IDENTIFICATION OF CONVEX FACTOR MODELS

CHRISTOPHER P. ADAMS,
FEDERAL TRADE COMMISSION

Abstract. This paper considers factor model problems where the observed outcomes are convex combinations of the hidden factor values. It presents identification results based on non-negative matrix factorization (Lee and Seung, 1999), orthonormality (Hotelling, 1933), and volume minimization (Fu et al., 2016). In the case of non-negative matrix factorization, the sparsity conditions of Huang et al. (2014) are necessary and sufficient (Adams, 2016). For the remaining cases, the paper presents new results. Under orthonormality, small factor models are identified and the sparsity conditions of Huang et al. (2014) on the weighting matrix, are sufficient for larger models. For exact factor models, the volume minimization approach finds the correct factors if and only if for each hidden factor there is one observed unit that places all of its weight on that factor. For approximate factor models, the paper presents sufficient conditions for using the volume minimization procedure to find the correct factors. The volume minimization approach to the approximate factor model is illustrated with a parametric maximum likelihood estimator. The estimator is run on simulated data and panel data analysis of a casino merger in Missouri.

The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Trade Commission. I'm grateful for continued discussions about this problem with, Stephane Bonhomme, Jeremy Sandford and Nathan Wilson. All errors are my own.

1. Introduction

Factor models date to at least 1933 and have found a wide variety of uses in econometrics and applied statistics including analysis of psychometric data, panel data, macroeconomics and document analysis (Hotelling, 1933; Bai and Ng, 2013; Lee and Seung, 1999). The non-uniqueness of the underlying matrix factorization is well-known (Hotelling, 1933).[1] This paper considers a restriction on the factor model where each observation is assumed to be a convex combination of the hidden factor values. The paper presents identification results under three alternative approaches. First, the paper presents necessary and sufficient conditions for uniqueness when the factor values are assumed to be non-negative (Lee and Seung, 1999). Second, the paper presents sufficient conditions for uniqueness when the matrix factor representing the hidden factor values is assumed to be orthonormal (Hotelling, 1933). Third, the paper presents conditions for uniqueness when the solution to the matrix factorization problem is found through volume minimization (Fu et al., 2016). The third approach suggests a maximum likelihood estimator. Results from the estimator are presented using simulated data and data from prior to the merger of casinos in Missouri.

Lee and Seung (1999) consider factor models where the factor values and the factor weights are assumed to be non-negative. The authors argue that such models are a good representation of document sets or sets of pictures. Huang et al. (2014) presents necessary, and separately, sufficient conditions for uniqueness under this restriction. Adams (2016) shows that when the observed values are a convex combination of the hidden factor values, the necessary condition of Huang et al. (2014) is also sufficient.

Hotelling (1933) considered an orthonormality restriction on the matrix factor representing the hidden factor values. That is, the author assumes that the matrix's inverse is equal to its transpose. Hotelling (1933) justifies the assumption by noting that if the number of "time periods" is large enough and the factor values are drawn iid from a standard normal distribution, the variance-covariance matrix is the identity matrix.[2] Bai and Ng (2013) present conditions on the factor weights that lead to

[1]Bai (2009) shows that strict identification of the factor model is not always necessary in treatment estimation, such as the problem considered in the last section. However, simulation results (not included) suggest that the estimator presented below out-performs the estimator suggested in Bai (2009) in finite samples.

[2]The analogy to time periods in the original paper's application is the individual test takers.

uniqueness. In particular, if for each factor there is some observed unit that only places weight on that factor, then the factorization is unique. In contrast, this paper considers combining the orthonormality assumption with the assumption that the observed values are convex combinations of the hidden factor values. The paper shows that if the number of factors is 2 or 3, then the model is uniquely determined. In addition, the paper shows that a sparsity condition (Huang et al., 2014) on the weighting matrix is sufficient for uniqueness.

Like this paper, Fu et al. (2016) considers factor models where the observed values are assumed to be convex combinations of the hidden factor values. The authors point out that the matrix factorization problem can be thought of as a volume minimization problem subject to constraints. The authors present and discuss various conditions on the problem such that the volume minimization problem provides the unique solution to an exact factor model. The authors present the sufficient condition in terms of how "spread out" the observations are. This paper takes a more direct approach. It shows if for each factor there is one observation that places all its weight on that factor, then this is necessary and sufficient for the volume minimization solution to be the unique solution to the factor model problem.[3] Obviously, if the observations are sufficiently "spread out" then the condition will hold.

It is unclear from Fu et al. (2016) whether the volume minimization approach can be applied to approximate factor models (Bai and Ng, 2002). The problem is that the volume minimization approach will tend to choose polyhedrons that are too large because of the "errant" observations. This paper shows that problem can be represented as a finite mixture model. Identification is based on two steps. In the first step, the distribution of the "mean values" can be determined following Adams (2016) and Efron (2010). In the second step, the results for the exact factor model are used to determine the hidden factors.

