

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

ESTIMATION OF RANDOM COEFFICIENTS LOGIT DEMAND MODELS WITH INTERACTIVE FIXED EFFECTS

Hyungsik Roger Moon
University of Southern California

Matthew Shum
California Institute of Technology

Martin Weidner
University of Southern California



SOCIAL SCIENCE WORKING PAPER 1325

May 2010

Estimation of Random Coefficients Logit Demand Models with Interactive Fixed Effects

Hyungsik Roger Moon

Matthew Shum

Martin Weidner

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.

JEL classification numbers: C23, C25

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

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

Hyungsik Roger Moon

Matthew Shum

Martin Weidner

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 (2009)), 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

*We are grateful for comments from the participants of the 2009 All-UC Econometrics Conference and of the econometrics seminar at Georgetown University. We also thank Han Hong, Sung Jae Jun and Jinyong Hahn for very helpful discussions. Moon acknowledges the NSF for financial support via SES 0920903.

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, to nonlinear models. Relative to the existing factor literature (for instance, Bai (2009), and Moon and Weidner (2009)), our model poses some estimation challenges. The usual principal components approach is inadequate due to the the nonlinearity of the model and the potential endogeneity of the regressors. The GMM estimation approach of BLP cannot be used due to the presence of the interactive fixed effects. Hence, we propose an alternative estimator which we call the *Least Squares-Minimum Distance (LS-MD)* estimator. The new estimator is calculated in two steps. The first step consists of a least squares fit, which includes the interactive fixed effects and instrumental variables as regressors. The second step minimizes the norm of the estimated IV coefficients of the first step. We show that our estimator is consistent and derive the limit distribution under an asymptotics where both the number of products and the number of markets goes to infinity. In practice, the estimator is simple and straightforward to compute, and shows a good small-sample performance in our Monte Carlo simulations.

Our work complements some recent papers in which alternative estimation approaches for 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 (2008), Harding and Hausman (2007), Bajari, Fox, Kim and Ryan (2008b), 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 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. Moreover, including the interactive fixed effects leads to estimates which are both more precise and (for the most part) larger in magnitude than estimates obtained from models without factors, and imply larger (in absolute value) price elasticities than the standard model.

The paper is organized as follows. Section 2 introduces the model. In Section 3 we

discuss the LS-MD estimation method. Consistency and asymptotic normality is discussed in Section 4. Section 5 contains Monte Carlo simulation results and Section 6 discusses the empirical example. Section 7 concludes. In appendix A we discuss the advantages of our estimation approach to a more standard GMM approach, in the presence of factors. The rest of the appendix lists our assumptions for the asymptotics and provides 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 $m \times n$ 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 ∇ 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 market¹ t is given by

$$u_{ijt} = \delta_{jt}^0 + \epsilon_{ijt} + X'_{jt} v_i, \quad (1)$$

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

$$\delta_{jt}^0 = X'_{jt} \beta^0 + \xi_{jt}^0, \quad (2)$$

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

¹The t subscript can also denote different time periods.

BLP, ξ_{jt}^0 denotes unobserved product characteristics of product j , which can vary across markets t . This is a “structural error”, in that it is observed by all consumers when they make their decisions, but is unobserved by the econometrician. In this paper, we focus on the case where these unobserved product characteristics vary across products and markets according to a factor structure:

$$\xi_{jt}^0 = \lambda_j^{0'} f_t^0 + e_{jt}^0, \quad (3)$$

where $\lambda_j^0 = (\lambda_{1j}^0, \dots, \lambda_{Rj}^0)'$ is a vector of factor loadings corresponding to the R factors $f_t^0 = (f_{1t}^0, \dots, f_{Rt}^0)'$, and e_{jt}^0 is a random component. Here $\lambda_j^{0'} f_t^0$ represent interactive fixed effects, in that both the factors f_t^0 and factor loadings λ_j^0 are unobserved to the econometrician, and can be correlated arbitrarily with the observed product characteristics X_{jt} . We assume that the number of factors R is known.² The superscript zero indicates the true parameters, and objects evaluated at the true parameters.

The factor structure in equation (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).³ 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 by the factor structure.

To begin with, we assume that the regressors X_{jt} are exogenous with respect to the errors e_{jt}^0 , i.e. X_{jt} and e_{jt}^0 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}^0 can still be endogenous with respect to the ξ_{jt}^0 , due to correlation with the factors and

²Known R is also assumed in Bai (2009) and Moon and Weidner (2009) 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 (2005), and Harding (2007).

³cf. *TV Dimensions* (1997).

factor loadings.⁴ 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}^0 . The errors e_{jt}^0 are assumed to be independent across j and t , but heteroscedasticity is allowed.

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 ϵ_{ijt} follows a marginal type I extreme value distribution iid across i and j (but not necessarily independent across t). 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:

$$\pi_{jt}(\delta_t, X_t, v_i) = \frac{\exp(\delta_{jt} + X'_{jt}v_i)}{1 + \sum_{l=1}^J \exp(\delta_{lt} + X'_{lt}v_i)}, \quad (4)$$

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

$$s_{jt}(\alpha^0, \delta_t, X_t) = \int \pi_{jt}(\delta_t, X_t, v) dG_{\alpha^0}(v), \quad (5)$$

where $G_{\alpha^0}(v)$ is the known distribution of consumer tastes v_i over the product characteristic, and α^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 Σ^0 is a $K \times K$ matrix of parameters, which can be subject to constraints (e.g. only one or a few regressors may have random coefficients, in which case the components of Σ^0 are only non-zero for these regressors), and α^0 consists of the independent parameters in Σ^0 .⁵ The distribution $dG_{\alpha^0}(v)$ could have a finite number of support points, but in any case we assume a continuum of agents i (at each support point, if their number is finite), in order to have a deterministic interpretation of the above market shares.

The observables in this model are the market shares s_{jt}^0 and the regressors X_{jt} .⁶ In addition, we need M instruments $Z_{jt} = (Z_{1,jt}, \dots, Z_{M,jt})'$ in order to estimate the

⁴An example of the case where price p_{jt} is an endogenous regressor with respect to the common factors but exogenous with respect to the error is as follows. Suppose that t denotes time. If p_{jt} is determined as a function of past unobserved product characteristic and some additional exogenous (wrt e_{jt}^0) variables Z_{jt} , i.e. $p_{jt} = p(Z_{jt}, \xi_{j,t-1}^0, \xi_{j,t-2}^0, \dots)$, then p_{jt} is endogenous wrt ξ_{jt}^0 (because ξ_{jt}^0 is correlated across time), but sequentially exogenous with respect to e_{jt}^0 . In this example the price endogeneity is completely captured by the factor structure and no instrument for price is required in our estimation procedure.

⁵We focus in this paper on the case where functional form of the distribution function G_{α} is known by the researcher. Recent papers have addressed estimation when this is not known; eg. Bajari, Fox, Kim and Ryan (2008b), (2008a).

⁶In the present paper we assume that the true market shares $s_{jt} = s_{jt}(\delta_t^0)$ are observed. Berry, Linton

parameters α , with $M \geq L$. These additional instruments can for example be chosen to be non-linear transformation of the $X_{k,jt}$. 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 our model would still be required to identify the covariance parameters in the random coefficients distribution.

The unknown parameters are α^0 , β^0 , λ^0 , and f^0 . Of these, the important parameters to estimate are β^0 and α^0 , in terms of which we can calculate the ultimate objects of interest (such as price elasticities) The factors and factor loadings λ^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 λ^0 and f^0 creates a Neyman and Scott (1948) incidental parameter problem: because the number of nuisance parameters grows with the sample size, the estimators for the parameters of interest become inconsistent. Following some recent panel data literature, 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 Estimation

Following BLP, one can assume (or under appropriate assumption on G_α and X_{jt} one can show) that equation (5) is invertible, i.e. for each market t the mean utilities $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})$ are unique functions of α , the market shares $s_t = (s_{1t}, \dots, s_{Jt})$, and the

and Pakes (2004) explicitly consider sampling error in the observed market shares in their asymptotic theory. Here, we abstract away from this additional complication and focus on the econometric issues introduced by the factor structure in ξ^0 .

regressors $X_t = (X_{1t}, \dots, X_{Jt})$.⁷ We denote these functions by $\delta_{jt}(\alpha, s_t, X_t)$. We have

$$\delta_{jt}^0 = \delta_{jt}(\alpha^0, s_t, X_t) = \sum_{k=1}^K \beta_k^0 X_{k,jt} + \sum_{r=1}^R \lambda_{jr}^0 f_{tr}^0 + e_{jt}^0. \quad (6)$$

If δ_{jt}^0 is known, then the above model reduces to the linear panel regression model with interactive fixed effects. Estimation of this model was discussed under fixed T asymptotics in 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 (2009).

In the previous section we introduced the random coefficient logit model, which provides one specification for the market shares as a function of mean utilities. Our analysis extends to other specifications as long as the relation between market shares and mean utilities is invertible, i.e. $\delta_{jt} = \delta_{jt}(\alpha^0, s_t, X_t)$ is well-defined, and the assumptions below are satisfied.

The computational challenge in estimating the model (6) lies in accommodating both the model parameters (α, β) , which in the existing literature has mainly been done in a GMM framework, as well as the nuisance elements λ_j, f_t , which in the existing literature have been treated using a principal components decomposition in a least-squares context (e.g., Bai (2009), and Moon and Weidner (2009)). Our estimation procedure combines both the GMM approach to demand estimation and the least squares (or QMLE) approach to the interactive fixed effect model.

Our least squares-minimum distance (LS-MD) estimators for α and β are defined by

Step 1 (least squares): for given α let

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

$$\left(\tilde{\beta}_\alpha, \tilde{\gamma}_\alpha, \tilde{\lambda}_\alpha, \tilde{f}_\alpha \right) = \underset{\{\beta \in \mathcal{B}_\beta, \gamma, \lambda, f\}}{\operatorname{argmin}} \left\| \delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m - \lambda f' \right\|_F^2,$$

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):

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

$$\left(\hat{\beta}, \hat{\lambda}, \hat{f} \right) = \underset{\{\beta \in \mathcal{B}_\beta, \lambda, f\}}{\operatorname{argmin}} \left\| \delta(\hat{\alpha}) - \sum_{k=1}^K \beta_k X_k - \lambda f' \right\|_F^2. \quad (7)$$

⁷Gandhi (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.

Here, $\delta(\alpha, s, X)$, X_k and Z_m are $J \times T$ matrices, λ is $J \times R$ and f is $T \times R$, W_{JT} is a positive definite $M \times M$ weight matrix, and $\mathcal{B}_\alpha \subset \mathbb{R}^L$ and $\mathcal{B}_\beta \subset \mathbb{R}^K$ are parameter sets for α and β . In step 1, we include the IV's Z_m as auxiliary regressors, with coefficients $\gamma \in \mathbb{R}^M$. Step 2 is based on imposing the exclusion restriction on the IV's, which requires that $\gamma = 0$, at the true value of α . Thus, we first estimate β , λ , f , and the instrument coefficients γ by least squares for fixed α , and subsequently we estimate α by minimizing the norm of $\tilde{\gamma}_\alpha$ with respect to α .

