehat{\alpha}}\right)=\min_{\beta,\gamma,\lambda,f}Q_{JT}\left(\widehat{\alpha},\beta,\gamma,\lambda,f\right)\leq\min_{\beta,\gamma}Q_{JT}\left(\widehat{\alpha},\beta,\gamma,\lambda^{0},f^{0}\right)\\
&=\min_{\beta,\gamma}\tfrac{1}{JT}\mathrm{Tr}\left[\left(\delta(\widehat{\alpha})-\beta\cdot X-\gamma\cdot Z-\lambda^{0}f^{0\prime}\right)'\left(\delta(\widehat{\alpha})-\beta\cdot X-\gamma\cdot Z-\lambda^{0}f^{0\prime}\right)\right]\\
&=\min_{\beta,\gamma}\tfrac{1}{JT}\mathrm{Tr}\Big[\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})-(\beta-\beta^{0})\cdot X-\gamma\cdot Z+e\right)'\\
&\qquad\qquad\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})-(\beta-\beta^{0})\cdot X-\gamma\cdot Z+e\right)\Big]\\
&=\tfrac{2}{JT}\mathrm{Tr}\Big[\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})+\tfrac{1}{2}e\right)'e\Big]+\min_{\beta,\gamma}\tfrac{1}{JT}\mathrm{Tr}\Big[\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})-(\beta-\beta^{0})\cdot X-\gamma\cdot Z\right)'\\
&\qquad\qquad\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})-(\beta-\beta^{0})\cdot X-\gamma\cdot Z\right)\Big]\\
&=\tfrac{2}{JT}\mathrm{Tr}\Big[\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})+\tfrac{1}{2}e\right)'e\Big]\\
&\qquad+\min_{\beta,\gamma}\tfrac{1}{JT}\left[\left(\Delta\xi_{\widehat{\alpha},\tilde{\beta}}-x(\beta-\tilde{\beta})-z\gamma\right)'\left(\Delta\xi_{\widehat{\alpha},\tilde{\beta}}-x(\beta-\tilde{\beta})-z\gamma\right)\right]\\
&=\tfrac{2}{JT}\mathrm{Tr}\Big[\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})+\tfrac{1}{2}e\right)'e\Big]+\tfrac{1}{JT}\left[\Delta\xi'_{\alpha,\tilde{\beta}}\Delta\xi_{\widehat{\alpha},\tilde{\beta}}\right]\\
&\qquad-\tfrac{1}{JT}\left[\Delta\xi'_{\widehat{\alpha},\tilde{\beta}}(x,z)\right]\left[(x,z)'(x,z)\right]^{-1}\left[(x,z)'\Delta\xi_{\widehat{\alpha},\tilde{\beta}}\right].
\end{aligned}
\tag{D.14}
$$

Combining this upper and lower bound we obtain

$$
\begin{aligned}
&\tfrac{1}{JT}\left[\Delta\xi'_{\widehat{\alpha},\tilde{\beta}}(x,z)\right]\left[(x,z)'(x,z)\right]^{-1}\left[(x,z)'\Delta\xi_{\widehat{\alpha},\tilde{\beta}}\right]\\
&\qquad\qquad-\tfrac{1}{JT}\left[\Delta\xi'_{\widehat{\alpha},\tilde{\beta}}\left(\mathbb{1}_{T}\otimes P_{(\tilde{\lambda},\lambda^{0})}\right)\Delta\xi_{\widehat{\alpha},\tilde{\beta}}\right]\leq o_{p}(\|\widehat{\alpha}-\alpha^{0}\|+\|\tilde{\beta}-\beta^{0}\|)+o_{p}(1).
\end{aligned}
\tag{D.15}
$$

Using Assumption 1(v) we thus obtain

$$
b\|\widehat{\alpha}-\alpha^{0}\|^{2}+b\|\tilde{\beta}-\beta^{0}\|^{2}\leq o_{p}(\|\widehat{\alpha}-\alpha^{0}\|+\|\tilde{\beta}-\beta^{0}\|)+o_{p}(1),
\tag{D.16}
$$

from which we can conclude that $\|\widehat{\alpha}-\alpha^{0}\|=o_{p}(1)$ and $\|\tilde{\beta}-\beta^{0}\|=o_{p}(1)$.