The rest of the paper proceeds as follows. Section 2 presents the notation. Section 3 presents the empirical model. Section 4 presents the main identification results. Section 5 presents the estimation results for the maximum likelihood estimator. Section 6 concludes.

[3]This is similar to the sufficient condition presented in Huang et al. (2014) and the assumption made in Bai and Ng (2013).

2. Notation

Let $\mathbf{B}$ denote a rectangular matrix. Let $\mathbf{B}'$ denote the transpose, $\mathbf{B}^{-1}$ denote the inverse of a square (full-rank) matrix, and $\mathbf{B}^{+}$ denote the generalized inverse of the matrix. Let $\mathbf{I}$ denote the identity matrix. $\mathbf{B} \geq 0$ means that each element of the matrix is non-negative. $\mathbf{1}$ is a vector of 1's. Here $vol(\mathbf{B})$ of a $I \times J$ matrix is the volume contained inside an I-dimensional convex polytope with J vertexes. The notation $\mathbf{B}_j$ refers to a vector, column j of the matrix, $\mathbf{B}_{ij}$ is an element of the vector. $\mathbf{B} = \mathbf{C}$ means that there is some arrangement of the columns of $\mathbf{B}$ such that the matrices are the same. The element B_{it} is a random scalar, while B is a fixed scalar, and $\mathbf{b}$ also refers to a vector.

3. Model

Consider an approximate factor model with observed outcome Y_{it} for unit i in time t. This outcome is assumed to be a weighted average over K hidden factor values and an error term.

$$Y_{it} = \sum_{k=1}^{K} \lambda_{ik} F_{kt} + \epsilon_{it} \tag{1}$$

where $i \in \{1, ..., I\}$ and $t \in \{1, ..., T\}$. In matrix notation, a model of a panel data set is as follows.

$$\mathbf{Y} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{E} \tag{2}$$

where $\mathbf{Y}$ is $T \times I$, $\mathbf{F}$ is $T \times K$, $\mathbf{\Lambda}$ is $I \times K$, and $\mathbf{E}$ is $T \times I$.

For ease of exposition, the paper will generally consider the "mean value" version of the model. However, in the latter sections the approximate version is considered explicitly.

$$\bar{\mathbf{Y}} = \mathbf{Y} - \mathbf{E} = \bar{\mathbf{F}}\bar{\mathbf{\Lambda}}' \tag{3}$$

The following assumption is maintained throughout the paper. It states that each observed outcome is a convex combination of the hidden factor values.[4]

Assumption 1. *$\sum_{k=1}^{K} \lambda_{ik} = 1$ for all $i \in \{1, ...I\}$ and $\lambda_{ik} \geq 0$ for all $i \in \{1, ..., I\}$ and $k \in \{1, ..., K\}$ or $\mathbf{1}'\mathbf{\Lambda}' = \mathbf{1}'$ and $\mathbf{\Lambda} \geq 0$.*

[4] In some literatures a matrix with this property is called "stochastic."

An immediate consequence of Assumption 1 is that it reduces the number of unknown parameters from K^2 to $(K-1)K$.

Lemma 1. *Let* $\mathbf{F} = \bar{\mathbf{F}}$ *and* $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$ *and let* $\mathbf{A}$ *be a* $K \times K$ *full-rank matrix such that*

$$\mathbf{Y} = \mathbf{F}\mathbf{A}\mathbf{A}^{-1}\mathbf{\Lambda}' \tag{4}$$

Given Assumption 1, if $\mathbf{\Lambda}$ *is rank* K*, then* $\mathbf{1}'\mathbf{A} = \mathbf{1}'$.

Proof. See Appendix. □

A standard assumption in the non-negative matrix factorization literature is that both matrix factors are non-negative (Lee and Seung, 1999). That is, each element of the matrix is either zero or positive.

Assumption 2. $F_{kt} \geq 0$ *for all* $k \in \{1, ..., K\}$ *and* $t \in \{1, ..., T\}$ *or* $\mathbf{F} \geq 0$.

Following Hotelling (1933), it is standard in the factor model literature to make some version of the following orthonormality assumption (Bai and Ng, 2013).

Assumption 3. $\mathbf{F}'\mathbf{F} = \mathbf{I}$

4. Identification

4.1. Non-Negative Matrix Factorization.

The non-negativity assumption is standard in the machine-learning literature following Lee and Seung (1999). The authors argue that it is a reasonable representation of documents or photos. As the matrix factorization problem is identical to the problem for finite mixture models, the result presented in Adams (2016) can be applied directly.

