

# 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 CWP08/12

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

Hyungsik Roger Moon<sup>‡</sup>    Matthew Shum<sup>§</sup>    Martin Weidner<sup>¶</sup>

First draft: October 2009; This draft: March 19, 2012

## 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 application 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, UC Davis, UC Irvine, UCLA, Univ. of Chicago Booth School, Univ. of Michigan, the 2009 California Econometrics Conference and the 2010 Econometric Society World Congress for helpful comments. Han Hong, Sung Jae Jun and Jinyong Hahn provided very helpful discussions. Moon acknowledges the NSF for financial support via SES 0920903.

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

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

<sup>¶</sup>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 modelling 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 Akerberg et. al. (2007)).

Taking a cue from recent developments in panel data econometrics (eg. Bai and Ng (2006), Bai (2009), and Moon and Weidner (2010a; 2010b)), 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. 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.

Our model represents one of the first applications of fixed effect factor modelling in panel data, which heretofore have mainly been considered in a linear setting with exogenous regressors, to nonlinear model with endogenous regressors. Relative to the existing panel factor literature (for instance, Bai (2009), and Moon and Weidner (2010a; 2010b)) that assume a linear regression with exogenous regressors, our nonlinear model poses identification and estimation challenges. For example, the usual principal components approach for linear factor models 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. To our knowledge, the current paper presents the first application of some recent developments in the econometrics of long panels to the workhorse BLP model in empirical IO.

To overcome these challenges, we propose an alternative identification and estimation scheme which we call the *Least Squares-Minimum Distance (LS-MD)* method.<sup>1</sup> 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. 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 an asymptotic where both the number of products and the number of markets goes to infinity. In practice, the estimator is simple and straightforward to compute. Our Monte Carlo simulations demonstrate that it has 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 (2008), Dube, Fox and Su (2012), Harding and Hausman (2007), Bajari, Fox, Kim and Ryan (2011), and Gandhi, Kim and Petrin (2010).

We implement 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 introduce the LS-MD estimation method. In Section 4 we discuss how to identify the model when valid instruments are available. 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 asymptotics and provide all technical derivations and proofs of the 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

---

<sup>1</sup>Recently, Chernozhukov and Hansen (2006) used a similar two stage estimation method for a class of instrumental quantile regressions.

$m \times n$  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  $P_A = A(A'A)^{-1}A'$  and  $M_A = \mathbb{1}_m - A(A'A)^{-1}A'$ , where  $\mathbb{1}_m$  is the  $m \times m$  identity matrix, and  $(A'A)^{-1}$  denotes some generalized inverse if  $A$  is not of full column rank. The vectorization of a  $m \times n$  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  $\nabla$  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<sup>2</sup>  $t$  is given by

$$u_{ijt} = \delta_{jt}^0 + \epsilon_{ijt} + X'_{jt} v_i, \quad (2.1)$$

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

$$\delta_{jt}^0 = X'_{jt} \beta^0 + \xi_{jt}^0, \quad (2.2)$$

where  $X_{jt} = (X_{1,jt}, \dots, X_{K,jt})'$  is a vector of  $K$  observed product characteristics (including price), and  $\beta^0 = (\beta_1^0, \dots, \beta_K^0)'$  is the corresponding vector of coefficients. Following BLP,  $\xi_{jt}^0$  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:

$$\xi_{jt}^0 = \lambda_j^{0'} f_t^0 + e_{jt}, \quad (2.3)$$

where  $\lambda_j^0 = (\lambda_{1j}^0, \dots, \lambda_{Rj}^0)'$  is a vector of factor loadings corresponding to the  $R$  factors<sup>3</sup>  $f_t^0 = (f_{1t}^0, \dots, f_{Rt}^0)'$ , and  $e_{jt}$  is a product and market specific error term. Here  $\lambda_j^{0'} f_t^0$

<sup>2</sup>The  $t$  subscript can also denote different time periods.

<sup>3</sup>Depending on the specific application one has in mind one may have different interpretations for  $\lambda_j$  and  $f_t$ . For example, in the case of national brands sold in different markets it seems more natural to interpret  $\lambda_j$  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  $\lambda_j$  as factor loadings throughout the whole paper, which is the typical naming convention in applications where  $t$  refers to time.

represent interactive fixed effects, in that both the factors  $f_t^0$  and factor loadings  $\lambda_j^0$  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.<sup>4</sup> The superscript zero indicates the true parameters, and objects evaluated at the true parameters.

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).<sup>5</sup> 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 (eg. 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  $e_{jt}$ , i.e.  $X_{jt}$  and  $e_{jt}$  are uncorrelated for given  $(j, t)$ . This assumption, however, is only made for ease of exposition, and in both section 3.1 below and the empirical application, 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  $\xi_{jt}^0$ , 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.

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

---

<sup>4</sup>Known  $R$  is also assumed in Bai (2009) and Moon and Weidner (2010a) 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. In pure factor models consistent inference procedures on the number of factors are known, e.g. Bai and Ng (2002), Onatski (2010), and Harding (2007).

<sup>5</sup>cf. *TV Dimensions* (1997).

of nationwide demand shock, where the  $\lambda_j$  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  $\lambda_j$  and  $f_t$ , the use of Hausman instruments in our model may be (at least partially) immune to the aforementioned criticism.

We assume that the distributions of  $\epsilon = (\epsilon_{ijt})$  and  $v = (v_i)$  are mutually independent, and are also independent of  $X = (X_{jt})$  and  $\xi^0 = (\xi_{jt}^0)$ . We also assume that  $\epsilon_{ijt}$  follows a marginal type I extreme value distribution iid across  $i$  and  $j$  (but not necessarily independent across  $t$ ).<sup>6</sup> For given preferences  $v_i$  and  $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})$ , the probability of agent  $i$  to choose product  $j$  in market  $t$  then takes the multinomial logit form:

$$\pi_{jt}(\delta_t, X_t, v_i) = \frac{\exp(\delta_{jt} + X'_{jt}v_i)}{1 + \sum_{l=1}^J \exp(\delta_{lt} + X'_{lt}v_i)}, \quad (2.4)$$

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

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

where  $G_{\alpha^0}(v)$  is the known distribution of consumer tastes  $v_i$  over the product characteristic, and  $\alpha^0$  is a  $L \times 1$  vector of parameters of this distribution. The most often used specification in this literature is to assume that the random coefficients are jointly multivariate normal distributed, corresponding to the assumptions that  $v \sim \mathcal{N}(0, \Sigma^0)$ , where  $\Sigma^0$  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  $\Sigma^0$  are only non-zero for these regressors), and  $\alpha^0$  consists of the independent parameters in  $\Sigma^0$ .<sup>7</sup>

The observables in this model are the market shares  $s_{jt}^0$  and the regressors  $X_{jt}$ .<sup>8</sup> In addition, we need  $M$  instruments  $Z_{jt} = (Z_{1,jt}, \dots, Z_{M,jt})'$  in order to estimate the parameters  $\alpha$ , with  $M \geq L$ . Note that these additional instruments are also needed in the usual BLP estimation procedure, even in the absence of the factor structure. There, even if all the  $X$ 's were exogenous with respect to  $\xi_{j,t}^0$ , instruments analogous to the  $Z$ 's in

---

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

<sup>7</sup>We focus in this paper on the case where the functional form of the distribution function  $G_{\alpha}$  is known by the researcher. Recent papers have addressed estimation when this is not known; eg. Bajari, Fox, Kim and Ryan (2011), (2012).

<sup>8</sup>In the present paper we assume that the true market shares  $s_{jt} = s_{jt}(\delta_t^0)$  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  $\xi^0$ .

our model would still be required to identify the covariance parameters in the random coefficients distribution.<sup>9</sup> In the presence of the unobserved factor structure, it is difficult to identify regression parameters of regressors that have a factor structure themselves, which includes product invariant and time invariant regressors, and all other regressors that when considered as a  $J \times T$  matrix have low rank.<sup>10</sup>

The unknown parameters are  $\alpha^0$ ,  $\beta^0$ ,  $\lambda^0$ , and  $f^0$ . Of these, the important parameters to estimate are  $\beta^0$  and  $\alpha^0$ , in terms of which we can calculate the ultimate objects of interest (such as price elasticities). The factors and factor loadings  $\lambda^0$  and  $f^0$  are not directly of interest, and are treated as nuisance parameters.

The existing literature on demand estimation usually considers asymptotics with either  $J$  large and  $T$  fixed, or  $T$  large and  $J$  fixed. Under these standard asymptotics, the estimation of the nuisance parameters  $\lambda^0$  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, e.g. Hahn and Kuersteiner (2002; 2004) and Hahn and Newey (2004), we handle this problem by considering asymptotics where both  $J$  and  $T$  become large. Under this alternative asymptotic, the incidental parameter problem is transformed into the issue of 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 asymptotics provides a good approximation of the properties of our estimator at finite sample sizes, as long as  $J$  and  $T$  are moderately large.

### 3 LS-MD Estimator

Following BLP, we assume in the rest of this paper that equation (2.5) is invertible, i.e. for each market  $t$  the mean utilities  $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})$  are unique functions of  $\alpha$ , the market

---

<sup>9</sup>The need for instruments even in the exogenous regressors case stems from an order condition. Namely, the dimension of the exogenous regressors  $X$  is the same as the dimension of  $\beta$ , but we need to estimate the parameters  $(\alpha, \beta)$  with dimensionality exceeding that of  $X$ . If no additional instruments are available this violates an order condition.

<sup>10</sup> 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 (2010a) for a further discussion of this topic.



shares  $s_t = (s_{1t}, \dots, s_{Jt})$ , and the regressors  $X_t = (X_{1t}, \dots, X_{Jt})$ .<sup>11</sup> We denote these functions by  $\delta_{jt}(\alpha, s_t, X_t)$ . We have

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

If  $\delta_{jt}^0$  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 e.g. Holtz-Eakin, Newey and Rosen (1988), and Ahn, Lee, Schmidt (2001), and for  $J, T \rightarrow \infty$  asymptotics in Bai (2009), and Moon and Weidner (2010a; 2010b).

The computational challenge in estimating the model (3.1) lies in accommodating both the model parameters  $(\alpha, \beta)$ , which in the existing literature has mainly been done in a GMM framework, as well as the nuisance elements  $\lambda_j, f_t$ , 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 (2010a; 2010b)). Our estimation procedure combines both the GMM approach to demand estimation and the least squares (or QMLE) approach to the interactive fixed effect model.

**Definition:** the *least squares-minimum distance (LS-MD)* estimators for  $\alpha$  and  $\beta$  are defined by

Step 1 (least squares): for given  $\alpha$  let

$$\begin{aligned} \delta(\alpha) &= \delta(\alpha, s, X), \\ (\tilde{\beta}_\alpha, \tilde{\gamma}_\alpha, \tilde{\lambda}_\alpha, \tilde{f}_\alpha) &= \underset{\{\beta, \gamma, \lambda, f\}}{\operatorname{argmin}} \|\delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda f'\|_F^2, \end{aligned}$$

Step 2 (minimum distance):

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

Step 3 (least squares):

$$\begin{aligned} \delta(\hat{\alpha}) &= \delta(\hat{\alpha}, s, X), \\ (\hat{\beta}, \hat{\lambda}, \hat{f}) &= \underset{\{\beta, \lambda, f\}}{\operatorname{argmin}} \|\delta(\hat{\alpha}) - \beta \cdot X - \lambda f'\|_F^2. \end{aligned} \quad (3.2)$$

Here,  $\beta \in \mathbb{R}^K$ ,  $\delta(\alpha, s, X)$ ,  $X_k$  and  $Z_m$  are  $J \times T$  matrices,  $\lambda$  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  $\alpha$ , we defined  $\beta \cdot X = \sum_{k=1}^K \beta_k X_k$  and  $\gamma \cdot Z = \sum_{m=1}^M \gamma_m Z_m$ , and  $\|\cdot\|_F$  refers to the Frobenius (least squares) norm that was defined in the introduction.