Part 3: Showing consistency of $\widehat{\beta}$ obtained from step 3 of the LS-MD estimation procedure is analogous to part 1 of the proof — one only needs to eliminate all γ variables from part 1 of the proof, which actually simplifies the proof. ■

D.3 Proof of Limiting Distribution

Lemma D.1. *Let Assumption 1 be satisfied and in addition let* $(JT)^{-1/2}\mathrm{Tr}(eX'_{k})=\mathcal{O}_{p}(1)$, *and* $(JT)^{-1/2}\mathrm{Tr}(eZ'_{m})=\mathcal{O}_{p}(1)$. *In the limit* $J,T\to\infty$ *with* $J/T\to\kappa^{2}$, $0<\kappa<\infty$, *we then have* $\sqrt{J}(\widehat{\alpha}-\alpha)=\mathcal{O}_{p}(1)$.

Proof. The proof is exactly analogous to the consistency proof. We know from Moon and Weidner (2015a; 2015b) that $\sqrt{J}\tilde{\gamma}_{\alpha^{0}}=\mathcal{O}_{p}(1)$. Applying the inequality (D.12) one thus finds $\sqrt{J}\tilde{\gamma}_{\widehat{\alpha}}=\mathcal{O}_{p}(1)$. With the additional assumptions in the lemma one can strengthen the result in (D.13) as follows

$$
\begin{aligned}
&Q_{JT}\left(\widehat{\alpha},\tilde{\beta},\tilde{\gamma},\tilde{\lambda},\tilde{f}\right)\\
&\geq\tfrac{1}{JT}\left[\Delta\xi'_{\widehat{\alpha},\tilde{\beta}}\Delta\xi_{\widehat{\alpha},\tilde{\beta}}\right]-\tfrac{1}{JT}\left[\Delta\xi'_{\widehat{\alpha},\tilde{\beta}}\left(\mathbb{1}_{T}\otimes P_{(\tilde{\lambda},\lambda^{0})}\right)\Delta\xi_{\widehat{\alpha},\tilde{\beta}}\right]\\
&\qquad+\tfrac{2}{JT}\mathrm{Tr}\Big[\left(\delta(\widehat{\alpha})-\delta(\alpha^{0})+\tfrac{1}{2}e\right)'e\Big]+\mathcal{O}_{p}\left(\sqrt{J}\|\widehat{\alpha}-\alpha^{0}\|+\sqrt{J}\|\tilde{\beta}-\beta^{0}\|\right)+\mathcal{O}_{p}(1/J).
\end{aligned}
\tag{D.17}
$$

Using this stronger result and following the steps in the consistency proof then yields $\sqrt{J}(\widehat{\alpha}-\alpha) = \mathcal{O}_p(1)$. ■

Proof of Theorem 5.2. Assumption 4 guarantees $(JT)^{-1/2}\text{Tr}(eX_k') = \mathcal{O}_p(1)$, and $(JT)^{-1/2}\text{Tr}(eZ_m') = \mathcal{O}_p(1)$, so that we can apply Lemma D.1 to conclude $\sqrt{J}(\widehat{\alpha}-\alpha) = \mathcal{O}_p(1)$.

The first step in the definition of the LS-MD estimator is equivalent to the linear regression model with interactive fixed effects, but with an error matrix that has an additional term $\Delta\delta(\alpha) \equiv \delta(\alpha) - \delta(\alpha^0)$, we write $\Psi(\alpha) \equiv e + \Delta\delta(\alpha)$ for this effective error term. Using $\widehat{\alpha} - \alpha^0 = o_p(1)$ and Assumption 1(i) we have $\|\Psi(\widehat{\alpha})\| = o_p(\sqrt{JT})$, so that the results in Moon and Weidner (2015a; 2015b) guarantee $\tilde{\beta}_{\widehat{\alpha}} - \beta^0 = o_p(1)$ and $\|\tilde{\gamma}_{\widehat{\alpha}}\| = o_p(1)$, which we already used in the consistency proof. Using $\sqrt{J}(\widehat{\alpha}-\alpha) = \mathcal{O}_p(1)$ and Assumption 4(i) we find $\|\Psi(\widehat{\alpha})\| = \mathcal{O}_p(\sqrt{J})$, which allows us to truncate the asymptotic likelihood expansion derived in Moon and Weidner (2015a; 2015b) at an appropriate order. Namely, applying their results we have