Theorem 1. *Let* $\mathbf{F} = \bar{\mathbf{F}}$ *and* $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$ *and let* $\mathbf{A}$ *be a* $K \times K$ *full-rank matrix such that Equation (4) holds. Define*

$$\begin{aligned} \mathcal{T}_k &= \{t \in \{1, ..., T\} | \mathbf{F}_{tk} \neq 0\} \\ \mathcal{I}_k &= \{i \in \{1, ..., I\} | \mathbf{\Lambda}_{ik} \neq 0\} \end{aligned} \tag{5}$$

Given Assumptions 1 and 2, then $\mathbf{A} = \mathbf{I}$ *(up to rearranging columns) if and only if there do not exist* $k_1, k_2 \in \{1, ..., K\}$, $k_1 \neq k_2$ *such that* $\mathcal{T}_{k_1} \subseteq \mathcal{T}_{k_2}$ *or* $\mathcal{I}_{k_1} \subseteq \mathcal{I}_{k_2}$.

Proof. See Adams (2016). □

Theorem 1 is based on the necessary conditions presented in Theorem 3 of Huang et al. (2014). The proof presented in Adams (2016) adjusts the proof of Huang et al. (2014) to account for the extra constraint provided by Assumption 1. The proof shows that the necessary condition is the same. This means that the minimum requirement for identification is the sparsity condition presented in Huang et al. (2014). The paper shows that the additional constraint of summing to one implies that this necessary condition is also sufficient.

4.2. **Orthonormality.** A more standard restriction in the economics literature is the orthonormality restriction of the matrix of factor values. Hotelling (1933) notes that for a large number of time periods, the matrix factor is related to the variance-covariance matrix. Thus, the author justifies this restriction by assuming the factor values are distributed standard normal and independent across factors.

Lemma 2. *Let* $\mathbf{F} = \bar{\mathbf{F}}$ *and* $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$ *and let* $\mathbf{A}$ *be a* $K \times K$ *full-rank matrix such that Equation (4) holds. Given Assumption 3,* $\mathbf{A}'\mathbf{A} = \mathbf{I}$ *providing* $\frac{K(K+1)}{2}$ *independent equations and* $\det(\mathbf{A}) = \pm 1$.

Proof. In the Appendix. □

The first result states that the orthonormal property of the matrix factor is passed on to the unknown matrix. As Hotelling (1933) points out, this result adds $\frac{K(K+1)}{2}$ constraints to the unknown parameters. In the original version of the model these additional restrictions are not enough for identification. Here, with the additional convexity restriction, the model is identified if the number of factors is either 2 or 3.

Theorem 2. *Let* $\mathbf{F} = \bar{\mathbf{F}}$ *and* $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$ *and let* $\mathbf{A}$ *be a* $K \times K$ *full-rank matrix such that Equation (4) holds. Given Assumptions 1 and 3 and* $K = 2$ *or* $K = 3$*, if* $\mathbf{\Lambda}$ *is rank* K*, then* $\mathbf{A} = \mathbf{I}$.

Proof. In the Appendix. □

The last result shows that the sparsity conditions of Huang et al. (2014), but limited to only the weighting matrix, are sufficient for identification.

Theorem 3. *Let* $\mathbf{F} = \bar{\mathbf{F}}$ *and* $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$ *and let* $\mathbf{A}$ *be a* $K \times K$ *full-rank matrix such that Equation (4) holds. Given Assumptions 1 and 3, if there do not exist* $k_1, k_2 \in \{1, ..., K\}$, $k_1 \neq k_2$ *such that* $\mathcal{I}_{k_1} \subseteq \mathcal{I}_{k_2}$ *then* $\mathbf{A} = \mathbf{I}$.

Proof. In the Appendix. □

4.3. **Volume Minimization with Exact Factor Model.** Fu et al. (2016) presents an alternative method for determining the parameters of the factor model. If the observed outcomes are a convex combination of the unobserved factor values, then the authors note that there exists a T-polyhedron that includes all the observations inside it. Moreover, the smallest such T-polyhedron has vertexes at hidden factor vectors.

Figure 1 represents the case where there are three hidden factors and two-time periods. The three factors are represented by the points F_1, F_2 and F_3 in the diagram. These are the factor value vectors for each of the three factors. The circles represent some set of observed outcome vectors for the units of observation. By assumption these observed outcomes must lie within the triangle. In this case, we may be able to identify the factor model by simply plotting out the data

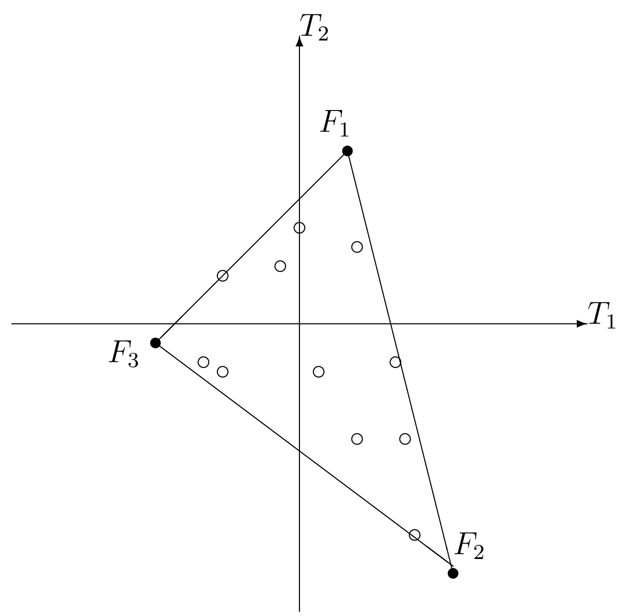

FIGURE 1. Three factor convex model.