---

<sup>11</sup>Gandhi (2008) shows this result under general conditions, and Berry and Haile (2009) and Chiappori and Komunjer (2009) utilize this inverse mapping in their nonparametric identification results.

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  $\alpha$ . Thus, we first estimate  $\beta$ ,  $\lambda$ ,  $f$ , and the instrument coefficients  $\gamma$  by least squares for fixed  $\alpha$ , and subsequently we estimate  $\alpha$  by minimizing the norm of  $\tilde{\gamma}_\alpha$  with respect to  $\alpha$ .

Step 3 in (3.2), which defines  $\hat{\beta}$ , is just a repetition of step 1, but with  $\alpha = \hat{\alpha}$  and  $\gamma = 0$ . One could also use the step 1 estimator  $\tilde{\beta}_{\hat{\alpha}}$  to estimate  $\beta$ . Under the assumptions for consistency of  $(\hat{\alpha}, \hat{\beta})$  presented below, this alternative estimator is also consistent for  $\beta^0$ . However, in general  $\tilde{\beta}_{\hat{\alpha}}$  has a larger variance than  $\hat{\beta}$ , since irrelevant regressors are included in the estimation of  $\tilde{\beta}_{\hat{\alpha}}$ .

For given  $\alpha$ ,  $\beta$  and  $\gamma$  the optimal factors and factor loadings in the least squares problems in step 1 (and step 3) of (3.2) turn out to be the principal components estimators for  $\lambda$  and  $f$ . These incidental parameters can therefore be concentrated out easily, and the remaining objective function for  $\beta$  and  $\gamma$  turns out to be given by an eigenvalue problem (see e.g. Moon and Weidner (2010a; 2010b) for details), namely

$$\left( \tilde{\beta}_\alpha, \tilde{\gamma}_\alpha \right) = \underset{\{\beta, \gamma\}}{\operatorname{argmin}} \sum_{t=R+1}^T \mu_t [(\delta(\alpha) - \beta \cdot X - \gamma \cdot Z)' (\delta(\alpha) - \beta \cdot X - \gamma \cdot Z)] . \quad (3.3)$$

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  $\beta$  and  $\gamma$ , not over  $\lambda$  and  $f$ .

The step 1 optimization problem in (3.2) 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 (2010a; 2010b) that (under their assumptions)  $\hat{\beta}_{\alpha^0}$  is  $\sqrt{JT}$ -consistent for  $\beta^0$  and asymptotically normal as  $J, T \rightarrow \infty$  with  $J/T \rightarrow \kappa^2$ ,  $0 < \kappa < \infty$ .

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  $\alpha$  is  $\hat{\alpha}_{2SLS} = (Y_2'P_ZY_2)^{-1}Y_2'P_Zy_1$ , where  $P_Z = Z(Z'Z)^{-1}Z'$ . In this set up, it is possible to show that  $\hat{\alpha}_{2SLS}$  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 regressors  $X$  and  $Z$  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  $\alpha$  gives  $\hat{\alpha}(W) = [Y_2'Z(Z'Z)^{-1}W(Z'Z)^{-1}Z'Y_2]^{-1}[Y_2'Z(Z'Z)^{-1}W(Z'Z)^{-1}Z'y_1]$ . Choosing  $W = Z'Z$  thus results in  $\hat{\alpha} = \hat{\alpha}(Z'Z) = \hat{\alpha}_{2SLS}$ . 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.<sup>12</sup>

### 3.1 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, e.g., when price  $p_{jt}$  is determined by  $\xi_{jt}$  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  $\beta^{\text{end}}$  denote the coefficients on these regressors, while  $\beta$  continues to denote the coefficients on the exogenous regressors. Correspondingly, we assume that  $M$ , the number of instruments, exceeds  $L + K_2$ .

**Definition:** the *least-squares minimum distance (LS-MD)* estimators for  $\alpha$  and  $\beta$  with endogenous regressors  $X^{\text{end}}$  is defined by:

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

$$\begin{aligned} \delta(\alpha) &= \delta(\alpha, s, X), \\ \left( \tilde{\beta}_{\alpha^{\text{end}}}, \tilde{\gamma}_{\alpha^{\text{end}}}, \tilde{\lambda}_{\alpha^{\text{end}}}, \tilde{f}_{\alpha^{\text{end}}} \right) &= \underset{\{\beta, \gamma, \lambda, f\}}{\operatorname{argmin}} \left\| \delta(\alpha) - \beta^{\text{end}} \cdot X^{\text{end}} - \beta \cdot X - \gamma \cdot Z - \lambda f' \right\|_F^2, \end{aligned}$$

step 2:

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

step 3:

$$\begin{aligned} \delta(\hat{\alpha}) &= \delta(\hat{\alpha}, s, X), \\ \left( \hat{\beta}, \hat{\lambda}, \hat{f} \right) &= \underset{\{\beta \in \mathbb{R}^K, \lambda, f\}}{\operatorname{argmin}} \left\| \delta(\hat{\alpha}) - \beta^{\text{end}} \cdot X^{\text{end}} - \beta \cdot X - \lambda f' \right\|_F^2, \end{aligned} \quad (3.4)$$

where  $\mathcal{B}_{\alpha}$  and  $\mathcal{B}_{\beta}^{\text{end}}$  are parameter sets for  $\alpha$  and  $\beta^{\text{end}}$ , and we use the notation  $\beta^{\text{end}} \cdot X^{\text{end}} = \sum_{k=1}^{K_2} \beta_k^{\text{end}} X_k^{\text{end}}$  and  $\beta \cdot X = \sum_{k=K_2+1}^K \beta_k X_k$ .

The difference between this estimator, and the previous one for which all the regressors were assumed exogenous, is that the estimation of  $\beta^{\text{end}}$ , the coefficients on the endogenous regressors  $\tilde{X}$ , has been moved to the second step. The structure of the estimation procedure

---

<sup>12</sup>Moreover, the presence of the factors makes it inappropriate to use the moment condition-based GMM approach proposed by BLP. We know of no way to handle the factors and factor loadings in a GMM moment condition setting such that the resulting estimator for  $\alpha$  and  $\beta$  is consistent.

in (3.4) is exactly equivalent to that of our original LS-MD estimator (3.2), only that  $\alpha$  is replaced by  $\alpha^{\text{end}}$ , and  $\delta(\alpha)$  is replaced by  $\delta(\alpha) - \beta^{\text{end}} \cdot X^{\text{end}}$ . Thus, all results below on the consistency, asymptotic distribution and bias correction of the LS-MD estimator (3.2) with only (sequentially) exogenous regressors directly generalize to the estimator (3.4) 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.  $R = 0$ ).

## 4 Identification

Given the non-linearity of the model, and the non-standard estimation procedure, questions regarding the adequacy of the LS-MD estimator to identify the model parameters of interest are naturally raised. In the following we provide conditions under which the parameters  $\alpha$  and  $\beta$  as well as the product  $\lambda f'$  are identified.<sup>13</sup> Since our focus is on estimating  $\alpha$  and  $\beta$ , we do not consider how to identify  $\lambda$  and  $f$  separately.

The identification scheme we use is based on an objective function. The identification argument turns out to be constructive, as it leads naturally to an argument regarding the consistency of the LS-MD parameter estimates for the true parameter values. Alternative identification schemes are possible, e.g. in Appendix E we provide an alternative identification result that requires different restrictions. In what follows, we first provide conditions under which the model parameters are identified (Section 4.1) and then discuss the required conditions (Section 4.2).

### 4.1 Identification Result

For the identification argument, we introduce a non-stochastic version of the LS-MD estimation problem. That is, we consider the following objective function

$$Q(\alpha, \beta, \gamma, \lambda, f) = \mathbb{E} \|\delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda f'\|_F^2,$$

which is the expected value of the first step objective function in (3.2), i.e. the expected sum of squares residuals in model (3.1) for a given set of parameters. Note that  $Q(\alpha, \beta, \gamma, \lambda, f)$ , and many other objects introduced below, depend on the panel dimensions  $J$  and  $T$ . We suppress this dependence, because no asymptotics is discussed in the current section, i.e. all statements in this section hold for given fixed values of  $J$  and  $T$ .

---

<sup>13</sup>The transformation  $\lambda \rightarrow \lambda S$  and  $f \rightarrow f S^{-1}$  gives observationally equivalent parameters for any non-degenerate  $R \times R$  matrix  $S$ .

Our goal is to identify the model parameters from the objective function  $Q(\alpha, \beta, \gamma, \lambda, f)$ . The distribution (or more precisely, the first and second moment structure) of a full  $J \times T$  panel of observables enters into the definition of  $Q(\alpha, \beta, \gamma, \lambda, f)$ , which means that 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. However, below we will employ an asymptotic where both  $J$  and  $T$  go to infinity and are thus treated relatively symmetrically, which makes the population distribution of a  $J \times T$  panel a natural starting point for identification in the present context.

In the LS-MD identification approach, a two-step procedure is performed. For fixed  $\alpha$ , we maximize  $Q(\alpha, \beta, \gamma, \lambda, f)$  with respect to  $\beta, \gamma, \lambda, f$ . Define<sup>14</sup>

$$\gamma(\alpha) \in \underset{\gamma}{\operatorname{argmin}} \min_{\beta, \lambda, f} Q(\alpha, \beta, \gamma, \lambda, f). \quad (4.1)$$

Here and in the rest of this section we always implicitly impose  $\alpha \in \mathcal{B}_\alpha$ ,  $\beta \in \mathbb{R}^K$ ,  $\gamma \in \mathbb{R}^M$ ,  $\lambda \in \mathbb{R}^{J \times R}$ , and  $f \in \mathbb{R}^{T \times R}$ . In the subsequent identification argument, we will first show that under appropriate assumptions  $Q(\alpha^0, \beta, \gamma, \lambda, f)$  is uniquely minimized at  $\beta = \beta^0$ ,  $\gamma = 0$ , and  $\lambda f' = \lambda^0 f^{0'}$ . Then, we show that  $\gamma(\alpha) = 0$  implies  $\alpha = \alpha^0$ .

To state our assumptions for identification, it is convenient to define the  $JT \times K$  matrix  $x$ , the  $JT \times M$  matrix  $z$ , and the  $JT \times 1$  vectors  $d(\alpha)$  and  $\Delta\xi_{\alpha, \beta}$  by

$$\begin{aligned} x_{.,k} &= \operatorname{vec}(X_k) , & z_{.,m} &= \operatorname{vec}(Z_m) , \\ d(\alpha) &= \operatorname{vec}(\delta(\alpha) - \delta(\alpha^0)) , & \Delta\xi_{\alpha, \beta} &= d(\alpha) - x(\beta - \beta^0). \end{aligned} \quad (4.2)$$

where  $k = 1, \dots, K$  and  $m = 1, \dots, M$ . Note that  $\Delta\xi_{\alpha, \beta}$  is simply the vectorized difference of the residual unobserved product characteristic at  $(\alpha, \beta)$  and  $(\alpha^0, \beta^0)$ , i.e.  $\Delta\xi_{\alpha, \beta} = \xi^{\operatorname{vec}}(\alpha, \beta) - \xi^{\operatorname{vec}}(\alpha^0, \beta^0)$ , where  $\xi^{\operatorname{vec}}(\alpha, \beta) = \operatorname{vec}(\delta(\alpha) - \beta \cdot X)$ . In the following the indices  $j$  and  $t$  run from 1 to  $J$  and  $t$  to  $T$ , respectively.