Step 3 in (7), 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 β . Under the assumptions for consistency of $(\hat{\alpha}, \hat{\beta})$ presented below, this alternative estimator is also consistent for β^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 α , β and γ the optimal factors and factor loadings in the least squares problems in step 1 (and step 3) of (7) turn out to be the principal components estimators for λ and f . These incidental parameters can therefore be concentrated out easily, and the remaining objective function for β and γ turns out to be given by an eigenvalue problem (see e.g. Moon and Weidner (2009) for details), namely

$$\left(\tilde{\beta}_\alpha, \tilde{\gamma}_\alpha \right) = \underset{\{\beta, \gamma\}}{\operatorname{argmin}} \sum_{t=R+1}^T \mu_t \left[\left(\delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m \right)' \left(\delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m \right) \right]. \quad (8)$$

This formulation greatly simplifies the numerical calculation of the estimator, since eigenvalues are easy and fast to compute, and we only need to perform numerical optimization over β and γ , not over λ and f .

The step 1 optimization problem in (7) 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 (2009) that (under their assumptions) $\hat{\beta}_{\alpha^0}$ is \sqrt{JT} -consistent for β^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 takes the reduced form specification $Y_2 = Z \delta + V$, and e and V are correlated. The two stage

least squares estimator of α is $\hat{\alpha}_{2SLS} = (Y_2' P_Z Y_2)^{-1} Y_2' P_Z y_1$, 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 α 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.⁸

Moreover, the presence of the factors makes it difficult to use the moment condition-based GMM approach proposed by BLP. Specifically, we know of no way to handle the factors and factor loadings in a GMM moment condition setting such that the resulting estimator for α and β is consistent. In appendix A we consider an alternative GMM estimator in which, rather than including the instruments Z as “extra” regressors in the first step, we estimate all the structural parameters of interest (α, β) by using GMM on the implied moment conditions of the model, after obtaining estimates of the factors λ and f via a preliminary principal components step. We show that in the absence of factors ($R = 0$) our LS-MD estimator is equivalent to the GMM estimator for an appropriate choice of weight matrix, but in the presence of factors the two estimators can be different and the GMM estimator may not be consistent as $J, T \rightarrow \infty$.

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^0 . However, this could be restrictive in some applications, e.g., when price p_{jt} is determined by ξ_{jt} contemporaneously. Hence, we consider here the possibility that the regressors X could also be correlated with e^0 . This is readily accommodated within our framework. Let $X^{\text{end}} \subset X$ denote the endogenous regressors, with $\dim(X^{\text{end}}) = K_2$. (Hence, the number of exogenous regressors equals $K - K_2$.) Similarly, let β^{end} denote the coefficients on these regressors, while β continues to denote the coefficients on the exogenous regressors. Correspondingly, we assume that M , the number of instruments, exceeds $L + K_2$. Then we define the following estimator,

⁸Recently, Chernozhukov and Hansen (2006) used a similar two stage estimation method for a class of instrumental quantile regressions.

which is a generalized version of our previous estimator:

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

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

$$\left(\tilde{\beta}_{\alpha^{\text{end}}}, \tilde{\gamma}_{\alpha^{\text{end}}}, \tilde{\lambda}_{\alpha^{\text{end}}}, \tilde{f}_{\alpha^{\text{end}}} \right) = \underset{\{\beta \in \mathcal{B}_{\beta}, \gamma, \lambda, f\}}{\text{argmin}} \left\| \delta(\alpha) - \sum_{k=1}^{K_2} \beta_k^{\text{end}} X_k^{\text{end}} - \sum_{k=K_2+1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m - \lambda f' \right\|_F^2,$$

step 2:

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

step 3:

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

$$\left(\hat{\beta}, \hat{\lambda}, \hat{f} \right) = \underset{\{\beta \in \mathcal{B}_{\beta}, \lambda, f\}}{\text{argmin}} \left\| \delta(\hat{\alpha}) - \sum_{k=1}^{K_2} \hat{\beta}_k^{\text{end}} X_k^{\text{end}} - \sum_{k=K_2+1}^K \beta_k X_k - \lambda f' \right\|_F^2, \quad (9)$$

where \mathcal{B}_{α} , \mathcal{B}_{β} and $\mathcal{B}_{\beta}^{\text{end}}$ are the parameter sets for α , β and β^{end} . The difference between this estimator, and the previous one for which all the regressors were assumed exogenous, is that the estimation of β^{end} , the coefficients on the endogenous regressors \tilde{X} , has been moved to the second step. The structure of the estimation procedure in (9) is exactly equivalent to that of our original LS-MD estimator (7), only that α is replaced by α^{end} , and $\delta(\alpha)$ is replaced by $\delta(\alpha) - \sum_{k=1}^{K_2} \beta_k^{\text{end}} X_k^{\text{end}}$. Thus, all results below on the consistency, asymptotic distribution and bias correction of the LS-MD estimator (7) with only sequentially exogenous regressors directly generalize to the estimator (9) 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 the factors are absent (i.e. $R = 0$).

4 Consistency and Asymptotic Distribution of $\hat{\alpha}$ and $\hat{\beta}$

In this section we present our results on the properties of the LS-MD estimator $\hat{\alpha}$ and $\hat{\beta}$ defined in (7) under the asymptotics $J, T \rightarrow \infty$. We define the $JT \times K$ matrix x^f , the $JT \times M$ matrix z^f , and the $JT \times 1$ vector $d(\alpha)$ by

$$x_{\cdot, k}^f = \text{vec}(X_k M_{f0}), \quad z_{\cdot, m}^f = \text{vec}(Z_m M_{f0}), \quad d(\alpha) = \text{vec}(\delta(\alpha) - \delta(\alpha^0)), \quad (10)$$

where $k = 1, \dots, K$ and $m = 1, \dots, M$. The key assumption that guarantees consistency of the LS-MD estimator is the following.

Assumption 1 *There exists a constant $c > 0$ such that for all $\alpha \in \mathcal{B}_\alpha$ we have*

$$\frac{1}{JT} \left[d'(\alpha) P_{(x^f, z^f)} d(\alpha) - \max_{\lambda} d'(\alpha) P_{(x^f, M_{f^0} \otimes \lambda)} d(\alpha) \right] \geq c \|\alpha - \alpha^0\|^2, \quad wpa1,$$

where we maximize over all $J \times R$ matrices λ .

This is a relevancy assumption on the instruments Z . It requires that the combination of $X_k M_{f^0}$ and $Z_m M_{f^0}$ has more explanatory power for $\delta(\alpha) - \delta(\alpha^0)$ than the combination of $X_k M_{f^0}$ and any $J \times R$ matrix λ , uniformly over α . The appearance of the projector M_{f^0} is not surprising in view of the interactive fixed effect specification in (6). For example, if $R = 1$ and $f_t^0 = 1$ for all t , then multiplying with M_{f^0} is equivalent to subtracting the time-mean for each cross-sectional unit, which is the standard procedure in a model with only individual fixed effects.⁹ The appearance of λ in Assumption 1 requires that, loosely speaking, the instruments must be more relevant for $\delta(\alpha) - \delta(\alpha^0)$ than any λ . Although λ can be chosen arbitrarily, it is time-invariant; it follows, then, that the instruments Z_m cannot all be chosen time-invariant without violating Assumption 1.¹⁰

The matrix valued function $\delta(\alpha) = \delta(\alpha, s, X)$ was introduced as the inverse of equation (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 should in principle be possible to analyze Assumption 1 further and to discuss validity and optimality of the instruments. Unfortunately, too little is known about the properties of $\delta(\alpha)$ to make such a further analysis feasible at the present time. Even if no endogeneity in the regressors is present, and even if $R = 0$, it is still difficult to prove that a given set of instruments satisfies our relevancy Assumption 1. This, however, is not only a feature of our approach, but is also true for BLP, and for Berry, Linton and Pakes (2004). If some regressors, in particular price, are treated as endogenous (wrt e^0), the discussion of relevance (and exogeneity) of the instruments becomes even more complicated.

The remaining assumptions 5 to 8 which are referred to in the following consistency theorem are presented in appendix B. These additional assumptions are only slight modifications of the ones used in Bai (2009) and Moon and Weidner (2009) for the linear models with interactive fixed effects, and we refer to those papers for a more

⁹Remember however, that both f^0 and λ^0 are unobserved in our model.

¹⁰Time-invariant instruments are also ruled out by Assumption 8 in the appendix, which requires instruments to be “high-rank”.

detailed discussion. The main contribution of the present paper is the generalization of the factor analysis to non-linear random coefficient discrete-choice demand models, and Assumption 1 is the key assumption needed for this generalization. Note also that Assumption 6 (in the appendix) requires $\frac{1}{JT} \text{Tr}(e^0 Z'_m) = o_p(1)$, i.e. exogeneity of the instruments with respect to e^0 , but instruments can be correlated with the factors and factor loadings and thus also with ξ^0 .

Theorem 2 *Let assumptions 1, 5, 6, 7 and 8 be satisfied, let $\alpha^0 \in \mathcal{B}_\alpha$ and $\beta^0 \in \mathcal{B}_\beta$, and let $\mathcal{B}_\alpha \subset \mathbb{R}^L$ and $\mathcal{B}_\beta \subset \mathbb{R}^K$ be bounded. 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 2 is given in the appendix. 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

$$x^{\lambda f}_{\cdot, k} = \text{vec}(M_{\lambda^0} X_k M_{f^0}) , \quad z^{\lambda f}_{\cdot, m} = \text{vec}(M_{\lambda^0} Z_m M_{f^0}) , \quad g_{\cdot, l} = -\text{vec}(\nabla_l \delta(\alpha^0)) , \quad (11)$$

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.¹¹ We introduce the $(L+K) \times (L+K)$ matrix G and the $(K+M) \times (K+M)$ matrix Ω as follows

$$G = \text{plim}_{J, T \rightarrow \infty} \frac{1}{JT} \begin{pmatrix} g' x^{\lambda f} & g' z^{\lambda f} \\ x^{\lambda f'} x^{\lambda f} & x^{\lambda f'} z^{\lambda f} \end{pmatrix}, \quad \Omega = \text{plim}_{J, T \rightarrow \infty} \frac{1}{JT} (x^{\lambda f}, z^{\lambda f})' \text{diag}(\Sigma_e^{\text{vec}}) (x^{\lambda f}, z^{\lambda f}), \quad (12)$$

where $\Sigma_e^{\text{vec}} = \text{vec} \left\{ \left[\mathbb{E} (e_{jt}^0)^2 \right]_{\substack{j=1, \dots, J \\ t=1, \dots, T}} \right\}$ is the JT -vector of vectorized variances of e_{jt}^0 .

Finally, we define the $(K+M) \times (K+M)$ weight matrix \mathcal{W} by

$$\mathcal{W} = \text{plim}_{J, T \rightarrow \infty} \left[\begin{pmatrix} \left(\frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} & 0_{K \times M} \\ 0_{M \times K} & 0_{M \times M} \end{pmatrix} + \begin{pmatrix} -(x^{\lambda f'} x^{\lambda f})^{-1} x^{\lambda f'} z^{\lambda f} \\ \mathbb{1}_M \end{pmatrix} \right. \\ \left. \times \begin{pmatrix} \frac{1}{JT} z' M_{x^{\lambda f} z} \end{pmatrix}^{-1} W_{JT} \begin{pmatrix} \frac{1}{JT} z' M_{x^{\lambda f} z} \end{pmatrix}^{-1} \begin{pmatrix} -(x^{\lambda f'} x^{\lambda f})^{-1} x^{\lambda f'} z^{\lambda f} \\ \mathbb{1}_M \end{pmatrix}' \right]. \quad (13)$$

Existence of these probability limits is imposed by Assumption 9 in the appendix. Some further regularity conditions are necessary to derive the limiting distribution of our LS-MD

¹¹We do not necessarily require that all $\delta_{jt}(\alpha)$ are differentiable. All we need is that $J \times T$ matrices $\nabla_l \delta(\alpha)$, $l = 1, \dots, L$, exist, which satisfy Assumption 10(i) in the appendix.

estimator, and those are summarized in Assumption 10 in the appendix. Assumption 10 is again a straightforward generalization of the assumptions imposed by Moon and Weidner (2009) for the linear model, except for part (i) of the assumption, which demands that $\delta(\alpha)$ can be linearly approximated around α^0 such that the Frobenius norm of the remainder term of the expansion is of order $o_p(\sqrt{JT}\|\alpha - \alpha^0\|)$ in any \sqrt{J} shrinking neighborhood of α^0 .