Lemma 3 states that the volume minimizing T-polyhedron is unique and its vertexes are observed outcomes from the model.

Lemma 3. *Let there be some data set* $\mathbf{Y}$*, where each column of the matrix represents an observed unit* i *and the value lies in* $\Re^T$*. Let* $\mathbf{G}^*$ *be a* $T \times K$ *matrix such that*

$$\begin{array}{ll} \mathbf{G}^* = & \arg\min_{\mathbf{G} \in \mathcal{G}} vol(\mathbf{G}) \\ s.t. & \mathbf{Y} = \mathbf{G}\mathbf{\Lambda}'_{\mathbf{G}} \\ & \mathbf{\Lambda}'_{\mathbf{G}} \geq 0 \\ & \mathbf{1}'\mathbf{\Lambda}'_{\mathbf{G}} = \mathbf{1}' \end{array} \tag{6}$$

where $\mathcal{G}$ *is the set matrices with* T *rows and some finite number of columns less than equal to* I*, then*

1. $\mathbf{G}^*$ *is unique,*
2. *There are* $N_K \geq K$ *elements associated with* $\mathbf{Y}$ *(columns of* $\mathbf{Y}$*) that lie exactly at the* K *vertexes of* $\mathbf{G}^*$*, and*
3. *All elements associated with* $\mathbf{Y}$ *(columns of* $\mathbf{Y}$*) either lie at the vertexes of* $\mathbf{G}^*$ *or strict convex combinations of two more vertexes of* $\mathbf{G}^*$*.*

Proof. In the Appendix. □

Theorem 4 states that if there is at least one observed outcome that is equal to each of the hidden factors, then that is both necessary and sufficient condition for the volume minimizing T-polyhedron to solve the factor model. If there is one observed outcome at each factor value, then the volume minimizing T-polyhedron will have vertexes at those observed outcomes. If not, then the volume minimizing T-polyhedron will have vertexes that are strict convex combinations of the true factors.

Theorem 4. *Let* $\mathbf{F} = \bar{\mathbf{F}}$ *and* $\mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$*. Given Assumption 1, there exists* $i_k \in \{1, ..., I\}$ *for each* $k \in \{1, ..., K\}$*, such that* $\mathbf{\Lambda}_{i_k k} = 1$ *if and only if* $\mathbf{F} = \mathbf{G}^*$*.*

Proof. In the Appendix. □

4.4. **Volume Minimization with Approximate Factor Model.** The volume minimization approach suggests a powerful tool for solving factor models. However, it is not robust to approximation error in the model. It is straightforward to see that if there are "outliers" then the volume minimization approach will find a T-polyhedron that is too large.

In general, we can re-write the factor model as a mixture model. Note that each observation is the vector of outcomes across the time periods for the observed unit i.

$$\Pr(\mathbf{y}_i = \mathbf{y}) = \int_{\bar{\mathbf{y}}} h(\mathbf{y} - \bar{\mathbf{y}}) g(\bar{\mathbf{y}}) \tag{7}$$

where $\bar{\mathbf{y}}_i = \mathbf{F}\mathbf{\Lambda}_i$ and $\mathbf{y}_i = \mathbf{F}\mathbf{\Lambda}_i + \mathbf{E}_i$, h is the distribution of the error term and g is the distribution of the "mean values."

Let the outcome set in $\Re^T$ be partitioned into M_1 sets and let the set of average outcomes in $\Re^T$ be partitioned into R_1 sets.

$$\mathbf{p} = \mathbf{H}_1 \mathbf{g} \tag{8}$$

where $\mathbf{p}$ is a $M_1 \times 1$ vector of probabilities over outcomes (in the T-dimensional reals), $\mathbf{H}_1$ is a $M_1 \times R_1$ matrix mapping from the mean value of the factor model to the observed probabilities over outcomes, and $\mathbf{g}$ is an $R_1 \times 1$ vector representing the probability distribution over mean values.

Given this representation, we can determine the distribution of mean values if the error term distribution is known.

Lemma 4. *If $\mathbf{H}_1$ is known and rank R_1, then $\mathbf{g} = \mathbf{H}_1^+ \mathbf{p}$.*

Proof. See Efron (2010). □

The identification result presented here, breaks the problem into two steps. In the first step, it considers the mixture model problem at the level of a unit-time observation. In the second step, the problem is aggregated back up to the unit level and the above result is used because the error distribution is known (found in the first step).

$$\Pr(Y_{it} = Y) = \int_{\bar{Y} \in \Re} h_t(Y|\bar{Y}) g_t(\bar{Y}) \tag{9}$$

The following assumption, that the error terms are distributed symmetrically across time, allows results from the mixture model literature to be used to show identification.