$$\sqrt{JT}\begin{pmatrix} \tilde{\beta}_\alpha - \beta^0 \\ \tilde{\gamma}_\alpha \end{pmatrix} = V_{JT}^{-1}\begin{pmatrix} \left[C^{(1)}\left(X_k,\Psi(\alpha)\right) + C^{(2)}\left(X_k,\Psi(\alpha)\right)\right]_{k=1,\ldots,K} \\ \left[C^{(1)}\left(Z_m,\Psi(\alpha)\right) + C^{(2)}\left(Z_m,\Psi(\alpha)\right)\right]_{m=1,\ldots,M} \end{pmatrix} + r^{\text{LS}}(\alpha), \tag{D.18}$$

where

$$\begin{aligned} V_{JT} &= \frac{1}{JT}\begin{pmatrix} \left[\text{Tr}(M_{f^0}X_{k_1}'M_{\lambda^0}X_{k_2})\right]_{k_1,k_2=1,\ldots,K} & \left[\text{Tr}(M_{f^0}X_k'M_{\lambda^0}Z_m)\right]_{k=1,\ldots,K;m=1,\ldots,M} \\ \left[\text{Tr}(M_{f^0}Z_m'M_{\lambda^0}X_k)\right]_{m=1,\ldots,M;k=1,\ldots,K} & \left[\text{Tr}(M_{f^0}Z_{m_1}'M_{\lambda^0}Z_{m_2})\right]_{m_1,m_2=1,\ldots,M} \end{pmatrix} \\ &= \frac{1}{JT}\left(x^{\lambda f}, z^{\lambda f}\right)'\left(x^{\lambda f}, z^{\lambda f}\right) \ , \end{aligned} \tag{D.19}$$

and for $\mathcal{X}$ either X_k or Z_m and $\Psi = \Psi(\alpha)$ we have

$$\begin{aligned} C^{(1)}\left(\mathcal{X},\,\Psi\right) &= \frac{1}{\sqrt{JT}}\,\text{Tr}\left[M_{f^0}\,\Psi'\,M_{\lambda^0}\,\mathcal{X}\right] \ , \\ C^{(2)}\left(\mathcal{X},\,\Psi\right) &= -\frac{1}{\sqrt{JT}}\Big[\text{Tr}\left(\Psi M_{f^0}\,\Psi'\,M_{\lambda^0}\,\mathcal{X}\,f^0\left(f^{0\prime}f^0\right)^{-1}\left(\lambda^{0\prime}\lambda^0\right)^{-1}\lambda^{0\prime}\right) \\ &\qquad + \text{Tr}\left(\Psi' M_{\lambda^0}\,\Psi\,M_{f^0}\,\mathcal{X}'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right) \\ &\qquad + \text{Tr}\left(\Psi' M_{\lambda^0}\,\mathcal{X}\,M_{f^0}\,\Psi'\,\lambda^0\left(\lambda^{0\prime}\lambda^0\right)^{-1}\left(f^{0\prime}f^0\right)^{-1}f^{0\prime}\right)\Big], \end{aligned} \tag{D.20}$$

and finally for the remainder we have

$$\begin{aligned} r^{\text{LS}}(\alpha) &= \mathcal{O}_p\left((JT)^{-3/2}\|\Psi(\alpha)\|^3\|X_k\|\right) + \mathcal{O}_p\left((JT)^{-3/2}\|\Psi(\alpha)\|^3\|Z_m\|\right) \\ &\quad + \mathcal{O}_p\left((JT)^{-1}\|\Psi(\alpha)\|\|X_k\|^2\|\|\tilde{\beta}_\alpha - \beta^0\|\right) + \mathcal{O}_p\left((JT)^{-1}\|\Psi(\alpha)\|\|Z_m\|^2\|\tilde{\gamma}_\alpha\|\right) \ , \end{aligned} \tag{D.21}$$

which holds uniformly over α. The first two terms in $r^{\text{LS}}(\alpha)$ stem from the bound on higher order terms in the score function ($C^{(3)}$, $C^{(4)}$, etc.), where $\Psi(\alpha)$ appears three times or more in the expansion, while the last two terms in $r^{\text{LS}}(\alpha)$ reflect the bound on higher order terms in the Hessian expansion, and beyond. Note that Assumption 1(iv) already guarantees that $V_{JT} > b > 0$, wpa1. Applying $\|X_k\| = \mathcal{O}_p(\sqrt{JT})$, $\|Z_m\| = \mathcal{O}_p(\sqrt{JT})$, and $\|\Psi(\alpha)\| = \mathcal{O}_p(\sqrt{J})$ within $\sqrt{J}\|\alpha - \alpha^0\| < c$, we find for all $c > 0$