Theorem 3 *Let the assumptions of Theorem 2 be satisfied, and in addition let Assumption 9 and 10 hold. In the limit $J, T \rightarrow \infty$ with $J/T \rightarrow \kappa^2$, $0 < \kappa < \infty$, we then have*

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

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

The proof of Theorem 3 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 κB_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 κB_2 are due to heteroscedasticity of the idiosyncratic error e_{jt}^0 across firms j and markets t , respectively. The first and last bias terms are proportional to κ , and thus are large when T is small compared to J , while the second bias terms is proportional to κ^{-1} , and thus is large when T is large compared to J . Note that no asymptotic bias is present if regressors and instruments are strictly exogenous and errors e_{jt}^0 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 (2009).

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. According to the discussion in appendix A, the LS-MD estimator is equivalent to the GMM estimator if no factors are present. In that case the weight matrix \mathcal{W} that appears in Theorem 3 is indeed just the probability limit of the GMM weight matrix¹² and our asymptotic variance

¹²See also equation (21) in the appendix.

matrix thus 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_{λ^0} in the formulas for G , Ω and \mathcal{W} . The presence of these projectors implies that those components of X_k and Z_m which are proportional to f^0 and λ^0 do not contribute to the asymptotic variance, 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^0) 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, Ω and \mathcal{W} . This is done explicitly in appendix B.4.

Theorem 4 *Let the assumption of Theorem 3 and Assumption 11 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 4 motivates the introduction of the bias corrected estimator

$$\begin{pmatrix} \hat{\alpha}^* \\ \hat{\beta}^* \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - \frac{1}{T} \hat{B}_0 - \frac{1}{J} \hat{B}_1 - \frac{1}{T} \hat{B}_2. \quad (14)$$

Under the assumptions of Theorem 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 α^0 and β^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}^0 are homoscedastic with variance σ_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

$$W_{JT} = \frac{1}{JT} z' M_{x^{\lambda f}} z. \quad (15)$$

Under heteroscedasticity of e_{jt}^0 there are in general not have 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. In our Monte Carlo simulations and in the empirical application we always choose W_{JT} according to equation (15). 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.

5 Monte Carlo Results

For our Monte Carlo simulation, we assume that there is one factor ($R = 1$), a constant regressor, one additional regressor X_{jt} , and we consider the following data generating process for mean utility and regressors

$$\begin{aligned}\delta_{jt} &= \beta_1^0 + \beta_2^0 X_{jt} + \lambda_j^0 f_t^0 + e_{jt}^0, \\ X_{jt} &= \tilde{X}_{jt} + \lambda_j^0 + f_t^0 + \lambda_j^0 f_t^0,\end{aligned}\tag{16}$$

where λ_j^0 , f_t^0 and \tilde{X}_{jt} are all independently and identically distributed across j and t as $\mathcal{N}(0, 1)$, and are also mutually independent. For the distribution of the error term e_{jt}^0 conditional on X_{jt} , λ_j^0 and f_t^0 we use two different specifications:

$$\begin{aligned}\text{specification 1: } & e_{jt}^0 |_{X, \lambda^0, f^0} \sim iid \mathcal{N}(0, 1), \\ \text{specification 2: } & e_{jt}^0 |_{X, \lambda^0, f^0} \sim iid \mathcal{N}(0, \sigma_{jt}^2), \quad \sigma_{jt}^2 = \frac{1}{1 + \exp(-\lambda_{jt})}.\end{aligned}\tag{17}$$

In the first specification, the error is distributed independently of the regressor, factor and factor loading. In the second specification there is heteroscedasticity conditional on the factor loading, namely the variance σ_{jt}^2 of e_{jt}^0 is an increasing function of λ_{jt} , which is chosen nonlinearly such that $\sigma_{jt}^2 \in (0, 1)$. According to our asymptotic analysis we expect this heterogeneity in e_{jt}^0 to result in a bias of the LS-MD estimator, which is accounted for in our bias corrected estimator. As values for the regression parameters we choose $\beta_1^0 = 0$ and $\beta_2^0 = 1$.

N,T	statistics	$\hat{\alpha}$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\alpha}^*$	$\hat{\beta}_2^*$	$\hat{\beta}_1^*$
20,20	bias	-0.0009	-0.0097	-0.0036	0.0040	-0.0044	-0.0033
	std	0.2438	0.2269	0.0630	0.3326	0.3091	0.0659
	rmse	0.2437	0.2270	0.0631	0.3325	0.3090	0.0659
50,50	bias	-0.0048	-0.0065	-0.0013	-0.0046	-0.0063	-0.0013
	std	0.0993	0.0960	0.0238	0.0990	0.0953	0.0239
	rmse	0.0993	0.0962	0.0239	0.0990	0.0955	0.0239
80,80	bias	-0.0017	-0.0024	-0.0001	-0.0018	-0.0026	-0.0001
	std	0.0614	0.0597	0.0133	0.0612	0.0595	0.0133
	rmse	0.0614	0.0597	0.0133	0.0612	0.0595	0.0133

Table 1: Simulation results for specification 1 (no heteroscedasticity), using 1000 repetitions. We report the bias, standard errors (std), and square roots of the mean square errors (rmse) of the LS-MD estimator $(\hat{\alpha}, \hat{\beta})$ and its bias corrected version $(\hat{\alpha}^*, \hat{\beta}^*)$.

The market shares are computed from the mean utilities according to equation (4) and (5), where we assume a normally distributed random coefficient on the regressor X_{jt} , i.e. $v \sim \mathcal{N}(0, \alpha^2)$, and we set $\alpha^0 = 1$. Although the regressors are strictly exogenous (wrt e_{jt}^0) in our simulation, we still need an instrument to identify α , and we choose $Z_{jt} = X_{jt}^2$, the square of the regressor X_{jt} . Both X_{jt} and Z_{jt} are therefore endogenous with respect to the total unobserved error $\lambda_j^0 f_t^0 + e_{jt}^0$, since the factors and factor loadings enter into their distributions.

Results of simulation runs with 1000 repetitions are reported in table 1 and 2. The bias corrected estimators $\hat{\alpha}^*$ and $\hat{\beta}^*$, whose summary statistics are reported, are the ones defined in (14), but without inclusion of \hat{B}_0 , since there is no pre-determined regressor in this setup.

In specification 1 there is no heteroscedasticity in e_{jt}^0 , and we thus expect no bias in the LS-MD estimator. The corresponding results in table 1 verify this expectation. We find that both the LS-MD estimator $(\hat{\alpha}, \hat{\beta})$ and its bias corrected version $(\hat{\alpha}^*, \hat{\beta}^*)$ have biases that are very small compared to their standard errors, and are not statistically significant in our sample of 1000 simulation runs. For $N = T = 50$ and $N = T = 80$ the standard errors of $(\hat{\alpha}, \hat{\beta})$ and of $(\hat{\alpha}^*, \hat{\beta}^*)$ are almost identical, but for $N = T = 20$ the standard errors for $(\hat{\alpha}^*, \hat{\beta}^*)$ are up to 37% larger than for $(\hat{\alpha}, \hat{\beta})$ in specification 1, i.e. at these small values for N and T the bias correction adds some noise to the estimator and thus increases their standard errors.

In specification 2 there is heteroscedasticity in e_{jt}^0 . Correspondingly, from table 2

N,T	statistics	$\hat{\alpha}$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\alpha}^*$	$\hat{\beta}_2^*$	$\hat{\beta}_1^*$
20,20	bias	-0.0326	-0.0460	-0.0004	-0.0040	-0.0277	-0.0036
	std	0.2068	0.1988	0.0475	0.5875	0.4523	0.0598
	rmse	0.2093	0.2039	0.0475	0.5872	0.4529	0.0599
50,50	bias	-0.0207	-0.0240	0.0016	-0.0056	-0.0063	0.0006
	std	0.0831	0.0852	0.0166	0.1325	0.1313	0.0167
	rmse	0.0856	0.0885	0.0167	0.1325	0.1314	0.0167
80,80	bias	-0.0116	-0.0138	0.0008	-0.0022	-0.0025	0.0001
	std	0.0452	0.0458	0.0096	0.0441	0.0444	0.0096
	rmse	0.0467	0.0478	0.0097	0.0441	0.0445	0.0096

Table 2: Simulation results for specification 2 (heteroscedasticity in e_{jt}^0), using 1000 repetitions. We report the bias, standard errors (std), and square roots of the mean square errors (rmse) of the LS-MD estimator $(\hat{\alpha}, \hat{\beta})$ and its bias corrected version $(\hat{\alpha}^*, \hat{\beta}^*)$.

we find that the LS-MD estimators $\hat{\alpha}$ and $\hat{\beta}_2$ have biases of magnitudes between 15% and 30% of their standard errors (the constant coefficient $\hat{\beta}_1$ is essentially unbiased). In contrast, the biases of the bias corrected estimators $\hat{\alpha}^*$ and $\hat{\beta}_2^*$ are much smaller, and are not statistically significant at 5% level in our sample of 1000 simulation runs. This shows that our bias correction formula adequately corrects for the bias due to heteroscedasticity in e_{jt}^0 . For $N = T = 80$ the standard errors of $(\hat{\alpha}, \hat{\beta})$ and $(\hat{\alpha}^*, \hat{\beta}^*)$ are almost identical (implying that the root mean square error of $(\hat{\alpha}^*, \hat{\beta}^*)$ is smaller, since its bias is smaller), which confirms our asymptotic results that the bias correction removes the bias but leaves the variance of the estimators unchanged as $N, T \rightarrow \infty$. However, as we already found for specification 1, at finite sample the bias correction also adds some noise to the estimators. For $N = T = 50$ the standard errors of $(\hat{\alpha}^*, \hat{\beta}^*)$ are up to 60% larger, and for $N = T = 20$ even up to 195% larger than the standard errors of $(\hat{\alpha}, \hat{\beta})$. Thus, at finite sample there is a trade-off between bias and variance of the estimators. $(\hat{\alpha}, \hat{\beta})$ has smaller variance, but larger bias, while $(\hat{\alpha}^*, \hat{\beta}^*)$ has smaller bias, but larger variance. Depending on the sample size, it may thus be advantageous in empirical application to ignore the bias due to heteroscedasticity of e_{jt}^0 and to simply use the LS-MD estimator without bias correction.

6 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.

For this empirical application, we use the same data as was used in BLP (1995), which are new automobile sales from 1971-1990.¹³ 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.¹⁴ 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. The unit for price is \$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} inch². The choice of units is rather arbitrary, we simply tried to avoid too small and too large decimal numbers.

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.

Results. Table 3 contains estimation results from four specifications of the model. In specification A, prices are considered exogenous (wrt e_{jt}^0), but one factor is present, which captures some degree of price endogeneity (wrt. ξ_{jt}). Specification B also contains one factor, but treats prices as endogenous, even conditional on the factor. Specification C corresponds to the BLP (1995) model, where prices are endogenous, but no factor is present. Finally, in specification D, we treat prices as exogenous, and do not allow for

¹³The data are available on the webpage of James Levinsohn.