Assumption 4. $h_t(Y|\bar{Y}) = h_{t'}(Y|\bar{Y}) = h(Y|\bar{Y})$ *for all* $t, t' \in \{1, ..., T\}$.

By Assumption 4, the distribution over the error term is the same for each time-period.

$$\Pr(Y_{it} = Y) = \int_{\bar{Y} \in \Re} h(Y|\bar{Y}) g_t(\bar{Y}) \tag{10}$$

Let the outcome set in $\Re$ be partitioned into M_2 sets and let the set of mean outcomes in $\Re$ be partitioned into R_2 sets.

$$\mathbf{P} = \mathbf{H}_2 \mathbf{\Gamma} \tag{11}$$

where $\mathbf{P}$ is a $M_2 \times T$ matrix, $\mathbf{H}_2$ is a $M_2 \times R_2$ matrix, and $\mathbf{\Gamma}$ is a $R_2 \times T$ matrix.

Note that the mean values ($\bar{\mathbf{Y}}$) represent the subset of the R_1 elements of $\mathbf{g}$, such that $\mathbf{g}_{r_1} > 0$. Define $\mathbf{G}^{**}$ as in Lemma 3, where $\mathbf{Y}$ is replaced with $\bar{\mathbf{Y}}$.

Theorem 5. *Define*

$$\begin{aligned} \mathcal{M}_r &= \{m \in \{1, ..., M\} | \mathbf{H}_{2mr} \neq 0\} \\ \mathcal{T}_r &= \{t \in \{1, ..., T\} | \mathbf{\Gamma}_{rt} \neq 0\} \end{aligned} \tag{12}$$

and $\mathbf{Q}$ *is a* $R_2 \times R_2$ *full-rank matrix such that*

$$\mathbf{P} = \mathbf{H}_2 \mathbf{Q} \mathbf{Q}^{-1} \mathbf{\Gamma} \tag{13}$$

Given Assumptions 1 and 4, if

1. *there do not exist* $r_1, r_2 \in \{1, ..., R_2\}$, $r_1 \neq r_2$ *such that* $\mathcal{M}_{r_1} \subseteq \mathcal{M}_{r_2}$ *or* $\mathcal{T}_{r_1} \subseteq \mathcal{T}_{r_2}$,
2. $\mathbf{H}_1$ *is rank* R_1, *and*
3. *there exists* $i_k \in \{1, ..., I\}$ *for each* $k \in \{1, ..., K\}$, *such that* $\mathbf{\Lambda}_{i_k k} = 1$,

then $\bar{\mathbf{Y}}$ *is identified and* $\mathbf{F} = \mathbf{G}^{**}$.

Theorem 5 presents sufficient conditions for the adjusted volume minimization approach to find the factor values. As the problem can be represented as a mixture model problem, the conditions presented in Huang et al. (2014) are sufficient. Once the distribution of error term is determined, it is possible to determine the distribution of the mean values. The result then uses the result for the exact factor model where the observed outcomes are replaced with outcomes that can occur with positive probability.

5. Estimation

5.1. Estimator. Bai and Ng (2013) present estimation results for matrix factorization problems similar to the ones based on non-negativity and orthonormality. To illustrate the volume minimization result consider a simple maximum likelihood estimator. Following the logic of the identification result, consider a two-step estimator in which the first step estimates the error distribution and the second step estimates the factor value vectors.

Step 1.

$$\{\hat{F}_{kt}, \hat{\sigma}\} = \arg\max \sum_{i=1}^{I} \sum_{t=1}^{T} \log \left(\int_{\lambda \in [0,1]^K} \phi \left(\frac{Y_{it} - \mathbf{F}_t \lambda}{\sigma} \right) \right) \text{ s.t. } \mathbf{1}'\lambda = 1 \tag{14}$$

Step 2.

$$\{\mathbf{F}_k\} = \arg\max \sum_{i=1}^{I} \log \left(\int_{\lambda \in [0,1]^K} \phi \left(\frac{||Y_i - \mathbf{F}\lambda||}{\hat{\sigma}} \right) \right) \text{ s.t. } \mathbf{1}'\lambda = 1 \tag{15}$$

Although the identification results are non-parametric, it is simpler to illustrate the volume minimization approach with a parametric estimator. Here the error term is normally distributed with the standard deviation to be estimated in the first step. The mean value distribution is uniform with the estimator determining the vertexes of the T-polyhedron.

5.2. Simulation. To illustrate the maximum-likelihood estimator described above, the simulated data is generated assuming that $Y_{it} = \sum_{k=1}^{3} \lambda_{ik} F_{kt} + \epsilon_{it}$, where $\epsilon_{it} \sim \mathcal{N}(0, \sigma)$, and $\sum_{k=1}^{3} \lambda_{ik} = 1$ for all $i \in 1, ..., 100$ and $t \in \{1, ..., 10\}$. Two cases are run with 100 simulated runs each. In the first case the standard deviation on the error term is 0.20. In the second case it is 0.10.