$$\sup_{\{\alpha:\,\sqrt{J}\|\alpha-\alpha^0\|<c\}} \frac{\left\|r^{\text{LS}}(\alpha)\right\|}{1+\sqrt{JT}\|\tilde{\beta}_\alpha - \beta^0\| + \sqrt{JT}\|\tilde{\gamma}_\alpha\|} = o_p(1) \ . \tag{D.22}$$

The inverse of the partitioned matrix V_{JT} is given by

$$V_{JT}^{-1} = JT \begin{pmatrix} \left(x^{\lambda f\prime} M_{z^{\lambda f}} x^{\lambda f}\right)^{-1} & -\left(x^{\lambda f\prime} M_{z^{\lambda f}} x^{\lambda f}\right)^{-1} \left(x^{\lambda f\prime} z^{\lambda f}\right) \left(z^{\lambda f\prime} z^{\lambda f}\right)^{-1} \\ -\left(z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \left(z^{\lambda f\prime} x^{\lambda f}\right) \left(x^{\lambda f\prime} x^{\lambda f}\right)^{-1} & \left(z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \end{pmatrix}. \tag{D.23}$$

Using $\sqrt{J}(\widehat{\alpha} - \alpha) = \mathcal{O}_p(1)$ and Assumption 4(i) we find

$$\begin{aligned} \begin{pmatrix} \left[C^{(1)}\left(X_k, \Psi(\widehat{\alpha})\right)\right]_{k=1,\ldots,K} \\ \left[C^{(1)}\left(Z_m, \Psi(\widehat{\alpha})\right)\right]_{m=1,\ldots,M} \end{pmatrix} &= \frac{1}{\sqrt{JT}} \left(x^{\lambda f}, z^{\lambda f}\right)' \varepsilon \\ &\quad - \left[\frac{1}{JT}\left(x^{\lambda f}, z^{\lambda f}\right)' g\right] \sqrt{JT}(\widehat{\alpha} - \alpha^0) + o_p(\sqrt{JT}\|\widehat{\alpha} - \alpha^0\|), \\ \begin{pmatrix} \left[C^{(2)}\left(X_k, \Psi(\widehat{\alpha})\right)\right]_{k=1,\ldots,K} \\ \left[C^{(2)}\left(Z_m, \Psi(\widehat{\alpha})\right)\right]_{m=1,\ldots,M} \end{pmatrix} &= \begin{pmatrix} c_x^{(2)} \\ c_z^{(2)} \end{pmatrix} + \mathcal{O}_p\left(\sqrt{J}\|\widehat{\alpha} - \alpha^0\|\right), \end{aligned} \tag{D.24}$$

where

$$c_x^{(2)} = \left[C^{(2)}\left(X_k, e\right)\right]_{k=1,\ldots,K}, \qquad c_z^{(2)} = \left[C^{(2)}\left(Z_m, e\right)\right]_{m=1,\ldots,M}. \tag{D.25}$$

From this one can conclude that $\sqrt{JT}\|\tilde{\beta}_{\widehat{\alpha}} - \beta^0\| = \mathcal{O}_p(1) + \mathcal{O}_p(\sqrt{JT}\|\widehat{\alpha} - \alpha^0\|)$ and $\sqrt{JT}\|\tilde{\gamma}_{\widehat{\alpha}}\| = \mathcal{O}_p(1) + \mathcal{O}_p(\sqrt{JT}\|\widehat{\alpha} - \alpha^0\|)$, so that we find $r^{\mathrm{LS}}(\widehat{\alpha}) = o_p(1) + o_p(\sqrt{JT}\|\widehat{\alpha} - \alpha^0\|)$. Combining the above results we obtain

$$\begin{aligned} \sqrt{JT}\,\tilde{\gamma}_{\widehat{\alpha}} = \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \Bigg[&\frac{1}{\sqrt{JT}} z^{\lambda f\prime} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - \left(z^{\lambda f\prime} x^{\lambda f}\right)\left(x^{\lambda f\prime} x^{\lambda f}\right)^{-1} c_x^{(2)} \\ &- \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} g\right) \sqrt{JT}\left(\widehat{\alpha} - \alpha^0\right) \Bigg] + o_p(1) + o_p(\sqrt{JT}\|\widehat{\alpha} - \alpha^0\|). \end{aligned} \tag{D.26}$$