¹⁴This resembles the treatment in Esteban and Shum’s (2007) study of the new and used car markets.

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

Table 3: Parameter estimates (and t-values) for four different model specifications (no factor $R = 0$ vs. one factor $R = 1$; exogenous price vs. endogenous price). α is the standard deviation of the random coefficient distribution (only price has a random coefficient), and the regressors are p (price), hp/weight (horse power per weight), mpd (miles per dollar), size (car length times car width), and a constant.

a factor. This final specification is clearly unrealistic, but is included for comparison with the other specifications. In table 3 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 it 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 α parameter, corresponding to the standard deviation of the random coefficient on price, is estimated to be 2.092. These point estimates imply that, roughly 97% of the time, the random price coefficient is negative, which is as we should expect.

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

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

Specifically, in Specification C, the key price coefficient and α are substantially smaller in magnitude; furthermore, the standard errors are large, so that none of the estimates are significant at usual significance levels. Moreover, the coefficient on *hp/weight* is negative, which is puzzling. In Specification D, which corresponds to a BLP model, but without price endogeneity, we see that the price coefficient is reduced dramatically relative to the other specifications, down to -0.308. This may be attributed to the usual attenuation bias. Altogether, these estimates seem less satisfactory than the ones obtained for Specifications A and B, where factors were included.

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.

Clearly, these differences in elasticities would have significant competitive implications. As a rough estimate, using an inverse-elasticity pricing rule as a benchmark, we see that Specification C, which corresponds to the standard BLP model, implies markups which are roughly twice the size of the markups from Specification B, which allows for

factors. This is an economically very significant difference, and suggests that not including the factors may lead to underestimation of the degree of competitiveness in this market.

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.

7 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.

Tables

	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)
α	1.613	(2.678)	1.633	(3.000)
const	-0.690	(-0.232)	-2.984	(-1.615)

Table 4: Parameter estimates (and t-values) for model specification A and B. Here we report the LS-MD estimators without bias correction, while in table 3 we report the bias corrected LS-MD estimators.

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

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

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

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

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

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

Appendix A Alternative GMM approach

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

$$\begin{aligned}\mathbb{E} [e_{jt}(\alpha^0, \beta^0, \lambda^0 f^{0'}) X_{k,jt}] &= 0, & k &= 1, \dots, K, \\ \mathbb{E} [e_{jt}(\alpha^0, \beta^0, \lambda^0 f^{0'}) Z_{m,jt}] &= 0, & m &= 1, \dots, M,\end{aligned}\quad (18)$$

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

$$\begin{aligned}m_k^X(\alpha, \beta, \lambda f') &= \frac{1}{JT} \text{Tr} (e(\alpha, \beta, \lambda f') X_k'), \\ m_m^Z(\alpha, \beta, \lambda f') &= \frac{1}{JT} \text{Tr} (e(\alpha, \beta, \lambda f') Z_m').\end{aligned}\quad (19)$$

We also define the sample moment vectors $m^X(\alpha, \beta, \lambda f') = (m_1^X, \dots, m_K^X)'$ and $m^Z(\alpha, \beta, \lambda f') = (m_1^Z, \dots, m_M^Z)'$. An alternative estimator for α, β, λ and f is then given by¹⁵

$$\begin{aligned}(\hat{\lambda}_{\alpha, \beta}, \hat{f}_{\alpha, \beta}) &= \underset{\{\lambda, f\}}{\text{argmin}} \sum_{j=1}^J \sum_{t=1}^T e_{jt}^2(\alpha, \beta, \lambda f'). \\ (\hat{\alpha}^{\text{GMM}}, \hat{\beta}^{\text{GMM}}) &= \underset{\{\alpha \in \mathcal{B}_{\alpha, \beta}\}}{\text{argmin}} \begin{pmatrix} m^X(\alpha, \beta, \hat{\lambda}_{\alpha, \beta} \hat{f}'_{\alpha, \beta}) \\ m^Z(\alpha, \beta, \hat{\lambda}_{\alpha, \beta} \hat{f}'_{\alpha, \beta}) \end{pmatrix}' \mathcal{W}_{JT} \begin{pmatrix} m^X(\alpha, \beta, \hat{\lambda}_{\alpha, \beta} \hat{f}'_{\alpha, \beta}) \\ m^Z(\alpha, \beta, \hat{\lambda}_{\alpha, \beta} \hat{f}'_{\alpha, \beta}) \end{pmatrix},\end{aligned}\quad (20)$$

where \mathcal{W}_{JT} is a positive definite $(K + M) \times (K + M)$ weight matrix. The main difference between this alternative estimator and our estimator (7) is that the least-squares step is used solely to recover estimates of the factors and factor loadings (principal components estimator), while the structural parameters (α, β) are estimated in the GMM second step. The relation between $\hat{\alpha}$ and $\hat{\beta}$ defined in (7) and $\hat{\alpha}^{\text{GMM}}$ and $\hat{\beta}^{\text{GMM}}$ defined in (20) is as follows

¹⁵The minimizing $\hat{\lambda}_{\alpha, \beta}$ and $\hat{f}_{\alpha, \beta}$ are the principal components estimators, e.g. $\hat{\lambda}_{\alpha, \beta}$ consists of the eigenvectors corresponding to the R largest eigenvalues of the $J \times J$ matrix

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

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

$$\mathcal{W}_{JT} = \begin{pmatrix} \left(\frac{1}{JT}x'x\right)^{-1} & 0_{K \times M} \\ 0_{M \times K} & 0_{M \times M} \end{pmatrix} + \begin{pmatrix} -(x'x)^{-1}x'z \\ \mathbb{1}_M \end{pmatrix} \left(\frac{1}{JT}z'M_xz\right)^{-1} \\ W_{JT} \left(\frac{1}{JT}z'M_xz\right)^{-1} \begin{pmatrix} -(x'x)^{-1}x'z \\ \mathbb{1}_M \end{pmatrix}', \quad (21)$$

where x is a $JT \times K$ matrix and z is a $JT \times M$ matrix, given by $x_{.,k} = \text{vec}(X_k)$, $k = 1, \dots, K$, and $z_{.,m} = \text{vec}(Z_m)$, $m = 1, \dots, M$. Then $\hat{\alpha}$ and $\hat{\beta}$ solve (7) with weight matrix W_{JT} if and only if they solve (20) with this weight matrix \mathcal{W}_{JT} ,¹⁶ i.e. in this case we have $(\hat{\alpha}, \hat{\beta}) = (\hat{\alpha}^{\text{GMM}}, \hat{\beta}^{\text{GMM}})$.

- (ii) Let $R > 0$ and $M = L$ (exactly identified case). Then a solution of (7) also is a solution of (20), but not every solution of (20) needs to be a solution of (7).
- (iii) For $M > L$ and $R > 0$ there is no straightforward characterization of the relationship between the estimators in (7) and (20).

We want to discuss the exactly identified case $M = L$ a bit further. The reason why in this case every solution of (7) also solves (20) is that the first order conditions (FOC's) wrt to β and γ of the first stage optimization in (7) read $m^X(\hat{\alpha}, \hat{\beta}, \hat{\lambda}_{\hat{\alpha}, \hat{\beta}} \hat{f}'_{\hat{\alpha}, \hat{\beta}}) = 0$ and $m^Z(\hat{\alpha}, \hat{\beta}, \hat{\lambda}_{\hat{\alpha}, \hat{\beta}} \hat{f}'_{\hat{\alpha}, \hat{\beta}}) = 0$, which implies that the GMM objective function of (20) is zero, i.e. minimized. The reverse statement is not true, because for $R > 0$ the first stage objective function in (7) is not a quadratic function of β and γ anymore once one concentrates out λ and f , and it can have multiple local minima that satisfy the FOC. Therefore, $\hat{\alpha}^{\text{GMM}}$ and $\hat{\beta}^{\text{GMM}}$ can be inconsistent, while $\hat{\alpha}$ and $\hat{\beta}$ are consistent, which is the main reason to consider the latter in this paper.

To illustrate this important difference between $\hat{\alpha}^{\text{GMM}}$, $\hat{\beta}^{\text{GMM}}$ and $\hat{\alpha}$, $\hat{\beta}$, we want to give a simple example for a linear model in which the QMLE objective function has multiple local minima. Consider a DGP where $Y_{jt} = \beta^0 X_{jt} + \lambda_j^0 f_t^0 + e_{jt}$, with $X_{jt} =$

¹⁶With this weight matrix \mathcal{W}_{JT} the second stage objective function in (20) becomes

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

where $d(\alpha) = \text{vec}(\delta(\alpha, s, X) - \delta(\alpha^0, s, X))$. Here, β only appears in the first term, and by choosing $\beta = \hat{\beta} = (x'x)^{-1}x'd(\alpha)$ this term becomes zero. Thus, we are left with the second term, which is exactly the second stage objective function in (7) in this case, since for $R = 0$ by the Frisch-Waugh theorem we have $\tilde{\gamma}_{\alpha} = (z'M_xz)^{-1}z'M_x d(\alpha)$.

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

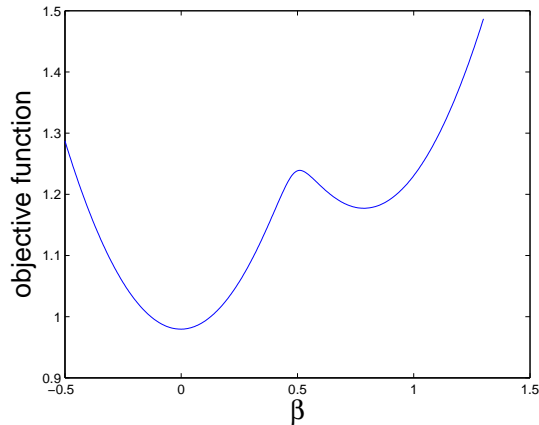


Figure 1: Example for multiple local minima in the least squares objective function $L(\beta)$. The global minimum can be found close to the true value $\beta^0 = 0$, but another local minimum exists around $\beta \approx 0.8$, which renders the FOC inappropriate for defining the estimator $\hat{\beta}$.

Appendix B Details for Section 4 (Consistency and Asymptotic Distribution)

B.1 Formulas for Asymptotic Bias Terms

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

$$\begin{aligned} \Sigma_{e,j}^{(1)} &= \frac{1}{T} \sum_{t=1}^T \mathbb{E} (e_{jt}^0)^2, & \Sigma_{e,t}^{(2)} &= \frac{1}{J} \sum_{j=1}^J \mathbb{E} (e_{jt}^0)^2, \\ \Sigma_{k,\tau}^{X,e} &= \frac{1}{J} \sum_{j=1}^J \mathbb{E} (X_{k,jt} e_{j\tau}^0), & \Sigma_{m,\tau}^{Z,e} &= \frac{1}{J} \sum_{j=1}^J \mathbb{E} (Z_{m,jt} e_{j\tau}^0), \end{aligned} \quad (22)$$

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

$$\begin{aligned} b_k^{(x,0)} &= \text{plim}_{J,T \rightarrow \infty} \text{Tr} \left(P_{f^0} \Sigma_k^{X,e} \right), \\ b_k^{(x,1)} &= \text{plim}_{J,T \rightarrow \infty} \text{Tr} \left[\text{diag} \left(\Sigma_e^{(1)} \right) M_{\lambda^0} X_k f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \right], \\ b_k^{(x,2)} &= \text{plim}_{J,T \rightarrow \infty} \text{Tr} \left[\text{diag} \left(\Sigma_e^{(2)} \right) M_{f^0} X_k' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} \right], \\ b_m^{(z,0)} &= \text{plim}_{J,T \rightarrow \infty} \text{Tr} \left(P_{f^0} \Sigma_m^{Z,e} \right), \\ b_m^{(z,1)} &= \text{plim}_{J,T \rightarrow \infty} \text{Tr} \left[\text{diag} \left(\Sigma_e^{(1)} \right) M_{\lambda^0} Z_m f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \right], \\ b_m^{(z,2)} &= \text{plim}_{J,T \rightarrow \infty} \text{Tr} \left[\text{diag} \left(\Sigma_e^{(2)} \right) M_{f^0} Z_m' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} \right], \end{aligned} \quad (23)$$

and we set $b^{(x,i)} = \left(b_1^{(x,i)}, \dots, b_K^{(x,i)} \right)'$ and $b^{(z,i)} = \left(b_1^{(z,i)}, \dots, b_M^{(z,i)} \right)'$, for $i = 0, 1, 2$. With these definitions we can now give the expression for the asymptotic bias terms which appear in Theorem 3, namely

$$B_i = - (G \mathcal{W} G')^{-1} G \mathcal{W} \begin{pmatrix} b^{(x,i)} \\ b^{(z,i)} \end{pmatrix}, \quad (24)$$

where $i = 0, 1, 2$.