Table 1 presents the average estimate for the standard deviation and the average root mean squared error for the model.

5.3. Merger of Casinos in Missouri. In 2013, the Federal Trade Commission agreed to allow Pinnacle Entertainment to merger with Ameristar Casinos, in a deal worth \$2.8 billion. The FTC required that the parties divest Lumiere Place Casino in St Louis MO. The FTC charged that without the divestiture the merger would reduce competition in St Louis MO. The merger was consummated in August 2013,

σ	Average	SD	Average RMSE	SD RMSE
0.20	0.17	0.05	1.46	0.47
0.10	0.12	0.05	1.48	0.52

Table 1. Average (and standard deviation) of the estimate of $\hat{\sigma}$ and the root mean squared error. Three factors (assumed) ($K = 3$), ten time periods ($T = 10$) and 100 individual units ($I = 100$). 100 simulations.

with the divestiture taking place 8 months later in April 2014. Osinski and Sandford (2017) use a standard difference-in-difference estimator to analyze the impact of both the merger and the divestiture. The authors find that the merger is associated with a small increase in output, suggestive that there were efficiencies due to the merger. The authors also find that there was a relatively large decrease in output associated with the divestiture, suggesting some issues with the owners of the divested casino. This paper uses the same data but the method suggested above to estimate the treatment effect of the merger and the divestiture on the St Louis MO casinos. The results are broadly similar to the difference in difference estimation, although they suggest more variation and uncertainty associated with the outcomes.

The analysis uses monthly data on the amount of money put into various slot machine games across casinos located in Missouri. The measure of "output" is log of "table drop," which is the amount of money put into a particular type of machine (1 cent, 5 cent, \$5 etc). A unit of observation is average log table drop per type of machine in each casino per month. The time period is from August 2011 to December 2016. The factor analysis is conducted on the pre-treatment period which is from August 2011 to August 2013. The analysis is limited to the penny slots, which has the largest revenue.

Figure 2 suggests that there are only two hidden factors. This is assumed in the maximum likelihood procedure. The first step estimates the standard deviation of the error term at 0.14. The second step assumes this is the true standard deviation and estimates the two factor value vectors. Table 2 presents the results from the second step.

A third step is to take the estimated factor values and determine the appropriate weights given the observed data. Table 3 presents the estimated weights. The table

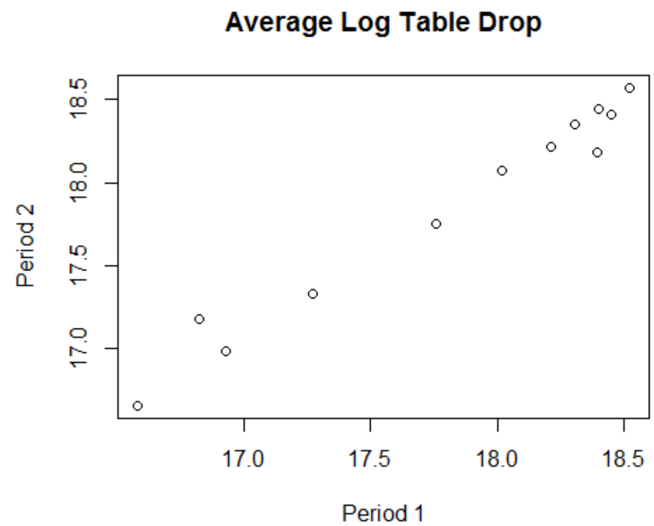

Figure 2. Average table drop per penny slot, splitting the pre-treatment into two periods and averaging.

also highlights the four "treated" firms. The table shows that all four treated units place all their weight on factor 1.

Figure 3 presents the treatment effect for each of the treated units. The line is the difference between the actual log table drop during the treated period and the estimated log table drop inferred from the factor model. Given the model estimates, the counter-factual log table drop is the same for each firm and it is a weighted average of the log table drop for Ameristar KC, Argosy, Harrah's NKC, Isle of Capri - KC, and St. Jo, with Ameristar KC getting most of the weight. The analysis does not account for sampling variance.

The results suggest that the merging firms had higher volumes most merger. This suggests that the merger was associated with efficiencies. One possibility is that the merger allowed the firms to better coordinate their marketing. The results also suggest that the divested firm under-performed relative to the other firms in the market and relative to its counter-factual performance.

	1	2
1	-1.36	-0.50
2	1.08	-0.76
3	0.48	-1.79
4	-0.38	-0.38
5	-0.08	0.12
6	-0.45	1.61
7	-0.50	0.05
8	0.10	0.57
9	-0.60	0.55
10	0.12	0.27
11	-0.70	-1.32
12	0.79	-0.47
$\hat{\sigma}$		0.14
Log Lik	Step 1	204.37
Log Lik	Step 2	20.72

Table 2. Factor values by time-period from Step 2 of the maximum likelihood estimator.

