Exploring intra-industry competitive heterogeneity
The identification of latent competitive groups

Wayne S. DeSarbo, Qiong Wang and Simon J. Blanchard
Department of Marketing, Pennsylvania State University, University Park, Pennsylvania, USA

Abstract
Purpose – The paper aims to examine the nature of competition within an industry by proposing and examining three separate sources of competitive heterogeneity: the strategies that industry members use, the performance that they obtain, and how effectively the strategies are utilized to obtain such performance results.

Design/methodology/approach – To do so, a restricted latent structure finite mixture model is devised that can quantify the contribution of these three potential sources of heterogeneity in the formulation of latent competitive groups within an industry. The paper illustrate this modeling framework with respect to COMPUSTAT strategy and performance data collected for public banks in the USA.

Findings – The paper shows how traditional conceptualizations via strategic or performance groups are inadequate to fully represent intra-industry heterogeneity.

Originality/value – This research paper proposes a new class of restricted finite mixture-based models, which fit a variety of alternative forms/models of heterogeneity. Information heuristics are developed to indicate “best model.”

Keywords Competitive strategy, Modelling, Public sector organizations, Banks, Economics, United States of America

Paper type Research paper

I. Introduction


There has been a considerable literature suggesting that these questions have become the driving forces behind much of the development of the field of management strategy and industrial organization/economics (Bower, 1986; Levinthal, 1995; Noda and Collis, 2001). This diversity in strategy and performance among firms in the same industry is called intra-industry competitive heterogeneity (Nelson, 1991; Noda and Collis, 2001). This phenomena has been documented in numerous empirical research studies (Bresnahan and Raff, 1991; Nelson, 1991; Noda and Collis, 2001; Hambrick et al., 2005). Additionally, Hambrick et al. (2005) find that intra-industry competitive heterogeneity has increased over time across a number of different major industries, and that increased intra-industry heterogeneity not only engendered greater intra-industry variety but also increased managerial discretion for the managers.
A number of plausible explanations have been provided for the existence and prevalence of intra-industry competitive heterogeneity. Early scholars suggested that every firm’s senior managers developed different corporate level strategic commitments to businesses that guided the firms’ actions, resulting in different performance outcomes (Andrews, 1971). Later, researchers adopted the principles of economics and shifted their attention away from internal corporate level commitments to the external analysis of a firm’s position relative to its competitors in a specific industry/business with respect to product and resource markets. For example, performance differences among firms in an industry were attributed to the firms’ pursuing unique product-market positions (Caves and Porter, 1977; Hatten and Schendel, 1977; McGee and Thomas, 1986) that were backed by product market heterogeneity and the inimitability of the necessary resources in the marketplace (Mehra and Floyd, 1998). Further, the resource-based view followed from this development and suggested that differential levels and/or effectiveness of resources (or capabilities) as a potential source of intra-industry competitive heterogeneity at the firm-level (Wernerfelt, 1984). Recently, building upon the resource-based-view, Hoopes et al. (2003) developed a broad theory of competitive heterogeneity to incorporate additional reasons to obtain competitive heterogeneity other than sticky resources (or capabilities). Whatever the particular underlying mechanism involved its formation, researchers have empirically exemplified intra-industry heterogeneity by deriving either strategic groups based on the diversity in various dimensions of strategy undertaken by firms in the same industry (Newman, 1978; Oster, 1982), or via performance groups based on the diversity in the various dimensions of performance realized by firms in the same industry (Wiggins and Ruefli, 1995). However, given the variety of the theoretical explanations of competitive heterogeneity, neither the strategic group nor performance group concepts alone can fully capture intra-industry competitive heterogeneity. Furthermore, when researchers tested the convergence of these two concepts, they found conflicting empirical evidence (Cool and Schendel, 1987; Fiegenbaum and Thomas, 1990; McNamara et al., 2003). That is, firms within the same strategic groups employing similar strategies often realized very different levels of performance. Similarly, researchers found that firms with similar levels of performance often employed very different strategies to get there. This suggests that representing competitive heterogeneity in a given industry may be more complex than what can be handled by current conceptual approaches and associated methodologies.