The above results holds not only for $\widehat{\alpha}$, but uniformly for all α in any $\sqrt{J}$ shrinking neighborhood of α^0 (we made this explicit in the bound on $r^{\mathrm{LS}}(\alpha)$ above; one could define similar remainder terms with corresponding bounds in all intermediate steps), i.e. we have

$$\begin{aligned} \sqrt{JT}\,\tilde{\gamma}_{\alpha} = \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \Bigg[&\frac{1}{\sqrt{JT}} z^{\lambda f\prime} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - \left(z^{\lambda f\prime} x^{\lambda f}\right)\left(x^{\lambda f\prime} x^{\lambda f}\right)^{-1} c_x^{(2)} \\ &- \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} g\right) \sqrt{JT}\left(\alpha - \alpha^0\right) \Bigg] + r^{\gamma}(\alpha), \end{aligned} \tag{D.27}$$

where for all $c > 0$

$$\sup_{\{\alpha:\, \sqrt{J}\|\alpha - \alpha^0\| < c\}} \frac{\|r^{\gamma}(\alpha)\|}{1 + \sqrt{JT}\|\alpha - \alpha^0\|} = o_p(1). \tag{D.28}$$

Therefore, the objective function for $\widehat{\alpha}$ reads

$$JT\,\tilde{\gamma}_{\alpha}' W_{JT}\,\tilde{\gamma}_{\alpha} = A_0 - 2\,A_1' \left[\sqrt{JT}\left(\alpha - \alpha^0\right)\right] + \left[\sqrt{JT}\left(\alpha - \alpha^0\right)\right]' A_2 \left[\sqrt{JT}\left(\alpha - \alpha^0\right)\right] + r^{\mathrm{obj}}(\alpha), \tag{D.29}$$

where A_0 is a scalar, A_1 is an $L \times 1$ vector, and A_2 is an $L \times L$ matrix defined by

$$\begin{aligned}
A_0 &= \left[\frac{1}{\sqrt{JT}}\, z^{\lambda f\prime}\, M_{x^{\lambda f}}\varepsilon + c_z^{(2)} - \left(z^{\lambda f\prime}x^{\lambda f}\right)\left(x^{\lambda f\prime}x^{\lambda f}\right)^{-1} c_x^{(2)}\right]' \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} W_{JT} \\
&\qquad \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \left[\frac{1}{\sqrt{JT}}\, z^{\lambda f\prime}\, M_{x^{\lambda f}}\varepsilon + c_z^{(2)} - \left(z^{\lambda f\prime}x^{\lambda f}\right)\left(x^{\lambda f\prime}x^{\lambda f}\right)^{-1} c_x^{(2)}\right] , \\
A_1 &= \left(\frac{1}{JT}\, g'\, M_{x^{\lambda f}} z^{\lambda f}\right)\left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} W_{JT} \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \\
&\qquad \left[\frac{1}{\sqrt{JT}}\, z^{\lambda f\prime}\, M_{x^{\lambda f}}\varepsilon + c_z^{(2)} - \left(z^{\lambda f\prime}x^{\lambda f}\right)\left(x^{\lambda f\prime}x^{\lambda f}\right)^{-1} c_x^{(2)}\right] , \\
A_2 &= \left(\frac{1}{JT}\, g'\, M_{x^{\lambda f}} z^{\lambda f}\right)\left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} W_{JT} \left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} \left(\frac{1}{JT}\, z^{\lambda f\prime}\, M_{x^{\lambda f}} g\right) ,
\end{aligned} \tag{D.30}$$

and the remainder term in the objective function satisfies

$$\sup_{\{\alpha:\, \sqrt{J}\|\alpha-\alpha^0\|<c\}} \frac{\|r^{\text{obj}}(\alpha)\|}{\left(1+\sqrt{JT}\|\alpha-\alpha^0\|\right)^2} = o_p(1) . \tag{D.31}$$