6. Conclusion

The paper considers a linear factor model where each observed outcome is assumed to be a convex combination of some finite set of hidden factor values. The paper considers three approaches to identifying this model. The first approach requires that the hidden factor values are non-negative. For this case, the paper presents a result from Adams (2016) providing the necessary and sufficient conditions for identification. The second approach requires that the matrix factor representing the hidden factors is orthonormal. The paper presents new results showing that the model is identified if the number of factors is 2 or 3. It shows that more generally, the condition of Huang et al. (2014) is sufficient for identification. The third approach considers the idea of finding the volume minimizing T-polyhedron subject to observations coming from a factor model. The paper presents necessary and sufficient conditions for the approach to find the correct hidden factors when the factor model is exact. It also considers approximate factor models and shows that the volume

	1	2	Treated
Ameristar KC 1 cent	1.00	0.00	
Ameristar SC 1 cent	1.00	0.00	Yes
Argosy 1 cent	0.74	0.26	
Harrah's NKC 1 cent	0.70	0.30	
Hollywood 1 cent	1.00	0.00	Yes
Isle - Boonville 1 cent	0.00	1.00	
Isle of Capri - KC 1 cent	0.32	0.68	
Lady Luck 1 cent	0.00	1.00	
Lumiere Place 1 cent	1.00	0.00	Yes
Mark Twain 1 cent	0.00	1.00	
River City 1 cent	1.00	0.00	Yes
St. Jo 1 cent	0.82	0.18	

TABLE 3. Factor weights by slot game

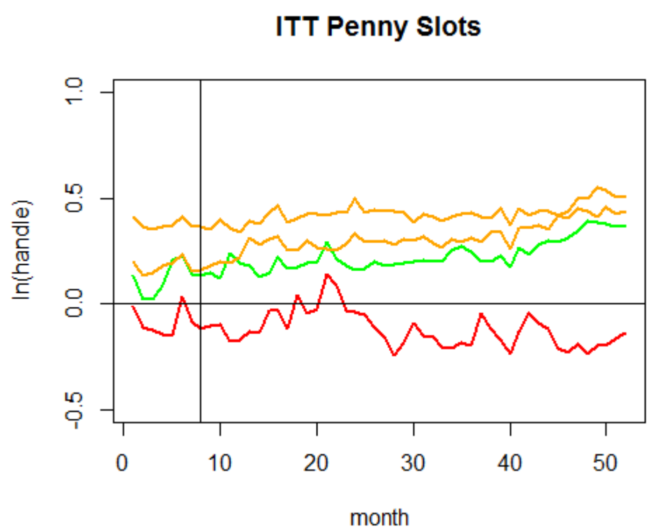

FIGURE 3. Treatment effect for the St Louis Casinos. In order from highest to lowest, they are the merging casinos Ameristar and River City, the non-merging firm Hollywood and the divested casino Lumiere Place.

minimization approach can be combined with a mixture model to determine the solution. The paper presents sufficient conditions for this approach to find the correct hidden factors. The last approach is illustrated with simulated data and analysis of a merger of casinos in Missouri.

7. Appendix

7.1. Proof of Lemma 1.

Proof. By Assumption 1, $\mathbf{1}'\mathbf{\Lambda}' = \mathbf{1}'$. By the condition of the lemma,

$$\mathbf{1}' = \mathbf{1}'(\mathbf{\Lambda}')^{+} \tag{16}$$

Also by definition of $\mathbf{A}$,

$$\mathbf{1}'\mathbf{A}^{-1}\mathbf{\Lambda}' = \mathbf{1}' \tag{17}$$

or

$$\mathbf{1}'\mathbf{A}^{-1} = \mathbf{1}'(\mathbf{\Lambda}')^{+} = \mathbf{1}' \tag{18}$$

□

7.2. Proof of Lemma 2.

Proof. By Assumption 3 and definition of $\mathbf{A}$,

$$\mathbf{A}'\mathbf{F}'\mathbf{F}\mathbf{A} = \mathbf{I} \tag{19}$$

Substituting $\mathbf{F}'\mathbf{F} = \mathbf{I}$ and rearranging

$$\mathbf{A}^{-1} = \mathbf{A}' \tag{20}$$

The rest follows from the orthonormal properties of the matrix $\mathbf{A}$. □

7.3. Proof of Theorem 2.

Proof. By Lemma 1 and Lemma 2, $\mathbf{1}'\mathbf{A}' = \mathbf{1}'$ and there are $(K-1)^2$ free-parameters.