B.2 Assumptions for Consistency

Assumption 5 *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 6

- (i) $\frac{1}{JT} \text{Tr}(e^0 X'_k) = o_p(1)$, for $k = 1, \dots, K$,
 $\frac{1}{JT} \text{Tr}(e^0 Z'_m) = o_p(1)$, for $m = 1, \dots, M$.
(ii) $\|e^0\| = \mathcal{O}_p(\max(J, T))$.

Assumption 7

- (i) $\sup_{\alpha \in \mathcal{B}_\alpha \setminus \alpha^0} \frac{\|\delta(\alpha) - \delta(\alpha^0)\|_F}{\|\alpha - \alpha^0\|} = \mathcal{O}_p(\sqrt{JT})$.
(ii) $W_{JT} \xrightarrow{p} W > 0$.

Assumption 8

- (a) Let $\Xi_{jt} = (X_{1,jt}, \dots, X_{K,jt}, Z_{1,jt}, \dots, Z_{M,jt})'$ be the $(K + M)$ -vectors of regressors and instruments that appear in step 1 of (7). We assume that the probability limit of the $(K + M) \times (K + M)$ matrix $(JT)^{-1} \sum_{j,t} \Xi_{jt} \Xi'_{jt}$ exists and is positive definite, i.e. $\text{plim}_{J,T \rightarrow \infty} \left[(JT)^{-1} \sum_{j=1}^J \sum_{t=1}^T \Xi_{jt} \Xi'_{jt} \right] > 0$.
- (b) We assume that the K regressors X can be decomposed into n “low-rank regressors” X_{low} and $K - n$ “high-rank regressors” X_{high} . The two types of regressors satisfy:
- (i) For $\rho \in \mathbb{R}^{K+M-n}$ define the $J \times T$ matrix $\Xi_{\text{high},\rho} = \sum_{m=1}^M \rho_m Z_m + \sum_{k=1}^{K-n} \rho_{M+k} X_{\text{high},k}$, which is a linear combination of high-rank regressors and instruments. We assume that there exists a constant $b > 0$ such that
- $$\min_{\{\rho \in \mathbb{R}^{K+M-n}, \|\rho\|=1\}} \sum_{t=2R+n+1}^T \mu_t \left(\frac{\Xi'_{\text{high},\rho} \Xi_{\text{high},\rho}}{JT} \right) \geq b \quad \text{wpa1.}$$
- (ii) For the low-rank regressors we assume $\text{rank}(\Xi_{\text{low},k}) = 1$, $k = 1, \dots, n$, i.e. they can be written as $X_k = w_k v'_k$ for $J \times 1$ vectors w_k and $T \times 1$ vectors v_k , and we define the $J \times n$ matrix $w = (w_1, \dots, w_n)$ and the $T \times n$ matrix $v = (v_1, \dots, v_n)$. We assume that there exists $B > 0$ such that $J^{-1} \lambda^0 M_v \lambda^0 > B \mathbb{I}_R$ wpa1, and $T^{-1} f^0 M_w f^0 > B \mathbb{I}_R$ wpa1.

Assumption 5 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 6 imposes (i) weak exogeneity of X_k and Z_m wrt e^0 , and (ii) a bound on the the spectral norm of e^0 , which

is satisfied as long as e_{jt}^0 has mean zero, has a uniformly bounded fourth moment (across j, t, J, T) and is weakly correlated across j and t . Assumption 7(i) demands that a bound on the Frobenius norm of $(\delta(\alpha) - \delta(\alpha^0))/\|\alpha - \alpha^0\|$ exists, which is satisfied as long as e.g. the elements $(\delta_{jt}(\alpha) - \delta_{jt}(\alpha^0))/\|\alpha - \alpha^0\|$ are uniformly bounded (across j, t, J, T). Assumption 7(ii) requires existence of a positive definite probability limit of the weight matrix W_{JT} .

Assumption 8(a) is the standard non-colinearity assumption on the regressors X_k and the instruments Z_m . As discussed in Bai (2009) and Moon and Weidner (2009) just assuming weak exogeneity and non-colinearity is not sufficient for consistency of the QMLE in the presence of factors, and the same is true here. In particular, in a model with factors one needs to distinguish so called “low-rank regressors” and “high-rank regressors” and treat them differently. This distinction is introduced in Assumption 8(b) and additional assumptions on the low- and high-rank regressors are imposed. Low-rank regressors are for example regressors that are constant over either markets t or products j , or more generally factor into a component that depends only on j and a component that depends only on t . All other regressors are usually high-rank regressors. Assumption 8 in this paper is equivalent to Assumption 4 in Moon and Weidner (2009), and some further discussion can be found there. If there are no low-rank regressors (if $n = 0$ in Assumption 8) then Theorem 2 holds even without imposing Assumption 5, i.e. also when factors are “weak”.

B.3 Additional Assumptions for Asymptotic Distribution and Bias Correction

Assumption 9 *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 $G\mathcal{W}G' > 0$ and $G\mathcal{W}\Omega\mathcal{W}G' > 0$.*

Assumption 10

(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}^0 are independent across j and t , they satisfy $\mathbb{E}e_{jt}^0 = 0$, and $\mathbb{E}(e_{jt}^0)^{8+\epsilon}$ is bounded uniformly across j, t and J, T , for some $\epsilon > 0$.
- (iv) The regressors X_k , $k = 1, \dots, K$, (both high- and low rank regressors) and the instruments Z_m , $m = 1, \dots, M$, can be decomposed as $X_k = X_k^{\text{str}} + X_k^{\text{weak}}$ and $Z_m = Z_m^{\text{str}} + Z_m^{\text{weak}}$. The components X_k^{str} and Z_m^{str} are strictly exogenous, i.e. $X_{k,jt}^{\text{str}}$ and $Z_{m,jt}^{\text{str}}$ are independent of $e_{i\tau}^0$ for all j, i, t, τ . The components X_k^{weak} and Z_m^{weak} are weakly exogenous, and we assume

$$X_{k,jt}^{\text{weak}} = \sum_{\tau=1}^{t-1} c_{k,j\tau} e_{j,t-\tau}^0, \quad Z_{m,jt}^{\text{weak}} = \sum_{\tau=1}^{t-1} d_{m,j\tau} e_{j,t-\tau}^0,$$

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 6 is implied by Assumption 10, so it is not necessary to impose it explicitly in Theorem 3. Part (ii), (iii) and (iv) of Assumption 10 are identical to Assumption 5 in Moon and Weidner (2009), 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 10 can for example be justified by assuming that within any \sqrt{J} -shrinking neighborhood of α^0 we have wpa1 that $\delta_{jt}(\alpha)$ is differentiable, that $|\nabla_l \delta_{jt}(\alpha)|$ is uniformly bounded across j, t, J and T , and that $\nabla_l \delta_{jt}(\alpha)$ is Lipschitz continuous with a Lipschitz constant that is uniformly bounded across j, t, J and T , for all $l = 1, \dots, L$. But since the assumption is only on the Frobenius norm of the gradient and remainder term, one can also conceive weaker sufficient conditions for Assumption 10(i).

Assumption 11 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.4 Bias and Variance Estimators

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

$$\hat{e} = \delta(\hat{\alpha}, s, X) - \sum_{k=1}^K \hat{\beta}_k X_k - \hat{\lambda} f'. \quad (25)$$

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

$$\hat{x}_{.,k}^{\lambda f} = \text{vec} \left(M_{\hat{\lambda}} X_k M_{\hat{f}} \right), \quad \hat{z}_{.,m}^{\lambda f} = \text{vec} \left(M_{\hat{\lambda}} Z_m M_{\hat{f}} \right), \quad \hat{g}_{.,l} = -\text{vec} \left(\nabla_l \delta(\hat{\alpha}) \right), \quad (26)$$

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 Σ_e^{vec} , $\Sigma_e^{(1)}$ and $\Sigma_e^{(2)}$, but with e replaced by \hat{e} . 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

$$\hat{\Sigma}_{k,t\tau}^{X,e} = \begin{cases} \frac{1}{J} \sum_{j=1}^J X_{k,jt} \hat{e}_{j\tau} & \text{for } 0 < t - \tau \leq h \\ 0 & \text{otherwise} \end{cases}$$

$$\hat{\Sigma}_{m,t\tau}^{Z,e} = \begin{cases} \frac{1}{J} \sum_{j=1}^J Z_{m,jt} \hat{e}_{j\tau} & \text{for } 0 < t - \tau \leq h \\ 0 & \text{otherwise} \end{cases} \quad (27)$$

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

$$\begin{aligned}
\hat{G} &= \frac{1}{JT} \begin{pmatrix} \hat{g}' \hat{x}^{\lambda f} & \hat{g}' \hat{z}^{\lambda f} \\ \hat{x}^{\lambda f'} \hat{x}^{\lambda f} & \hat{x}^{\lambda f'} \hat{z}^{\lambda f} \end{pmatrix}, \\
\hat{\Omega} &= \frac{1}{JT} (\hat{x}^{\lambda f}, \hat{z}^{\lambda f})' \text{diag}(\hat{\Sigma}_e^{\text{vec}}) (\hat{x}^{\lambda f}, \hat{z}^{\lambda f}), \\
\hat{b}_k^{(x,0)} &= \text{Tr} \left(P_{\hat{f}} \hat{\Sigma}_k^{X,e} \right), \\
\hat{b}_k^{(x,1)} &= \text{Tr} \left[\text{diag} \left(\hat{\Sigma}_e^{(1)} \right) M_{\hat{\lambda}} X_k \hat{f} (\hat{f}' \hat{f})^{-1} (\hat{\lambda}' \hat{\lambda})^{-1} \hat{\lambda}' \right], \\
\hat{b}_k^{(x,2)} &= \text{Tr} \left[\text{diag} \left(\hat{\Sigma}_e^{(2)} \right) M_{\hat{f}} X'_k \hat{\lambda} (\hat{\lambda}' \hat{\lambda})^{-1} (\hat{f}' \hat{f})^{-1} \hat{f}' \right], \\
\hat{b}_m^{(z,0)} &= \text{Tr} \left(P_{\hat{f}} \hat{\Sigma}_m^{Z,e} \right), \\
\hat{b}_m^{(z,1)} &= \text{Tr} \left[\text{diag} \left(\hat{\Sigma}_e^{(1)} \right) M_{\hat{\lambda}} Z_m \hat{f} (\hat{f}' \hat{f})^{-1} (\hat{\lambda}' \hat{\lambda})^{-1} \hat{\lambda}' \right], \\
\hat{b}_m^{(z,2)} &= \text{Tr} \left[\text{diag} \left(\hat{\Sigma}_e^{(2)} \right) M_{\hat{f}} Z'_m \hat{\lambda} (\hat{\lambda}' \hat{\lambda})^{-1} (\hat{f}' \hat{f})^{-1} \hat{f}' \right], \tag{28}
\end{aligned}$$