Under our assumptions one can show that $\|A_1\| = \mathcal{O}_p(1)$ and $\text{plim}_{J,T\to\infty} A_2 > 0$. Combining the expansion of the objective function with the results of $\sqrt{J}$-consistency of $\widehat{\alpha}$ we can thus conclude that

$$\sqrt{JT}\, \left(\widehat{\alpha} - \alpha^0\right) = A_2^{-1}\, A_1 + o_p(1) . \tag{D.32}$$

Analogous to equation (D.18) for the first step, we can apply the results in Moon and Weidner (2015a; 2015b) to the third step of the LS-MD estimator to obtain

$$\begin{aligned}
\sqrt{JT}(\widehat{\beta} - \beta^0) &= \left(\frac{1}{JT} x^{\lambda f\prime} x^{\lambda f}\right)^{-1} \left[C^{(1)}\left(X_k, \Psi(\widehat{\alpha})\right) + C^{(2)}\left(X_k, \Psi(\widehat{\alpha})\right)\right]_{k=1,\ldots,K} + o_p(1) \\
&= \left(\frac{1}{JT} x^{\lambda f\prime} x^{\lambda f}\right)^{-1} \left[\frac{1}{\sqrt{JT}} x^{\lambda f\prime}\, \varepsilon - \left(\frac{1}{JT} x^{\lambda f\prime}\, g\right) \sqrt{JT}\, (\widehat{\alpha} - \alpha^0) + c_x^{(2)}\right] + o_p(1) .
\end{aligned} \tag{D.33}$$

Here, the remainder term $o_p(\sqrt{JT}\|\widehat{\alpha} - \alpha^0\|)$ is already absorbed into the $o_p(1)$ term, since (D.32) already shows $\sqrt{JT}$-consistency of $\widehat{\alpha}$. Let G_{JT} and $\mathcal{W}_{JT}$ be the expressions in equation (C.2) and (C.3) before taking the probability limits, i.e. $G = \text{plim}_{J,T\to\infty} G_{JT}$ and $\mathcal{W} = \text{plim}_{J,T\to\infty} \mathcal{W}_{JT}$. One can show that

$$G_{JT} W_{JT} G_{JT}' = \frac{1}{JT}\, (g, x)'\, P_{x^{\lambda f}}\, (g, x) + \begin{pmatrix} A_2 & 0_{L\times K} \\ 0_{K\times L} & 0_{K\times K} \end{pmatrix} . \tag{D.34}$$

Using this, one can rewrite equation (D.32) and (D.33) as follows

$$\begin{aligned}
&G_{JT} \mathcal{W}_{JT} G_{JT}' \sqrt{JT} \begin{pmatrix} \widehat{\alpha} - \alpha^0 \\ \widehat{\beta} - \beta^0 \end{pmatrix} \\
&\quad = \frac{1}{\sqrt{JT}}\, (g, x)'\, P_{x^{\lambda f}} \varepsilon + \begin{pmatrix} A_1 + \left(\frac{1}{JT} g' x^{\lambda f}\right)\left(\frac{1}{JT} x^{\lambda f\prime} x^{\lambda f}\right)^{-1} c_x^{(2)} \\ c_x^{(2)} \end{pmatrix} + o_p(1) ,
\end{aligned} \tag{D.35}$$

and therefore

$$\begin{aligned}
&\sqrt{JT}\begin{pmatrix} \widehat{\alpha}-\alpha^0 \\ \widehat{\beta}-\beta^0 \end{pmatrix} \\
&\quad = \left(G_{JT}\mathcal{W}_{JT}G_{JT}'\right)^{-1} G_{JT}\mathcal{W}_{JT}\left[\frac{1}{\sqrt{JT}}\left(x^{\lambda f}, z^{\lambda f}\right)'\varepsilon\right] \\
&\qquad + \left(G_{JT}\mathcal{W}_{JT}G_{JT}'\right)^{-1}\begin{pmatrix} A_3 c_z^{(2)} + \left[\left(g' x^{\lambda f}\right) - A_3\left(z^{\lambda f\prime} x^{\lambda f}\right)\right]\left(x^{\lambda f\prime} x^{\lambda f}\right)^{-1} c_x^{(2)} \\ c_x^{(2)} \end{pmatrix} + o_p(1) \\
&\quad = \left(G\mathcal{W}G'\right)^{-1} G\mathcal{W}\left[\frac{1}{\sqrt{JT}}\left(x^{\lambda f}, z^{\lambda f}\right)'\varepsilon + \begin{pmatrix} c_x^{(2)} \\ c_z^{(2)} \end{pmatrix}\right] + o_p(1), \qquad \text{(D.36)}
\end{aligned}$$