**Assumption ID (Assumptions for Identification).**

- (i) The second moments of  $\delta_{jt}(\alpha)$ ,  $X_{jt}$  and  $Z_{jt}$  exist for all  $\alpha$ , and all  $j, t$ .
- (ii)  $\mathbb{E}(e_{jt}) = 0$ .

---

<sup>14</sup>Assumption (i) below guarantees existence of  $Q(\alpha, \beta, \gamma, \lambda, f)$  for all values of the parameters. As discussed in the previous section, for given  $\alpha, \beta, \gamma$  the minimization of  $Q(\alpha, \beta, \gamma, \lambda, f)$  over  $\lambda$  and  $f$  is an eigenvalue problem, and existence of the minimizing  $\lambda$  and  $f$  is therefore guaranteed. Under Assumption (iii) we furthermore have  $\min_{\lambda, f} Q(\alpha, \beta, \gamma, \lambda, f) > g_1(\alpha) + g_2(\alpha)(\|\beta\| + \|\gamma\|) + b(\|\beta\|^2 + \|\gamma\|^2)$ , for  $b > 0$  and some  $g_1(\alpha)$  and  $g_2(\alpha)$ . This shows that the norm of the minimizing  $\beta$  and  $\gamma$  is bounded, and by again referring to the eigenvalue problem one finds that the same needs to be true for  $\lambda f'$ . Since  $Q(\alpha, \beta, \gamma, \lambda, f)$  is a continuous function, boundedness of the optimal  $\beta, \gamma, \lambda f'$  implies existence of  $\gamma(\alpha)$ , although it may not necessarily be unique.

- (iii)  $\mathbb{E}(X_{jt}e_{jt}) = 0$ ,  $\mathbb{E}(Z_{jt}e_{jt}) = 0$ , for all  $j$ ,  $t$ .
- (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  $\lambda \in \mathbb{R}^{J \times R}$ .
- (v) For all  $(\alpha, \beta) \neq (\alpha^0, \beta^0)$ , and all  $\lambda \in \mathbb{R}^{J \times R}$  we assume that
 
$$\mathbb{E}[\Delta \xi'_{\alpha, \beta}(x, z)] \mathbb{E}[(x, z)'(x, z)]^{-1} \mathbb{E}[(x, z)' \Delta \xi_{\alpha, \beta}] > \mathbb{E}[\Delta \xi'_{\alpha, \beta}(\mathbb{1}_T \otimes P_{(\lambda, \lambda^0)}) \Delta \xi_{\alpha, \beta}].$$

**Theorem 4.1 (Identification).**

- (i) Suppose that Assumption ID(i)-(iv) are satisfied. Then, the minima of the function  $Q(\alpha^0, \beta, \gamma, \lambda, f)$  over  $(\beta, \gamma, \lambda, f)$  satisfy  $\beta = \beta^0$ ,  $\gamma = 0$ , and  $\lambda f' = \lambda^0 f^{0'}$ .
- (ii) Suppose that Assumption ID(i)-(v) are satisfied. Then,  $\gamma(\alpha) = 0$  implies  $\alpha = \alpha^0$ .

The proof of Theorem 4.1 is given in the appendix. Note that the definition of  $\gamma(\alpha)$  in (4.1) corresponds to step 1 of the LS-MD estimation method in (3.2). The theorem implies that under the conditions of Assumption ID we have  $\gamma(\alpha) = 0$  if and only if  $\alpha = \alpha^0$ , i.e.  $\alpha^0$  is identified from the condition  $\gamma(\alpha^0) = 0$ , which motivates step 2 of the LS-MD estimation procedure. The first part of the theorem also implies that the function  $\min_{\lambda, f} Q(\alpha^0, \beta, 0, \lambda, f)$  is uniquely minimized at  $\beta = \beta^0$ , which corresponds to step 3 of the LS-MD estimation procedure. Thus, the statement of the theorem is closely tied to our estimation approach.

If the true parameter  $\alpha^0$  is known or given, then model (3.1) reduces to a linear regression model with interactive fixed effects. In this case, the first and third step of the LS-MD estimation method in (3.2) correspond exactly to the least squares estimation procedure in Bai (2009), and Moon and Weidner (2010a; 2010b). The first part of Theorem 4.1 can be shown by following the logic in these existing papers.<sup>15</sup>

## 4.2 Discussion of the Identification Conditions

In this section we discuss the conditions of the identification theorem. Notice that Theorem 4.1(i) is about identification of  $\beta^0$  and  $\lambda^0 f^{0'}$  when  $\alpha^0$  is already identified. This result requires the conditions (i) to (iv) of Assumption ID. These conditions are quite typical regularity conditions for identification of a linear regression model with a modification only required in condition (iv) to accommodate the interactive fixed effects. 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  $e_{jt}$ .

---

<sup>15</sup>These papers do not explicitly discuss identification, but provide all key ideas to show identification within their consistency proof.

Apart from the term  $M_{(\lambda, \lambda^0)}$ , Assumption ID( $v$ ) is a standard non-collinearity condition on the product characteristics and the instruments – which jointly appear as regressors in the first step of (3.2). 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  $\lambda^0$  and to any other possible factor loadings  $\lambda$ . A sufficient condition for this assumption is the 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.

The key additional assumption that we need for identification of  $\alpha^0$  in Theorem 4.1(ii) is Assumption ID( $v$ ). Note that  $\Delta \xi_{\alpha^0, \beta^0} = 0$ , i.e. 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  $\lambda^0$  and an arbitrary other factor loading  $\lambda$ . 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  $\lambda$  and  $\lambda^0$  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  $\forall (\alpha, \beta) \neq (\alpha^0, \beta^0)$ :

$$\mathbb{E} [\Delta \xi'_{\alpha, \beta}(x, z)] \mathbb{E} [(x, z)'(x, z)]^{-1} \mathbb{E} [(x, z)' \Delta \xi_{\alpha, \beta}] > 0. \quad (4.3)$$

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

$$\mathbb{E} [d(\alpha)'(x, z)] \mathbb{E} [(x, z)'(x, z)]^{-1} \mathbb{E} [(x, z)'d(\alpha)] > \mathbb{E} [d(\alpha)'x] \mathbb{E} (x'x)^{-1} \mathbb{E} [x'd(\alpha)]. \quad (4.4)$$

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.<sup>16</sup> For this reason, in our Monte Carlo simulations in section 6 below,

---

<sup>16</sup>This is a problem not only with our approach, but also with the estimators in BLP, and for Berry, Linton

we provide both analytical and numerical verifications for Assumption ID(v) for the specific setup there.

## 5 Consistency and Asymptotic Distribution

In this section we present our results on the properties of the LS-MD estimator  $\hat{\alpha}$  and  $\hat{\beta}$  defined in (3.2) under the asymptotics  $J, T \rightarrow \infty$ .

**Assumption 1 (Assumptions for Consistency).**

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

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

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

$$(iii) \quad \frac{1}{JT} \text{Tr}(X_k e') = o_p(1), \quad \frac{1}{JT} \text{Tr}(Z_m e') = o_p(1), \quad \text{for } k = 1, \dots, K \text{ and } m = 1, \dots, M.$$

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

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

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

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

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

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 4.2 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  $\alpha^0$ , 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

---

and Pakes (2004).



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  $t$ .<sup>17</sup> The assumption is therefore the analog of Assumption ID(ii).

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  $\Xi$  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.<sup>18</sup> 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 (2010a) for a further discussion.

Next, we present results on the limiting distribution of  $\hat{\alpha}$  and  $\hat{\beta}$ . This requires some additional notation. 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

$$x_{.,k}^{\lambda f} = \text{vec}(M_{\lambda^0} X_k M_{f^0}), \quad z_{.,m}^{\lambda f} = \text{vec}(M_{\lambda^0} Z_m M_{f^0}), \quad g_{.,l} = -\text{vec}(\nabla_l \delta(\alpha^0)), \quad (5.1)$$

where  $k = 1, \dots, K$ ,  $m = 1, \dots, M$ , and  $l = 1, \dots, L$ . 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  $\Omega$  as follows

$$G = \text{plim}_{J,T \rightarrow \infty} \frac{1}{JT} \begin{pmatrix} g' x^{\lambda f} & g' z^{\lambda f} \\ x^{\lambda f'} x^{\lambda f} & x^{\lambda f'} z^{\lambda f} \end{pmatrix}, \quad \Omega = \text{plim}_{J,T \rightarrow \infty} \frac{1}{JT} \begin{pmatrix} x^{\lambda f} & z^{\lambda f} \end{pmatrix}' \text{diag}(\Sigma_e^{\text{vec}}) \begin{pmatrix} x^{\lambda f} & z^{\lambda f} \end{pmatrix}, \quad (5.2)$$

---

<sup>17</sup>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 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 (2010a) for other examples.

<sup>18</sup>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 (2010a).

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

$$\begin{aligned} \mathcal{W} = & \text{plim}_{J, T \rightarrow \infty} \left[ \left( \begin{array}{cc} \left( \frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} & 0_{K \times M} \\ 0_{M \times K} & 0_{M \times M} \end{array} \right) + \left( \begin{array}{c} -(x^{\lambda f'} x^{\lambda f})^{-1} x^{\lambda f'} z^{\lambda f} \\ \mathbb{1}_M \end{array} \right) \right. \\ & \left. \times \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left( \begin{array}{c} -(x^{\lambda f'} x^{\lambda f})^{-1} x^{\lambda f'} z^{\lambda f} \\ \mathbb{1}_M \end{array} \right)' \right]. \end{aligned} \quad (5.3)$$

Existence of these probability limits is imposed by Assumption 3 in the appendix. 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 (2010a; 2010b) for the linear model, except for part (i) of Assumption 4, which demands that  $\delta(\alpha)$  can be linearly approximated around  $\alpha^0$  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  $\alpha^0$ .

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

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

with the formulas for  $B_0$ ,  $B_1$  and  $B_2$  given in the appendix B.1.

The proof of Theorem 5.2 is provided in the appendix. Analogous to the QMLE 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  $\kappa B_0$  is only present if regressors or instruments are pre-determined, i.e. 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 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  $\kappa B_2$  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  $\kappa$ , and thus are large when  $T$  is small compared to  $J$ , while the second bias terms is proportional to  $\kappa^{-1}$ , 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 (2010a).

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  $\hat{\alpha}$  and  $\hat{\beta}$  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. 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  $M_{\lambda^0}$  in the formulas for  $G$ ,  $\Omega$  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  $\lambda^0$  do not contribute to the asymptotic variance, i.e. do not help in the estimation of  $\hat{\alpha}$  and  $\hat{\beta}$ . 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  $(\alpha^0, \beta^0, \lambda^0, f^0)$  by the estimated parameters  $(\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{f})$ , the error term  $(e)$  by the residuals  $(\hat{e})$ , and population values by sample values it is easy to define estimators  $\hat{B}_0, \hat{B}_1, \hat{B}_2, \hat{G}, \hat{\Omega}$  and  $\hat{\mathcal{W}}$  for  $B_0, B_1, B_2, G, \Omega$  and  $\mathcal{W}$ . This is done explicitly in appendix B.3.

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

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

$$\begin{pmatrix} \hat{\alpha}^* \\ \hat{\beta}^* \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \frac{1}{T} \hat{B}_0 - \frac{1}{J} \hat{B}_1 - \frac{1}{T} \hat{B}_2. \quad (5.4)$$

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  $(\hat{G}\hat{\mathcal{W}}\hat{G}')^{-1}\hat{G}\hat{\mathcal{W}}\hat{\Omega}\hat{\mathcal{W}}\hat{G}'(\hat{G}\hat{\mathcal{W}}\hat{G}')^{-1}$ . These results allow inference on  $\alpha^0$  and  $\beta^0$ .