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

$$\begin{aligned}
\widehat{\mathcal{W}} &= \begin{pmatrix} \left(\frac{1}{JT} \hat{x}^{\lambda f'} \hat{x}^{\lambda f} \right)^{-1} & 0_{K \times M} \\ 0_{M \times K} & 0_{M \times M} \end{pmatrix} + \begin{pmatrix} -(\hat{x}^{\lambda f'} \hat{x}^{\lambda f})^{-1} \hat{x}^{\lambda f'} \hat{z}^{\lambda f} \\ \mathbb{1}_M \end{pmatrix} \left(\frac{1}{JT} \hat{z}^{\lambda f'} M_{\hat{x}^{\lambda f}} \hat{z}^{\lambda f} \right)^{-1} \\
&\quad W_{JT} \left(\frac{1}{JT} \hat{z}^{\lambda f'} M_{\hat{x}^{\lambda f}} \hat{z}^{\lambda f} \right)^{-1} \begin{pmatrix} -(\hat{x}^{\lambda f'} \hat{x}^{\lambda f})^{-1} \hat{x}^{\lambda f'} \hat{z}^{\lambda f} \\ \mathbb{1}_M \end{pmatrix}'. \tag{29}
\end{aligned}$$

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

$$\hat{B}_i = - \left(\hat{G} \widehat{\mathcal{W}} \hat{G}' \right)^{-1} \hat{G} \widehat{\mathcal{W}} \begin{pmatrix} \hat{b}^{(x,i)} \\ \hat{b}^{(z,i)} \end{pmatrix}. \tag{30}$$

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^0 , 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.

Appendix C Proofs

In addition to the vectorizations x , x^f , $x^{\lambda f}$, z , z^f , $z^{\lambda f}$, g , and $d(\alpha)$, which were already defined above, we also introduce the $JT \times 1$ vector $\varepsilon = \text{vec}(e^0)$.

C.1 Proof of Consistency

Proof of Theorem 2: From Bai (2009) and Moon and Weidner (2009) we know that for $\alpha = \alpha^0$ one has $\tilde{\gamma}_{\alpha^0} = o_p(1)$. Since the optimal choice $\hat{\alpha}$ minimizes $\tilde{\gamma}'_{\hat{\alpha}} W_{JT} \tilde{\gamma}_{\hat{\alpha}}$ we have

$$\tilde{\gamma}'_{\hat{\alpha}} W_{JT} \tilde{\gamma}_{\hat{\alpha}} \leq \tilde{\gamma}'_{\alpha^0} W_{JT} \tilde{\gamma}_{\alpha^0} = o_p(1), \quad (31)$$

and therefore $\tilde{\gamma}_{\hat{\alpha}} = o_p(1)$, since W_{JT} converges to a positive definite matrix in probability. After minimization over λ and f , the objective function for step 1 optimization reads

$$\begin{aligned} L_\alpha(\beta, \gamma) &= \min_{\lambda, f} \frac{1}{JT} \text{Tr} \left[\left(\delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m - \lambda f' \right) \right. \\ &\quad \left. \left(\delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m - \lambda f' \right)' \right] \\ &= \min_f \frac{1}{JT} \text{Tr} \left[\left(\delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m \right) M_f \left(\delta(\alpha) - \sum_{k=1}^K \beta_k X_k - \sum_{m=1}^M \gamma_m Z_m \right)' \right] \\ &= \min_f \frac{1}{JT} \text{Tr} \left[\left(\lambda^0 f^{0'} + \delta(\alpha) - \delta(\alpha^0) - \sum_{k=1}^K (\beta_k - \beta^0) X_k - \sum_{m=1}^M \gamma_m Z_m + e^0 \right) M_f \right. \\ &\quad \left. \left(\lambda^0 f^{0'} + \delta(\alpha) - \delta(\alpha^0) - \sum_{k=1}^K (\beta_k - \beta^0) X_k - \sum_{m=1}^M \gamma_m Z_m + e^0 \right)' \right]. \quad (32) \end{aligned}$$

Defining

$$\begin{aligned} L_\alpha^{\text{up}}(\beta, \gamma) &= \frac{1}{JT} \text{Tr} \left[\left(\delta(\alpha) - \delta(\alpha^0) - \sum_{k=1}^K (\beta_k - \beta^0) X_k - \sum_{m=1}^M \gamma_m Z_m + e^0 \right) M_{f^0} \right. \\ &\quad \left. \left(\delta(\alpha) - \delta(\alpha^0) - \sum_{k=1}^K (\beta_k - \beta^0) X_k - \sum_{m=1}^M \gamma_m Z_m + e^0 \right)' \right], \\ L_\alpha^{\text{low}}(\beta, \gamma) &= \min_\lambda \frac{1}{JT} \text{Tr} \left[M_\lambda \left(\delta(\alpha) - \delta(\alpha^0) - \sum_{k=1}^K (\beta_k - \beta^0) X_k - \sum_{m=1}^M \gamma_m Z_m + e^0 \right) M_{f^0} \right. \\ &\quad \left. \left(\delta(\alpha) - \delta(\alpha^0) - \sum_{k=1}^K (\beta_k - \beta^0) X_k - \sum_{m=1}^M \gamma_m Z_m + e^0 \right)' \right], \quad (33) \end{aligned}$$

we have for all β, γ

$$L_\alpha^{\text{low}}(\beta, \gamma) \leq L_\alpha(\beta, \gamma) \leq L_\alpha^{\text{up}}(\beta, \gamma). \quad (34)$$

Here, for the upper bound we simply choose $f = f^0$ in the minimization problem of $L_\alpha(\beta, \gamma)$. We arrive at the lower bound by starting with the dual formulation of $L_\alpha(\beta, \gamma)$ — in which we minimize over λ —

and subtract the term with P_{f^0} , which can be written as the trace of a positive definite matrix. Due to the projection with M_{f^0} the $\lambda^0 f^{0'}$ term drops out of both $L_\alpha^{\text{low}}(\beta, \gamma)$ and $L_\alpha^{\text{up}}(\beta, \gamma)$.

Let $(\tilde{\beta}_\alpha, \tilde{\gamma}_\alpha)$ be the minimizing parameters of $L_\alpha(\beta, \gamma)$ given α ; let $(\tilde{\beta}_\alpha^{\text{up}}, \tilde{\gamma}_\alpha^{\text{up}})$ be the minimizing parameters of $L_\alpha^{\text{up}}(\beta, \gamma)$ given α ; and let $\tilde{\beta}_{\alpha, \gamma}^{\text{low}}$ be the minimizing parameter of $L_\alpha^{\text{low}}(\beta, \gamma)$ given α and γ . We then have

$$L_\alpha^{\text{low}}(\tilde{\beta}_{\alpha, \gamma}^{\text{low}}, \tilde{\gamma}_\alpha) \leq L_\alpha^{\text{low}}(\tilde{\beta}_\alpha, \tilde{\gamma}_\alpha) \leq L_\alpha(\tilde{\beta}_\alpha, \tilde{\gamma}_\alpha) \leq L_\alpha(\tilde{\beta}_\alpha^{\text{up}}, \tilde{\gamma}_\alpha^{\text{up}}) \leq L_\alpha^{\text{up}}(\tilde{\beta}_\alpha^{\text{up}}, \tilde{\gamma}_\alpha^{\text{up}}). \quad (35)$$

Using the vectorizations of X_k , Z_m and e^0 , we can rewrite the lower and upper bound in vector notation:

$$\begin{aligned} L_\alpha^{\text{up}}(\beta, \gamma) &= \frac{1}{JT} (d(\alpha) - x(\beta - \beta^0) - z\gamma + \varepsilon)' (M_{f^0} \otimes \mathbb{1}_J) (d(\alpha) - x(\beta - \beta^0) - z\gamma + \varepsilon), \\ L_\alpha^{\text{low}}(\tilde{\beta}_{\alpha, \gamma}^{\text{low}}, \gamma) &= \min_{\beta, \lambda} \frac{1}{JT} (d(\alpha) - x(\beta - \beta^0) - z\gamma + \varepsilon)' (M_{f^0} \otimes M_\lambda) (d(\alpha) - x(\beta - \beta^0) - z\gamma + \varepsilon). \end{aligned} \quad (36)$$

For given λ , let $\tilde{\beta}_{\alpha, \gamma, \lambda}^{\text{low}}$ be the optimal β in the last equation. We have

$$\begin{aligned} \begin{pmatrix} \tilde{\beta}_\alpha^{\text{up}} \\ \tilde{\gamma}_\alpha^{\text{up}} \end{pmatrix} &= [(x^f, z^f)'(x^f, z^f)]^{-1} (x^f, z^f)' (d(\alpha) + \varepsilon), \\ \tilde{\beta}_{\alpha, \gamma, \lambda}^{\text{low}} &= [x' (M_{f^0} \otimes M_\lambda) x]^{-1} x' (M_{f^0} \otimes M_\lambda) (d(\alpha) - z\gamma + \varepsilon), \end{aligned} \quad (37)$$

and therefore

$$\begin{aligned} L_\alpha^{\text{up}}(\tilde{\beta}_\alpha^{\text{up}}, \tilde{\gamma}_\alpha^{\text{up}}) &= \frac{1}{JT} (d^f(\alpha) + \varepsilon^f)' M_{(x^f, z^f)} (d^f(\alpha) + \varepsilon^f) \\ &= \frac{1}{JT} (d^f(\alpha) + \varepsilon^f)' (d^f(\alpha) + \varepsilon^f) - \frac{1}{JT} d'(\alpha) P_{(x^f, z^f)} d(\alpha) - R_1(\alpha), \\ L_\alpha^{\text{low}}(\tilde{\beta}_{\alpha, \gamma}^{\text{low}}, \gamma) &= \min_{\lambda} \frac{1}{JT} (d^f(\alpha) - z^f \gamma + \varepsilon^f)' M_{(x^f, M_{f^0} \otimes \lambda)} (d^f(\alpha) - z^f \gamma + \varepsilon^f) \\ &= \frac{1}{JT} (d^f(\alpha) - z^f \gamma + \varepsilon^f)' (d^f(\alpha) - z^f \gamma + \varepsilon^f) \\ &\quad - \max_{\lambda} \frac{1}{JT} (d(\alpha) - z\gamma + \varepsilon)' P_{(x^f, M_{f^0} \otimes \lambda)} (d(\alpha) - z\gamma + \varepsilon) \\ &= \frac{1}{JT} (d^f(\alpha) + \varepsilon^f)' (d^f(\alpha) + \varepsilon^f) \\ &\quad - \max_{\lambda} \left[\frac{1}{JT} (d(\alpha) - z\gamma)' P_{(x^f, M_{f^0} \otimes \lambda)} (d(\alpha) - z\gamma) + R_2(\alpha, \gamma, \lambda) \right], \end{aligned} \quad (38)$$

where $\varepsilon^f(\alpha) = \text{vec}(e^0 M_{f^0})$, $d^f(\alpha) = \text{vec}((\delta(\alpha) - \delta(\alpha^0)) M_{f^0})$, and the remainder terms $R_1(\alpha)$ and $R_2(\alpha, \gamma, \lambda)$ are given by

$$\begin{aligned} R_1(\alpha) &= \frac{2}{JT} d'(\alpha) P_{(x^f, z^f)} \varepsilon + \frac{1}{JT} \varepsilon' P_{(x^f, z^f)} \varepsilon, \\ R_2(\alpha, \gamma, \lambda) &= \frac{2}{JT} (d(\alpha) - z\gamma)' P_{(x^f, M_{f^0} \otimes \lambda)} \varepsilon + \frac{1}{JT} \varepsilon' P_{(x^f, M_{f^0} \otimes \lambda)} \varepsilon + \frac{1}{JT} (d^f(\alpha) + \varepsilon^f)' z\gamma. \end{aligned} \quad (39)$$