where $A_3 = \left(\frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f}\right)\left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1} W_{JT}\left(\frac{1}{JT} z^{\lambda f\prime} M_{x^{\lambda f}} z^{\lambda f}\right)^{-1}$. Having equation (D.36), all that is left to do is to derive the asymptotic distribution of $c_x^{(2)}$, $c_z^{(2)}$ and $\frac{1}{\sqrt{JT}}\left(x^{\lambda f}, z^{\lambda f}\right)'\varepsilon$. This was done in Moon and Weidner (2015a; 2015b) under the same assumptions that we impose here. They show that

$$c_x^{(2)} = -\kappa^{-1} b^{(x,1)} - \kappa\, b^{(x,2)} + o_p(1)\,, \qquad c_z^{(2)} = -\kappa^{-1} b^{(z,1)} - \kappa\, b^{(z,2)} + o_p(1)\,, \qquad \text{(D.37)}$$

and

$$\frac{1}{\sqrt{JT}}\left(x^{\lambda f}, z^{\lambda f}\right)'\varepsilon \;\underset{d}{\longrightarrow}\; \mathcal{N}\left[-\kappa\begin{pmatrix} b^{(x,0)} \\ b^{(z,0)} \end{pmatrix}, \Omega\right]\,. \qquad \text{(D.38)}$$

Plugging this into (D.36) gives the result on the limiting distribution of $\widehat{\alpha}$ and $\widehat{\beta}$ which is stated in the theorem. ■

D.4 Consistency of Bias and Variance Estimators

Proof of Theorem 5.3. From Moon and Weidner (2015a; 2015b) we already know that under our assumptions we have $\widehat{\Omega} = \Omega + o_p(1)$, $\widehat{b}^{(x,i)} = b^{(x,i)} + o_p(1)$ and $\widehat{b}^{(z,i)} = b^{(z,i)} + o_p(1)$, for $i = 0, 1, 2$. They also show that $\|M_{\widehat{\lambda}} - M_{\lambda^0}\| = \mathcal{O}_p(J^{-1/2})$ and $\|M_{\widehat{f}} - M_{f^0}\| = \mathcal{O}_p(J^{-1/2})$, from which we can conclude that $\widehat{\mathcal{W}} = \mathcal{W} + o_p(1)$. These results on $M_{\widehat{\lambda}}$ and $M_{\widehat{f}}$ together with $\sqrt{JT}$-consistency of $\widehat{\alpha}$ and Assumption 5 are also sufficient to conclude $\widehat{G} = G + o_p(1)$. It follows that $\widehat{B}_i = B_i + o_p(1)$, for $i = 0, 1, 2$. ■

E Additional details on numerical verification of instrument relevance condition

Here we present some additional details related to the numerical verification of the instrument relevance condition, which was discussed in Section 6.1 of the main text.

For the numerator of $\rho_F(\alpha,\beta)$ one finds

$$\begin{aligned}
&\max_{\lambda\in\mathbb{R}^{J\times R}}\left[\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{(\lambda,\lambda^0)}\right)\Delta\xi_{\alpha,\beta}\right]\\
&=\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{\lambda^0}\right)\Delta\xi_{\alpha,\beta}+\max_{\lambda\in\mathbb{R}^{J\times R}}\left[\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes M_{\lambda^0}P_\lambda M_{\lambda^0}\right)\Delta\xi_{\alpha,\beta}\right]\\
&=\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{\lambda^0}\right)\Delta\xi_{\alpha,\beta}+\max_{\lambda\in\mathbb{R}^{J\times R}}\mathrm{Tr}\left[(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)'M_{\lambda^0}P_\lambda M_{\lambda^0}(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)\right]\\
&=\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{\lambda^0}\right)\Delta\xi_{\alpha,\beta}+\sum_{r=1}^{R}\mu_r\left[(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)'M_{\lambda^0}(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)\right].
\end{aligned}$$

In the first step we used $P_{(\lambda,\lambda^0)}=P_{\lambda^0}+M_{\lambda^0}P_{M_{\lambda^0}\lambda}M_{\lambda^0}$. The optimal value of λ in the second line always satisfies $\lambda=M_{\lambda^0}\lambda$, so we could write P_λ instead of $P_{M_{\lambda^0}\lambda}$. In the second step we plugged in the definition of $\Delta\xi_{\alpha,\beta}$. In the final step we used the characterization of the eigenvalues in terms of a maximization problem, and the fact that the non-zero eigenvalues of the matrices $(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)'M_{\lambda^0}(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)$ and $M_{\lambda^0}(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)'M_{\lambda^0}$ are identical.