From the standard GMM analysis it is known 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  $\sigma_e^2$  we have  $\Omega = \sigma_e^2 \text{plim}_{J,T \rightarrow \infty} \frac{1}{JT} (x^{\lambda f}, z^{\lambda f})' (x^{\lambda f}, z^{\lambda f})$ , and in this case it is straightforward to show that the optimal  $\mathcal{W} = \sigma_e^2 \Omega^{-1}$  is attained by choosing

$$W_{JT} = \frac{1}{JT} z' M_{x^{\lambda f}} z . \quad (5.5)$$

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, i.e. a different alternative estimator could theoretically have even lower asymptotic variance.

Note that the unobserved factor loading  $\lambda^0$  and factor  $f^0$  enter into the definition of  $x^{\lambda f}$  and thus also into the optimal  $W_{JT}$  in (5.5). A consistent estimator for the optimal  $W_{JT}$  can be obtained by estimating  $\lambda^0$  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

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

where  $\lambda_j^0$ ,  $f_t^0$ ,  $e_{jt}$  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,  $R_{\text{EST}}$ , 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  $\alpha$ . We choose  $Z_{jt} = p_{jt}^2$ , the squared price, i.e. 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  $\hat{\alpha}$  and  $\hat{\beta}$  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  $\xi_{jt}^0$  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  $\hat{\alpha}$  and  $\hat{\beta}$  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  $J = T = 80$  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  $R_{\text{EST}} = 1$ , i.e. overestimating the number of factors does not affect the estimation quality much in our simulation, at least as long as  $R_{\text{EST}}$  is small relative to the sample size  $J, T$ . 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 (2010b) for the linear model with interactive fixed effects.

We have chosen a data generating process for our simulation where regressors and

J,T	statistics	$R_{\text{EST}} = 0$		$R_{\text{EST}} = 1$		$R_{\text{EST}} = 2$	
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
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  $(\hat{\alpha}, \hat{\beta})$ . The true number of factors in the process is  $R = 1$ , but we use  $R_{\text{EST}} = 0, 1$ , and 2 in the estimation.

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  $\hat{\alpha}$  and  $\hat{\beta}$ , which is consistent with the results in Table 1. The simulation results for the bias corrected estimators  $\hat{\alpha}^*$  and  $\hat{\beta}^*$  are reported in Table 3 in the appendix, but there is virtually no effect from bias correction here, i.e. the results in Table 1 and Table 3 are almost identical.

## 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

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

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

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

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  $\beta$  negative and large (in absolute value) relative to  $\alpha^2$ . Then  $\delta_{jt}$  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 above 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

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

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  $\alpha$ . 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  $Y_{it} = \beta p_{jt} + \alpha^2 p_{jt}^2 + \lambda'_j f_t + e_{jt}$ . Consistency of this least squared estimator for  $\beta$  and  $\alpha^2$  in the presence of the parameters  $\lambda_j$  and  $f_t$  is discussed in Bai (2009) and Moon and Weidner (2010a).

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  $\nu_{jt}$  in that setup one typically finds values much larger than one. Therefore, we next confirm by numerical methods that  $p_{jt}^2$  is also 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

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

$\rho_{IV}(\alpha, \beta)$  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.  $\rho_F(\alpha, \beta)$  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. Note that  $\rho_{IV}(\alpha, \beta)$  and  $\rho_F(\alpha, \beta)$  take values between 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  $\rho_{IV}(\alpha, \beta)$ ,  $\rho_F(\alpha, \beta)$  and  $\Delta \rho(\alpha, \beta)$  as a function of  $\alpha$  and  $\beta$  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.<sup>19</sup> Although the data generating process only contains one factors, we used  $R = 2$  factors in the calculation of  $\rho_F(\alpha, \beta)$  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  $\rho_F(\alpha, \beta)$  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  $\rho_{IV}(\alpha, \beta)$  varies between 0.69 and 1.00,  $\rho_F(\alpha, \beta)$  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.

Note that the variation in  $\Delta \rho(\alpha, \beta)$  in this example is mostly driven by the variation in  $\rho_F(\alpha, \beta)$ , since  $\rho_{IV}(\alpha, \beta)$  for the most part is quite close to one, i.e. the explanatory power of the instruments and regressors is very large. Note that the analytical approximation above showed that for  $\nu_{jt} \ll 1$  the regressor  $p_{jt}$  and the instrument  $p_{jt}^2$  perfectly predict  $\Delta \xi_{\alpha, \beta}$ , i.e.  $\rho_{IV}(\alpha, \beta) \approx 1$  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.

---

<sup>19</sup>The appendix contains additional details on the numerical calculation of  $\rho_F(\alpha, \beta)$ .



## 7 Empirical application: demand for new automobiles, 1973-1988

As an illustration of our procedure, we estimate an aggregate random coefficients logit model of demand for new automobiles, modelled 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.<sup>20</sup>

For this empirical application, we use the same data as was used in BLP (1995), which are new automobile sales from 1971-1990.<sup>21</sup> 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.<sup>22</sup> Furthermore, we also reduce 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  $10^{-4}$  inch<sup>2</sup>.

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  $R = 0$ .<sup>23</sup>

---

<sup>20</sup>In such a setting, where we have a single national market evolving over time, we can interpret  $\lambda_j$  as (un-observed) 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.

<sup>21</sup>The data are available on the webpage of James Levinsohn.

<sup>22</sup>This resembles the treatment in Esteban and Shum’s (2007) empirical study of the new and used car markets, which likewise required a balanced panel.

<sup>23</sup>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

	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)
$\alpha$	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).  $\alpha$  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.

**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.  $\xi_{jt}$ ). 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  $\hat{B}_0$ ). 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 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  $\alpha$  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

---

(“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.

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  $\alpha$  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 computed the matrices of own- and cross-price elasticities for Specifications B (in Table (6)) and C (in Table (7)). In both these matrices, 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 application 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 considered 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.

We apply our estimator to US automobile demand. Significantly, we find that, once factors are included in the specification, the results assuming that price is exogenous or endogenous are quite similar, suggesting that the factors are indeed capturing much of the unobservable product and time effects leading to price endogeneity.

The model in this paper is, to our knowledge, the first application of factor-modelling 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-modelling 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

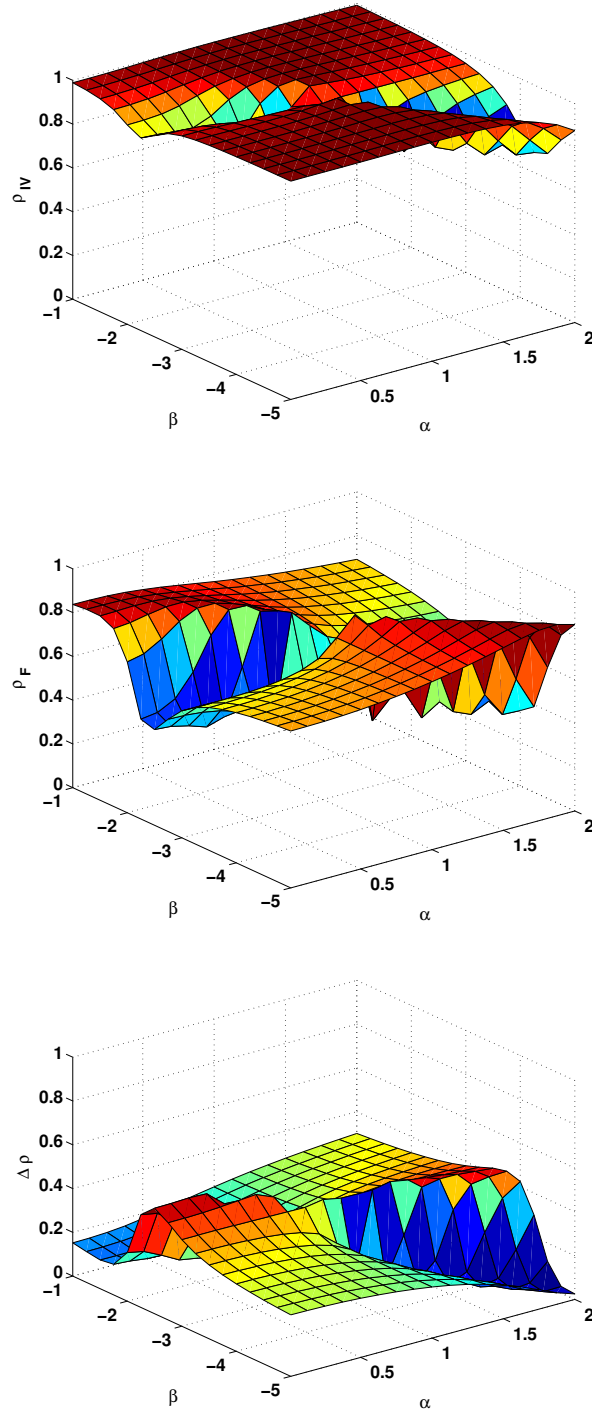


Figure 1: For one draw of the data generating process used in the Monte Carlo design with  $J = T = 80$  we plot  $\rho_{IV}(\alpha, \beta)$ ,  $\rho_F(\alpha, \beta)$  and  $\Delta\rho(\alpha, \beta)$  defined in (6.5) as a function of  $\alpha$  and  $\beta$ . The number of factors used in the calculation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

J,T	statistics	$R_{\text{EST}} = 0$		$R_{\text{EST}} = 1$		$R_{\text{EST}} = 2$	
		$\hat{\alpha}^*$	$\hat{\beta}^*$	$\hat{\alpha}^*$	$\hat{\beta}^*$	$\hat{\alpha}^*$	$\hat{\beta}^*$
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
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
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

Table 3: 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 bias corrected LS-MD estimator ( $\hat{\alpha}^*, \hat{\beta}^*$ ). The true number of factors in the process is  $R = 1$ , but we use  $R_{\text{EST}} = 0, 1$ , and 2 in the estimation.

	Specifications:			
	A: $R = 1$		B: $R = 1$	
	exogenous p		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)
$\alpha$	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 1	OD s	OD 1	PT s	PT 1	BK	CD	FD s	FD 1	MC s	MC 1	LC	PL s	PL 1	DG s	DG 1	TY	VWDT/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 1	1.50-34.54	0.72	2.02	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 1	0.98	0.64	0.65-35.80	0.59	0.13	3.37	2.09	0.64	4.34	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	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 1	2.17	0.81	0.68	1.55	0.98-26.85	2.53	0.06	1.40	1.97	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	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	0.00	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	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 1	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	0.90	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	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 1	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	0.20	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	0.00	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	0.21	0.49	0.30	1.66	0.63	1.67	0.95	0.16	4.42
PL 1	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	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 1	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	0.31	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 1	OD s	OD 1	PT s	PT 1	BK	CD	FD s	FD 1	MC s	MC 1	LC	PL s	PL 1	DG s	DG 1	TY	VWDT/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 1	0.43-15.20	0.49	0.53	0.45	0.41	0.53	2.46	0.43	0.60	0.46	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 1	0.38	0.45	0.48-16.57	0.41	0.35	0.53	3.40	0.38	0.63	0.44	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	0.44	0.45	0.44	0.47	0.45	0.44	0.58
PT 1	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	0.44	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	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	0.07	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	0.46	0.47	0.52
FD 1	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	0.30	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 1	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	0.03	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	0.47	0.39
PL 1	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	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 1	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	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	0.46	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	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	0.44	0.41	0.61	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	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	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 Details for Theorems 5.2 and 5.3

### B.1 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  $\Sigma_k^{X,e}$ ,  $k = 1, \dots, K$ , and  $\Sigma_m^{Z,e}$ ,  $m = 1, \dots, M$ , be defined by

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

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

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

and we set  $b^{(x,i)} = (b_1^{(x,i)}, \dots, b_K^{(x,i)})'$  and  $b^{(z,i)} = (b_1^{(z,i)}, \dots, b_M^{(z,i)})'$ , 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

$$B_i = - (G W G')^{-1} G W \begin{pmatrix} b^{(x,i)} \\ b^{(z,i)} \end{pmatrix}, \quad (\text{B.3})$$

where  $i = 0, 1, 2$ .