The inequality $L_\alpha^{\text{low}}(\tilde{\beta}_{\alpha, \gamma}^{\text{low}}, \tilde{\gamma}_\alpha) \leq L_\alpha^{\text{up}}(\tilde{\beta}_\alpha^{\text{up}}, \tilde{\gamma}_\alpha^{\text{up}})$ evaluated at $\alpha = \hat{\alpha}$ thus gives

$$\begin{aligned} R_2(\hat{\alpha}, \tilde{\gamma}_{\hat{\alpha}}, \tilde{\lambda}) - R_1(\hat{\alpha}) &\geq \left(\frac{1}{JT} d'(\hat{\alpha}) P_{(x^f, z^f)} d(\hat{\alpha}) - \max_{\lambda} \frac{1}{JT} d'(\hat{\alpha}) P_{(x^f, M_{f^0} \otimes \lambda)} d(\hat{\alpha}) \right) \\ &\geq c \|\hat{\alpha} - \alpha^0\|^2, \quad \text{wpa1,} \end{aligned} \quad (40)$$

where $\tilde{\lambda}$ is the optimal choice of λ in $L_{\tilde{\alpha}}^{\text{low}}(\tilde{\beta}_{\tilde{\alpha}, \tilde{\gamma}_{\tilde{\alpha}}}^{\text{low}}, \tilde{\gamma}_{\tilde{\alpha}})$, and we used Assumption 1. Assumption 8(i) implies $\|X_k\| = \mathcal{O}_p(\sqrt{JT})$ and $\|Z_m\| = \mathcal{O}_p(\sqrt{JT})$. Using this and Assumption 6 we find $(x^f, z^f)' \varepsilon = o_p(JT)$ and therefore

$$\|P_{(x^f, z^f)} \varepsilon\|_F = \left\| (x^f, z^f) [(x^f, z^f)'(x^f, z^f)]^{-1} (x^f, z^f)' \varepsilon \right\|_F = o_p(\sqrt{JT}), \quad (41)$$

where we also used that Assumption 8 guarantees $[(x^f, z^f)'(x^f, z^f)]^{-1} = \mathcal{O}_p(1/\sqrt{JT})$. Below we also show that $\|P_{(x^f, M_{f^0} \otimes \lambda)} \varepsilon\|_F = o_p(\sqrt{JT})$. Using these results, Assumption 7(i) and the fact that $\tilde{\gamma}_{\tilde{\alpha}} = o_p(1)$ one obtains

$$R_2(\hat{\alpha}, \tilde{\gamma}_{\hat{\alpha}}, \tilde{\lambda}) = o_p(1) + o_p(\|\hat{\alpha} - \alpha^0\|), \quad R_1(\hat{\alpha}) = o_p(1) + o_p(\|\hat{\alpha} - \alpha^0\|). \quad (42)$$

Therefore we have

$$o_p(1) + o_p(\|\hat{\alpha} - \alpha^0\|) \geq c\|\hat{\alpha} - \alpha^0\|^2, \quad (43)$$

which implies $\hat{\alpha} - \alpha^0 = o_p(1)$.

What is left to show is that $\|P_{(x^f, M_{f^0} \otimes \lambda)} \varepsilon\|_F = o_p(\sqrt{JT})$. For $A = x^f$ and $B = M_{f^0} \otimes \lambda$ we use the general formula $P_{(A, B)} = P_B + M_B + P_{(M_B A)} M_B$ and the fact that $M_{(M_{f^0} \otimes \lambda)} x^f = M_{\mathbb{1}_T \otimes \lambda} x^f$ to obtain

$$P_{(x^f, M_{f^0} \otimes \lambda)} = P_{(M_{f^0} \otimes \lambda)} + M_{(M_{f^0} \otimes \lambda)} P_{[M_{(\mathbb{1}_T \otimes \lambda)} x^f]} - M_{(M_{f^0} \otimes \lambda)} P_{[M_{(\mathbb{1}_T \otimes \lambda)} x^f]} P_{(M_{f^0} \otimes \lambda)}, \quad (44)$$

and therefore

$$\left\| P_{(x^f, M_{f^0} \otimes \lambda)} \varepsilon \right\|_F \leq 2 \left\| P_{(M_{f^0} \otimes \lambda)} \varepsilon \right\|_F + \left\| P_{[M_{(\mathbb{1}_T \otimes \lambda)} x^f]} \varepsilon \right\|_F \quad (45)$$

Note that the $JT \times K$ matrix $M_{(\mathbb{1}_T \otimes \lambda)} x^f$ is simply $\text{vec}(M_{\lambda} X_k M_{f^0})_{k=1, \dots, K}$. We have

$$\left\| P_{(M_{f^0} \otimes \lambda)} \varepsilon \right\|_F = \|P_{\lambda} e^0 M_{f^0}\|_F \leq R \|e^0\| = o_p(\sqrt{JT}), \quad (46)$$

and thus also $x^{f'} M_{(\mathbb{1}_T \otimes \lambda)} \varepsilon = o_p(JT)$, which analogous to (41) implies $\left\| P_{[M_{(\mathbb{1}_T \otimes \lambda)} x^f]} \varepsilon \right\|_F = o_p(\sqrt{JT})$, so that the required result follows.

Having $\hat{\alpha} = \alpha^0 + o_p(1)$, the proof of $\hat{\beta} = \beta^0 + o_p(1)$ is straightforward using the methods in Bai (2009) and Moon and Weidner (2009) — the only additional term appearing here is $\delta(\hat{\alpha}) - \delta(\alpha^0)$, which for $\|\hat{\alpha} - \alpha^0\| = o_p(1)$ has no effect on the consistency of $\hat{\beta}$. \blacksquare

C.2 Proof of Limiting Distribution

Lemma 12 *Let the assumptions of Theorem 2 (consistency) be satisfied and in addition let $(JT)^{-1/2} \text{Tr}(e^0 X_k') = \mathcal{O}_p(1)$, and $(JT)^{-1/2} \text{Tr}(e^0 Z_m') = \mathcal{O}_p(1)$. 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 analogous to the consistency proof. We know from Moon and Weidner (2009) that $\sqrt{J} \tilde{\gamma}_{\alpha^0} = \mathcal{O}_p(1)$ and therefore $\sqrt{J} \tilde{\gamma}_{\hat{\alpha}} = \mathcal{O}_p(1)$, applying the same logic as in equation (31).

With the additional assumptions in the lemma we furthermore obtain $\|P_{(x^f, z^f)}\varepsilon\|_F = \mathcal{O}_p(\sqrt{J})$ and $\|P_{(x^f, M_{f^0} \otimes \lambda)}\varepsilon\|_F = \mathcal{O}_p(\sqrt{J})$ and can conclude

$$R_2(\hat{\alpha}, \tilde{\gamma}_{\hat{\alpha}}, \tilde{\lambda}) - R_1(\hat{\alpha}) = \mathcal{O}_p(J^{-1}) + \mathcal{O}_p(J^{-1/2} \|\hat{\alpha} - \alpha^0\|). \quad (47)$$

This implies

$$\mathcal{O}_p(J^{-1}) + \mathcal{O}_p(J^{-1/2} \|\hat{\alpha} - \alpha^0\|) \geq c \|\hat{\alpha} - \alpha^0\|^2, \quad (48)$$

so that we obtain $\sqrt{J}(\hat{\alpha} - \alpha) = \mathcal{O}_p(1)$. \blacksquare

Proof of Theorem 3: Assumption 10 guarantees $(JT)^{-1/2}\text{Tr}(e^0 X'_k) = \mathcal{O}_p(1)$, and $(JT)^{-1/2}\text{Tr}(e^0 Z'_m) = \mathcal{O}_p(1)$, so that we can apply Lemma 12 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 $\Delta\delta(\alpha) \equiv \delta(\alpha) - \delta(\alpha^0)$, namely $E(\alpha) \equiv e + \Delta\delta(\alpha)$. Using $\hat{\alpha} - \alpha^0 = o_p(1)$ and Assumption 7(i) we have $\|E(\hat{\alpha})\| = o_p(\sqrt{JT})$, so that the results in Moon and Weidner (2009) 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 10(i) we find $\|E(\hat{\alpha})\| = \mathcal{O}_p(\sqrt{J})$, which allows us to truncate the asymptotic likelihood expansion derived in Moon and Weidner (2009) at an appropriate order. Namely, applying their results we have

$$\sqrt{JT} \begin{pmatrix} \tilde{\beta}_{\hat{\alpha}} - \beta^0 \\ \tilde{\gamma}_{\hat{\alpha}} \end{pmatrix} = V_{JT}^{-1} \begin{pmatrix} [C^{(1)}(X_k, E(\alpha)) + C^{(2)}(X_k, E(\alpha))]_{k=1, \dots, K} \\ [C^{(1)}(Z_m, E(\alpha)) + C^{(2)}(Z_m, E(\alpha))]_{m=1, \dots, M} \end{pmatrix} + r^{\text{QMLE}}(\alpha), \quad (49)$$

where

$$\begin{aligned} V_{JT} &= \frac{1}{JT} \begin{pmatrix} [\text{Tr}(M_{f^0} X'_{k_1} M_{\lambda^0} X_{k_2})]_{k_1, k_2=1, \dots, K} & [\text{Tr}(M_{f^0} X'_k M_{\lambda^0} Z_m)]_{k=1, \dots, K; m=1, \dots, M} \\ [\text{Tr}(M_{f^0} Z'_m M_{\lambda^0} X_k)]_{m=1, \dots, M; k=1, \dots, K} & [\text{Tr}(M_{f^0} Z'_{m_1} M_{\lambda^0} Z_{m_2})]_{m_1, m_2=1, \dots, M} \end{pmatrix} \\ &= \frac{1}{JT} (x^{\lambda^f}, z^{\lambda^f})' (x^{\lambda^f}, z^{\lambda^f}), \end{aligned} \quad (50)$$