Case 1. $K = 2$. From Lemma 1 and above,

$$\mathbf{A} = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix} \tag{21}$$

By Lemma 2

$$1 + 2a = \pm 1 \tag{22}$$

Case 2. $K = 3$. From Lemma 1 and above,

$$\mathbf{A} = \begin{bmatrix} 1+a+b & -a-d & -b+d \\ -b-c & 1+a+b+c & -a \\ -a+c & -b-c+d & 1+a+b-d \end{bmatrix} \tag{23}$$

where $\{a, b, c, d\} \in \Re^4$. From orthonormality we have

$$-a-d = (-b-c)(1+a+b-d) - (-a)(-a+c) \tag{24}$$

or

$$-a+b+c-d+ab+bc-bd-cd+b^2+a^2 = 0 \tag{25}$$

and in addition we have

$$-a+c = (-a-d)(-a) - (-b+d)(1+a+b+c) \tag{26}$$

or

$$-a-b+c+d-ab-bc+bd+cd-a^2-b^2 = 0 \tag{27}$$

Adding Equations (25) and (27),

$$-2a+2c = 0 \tag{28}$$

or $a = c$. Also from orthonormality we have

$$-b-c+d = (1+a+b)(-a) - (-b+d)(-b-c) \tag{29}$$

or

$$a-b-c+d+ab+bc-bd-cd+a^2+b^2 = 0 \tag{30}$$

Adding Equation (27) and (30)

$$-2b+2d = 0 \tag{31}$$

or $b = d$. These results allow us to simplify $\mathbf{A}$

$$\mathbf{A} = \begin{bmatrix} 1+a+b & -a-b & 0 \\ -b-a & 1+2a+b & -a \\ 0 & -a & 1+a \end{bmatrix} \tag{32}$$

Again by orthonormality

$$0 = (-a-b)(-a) \tag{33}$$

For $a = 0$ or $a = -b$ see Case (1). □

7.4. Proof of Theorem 3.

Proof. By Assumption 1, $\mathbf{A}^{-1}\mathbf{\Lambda}' \geq 0$. By Lemma 2, $\mathbf{A}'\mathbf{\Lambda}' \geq 0$. Taking the transpose, $\mathbf{\Lambda}\mathbf{A} \geq 0$. The rest follows from Theorem 1, where $\mathbf{\Lambda}$ is labeled $\mathbf{F}$.

□

7.5. Proof of Lemma 3.

Proof. (3). From the constraint of the optimization problem.

(2). Suppose not. There is at least one vertex ($\mathbf{G}_k^*$) which does not correspond to any elements of $\mathbf{Y}$. Therefore there exists two elements of $\mathbf{Y}$ such that the three points form a T-polyhedron with no elements of $\mathbf{Y}$ strictly inside the volume. Therefore the volume has not been minimized subject to the constraints. A contradiction.

(1). Suppose not. Case 1. There is some other solution with vertexes that are not vertexes of $\mathbf{G}^*$. This contradicts (2). Case 2. There is some other solution where the vertexes of $\mathbf{G}^*$ are not vertexes of the alternative solution. This contradicts (2). □

7.6. Proof of Theorem 4.

Proof. $\Rightarrow$ By the condition, there exists a subset of the columns of $\mathbf{Y}$, such that $\mathbf{Y}_K = \mathbf{F}\mathbf{I}$. By Lemma 3, $\mathbf{Y}_K = \mathbf{G}^*$.

$\Leftarrow$ By the condition, $\mathbf{F} = \mathbf{G}^*$. By Lemma 3, $\mathbf{G}^* = \mathbf{Y}_K$. So there exists a subset of the columns of $\mathbf{Y}$ such that $\mathbf{Y}_K = \mathbf{F}\mathbf{I}$.

□

7.7. Proof of Theorem 5.

Proof. By Assumption 4 and Theorem 1, then if the condition (1) of the theorem holds, $\mathbf{H}_2$ is uniquely determined from the data. So $\mathbf{H}_1$ is known. By condition (2) and Lemma 4, $\mathbf{g}$ is determined. $\bar{\mathbf{Y}}$ is determined by definition. By definition of $\mathbf{G}^{**}$ and Lemma 3, $\mathbf{G}^{**}$ is uniquely determined. By Assumption 1 and condition (3), $\mathbf{F} = \mathbf{G}^{**}$.

□

References

Christopher P. Adams. Finite mixture models with one exclusion restriction. *The Econometrics Journal*, 19:150–165, 2016.

Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77: 1229–1279, 2009.

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

Jushan Bai and Serena Ng. Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29, September 2013. doi:10.1016/j.jeconom.2013.03.007.

Bradley Efron. *Large-Scale Inference: Empirical Bayesian methods for estimation, testing and prediction*. Stanford University, Stanford CA, 2010.

Xiao Fu, Kejun Huang, Bo Yang, and Nicholas D. Sidiropoulos. Robust volume minimization-based matrix factorization for remote sensing and document clustering. *IEEE Transactions on Signal Processing*, 64(23):6254–6268, December 2016.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, September 1933. http://dx.doi.org/10.1037/h0071325.

Kejun Huang, Nicholas D Sidiropoulos, and Ananthram Swami. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Transactions on Signal Processing*, 2014.

D D Lee and H S Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

F. David Osinski and Jeremy A. Sandford. Merger remedies: A retrospective?analysis of pinnacle/ameristar. Federal Trade Commission, Februar 2017.

E-mail address: `cadams@ftc.gov`