Because of this, $\rho_F(\alpha,\beta)$ is equal to

$$\begin{aligned}
\rho_{\mathrm{F}}(\alpha,\beta)=&\frac{\Delta\xi'_{\alpha,\beta}\left(\mathbb{1}_T\otimes P_{\lambda^0}\right)\Delta\xi_{\alpha,\beta}}{\Delta\xi'_{\alpha,\beta}\Delta\xi_{\alpha,\beta}}\\
&+\frac{\sum_{r=1}^{R}\mu_r\left[(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)'M_{\lambda^0}(\delta(\alpha)-\delta(\alpha^0)-\beta\cdot X)\right]}{\Delta\xi'_{\alpha,\beta}\Delta\xi_{\alpha,\beta}}.
\end{aligned}$$

Thus, computation of ρ_F only involves the numerical calculation of the first R eigenvalues μ_r of a $T\times T$ matrix, which can be done very quickly even for relatively large values of T.

References

Ackerberg, D., Benkard, L., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics, Vol. 6A*. North-Holland.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255.

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1):1–14.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

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

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bajari, P., Fox, J., Kim, K., and Ryan, S. (2011). A simple nonparametric estimator for the distribution of random coefficients. *Quantitative Economics*, 2:381–418.

Bajari, P., Fox, J., Kim, K., and Ryan, S. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, 166:204–212.

Berry, S., Gandhi, A., and Haile, P. (2013). Connected Substitutes and Invertibility of Demand. *Econometrica*, 81:2087–2111.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S., Linton, O. B., and Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Berry, S. T. and Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797.

Besanko, D. and Doraszelski, U. (2004). Capacity dynamics and endogenous asymmetries in firm size. *RAND Journal of Economics*, 35:23–49.

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.

Chiappori, P. and Komunjer, I. (2009). On the nonparametric identification of multiple choice models. Manuscript, Columbia University.

Dube, J. P., Fox, J., and Su, C. (2012). Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation. *Econometrica*, 80:2231–2267.

Esteban, S. and Shum, M. (2007). Durable Goods Oligopoly with Secondary Markets: the Case of Automobiles. *RAND Journal of Economics*, 38:332–354.

Gandhi, A., Kim, K., and Petrin, A. (2010). Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error. Manuscript.

Gandhi, A., Lu, Z., and Shi, X. (2013). Estimating Demand for Differentiated Products with Error in Market Shares. Working paper, University of Wisconsin.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both "n" and "T" are large. *Econometrica*, 70(4):1639–1657.

Hahn, J. and Kuersteiner, G. (2004). Bias reduction for dynamic nonlinear panel models with fixed effects. Manuscript.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Harding, M. (2007). Structural estimation of high-dimensional factor models. Manuscript.

Harding, M. and Hausman, J. (2007). Using a Laplace Approximation to Estimate the Random Coefficients Logit Model By Nonlinear Least Squares. *International Economic Review*, 48:1311–1328.

Hausman, J. (1997). Valuation of new goods under perfect and imperfect competition. In *The Economics of New Goods*. University of Chicago Press.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6):1371–95.

Knittel, C. and Metaxoglou, K. (2014). Estimation of random coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics*, 96:34–59.

Matzkin, R. (2013). Nonparametric identification in structural economic models. In *Annual Review of Economics*, volume 5.

Media Dynamics, Inc. (1997). *TV Dimensions*. Media Dynamics, Inc. annual publication.

Moon, H. and Weidner, M. (2015a). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. *Forthcoming in Econometric Theory*.

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

Nevo, A. (2001). Measuring market power in the ready-to-eat cereals industry. *Econometrica*, 69:307–342.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

Petrin, A. (2002). Quantifying the benefits of new products: the case of the minivan. *Journal of Political Economy*, 110:705–729.

Villas-Boas, J. and Winer, R. (1999). Endogeneity in brand choice models. *Management Science*, 45:1324–1338.