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

$$\begin{aligned} C^{(1)}(\mathcal{X}, E) &= \frac{1}{\sqrt{JT}} \text{Tr} [M_{f^0} E' M_{\lambda^0} \mathcal{X}], \\ C^{(2)}(\mathcal{X}, E) &= -\frac{1}{\sqrt{JT}} \left[\text{Tr} (EM_{f^0} E' M_{\lambda^0} \mathcal{X} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'}) \right. \\ &\quad + \text{Tr} (E' M_{\lambda^0} E M_{f^0} \mathcal{X}' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'}) \\ &\quad \left. + \text{Tr} (E' M_{\lambda^0} \mathcal{X} M_{f^0} E' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'}) \right], \end{aligned} \quad (51)$$

and finally for the remainder we have

$$\begin{aligned} r^{\text{QMLE}}(\alpha) &= \mathcal{O}_p \left((JT)^{-3/2} \|E(\alpha)\|^3 \|X_k\| \right) + \mathcal{O}_p \left((JT)^{-3/2} \|E(\alpha)\|^3 \|Z_m\| \right) \\ &\quad + \mathcal{O}_p \left((JT)^{-1} \|E(\alpha)\| \|X_k\|^2 \|\tilde{\beta}_{\hat{\alpha}} - \beta^0\| \right) + \mathcal{O}_p \left((JT)^{-1} \|E(\alpha)\| \|Z_m\|^2 \|\tilde{\gamma}_{\hat{\alpha}}\| \right), \end{aligned} \quad (52)$$

which holds uniformly over α . 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 $E(\alpha)$ 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 8 already guarantees that $V_{JT} > b > 0$, wpa1. Applying $\|X_k\| = \mathcal{O}_p(\sqrt{JT})$, $\|Z_m\| = \mathcal{O}_p(\sqrt{JT})$, and $\|E(\alpha)\| = \mathcal{O}_p(\sqrt{J})$ within $\sqrt{J}\|\alpha - \alpha^0\| < c$, we find for all $c > 0$

$$\sup_{\{\alpha: \sqrt{J}\|\alpha - \alpha^0\| < c\}} \frac{\|r^{\text{QMLE}}(\alpha)\|}{1 + \sqrt{JT}\|\tilde{\beta}_\alpha - \beta^0\| + \sqrt{JT}\|\tilde{\gamma}_\alpha\|} = o_p(1). \quad (53)$$

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

$$V_{JT}^{-1} = JT \begin{pmatrix} (x^{\lambda f'} M_{z^{\lambda f}} x^{\lambda f})^{-1} & - (x^{\lambda f'} M_{z^{\lambda f}} x^{\lambda f})^{-1} (x^{\lambda f'} z^{\lambda f}) (z^{\lambda f'} z^{\lambda f})^{-1} \\ - (z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f})^{-1} (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} & (z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f})^{-1} \end{pmatrix}. \quad (54)$$

Using $\sqrt{J}(\hat{\alpha} - \alpha) = \mathcal{O}_p(1)$ and Assumption 10(i) we find

$$\begin{pmatrix} [C^{(1)}(X_k, E(\hat{\alpha}))]_{k=1, \dots, K} \\ [C^{(1)}(Z_m, E(\hat{\alpha}))]_{m=1, \dots, M} \end{pmatrix} = \frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon - \left[\frac{1}{JT} (x^{\lambda f}, z^{\lambda f})' g \right] \sqrt{JT}(\hat{\alpha} - \alpha^0) + o_p(\sqrt{JT}\|\hat{\alpha} - \alpha^0\|), \\ \begin{pmatrix} [C^{(2)}(X_k, E(\hat{\alpha}))]_{k=1, \dots, K} \\ [C^{(2)}(Z_m, E(\hat{\alpha}))]_{m=1, \dots, M} \end{pmatrix} = \begin{pmatrix} c_x^{(2)} \\ c_z^{(2)} \end{pmatrix} + \mathcal{O}_p(\sqrt{J}\|\hat{\alpha} - \alpha^0\|), \quad (55)$$

where

$$c_x^{(2)} = [C^{(2)}(X_k, e)]_{k=1, \dots, K}, \quad c_z^{(2)} = [C^{(2)}(Z_m, e)]_{m=1, \dots, M}. \quad (56)$$

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

$$\sqrt{JT} \tilde{\gamma}_{\hat{\alpha}} = \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left[\frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} - \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} g \right) \sqrt{JT}(\hat{\alpha} - \alpha^0) \right] + o_p(1) + o_p(\sqrt{JT}\|\hat{\alpha} - \alpha^0\|). \quad (57)$$

The above results holds not only for $\hat{\alpha}$, but uniformly for all α in any \sqrt{J} shrinking neighborhood of α^0 (we still made this explicit in the bound on $r^{\text{QMLE}}(\alpha)$ above, but found it too tedious to define remainder terms in all intermediate steps), i.e. we have

$$\sqrt{JT} \tilde{\gamma}_\alpha = \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left[\frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} - \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} g \right) \sqrt{JT}(\alpha - \alpha^0) \right] + r^\gamma(\alpha), \quad (58)$$

where for all $c > 0$

$$\sup_{\{\alpha: \sqrt{J}\|\alpha - \alpha^0\| < c\}} \frac{\|r^\gamma(\alpha)\|}{1 + \sqrt{JT}\|\alpha - \alpha^0\|} = o_p(1). \quad (59)$$

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

$$JT \tilde{\gamma}'_{\alpha} W_{JT} \tilde{\gamma}_{\alpha} = A_0 - 2 A_1' \left[\sqrt{JT} (\alpha - \alpha^0) \right] + \left[\sqrt{JT} (\alpha - \alpha^0) \right]' A_2 \left[\sqrt{JT} (\alpha - \alpha^0) \right] + r^{\text{obj}}(\alpha), \quad (60)$$

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

$$\begin{aligned} A_0 &= \left[\frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right]' \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \\ &\quad \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left[\frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right], \\ A_1 &= \left(\frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f} \right) \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \\ &\quad \left[\frac{1}{\sqrt{JT}} z^{\lambda f'} M_{x^{\lambda f}} \varepsilon + c_z^{(2)} - (z^{\lambda f'} x^{\lambda f}) (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \right], \\ A_2 &= \left(\frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f} \right) \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} g \right), \end{aligned} \quad (61)$$

and the remainder term in the objective function satisfies

$$\sup_{\{\alpha: \sqrt{J}\|\alpha - \alpha^0\| < c\}} \frac{\|r^{\text{obj}}(\alpha)\|}{\left(1 + \sqrt{JT}\|\alpha - \alpha^0\|\right)^2} = o_p(1). \quad (62)$$

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

$$\sqrt{JT} (\hat{\alpha} - \alpha^0) = A_2^{-1} A_1 + o_p(1). \quad (63)$$

Analogous to equation (49) for the first step, we can apply the results in Moon and Weidner (2009) to the third step of the LS-MD estimator to obtain

$$\begin{aligned} \sqrt{JT}(\hat{\beta} - \beta^0) &= \left(\frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} \left[C^{(1)}(X_k, E(\hat{\alpha})) + C^{(2)}(X_k, E(\hat{\alpha})) \right]_{k=1, \dots, K} + o_p(1) \\ &= \left(\frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} \left[\frac{1}{\sqrt{JT}} x^{\lambda f'} \varepsilon - \left(\frac{1}{JT} x^{\lambda f'} g \right) \sqrt{JT} (\hat{\alpha} - \alpha^0) + c_x^{(2)} \right] + o_p(1). \end{aligned} \quad (64)$$

Here, the remainder term $o_p(\sqrt{JT}\|\hat{\alpha} - \alpha^0\|)$ is already absorbed into the $o_p(1)$ term, since (63) already shows \sqrt{JT} -consistency of $\hat{\alpha}$. Let G_{JT} and \mathcal{W}_{JT} be the expressions in equation (12) and (13) before taking the probability limits, i.e. $G = \text{plim}_{J,T \rightarrow \infty} G_{JT}$ and $\mathcal{W} = \text{plim}_{J,T \rightarrow \infty} \mathcal{W}_{JT}$. One can show that

$$G_{JT} \mathcal{W}_{JT} G'_{JT} = \frac{1}{JT} (g, x)' P_{x^{\lambda f}} (g, x) + \begin{pmatrix} A_2 & 0_{L \times K} \\ 0_{K \times L} & 0_{K \times K} \end{pmatrix}. \quad (65)$$

Using this, one can rewrite equation (63) and (64) as follows

$$\begin{aligned} &G_{JT} \mathcal{W}_{JT} G'_{JT} \sqrt{JT} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\beta} - \beta^0 \end{pmatrix} \\ &= \frac{1}{\sqrt{JT}} (g, x)' P_{x^{\lambda f}} \varepsilon + \begin{pmatrix} A_1 + \left(\frac{1}{JT} g' x^{\lambda f} \right) \left(\frac{1}{JT} x^{\lambda f'} x^{\lambda f} \right)^{-1} c_x^{(2)} \\ c_x^{(2)} \end{pmatrix} + o_p(1), \end{aligned} \quad (66)$$

and therefore

$$\begin{aligned}
& \sqrt{JT} \begin{pmatrix} \hat{\alpha} - \alpha^0 \\ \hat{\beta} - \beta^0 \end{pmatrix} \\
&= (G_{JT} \mathcal{W}_{JT} G'_{JT})^{-1} G_{JT} \mathcal{W}_{JT} \left[\frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon \right] \\
&\quad + (G_{JT} \mathcal{W}_{JT} G'_{JT})^{-1} \begin{pmatrix} A_3 c_z^{(2)} + [(g' x^{\lambda f}) - A_3 (z^{\lambda f'} x^{\lambda f})] (x^{\lambda f'} x^{\lambda f})^{-1} c_x^{(2)} \\ c_x^{(2)} \end{pmatrix} + o_p(1) \\
&= (G \mathcal{W} G')^{-1} G \mathcal{W} \left[\frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon + c_x^{(2)} c_z^{(2)} \right] + o_p(1), \tag{67}
\end{aligned}$$

where $A_3 = \left(\frac{1}{JT} g' M_{x^{\lambda f}} z^{\lambda f} \right) \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1} W_{JT} \left(\frac{1}{JT} z^{\lambda f'} M_{x^{\lambda f}} z^{\lambda f} \right)^{-1}$. Having equation (67), 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 (2009) under the same assumptions that we impose here. They show that

$$c_x^{(2)} = -\kappa^{-1} b^{(x,1)} - \kappa b^{(x,2)} + o_p(1), \quad c_z^{(2)} = -\kappa^{-1} b^{(z,1)} - \kappa b^{(z,2)} + o_p(1), \tag{68}$$

and

$$\frac{1}{\sqrt{JT}} (x^{\lambda f}, z^{\lambda f})' \varepsilon \xrightarrow{d} \mathcal{N} \left[-\kappa b^{(x,0)} b^{(z,0)}, \Omega \right]. \tag{69}$$

Plugging this into (67) gives the result on the limiting distribution of $\hat{\alpha}$ and $\hat{\beta}$ which is stated in the theorem. \blacksquare

C.3 Consistency of Bias and Variance Estimators

Proof of Theorem 4: From Moon and Weidner (2009) 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 $\widehat{W} = W + o_p(1)$. These results on $M_{\hat{\lambda}}$ and $M_{\hat{f}}$ together with \sqrt{JT} -consistency of $\hat{\alpha}$ and Assumption 11 are also sufficient to conclude $\hat{G} = G + o_p(1)$. It follows that $\hat{B}_i = B_i + o_p(1)$, for $i = 0, 1, 2$. \blacksquare

References

- Ackerberg, D., Benkard, L., Berry, S., and Pakes, A. (2007). Econometric tools for analyzing market outcomes. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics, Vol. 6A*. North-Holland.
- Ahn, S. C., 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. (2008a). The random coefficients logit model is identified. Manuscript, University of Minnesota.
- Bajari, P., Fox, J., Kim, K., and Ryan, S. (2008b). A simple nonparametric estimator for the distribution of random coefficients. Manuscript, University of Minnesota.
- 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. (2008). Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation. Manuscript.
- 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). The Interaction of Observed and Unobserved Factors in Discrete Choice Demand Models. 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.
- 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. (2009). Likelihood Expansion for Panel Regression Models with Factors. 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. (2005). Determining the number of factors from empirical distribution of eigenvalues. Discussion Papers 0405-19, Columbia University, Department of Economics.
- 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.