### B.2 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 probability limits of  $\lambda^{0'}\lambda^0/J$  and  $f^{0'}f^0/T$  are finite and have full rank, i.e. (a)  $\text{plim}_{J,T \rightarrow \infty} (\lambda^{0'}\lambda^0/J) > 0$ , (b)  $\text{plim}_{J,T \rightarrow \infty} (f^{0'}f^0/T) > 0$ .*

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  $\hat{\lambda}$  and  $\hat{f}$  can pick up the correct factor loadings and factors.

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

**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

$$\begin{aligned} \frac{1}{\sqrt{JT}} \|\nabla_l \delta(\alpha^0)\|_F &= \mathcal{O}_p(1), & \text{for } l = 1, \dots, L, \\ \sup_{\{\alpha: \sqrt{J}\|\alpha - \alpha^0\| < c, \alpha \neq \alpha^0\}} \frac{\frac{1}{\sqrt{JT}} \|r^\Delta(\alpha)\|_F}{\|\alpha - \alpha^0\|} &= o_p(1), & \text{for all } c > 0. \end{aligned}$$

(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  $X_k^{\text{str}}$  and  $Z_m^{\text{str}}$  are strictly exogenous, i.e.  $X_{k,jt}^{\text{str}}$  and  $Z_{m,jt}^{\text{str}}$  are independent of  $e_{i\tau}$  for all  $j, i, t, \tau$ . The components  $X_k^{\text{weak}}$  and  $Z_m^{\text{weak}}$  are weakly exogenous, and we assume*

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

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

$$|c_{k,j\tau}| < \alpha^\tau, \quad |d_{m,j\tau}| < \alpha^\tau,$$

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 (2010a; 2010b), 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  $\sqrt{JT}$ -shrinking neighborhood of  $\alpha^0$  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.

### B.3 Bias and Variance Estimators

Here we present consistent estimators for the matrices  $G$ ,  $\Omega$ , 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  $\hat{\alpha}$  and  $\hat{\beta}$ , we can define the residuals

$$\hat{e} = \delta(\hat{\alpha}, s, X) - \sum_{k=1}^K \hat{\beta}_k X_k - \hat{\lambda} \hat{f}' . \quad (\text{B.4})$$

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

$$\hat{x}_{.,k}^{\lambda f} = \text{vec} \left( M_{\hat{\lambda}} X_k M_{\hat{f}} \right) , \quad \hat{z}_{.,m}^{\lambda f} = \text{vec} \left( M_{\hat{\lambda}} Z_m M_{\hat{f}} \right) , \quad \hat{g}_{.,l} = -\text{vec} \left( \nabla_l \delta(\hat{\alpha}) \right) , \quad (\text{B.5})$$

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

matrices  $\hat{\Sigma}_k^{X,e}$ ,  $k = 1, \dots, K$ , and  $\hat{\Sigma}_m^{Z,e}$ ,  $m = 1, \dots, M$ , are defined by

$$\begin{aligned}\hat{\Sigma}_{k,t\tau}^{X,e} &= \begin{cases} \frac{1}{J} \sum_{j=1}^J X_{k,jt} \hat{e}_{j\tau} & \text{for } 0 < t - \tau \leq h \\ 0 & \text{otherwise} \end{cases} \\ \hat{\Sigma}_{m,t\tau}^{Z,e} &= \begin{cases} \frac{1}{J} \sum_{j=1}^J Z_{m,jt} \hat{e}_{j\tau} & \text{for } 0 < t - \tau \leq h \\ 0 & \text{otherwise} \end{cases}\end{aligned}\quad (\text{B.6})$$

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

$$\begin{aligned}\hat{G} &= \frac{1}{JT} \begin{pmatrix} \hat{g}' \hat{x}^{\lambda f} & \hat{g}' \hat{z}^{\lambda f} \\ \hat{x}^{\lambda f'} \hat{x}^{\lambda f} & \hat{x}^{\lambda f'} \hat{z}^{\lambda f} \end{pmatrix}, \\ \hat{\Omega} &= \frac{1}{JT} \left( \hat{x}^{\lambda f}, \hat{z}^{\lambda f} \right)' \text{diag}(\hat{\Sigma}_e^{\text{vec}}) \left( \hat{x}^{\lambda f}, \hat{z}^{\lambda f} \right), \\ \hat{b}_k^{(x,0)} &= \text{Tr} \left( P_{\hat{f}} \hat{\Sigma}_k^{X,e} \right), \\ \hat{b}_k^{(x,1)} &= \text{Tr} \left[ \text{diag} \left( \hat{\Sigma}_e^{(1)} \right) M_{\hat{\lambda}} X_k \hat{f} (\hat{f}' \hat{f})^{-1} (\hat{\lambda}' \hat{\lambda})^{-1} \hat{\lambda}' \right], \\ \hat{b}_k^{(x,2)} &= \text{Tr} \left[ \text{diag} \left( \hat{\Sigma}_e^{(2)} \right) M_{\hat{f}} X'_k \hat{\lambda} (\hat{\lambda}' \hat{\lambda})^{-1} (\hat{f}' \hat{f})^{-1} \hat{f}' \right], \\ \hat{b}_m^{(z,0)} &= \text{Tr} \left( P_{\hat{f}} \hat{\Sigma}_m^{Z,e} \right), \\ \hat{b}_m^{(z,1)} &= \text{Tr} \left[ \text{diag} \left( \hat{\Sigma}_e^{(1)} \right) M_{\hat{\lambda}} Z_m \hat{f} (\hat{f}' \hat{f})^{-1} (\hat{\lambda}' \hat{\lambda})^{-1} \hat{\lambda}' \right], \\ \hat{b}_m^{(z,2)} &= \text{Tr} \left[ \text{diag} \left( \hat{\Sigma}_e^{(2)} \right) M_{\hat{f}} Z'_m \hat{\lambda} (\hat{\lambda}' \hat{\lambda})^{-1} (\hat{f}' \hat{f})^{-1} \hat{f}' \right],\end{aligned}\quad (\text{B.7})$$

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

$$\begin{aligned}\widehat{\mathcal{W}} &= \begin{pmatrix} \left( \frac{1}{JT} \hat{x}^{\lambda f'} \hat{x}^{\lambda f} \right)^{-1} & 0_{K \times M} \\ 0_{M \times K} & 0_{M \times M} \end{pmatrix} + \begin{pmatrix} -(\hat{x}^{\lambda f'} \hat{x}^{\lambda f})^{-1} \hat{x}^{\lambda f'} \hat{z}^{\lambda f} \\ \mathbb{1}_M \end{pmatrix} \left( \frac{1}{JT} \hat{z}^{\lambda f'} M_{\hat{x}^{\lambda f}} \hat{z}^{\lambda f} \right)^{-1} \\ &W_{JT} \left( \frac{1}{JT} \hat{z}^{\lambda f'} M_{\hat{x}^{\lambda f}} \hat{z}^{\lambda f} \right)^{-1} \begin{pmatrix} -(\hat{x}^{\lambda f'} \hat{x}^{\lambda f})^{-1} \hat{x}^{\lambda f'} \hat{z}^{\lambda f} \\ \mathbb{1}_M \end{pmatrix}'.\end{aligned}\quad (\text{B.8})$$

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

$$\hat{B}_i = - \left( \hat{G} \widehat{\mathcal{W}} \hat{G}' \right)^{-1} \hat{G} \widehat{\mathcal{W}} \begin{pmatrix} \hat{b}^{(x,i)} \\ \hat{b}^{(z,i)} \end{pmatrix}.\quad (\text{B.9})$$

The only subtlety here lies in the definition of  $\hat{\Sigma}_k^{X,e}$  and  $\hat{\Sigma}_m^{Z,e}$ , where we explicitly impose the constraint that  $\hat{\Sigma}_{k,t\tau}^{X,e} = \hat{\Sigma}_{m,t\tau}^{Z,e} = 0$  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 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  $\varepsilon$  by

$$x_{:,k}^f = \text{vec}(X_k M_{f^0}), \quad z_{:,m}^f = \text{vec}(Z_m M_{f^0}), \quad \varepsilon = \text{vec}(e),$$

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

### C.1 Proof of Identification

**Proof of Theorem 4.1.** # Proof of part (i): We show that the minima of the function  $Q(\alpha^0, \beta, \gamma, \lambda, f)$  over  $\beta, \gamma, \lambda, f$  satisfy  $\beta = \beta^0$ ,  $\gamma = 0$ , and  $\lambda f' = \lambda^0 f^{0'}$ . Using model (3.1) and Assumption ID(ii) and (iii) we find

$$\begin{aligned} Q(\alpha^0, \beta, \gamma, \lambda, f) &= \mathbb{E} \text{Tr} \{ [\delta(\alpha^0) - \beta \cdot X - \gamma \cdot Z - \lambda f']' [\delta(\alpha^0) - \beta \cdot X - \gamma \cdot Z - \lambda f'] \} \\ &= \mathbb{E} \text{Tr} \{ [(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0'} - \lambda f' + e]' [(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0'} - \lambda f' + e] \} \\ &= \mathbb{E} \text{Tr}(e'e) + \underbrace{\mathbb{E} \text{Tr} \{ [(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0'} - \lambda f']' [(\beta^0 - \beta) \cdot X - \gamma \cdot Z + \lambda^0 f^{0'} - \lambda f'] \}}_{=Q^*(\beta, \gamma, \lambda, f)}. \end{aligned} \tag{C.1}$$

Note that  $Q^*(\beta, \gamma, \lambda, f) \geq 0$  and that  $Q^*(\beta^0, 0, \lambda^0, f^0) = 0$ . Thus, the minimum value of  $Q(\alpha^0, \beta, \gamma, \lambda, f)$  equals  $\mathbb{E} \text{Tr}(e'e)$  and all parameters that minimize  $Q(\alpha^0, \beta, \gamma, \lambda, f)$  must satisfy  $Q^*(\beta, \gamma, \lambda, f) = 0$ .

We have for any  $\lambda$  and  $f$

$$\begin{aligned} Q^*(\beta, \gamma, \lambda, f) &\geq \mathbb{E} \text{Tr} \{ [(\beta^0 - \beta) \cdot X - \gamma \cdot Z]' M_{(\lambda, \lambda^0)} [(\beta^0 - \beta) \cdot X - \gamma \cdot Z] \} \\ &= [(\beta^0 - \beta)', \gamma'] \mathbb{E} [(x, z)' (\mathbb{1}_T \otimes M_{(\lambda, \lambda^0)})(x, z)] [(\beta^0 - \beta)', \gamma'] \\ &\geq b (\|\beta - \beta^0\|^2 + \|\gamma\|^2)^2, \end{aligned} \tag{C.2}$$

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

# Proof of part (ii): We show that  $\gamma(\alpha) = 0$  implies  $\alpha = \alpha^0$ . From part (i) we already know that  $\gamma(\alpha^0) = 0$ . The proof proceeds by contradiction. Assume that  $\gamma(\alpha) = 0$  for  $\alpha \neq \alpha^0$ . This implies that there exist  $\tilde{\beta}$ ,  $\tilde{\lambda}$  and  $\tilde{f}$  such that  $Q(\alpha, \tilde{\beta}, 0, \tilde{\lambda}, \tilde{f}) \leq \min_{\beta, \gamma, \lambda, f} Q(\alpha, \beta, \gamma, \lambda, f)$ . Using

model (3.1) and our assumptions we obtain the following lower bound for the lhs of this inequality

$$\begin{aligned}
Q(\alpha, \tilde{\beta}, 0, \tilde{\lambda}, \tilde{f}) &= \mathbb{E} \text{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} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X + \lambda^0 f^0 - \tilde{\lambda} \tilde{f}' + e \right)' \right. \\
&\quad \left. \left( \delta(\alpha) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X + \lambda^0 f^0 - \tilde{\lambda} \tilde{f}' + e \right) \right] \\
&= 2 \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X + \lambda^0 f^0 - \tilde{\lambda} \tilde{f}' \right)' \right. \\
&\quad \left. \left( \delta(\alpha) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X + \lambda^0 f^0 - \tilde{\lambda} \tilde{f}' \right) \right] \\
&\geq 2 \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X \right)' M_{(\tilde{\lambda}, \lambda^0)} \right. \\
&\quad \left. \left( \delta(\alpha) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X \right) \right] \\
&= 2 \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + \mathbb{E} \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} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + \mathbb{E} \left[ \Delta \xi'_{\alpha, \tilde{\beta}} \Delta \xi_{\alpha, \tilde{\beta}} \right] - \mathbb{E} \left[ \Delta \xi'_{\alpha, \tilde{\beta}} \left( \mathbb{1}_T \otimes P_{(\tilde{\lambda}, \lambda^0)} \right) \Delta \xi_{\alpha, \tilde{\beta}} \right]. \quad (\text{C.3})
\end{aligned}$$

Similarly, we obtain the following upper bound for the rhs of the above inequality

$$\begin{aligned}
\min_{\beta, \gamma, \lambda, f} Q(\alpha, \beta, \gamma, \lambda, f) &\leq \min_{\beta, \gamma} Q(\alpha, \beta, \gamma, \lambda^0, f^0) \\
&= \min_{\beta, \gamma} \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda^0 f^{0'} \right)' \left( \delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda^0 f^{0'} \right) \right] \\
&= \min_{\beta, \gamma} \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z + e \right)' \right. \\
&\quad \left. \left( \delta(\alpha) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z + e \right) \right] \\
&= 2 \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + \min_{\beta, \gamma} \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z \right)' \right. \\
&\quad \left. \left( \delta(\alpha) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z \right) \right] \\
&= 2 \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] \\
&\quad + \min_{\beta, \gamma} \mathbb{E} \left[ \left( \Delta \xi_{\alpha, \tilde{\beta}} - x(\beta - \tilde{\beta}) - z\gamma \right)' \left( \Delta \xi_{\alpha, \tilde{\beta}} - x(\beta - \tilde{\beta}) - z\gamma \right) \right] \\
&= 2 \mathbb{E} \text{Tr} \left[ \left( \delta(\alpha) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + \mathbb{E} \left[ \Delta \xi'_{\alpha, \tilde{\beta}} \Delta \xi_{\alpha, \tilde{\beta}} \right] \\
&\quad - \mathbb{E} \left[ \Delta \xi'_{\alpha, \tilde{\beta}}(x, z) \right] \mathbb{E} \left[ (x, z)'(x, z) \right]^{-1} \mathbb{E} \left[ (x, z)' \Delta \xi_{\alpha, \tilde{\beta}} \right]. \quad (\text{C.4})
\end{aligned}$$

Plugging these bounds in the original inequality we obtain

$$\mathbb{E} \left[ \Delta \xi'_{\alpha, \tilde{\beta}}(x, z) \right] \mathbb{E} \left[ (x, z)'(x, z) \right]^{-1} \mathbb{E} \left[ (x, z)' \Delta \xi_{\alpha, \tilde{\beta}} \right] \leq \mathbb{E} \left[ \Delta \xi'_{\alpha, \tilde{\beta}} \left( \mathbb{1}_T \otimes P_{(\tilde{\lambda}, \lambda^0)} \right) \Delta \xi_{\alpha, \tilde{\beta}} \right], \quad (\text{C.5})$$

which is a contradiction to Assumption ID(v). We have thus shown that  $\gamma(\alpha) = 0$  implies  $\alpha = \alpha^0$ . ■

## C.2 Proof of Consistency

**Proof of Theorem 5.1.** # Part 1: We show that for any consistent estimator  $\hat{\alpha}$  (not necessarily the LS-MD estimator) we have  $\tilde{\beta}_{\hat{\alpha}} = \beta^0 + o_p(1)$  and  $\tilde{\gamma}_{\hat{\alpha}} = o_p(1)$ . Thus, for this part of the proof

assume that  $\hat{\alpha} = \alpha^0 + o_p(1)$ . This part of the proof is a direct extension of the consistency proof in Moon and Weidner (2010b). We denote the least square objective function by  $Q_{JT}(\alpha, \beta, \gamma, \lambda, f) = \frac{1}{JT} \|\delta(\alpha) - \beta \cdot X - \gamma \cdot Z - \lambda f'\|_F^2$ . We first establish a lower bound on  $Q_{JT}(\hat{\alpha}, \beta, \gamma, \lambda, f)$ . We have for all  $\lambda, f$ :

$$\begin{aligned}
Q_{JT}(\hat{\alpha}, \beta, \gamma, \lambda, f) &= \frac{1}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f')' (\delta(\hat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f') \right] \\
&\geq \frac{1}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f')' M_{(\lambda, \lambda^0)} (\delta(\hat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda f') \right] \\
&= \frac{1}{JT} \text{Tr} \left[ ((\delta(\hat{\alpha}) - \delta(\alpha^0)) + e - (\beta - \beta^0) \cdot X - \gamma \cdot Z)' M_{(\lambda, \lambda^0)} \right. \\
&\quad \left. ((\delta(\hat{\alpha}) - \delta(\alpha^0)) + e - (\beta - \beta^0) \cdot X - \gamma \cdot Z) \right] \\
&\geq b \|\beta - \beta^0\|^2 + b \|\gamma\|^2 + o_p(\|\beta - \beta^0\| + \|\gamma - \gamma^0\|) + \frac{1}{JT} \text{Tr}(ee') + o_p(1).
\end{aligned} \tag{C.6}$$

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:

$$\begin{aligned}
&\frac{1}{JT} \text{Tr} \left[ ((\beta - \beta^0) \cdot X - \gamma \cdot Z)' M_{(\lambda, \lambda^0)} ((\beta - \beta^0) \cdot X - \gamma \cdot Z) \right] \\
&= (\beta', \gamma') \left[ \frac{1}{JT} (x, z)' (\mathbb{1}_T \otimes M_{(\lambda, \lambda^0)})(x, z) \right] (\beta', \gamma')' \\
&\geq b(\beta', \gamma')(\beta', \gamma')' = b \|\beta - \beta^0\|^2 + b \|\gamma\|^2, \\
&\left| \frac{1}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0))' M_{(\lambda, \lambda^0)} ((\beta - \beta^0) \cdot X) \right] \right| \\
&\leq \frac{1}{JT} \|\delta(\hat{\alpha}) - \delta(\alpha^0)\|_F \|M_{(\lambda, \lambda^0)} ((\beta - \beta^0) \cdot X)\|_F \\
&\leq \frac{1}{JT} \|\delta(\hat{\alpha}) - \delta(\alpha^0)\|_F \|(\beta - \beta^0) \cdot X\|_F \\
&= \mathcal{O}_p(1) \|\hat{\alpha} - \alpha^0\| \|\beta - \beta^0\| = o_p(\|\beta - \beta^0\|), \\
&\left| \frac{1}{JT} \text{Tr} \left[ e' M_{(\lambda, \lambda^0)} ((\beta - \beta^0) \cdot X) \right] \right| \\
&= \left| \frac{1}{JT} \text{Tr} \left[ e' ((\beta - \beta^0) \cdot X) \right] \right| + \left| \frac{1}{JT} \text{Tr} \left[ e' P_{(\lambda, \lambda^0)} ((\beta - \beta^0) \cdot X) \right] \right| \\
&\leq o_p(1) \|\beta - \beta^0\| + \frac{R}{JT} \|e\| \|(\beta - \beta^0) \cdot X\| \\
&\leq o_p(1) \|\beta - \beta^0\| + \frac{R}{JT} \|e\| \|(\beta - \beta^0) \cdot X\|_F = o_p(\|\beta - \beta^0\|).
\end{aligned} \tag{C.7}$$

See the supplementary material in Moon and Weidner (2010a) for further details regarding the algebra here. Applying the same methods, we also obtain

$$Q_{JT}(\hat{\alpha}, \beta^0, 0, \lambda^0, f^0) = \frac{1}{JT} \text{Tr}(ee') + o_p(1). \tag{C.8}$$

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  $\hat{\alpha}$  need to satisfy  $Q_{JT}(\hat{\alpha}, \tilde{\beta}_{\hat{\alpha}}, \tilde{\gamma}_{\hat{\alpha}}, \tilde{\lambda}_{\hat{\alpha}}, \tilde{f}_{\hat{\alpha}}) \leq Q_{JT}(\hat{\alpha}, \beta^0, 0, \lambda^0, f^0)$ . Using the above results thus gives

$$b \|\tilde{\beta}_{\hat{\alpha}} - \beta^0\|^2 + b \|\tilde{\gamma}_{\hat{\alpha}}\|^2 + o_p(\|\tilde{\beta}_{\hat{\alpha}} - \beta^0\| + \|\tilde{\gamma}_{\hat{\alpha}} - \gamma^0\|) + o_p(1) \leq 0. \tag{C.9}$$



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

# Part 2: Now, let  $\hat{\alpha}$  be the LS-MD estimator. We want to show that  $\hat{\alpha} - \alpha^0 = o_p(1)$ . 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  $\hat{\alpha}$  minimizes  $\tilde{\gamma}'_{\hat{\alpha}} W_{JT} \tilde{\gamma}_{\hat{\alpha}}$ , which implies that

$$\tilde{\gamma}'_{\hat{\alpha}} W_{JT} \tilde{\gamma}_{\hat{\alpha}} \leq \tilde{\gamma}'_{\alpha^0} W_{JT} \tilde{\gamma}_{\alpha^0} = o_p(1), \quad (\text{C.10})$$

and therefore  $\tilde{\gamma}_{\hat{\alpha}} = o_p(1)$ . 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  $Q_{JT}(\hat{\alpha}, \tilde{\beta}_{\hat{\alpha}}, \tilde{\gamma}_{\hat{\alpha}}, \tilde{\lambda}_{\hat{\alpha}}, \tilde{f}_{\hat{\alpha}})$ . In the rest of this proof we drop the subscript  $\hat{\alpha}$  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}(\hat{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\lambda}, \tilde{f}) &= \frac{1}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \tilde{\beta} \cdot X - \tilde{\gamma} \cdot Z - \tilde{\lambda} \tilde{f}' \right)' \left( \delta(\hat{\alpha}) - \tilde{\beta} \cdot X - \tilde{\gamma} \cdot Z - \tilde{\lambda} \tilde{f}' \right) \right] \\ &= \frac{1}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + \lambda^0 f^0 - \tilde{\lambda} \tilde{f}' + e \right)' \right. \\ &\quad \left. \left( \delta(\hat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + \lambda^0 f^0 - \tilde{\lambda} \tilde{f}' + e \right) \right] \\ &\geq \frac{1}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + e \right)' M_{(\tilde{\lambda}, \lambda^0)} \right. \\ &\quad \left. \left( \delta(\hat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X - \tilde{\gamma} \cdot Z + e \right) \right] \\ &= \frac{1}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X \right)' M_{(\tilde{\lambda}, \lambda^0)} \left( \delta(\hat{\alpha}) - \delta(\alpha^0) - (\tilde{\beta} - \beta^0) \cdot X \right) \right] \\ &\quad + \frac{2}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + o_p(\|\hat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1) \\ &= \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \left( \mathbb{1}_T \otimes M_{(\tilde{\lambda}, \lambda^0)} \right) \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] \\ &\quad + \frac{2}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + o_p(\|\hat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1) \\ &= \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] - \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \left( \mathbb{1}_T \otimes P_{(\tilde{\lambda}, \lambda^0)} \right) \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] \\ &\quad + \frac{2}{JT} \text{Tr} \left[ \left( \delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2} e \right)' e \right] + o_p(\|\hat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1). \end{aligned} \quad (\text{C.11})$$

The bounds used here are analogous to those in (C.7), and we again refer to the supplementary material in Moon and Weidner (2010a).

Similarly, we obtain the following upper bound

$$\begin{aligned}
Q_{JT}(\hat{\alpha}, \tilde{\beta}_{\hat{\alpha}}, \tilde{\gamma}_{\hat{\alpha}}, \tilde{\lambda}_{\hat{\alpha}}, \tilde{f}_{\hat{\alpha}}) &= \min_{\beta, \gamma, \lambda, f} Q_{JT}(\hat{\alpha}, \beta, \gamma, \lambda, f) \leq \min_{\beta, \gamma} Q_{JT}(\hat{\alpha}, \beta, \gamma, \lambda^0, f^0) \\
&= \min_{\beta, \gamma} \frac{1}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda^0 f^0)' (\delta(\hat{\alpha}) - \beta \cdot X - \gamma \cdot Z - \lambda^0 f^0) \right] \\
&= \min_{\beta, \gamma} \frac{1}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z + e)' \right. \\
&\quad \left. (\delta(\hat{\alpha}) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z + e) \right] \\
&= \frac{2}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2}e)' e \right] + \min_{\beta, \gamma} \frac{1}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z)' \right. \\
&\quad \left. (\delta(\hat{\alpha}) - \delta(\alpha^0) - (\beta - \beta^0) \cdot X - \gamma \cdot Z) \right] \\
&= \frac{2}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2}e)' e \right] \\
&\quad + \min_{\beta, \gamma} \frac{1}{JT} \left[ (\Delta \xi_{\hat{\alpha}, \tilde{\beta}} - x(\beta - \tilde{\beta}) - z\gamma)' (\Delta \xi_{\hat{\alpha}, \tilde{\beta}} - x(\beta - \tilde{\beta}) - z\gamma) \right] \\
&= \frac{2}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2}e)' e \right] + \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] \\
&\quad - \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}}(x, z) \right] [(x, z)'(x, z)]^{-1} [(x, z)' \Delta \xi_{\hat{\alpha}, \tilde{\beta}}]. \tag{C.12}
\end{aligned}$$

Combining this upper and lower bound we obtain

$$\begin{aligned}
&\frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}}(x, z) \right] [(x, z)'(x, z)]^{-1} [(x, z)' \Delta \xi_{\hat{\alpha}, \tilde{\beta}}] \\
&\quad - \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \left( \mathbf{1}_T \otimes P_{(\tilde{\lambda}, \lambda^0)} \right) \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] \leq o_p(\|\hat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1). \tag{C.13}
\end{aligned}$$

Using Assumption 1(v) we thus obtain

$$b\|\hat{\alpha} - \alpha^0\|^2 + b\|\tilde{\beta} - \beta^0\|^2 \leq o_p(\|\hat{\alpha} - \alpha^0\| + \|\tilde{\beta} - \beta^0\|) + o_p(1), \tag{C.14}$$

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

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

### C.3 Proof of Limiting Distribution

**Lemma C.1.** *Let Assumption 1 be satisfied and in addition let  $(JT)^{-1/2} \text{Tr}(eX'_k) = \mathcal{O}_p(1)$ , and  $(JT)^{-1/2} \text{Tr}(eZ'_m) = \mathcal{O}_p(1)$ . In the limit  $J, T \rightarrow \infty$  with  $J/T \rightarrow \kappa^2$ ,  $0 < \kappa < \infty$ , we then have  $\sqrt{J}(\hat{\alpha} - \alpha) = \mathcal{O}_p(1)$ .*

**Proof.** The proof is exactly analogous to the consistency proof. We know from Moon and Weidner (2010a; 2010b) that  $\sqrt{J}\tilde{\gamma}_{\alpha^0} = \mathcal{O}_p(1)$ . Applying the inequality (C.10) one thus finds  $\sqrt{J}\tilde{\gamma}_{\hat{\alpha}} = \mathcal{O}_p(1)$ . With the additional assumptions in the lemma one can strengthen the result in (C.11) as follows

$$\begin{aligned}
&Q_{JT}(\hat{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\lambda}, \tilde{f}) \\
&\geq \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] - \frac{1}{JT} \left[ \Delta \xi'_{\hat{\alpha}, \tilde{\beta}} \left( \mathbf{1}_T \otimes P_{(\tilde{\lambda}, \lambda^0)} \right) \Delta \xi_{\hat{\alpha}, \tilde{\beta}} \right] \\
&\quad + \frac{2}{JT} \text{Tr} \left[ (\delta(\hat{\alpha}) - \delta(\alpha^0) + \frac{1}{2}e)' e \right] + \mathcal{O}_p \left( \sqrt{J}\|\hat{\alpha} - \alpha^0\| + \sqrt{J}\|\tilde{\beta} - \beta^0\| \right) + \mathcal{O}_p(1/J). \tag{C.15}
\end{aligned}$$

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

**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 C.1 to conclude  $\sqrt{J}(\hat{\alpha} - \alpha) = \mathcal{O}_p(1)$ .

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  $\hat{\alpha} - \alpha^0 = o_p(1)$  and Assumption 1(i) we have  $\|\Psi(\hat{\alpha})\| = o_p(\sqrt{JT})$ , so that the results in Moon and Weidner (2010a; 2010b) guarantee  $\tilde{\beta}_{\hat{\alpha}} - \beta^0 = o_p(1)$  and  $\|\tilde{\gamma}_{\hat{\alpha}}\| = o_p(1)$ , which we already used in the consistency proof. Using  $\sqrt{J}(\hat{\alpha} - \alpha) = \mathcal{O}_p(1)$  and Assumption 4(i) we find  $\|\Psi(\hat{\alpha})\| = \mathcal{O}_p(\sqrt{J})$ , which allows us to truncate the asymptotic likelihood expansion derived in Moon and Weidner (2010a; 2010b) at an appropriate order. Namely, applying their results we have

$$\sqrt{JT} \begin{pmatrix} \tilde{\beta}_{\hat{\alpha}} - \beta^0 \\ \tilde{\gamma}_{\hat{\alpha}} \end{pmatrix} = V_{JT}^{-1} \begin{pmatrix} [C^{(1)}(X_k, \Psi(\alpha)) + C^{(2)}(X_k, \Psi(\alpha))]_{k=1, \dots, K} \\ [C^{(1)}(Z_m, \Psi(\alpha)) + C^{(2)}(Z_m, \Psi(\alpha))]_{m=1, \dots, M} \end{pmatrix} + r^{\text{QMLE}}(\alpha), \quad (\text{C.16})$$

where

$$\begin{aligned} V_{JT} &= \frac{1}{JT} \begin{pmatrix} [\text{Tr}(M_{f^0} X'_{k_1} M_{\lambda^0} X_{k_2})]_{k_1, k_2=1, \dots, K} & [\text{Tr}(M_{f^0} X'_k M_{\lambda^0} Z_m)]_{k=1, \dots, K; m=1, \dots, M} \\ [\text{Tr}(M_{f^0} Z'_m M_{\lambda^0} X_k)]_{m=1, \dots, M; k=1, \dots, K} & [\text{Tr}(M_{f^0} Z'_{m_1} M_{\lambda^0} Z_{m_2})]_{m_1, m_2=1, \dots, M} \end{pmatrix} \\ &= \frac{1}{JT} (x^{\lambda f}, z^{\lambda f})' (x^{\lambda f}, z^{\lambda f}), \end{aligned} \quad (\text{C.17})$$

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

$$\begin{aligned} C^{(1)}(\mathcal{X}, \Psi) &= \frac{1}{\sqrt{JT}} \text{Tr} [M_{f^0} \Psi' M_{\lambda^0} \mathcal{X}], \\ C^{(2)}(\mathcal{X}, \Psi) &= -\frac{1}{\sqrt{JT}} \left[ \text{Tr} (\Psi M_{f^0} \Psi' M_{\lambda^0} \mathcal{X} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'}) \right. \\ &\quad + \text{Tr} (\Psi' M_{\lambda^0} \Psi M_{f^0} \mathcal{X}' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'}) \\ &\quad \left. + \text{Tr} (\Psi' M_{\lambda^0} \mathcal{X} M_{f^0} \Psi' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'}) \right], \end{aligned} \quad (\text{C.18})$$

and finally for the remainder we have

$$\begin{aligned} r^{\text{QMLE}}(\alpha) &= \mathcal{O}_p \left( (JT)^{-3/2} \|\Psi(\alpha)\|^3 \|X_k\| \right) + \mathcal{O}_p \left( (JT)^{-3/2} \|\Psi(\alpha)\|^3 \|Z_m\| \right) \\ &\quad + \mathcal{O}_p \left( (JT)^{-1} \|\Psi(\alpha)\| \|X_k\|^2 \|\tilde{\beta}_{\hat{\alpha}} - \beta^0\| \right) + \mathcal{O}_p \left( (JT)^{-1} \|\Psi(\alpha)\| \|Z_m\|^2 \|\tilde{\gamma}_{\hat{\alpha}}\| \right), \end{aligned} \quad (\text{C.19})$$

which holds uniformly over  $\alpha$ . The first two terms in  $r^{\text{QMLE}}(\alpha)$  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  $r^{\text{QMLE}}(\alpha)$  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  $c > 0$

$$\sup_{\{\alpha: \sqrt{J}\|\alpha - \alpha^0\| < c\}} \frac{\|r^{\text{QMLE}}(\alpha)\|}{1 + \sqrt{JT}\|\tilde{\beta}_{\hat{\alpha}} - \beta^0\| + \sqrt{JT}\|\tilde{\gamma}_{\hat{\alpha}}\|} = o_p(1). \quad (\text{C.20})$$

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

$$V_{JT}^{-1} = JT \begin{pmatrix} (x^{\lambda f'} M_{z^{\lambda f}} x^{\lambda f})^{-1} & - (x^{\lambda f'} M_{z^{\lambda f}} x^{\lambda f})^{-1} (x^{\lambda f'} z^{\lambda f}) (z^{\lambda f'} z^{\lambda f})^{-1} \\ - (z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f})^{-1} (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} & (z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f})^{-1} \end{pmatrix}. \quad (\text{C.21})$$

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

$$\begin{pmatrix} [C^{(1)}(X_k, \Psi(\hat{\alpha}))]_{k=1, \dots, K} \\ [C^{(1)}(Z_m, \Psi(\hat{\alpha}))]_{m=1, \dots, M} \end{pmatrix} = \frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon - \left[ \frac{1}{JT} (x^{\lambda f}, z^{\lambda f})' g \right] \sqrt{JT}(\hat{\alpha} - \alpha^0) + o_p(\sqrt{JT}\|\hat{\alpha} - \alpha^0\|),$$

$$\begin{pmatrix} [C^{(2)}(X_k, \Psi(\hat{\alpha}))]_{k=1, \dots, K} \\ [C^{(2)}(Z_m, \Psi(\hat{\alpha}))]_{m=1, \dots, M} \end{pmatrix} = \begin{pmatrix} c_x^{(2)} \\ c_z^{(2)} \end{pmatrix} + \mathcal{O}_p(\sqrt{J}\|\hat{\alpha} - \alpha^0\|), \quad (\text{C.22})$$

where

$$c_x^{(2)} = [C^{(2)}(X_k, e)]_{k=1, \dots, K}, \quad c_z^{(2)} = [C^{(2)}(Z_m, e)]_{m=1, \dots, M}. \quad (\text{C.23})$$

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

$$\begin{aligned} \sqrt{JT} \tilde{\gamma}_{\hat{\alpha}} &= \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left[ \frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right. \\ &\quad \left. - \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} g \right) \sqrt{JT} (\hat{\alpha} - \alpha^0) \right] + o_p(1) + o_p(\sqrt{JT}\|\hat{\alpha} - \alpha^0\|). \end{aligned} \quad (\text{C.24})$$

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

$$\begin{aligned} \sqrt{JT} \tilde{\gamma}_{\alpha} &= \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left[ \frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right. \\ &\quad \left. - \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} g \right) \sqrt{JT} (\alpha - \alpha^0) \right] + r^\gamma(\alpha), \end{aligned} \quad (\text{C.25})$$

where for all  $c > 0$

$$\sup_{\{\alpha: \sqrt{J}\|\alpha - \alpha^0\| < c\}} \frac{\|r^\gamma(\alpha)\|}{1 + \sqrt{JT}\|\alpha - \alpha^0\|} = o_p(1). \quad (\text{C.26})$$

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

$$JT \tilde{\gamma}'_{\alpha} W_{JT} \tilde{\gamma}_{\alpha} = A_0 - 2 A_1' \left[ \sqrt{JT} (\alpha - \alpha^0) \right] + \left[ \sqrt{JT} (\alpha - \alpha^0) \right]' A_2 \left[ \sqrt{JT} (\alpha - \alpha^0) \right] + r^{\text{obj}}(\alpha), \quad (\text{C.27})$$

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

$$\begin{aligned} A_0 &= \left[ \frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right]' \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \\ &\quad \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left[ \frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right], \\ A_1 &= \left( \frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f} \right) \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \\ &\quad \left[ \frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right], \\ A_2 &= \left( \frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f} \right) \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} g \right), \end{aligned} \quad (\text{C.28})$$

and the remainder term in the objective function satisfies

$$\sup_{\{\alpha: \sqrt{J} \|\alpha - \alpha^0\| < c\}} \frac{\|r^{\text{obj}}(\alpha)\|}{\left(1 + \sqrt{JT} \|\alpha - \alpha^0\|\right)^2} = o_p(1). \quad (\text{C.29})$$

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

$$\sqrt{JT} (\hat{\alpha} - \alpha^0) = A_2^{-1} A_1 + o_p(1). \quad (\text{C.30})$$

Analogous to equation (C.16) for the first step, we can apply the results in Moon and Weidner (2010a; 2010b) to the third step of the LS-MD estimator to obtain

$$\begin{aligned} \sqrt{JT} (\hat{\beta} - \beta^0) &= \left( \frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} \left[ C^{(1)}(X_k, \Psi(\hat{\alpha})) + C^{(2)}(X_k, \Psi(\hat{\alpha})) \right]_{k=1, \dots, K} + o_p(1) \\ &= \left( \frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} \left[ \frac{1}{\sqrt{JT}} x^{\lambda f'} \varepsilon - \left( \frac{1}{JT} x^{\lambda f'} g \right) \sqrt{JT} (\hat{\alpha} - \alpha^0) + c_x^{(2)} \right] + o_p(1). \end{aligned} \quad (\text{C.31})$$

Here, the remainder term  $o_p(\sqrt{JT} \|\hat{\alpha} - \alpha^0\|)$  is already absorbed into the  $o_p(1)$  term, since (C.30) already shows  $\sqrt{JT}$ -consistency of  $\hat{\alpha}$ . Let  $G_{JT}$  and  $W_{JT}$  be the expressions in equation (5.2) and (5.3) before taking the probability limits, i.e.  $G = \text{plim}_{J,T \rightarrow \infty} G_{JT}$  and  $W = \text{plim}_{J,T \rightarrow \infty} W_{JT}$ . One can show that

$$G_{JT} W_{JT} G'_{JT} = \frac{1}{JT} (g, x)' P_{x^{\lambda f}} (g, x) + \begin{pmatrix} A_2 & 0_{L \times K} \\ 0_{K \times L} & 0_{K \times K} \end{pmatrix}. \quad (\text{C.32})$$

Using this, one can rewrite equation (C.30) and (C.31) as follows

$$\begin{aligned} & G_{JT} \mathcal{W}_{JT} G'_{JT} \sqrt{JT} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\beta} - \beta^0 \end{pmatrix} \\ &= \frac{1}{\sqrt{JT}} (g, x)' P_{x^{\lambda f}} \varepsilon + \begin{pmatrix} A_1 + \left( \frac{1}{JT} g' x^{\lambda f} \right) \left( \frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} c_x^{(2)} \\ c_x^{(2)} \end{pmatrix} + o_p(1), \end{aligned} \quad (\text{C.33})$$

and therefore

$$\begin{aligned} & \sqrt{JT} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\beta} - \beta^0 \end{pmatrix} \\ &= (G_{JT} \mathcal{W}_{JT} G'_{JT})^{-1} G_{JT} \mathcal{W}_{JT} \left[ \frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon \right] \\ &+ (G_{JT} \mathcal{W}_{JT} G'_{JT})^{-1} \begin{pmatrix} A_3 c_z^{(2)} + [(g' x^{\lambda f}) - A_3 (z^{\lambda f'} x^{\lambda f})] (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \\ c_x^{(2)} \end{pmatrix} + o_p(1) \\ &= (GWG')^{-1} GW \left[ \frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon + \begin{pmatrix} c_x^{(2)} \\ c_z^{(2)} \end{pmatrix} \right] + o_p(1), \end{aligned} \quad (\text{C.34})$$

where  $A_3 = \left( \frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f} \right) \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left( \frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1}$ . Having equation (C.34), all that is left to do is to derive the asymptotic distribution of  $c_x^{(2)}$ ,  $c_z^{(2)}$  and  $\frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon$ . This was done in Moon and Weidner (2010a; 2010b) under the same assumptions that we impose here. They show that

$$c_x^{(2)} = -\kappa^{-1} b^{(x,1)} - \kappa b^{(x,2)} + o_p(1), \quad c_z^{(2)} = -\kappa^{-1} b^{(z,1)} - \kappa b^{(z,2)} + o_p(1), \quad (\text{C.35})$$

and

$$\frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon \xrightarrow{d} \mathcal{N} \left[ -\kappa \begin{pmatrix} b^{(x,0)} \\ b^{(z,0)} \end{pmatrix}, \Omega \right]. \quad (\text{C.36})$$

Plugging this into (C.34) gives the result on the limiting distribution of  $\hat{\alpha}$  and  $\hat{\beta}$  which is stated in the theorem. ■

## C.4 Consistency of Bias and Variance Estimators

**Proof of Theorem 5.3.** From Moon and Weidner (2010a; 2010b) we already know that under our assumptions we have  $\hat{\Omega} = \Omega + o_p(1)$ ,  $\hat{b}^{(x,i)} = b^{(x,i)} + o_p(1)$  and  $\hat{b}^{(z,i)} = b^{(z,i)} + o_p(1)$ , for  $i = 0, 1, 2$ . They also show that  $\|M_{\hat{\lambda}} - M_{\lambda^0}\| = \mathcal{O}_p(J^{-1/2})$  and  $\|M_{\hat{f}} - M_{f^0}\| = \mathcal{O}_p(J^{-1/2})$ , from which we can conclude that  $\widehat{\mathcal{W}} = \mathcal{W} + o_p(1)$ . These results on  $M_{\hat{\lambda}}$  and  $M_{\hat{f}}$  together with  $\sqrt{JT}$ -consistency of  $\hat{\alpha}$  and Assumption 5 are also sufficient to conclude  $\hat{G} = G + o_p(1)$ . It follows that  $\hat{B}_i = B_i + o_p(1)$ , for  $i = 0, 1, 2$ . ■

## D 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

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

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  $\lambda$  in the second line always satisfies  $\lambda = M_{\lambda^0} \lambda$ , so we could write  $P_{\lambda}$  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

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

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

## E Alternative Identification Argument

Here we present an alternative identification statement for the BLP model with interactive fixed effects. In contrast to Theorem ID in the main text the following theorem does not employ the LS-MD approach to obtain identification of the model parameters.

**Theorem E.1.** *Assume that  $\mathbb{E}(e_{jt}) = 0$ ,  $\mathbb{E}(X_{it} e_{jt}) = 0$ , and  $\mathbb{E}(Z_{it} e_{jt}) = 0$  for all  $i, j = 1, \dots, J$  and all  $t = 1, \dots, T$ . Furthermore assume that for all  $\alpha \neq \alpha^0$  and all  $\lambda \in \mathbb{R}^{J \times R}$  the matrix  $\mathbb{E}[(d(\alpha), x)' M_{(\lambda, \lambda^0)}(x, z)]$  has full rank (equal to  $K + 1$ ). Then, the parameters  $\alpha$ ,  $\beta$  and  $\lambda f'$  are identified.*

The proof of the theorem is available upon request. The exogeneity condition here is stronger than the one imposed in Theorem ID. Namely, in each market the characteristics and instruments

for all products are assumed uncorrelated of all the product specific shocks, i.e. in each market the theorem assumes strict exogeneity. By plugging in model into the conditions  $\mathbb{E}(e_{jt}) = 0$ ,  $\mathbb{E}(X_{it}e_{jt}) = 0$ ,  $\mathbb{E}(Z_{it}e_{jt}) = 0$ , one obtains moment conditions, and the theorem states that these moment conditions are sufficient to identify the parameters of the model, as long as the additional rank condition is satisfied. One could therefore use these moment conditions to do inference on the model parameters. However, in particular when  $J$  and  $T$  are relatively large, one has to overcome both a computational problem and an incidental parameter problem here, because we need to estimate  $JR$  parameters for  $\lambda$  and  $TR$  parameters for  $f$ . We do not follow this approach further in the this paper.



## References

- Akerberg, 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., 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.
- 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. and Haile, P. (2009). Identification of discrete choice demand from market-level data. Manuscript, Yale University.
- 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.
- 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*. forthcoming.
- 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. (2008). On the Nonparametric Foundations of Discrete-Choice Demand Estimation. Manuscript, University of Wisconsin.
- Gandhi, A., Kim, K., and Petrin, A. (2010). Identification and Estimation in Discrete Choice Demand Models when Endogenous Variables Interact with the Error. Manuscript.
- 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. (2008). Estimation of random coefficient demand models: Challenges, difficulties and warnings. Manuscript.
- Media Dynamics, Inc. (1997). *TV Dimensions*. Media Dynamics, Inc. annual publication.
- Moon, H. and Weidner, M. (2010a). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. *Manuscript*.
- Moon, H. and Weidner, M. (2010b). Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects. *Manuscript*.
- 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.