The purpose of this paper is to provide a new conceptualization and quantitative methodology to more fully capture the nature of intra-industry competitive heterogeneity in an empirically diagnostic manner as an attempt to resolve the conflicting evidence in the extant literature concerning this empirically observed asymmetry regarding strategic and performance groups. We first present a simple conceptual framework of intra-industry heterogeneity that is shown in Figure 1. As the traditional strategy-performance paradigm (related to the famous structure-conduct-performance framework in industrial economics) in Figure 1 shows, we suggest that there can be three distinct sources of competitive heterogeneity present within any given industry. First, there can be strategic heterogeneity defined as the diversity in the various dimensions of strategy undertaken by firms in the same industry. Second, there can be performance heterogeneity defined as the diversity in the various dimensions of performance realized by firms in the same industry. Finally, there can be impact heterogeneity, the diversity in the impact or effect that different strategies have on performance realized for firms in the same industry.
Note how this figure can also be employed to summarize much of the research performed in various related aspects of managerial strategy. For example, the research conducted in the area of strategic groups (McGee and Thomas, 1986) can be seen as attempts to investigate for the presence of strategic heterogeneity. The more recent work on performance groups (Wiggins and Ruefli, 1995) can be characterized as attempts to explore performance heterogeneity. Likewise, attempts to examine differences in the effects of strategy on resulting performance (DeSarbo et al., 2007) can be categorized as studying impact heterogeneity. Thus, Figure 1 shows a concise summary of these different components of intra-industry competitive heterogeneity. Later, we provide a mathematical derivation as the basis for this figure.

Our primary objective is to construct a general modeling framework for determining if any subset of these three types of intra-industry competitive heterogeneity is present in a given industry. In other words, we are interested in answering three simple questions. Given data in a specific industry, is there evidence for the presence of strategic heterogeneity? Is there evidence for the presence of performance heterogeneity? Is there evidence for the presence of impact heterogeneity? On the basis of the corresponding answers, we then wish to capture such heterogeneity by optimally devising latent competitive groups present in the industry. We thus wish to devise an empirical, finite mixture-based methodology, which can detect and calibrate the magnitudes of each of these three separate sources of heterogeneity (as well as any combination of them) for any specified industry. We devise model selection tests/heuristics so that an optimal model is selected which informs the researcher as to which subset of these effects is most salient in a specific application, as well as determine the number, composition, and size of the derived latent competitive groups that are simultaneously created.

We argue that to explore intra-industry competitive heterogeneity through the identification of latent competitive groups, all three types of previously mentioned heterogeneity must be considered and evidence of their presence or non-presence in the data estimated. Note, it is unknown a priori, which type heterogeneity (if any) is present in the data, nor how many latent competitive groups there are. We propose a model that identifies latent competitive groups whose sources of heterogeneity may or may not include strategic, performance, and impact heterogeneity.

Heterogeneity is represented via the estimation of model parameters by latent competitive group (as contrasted to aggregate sample estimates). Our methodology thus permits the identification of the number competitive groups in a given industry, their composition and size, and the types of heterogeneity that actually characterize the differences between these derived latent competitive groups. Nested forms of the general
model identifies industries in which the heterogeneity comes from strategic groups only, performance groups only, impact groups only, or various combinations of them which can be estimated and statistically compared to determine the most appropriate representation of intra-industry heterogeneity for a given empirical application.

The next section of the paper reviews the literature on strategic groups and performance groups, and develops our arguments for the importance of identifying latent competitive groups. Then, we present the technical details of the proposed finite mixture model complete with estimation algorithm and details concerning model selection heuristics. Section III presents an application of this new methodology to US public banks where a number of different alternative model solutions are estimated and tested for optimality. The final section provides a summary and discussion of the results, as well as directions for future research.

II. Conceptual background
The extant literature has represented competitive heterogeneity primarily through the functionalities of strategic groups (McGee and Thomas, 1986) or performance groups (Wiggins and Rueffli, 1995). Since the seminal work of Caves and Porter (1977) in the economics literature that proposed the theory of strategic groups, there has been increasing recognition that considerable differences exist between firms within a specified market or industry. Strategic groups represent sets of firms in an industry similar on some dimension such as cost structures, product diversification, formal organization, resources profiles, performance, and/or strategic variables (Amel and Rhoades, 1988; Baik and Lee, 2001; Caves, 1984; De Bondt et al., 1988; Kumar, 1990; Wijnberg, 1995). In fact, aside from economics, business researchers have also recognized and studied the notion of strategic groups as providing a viable means of assessing market structures (for review see Ketchen et al., 1997; McGee and Thomas, 1986). This research in economics and business shows that strategic groups exist:

- in diverse industries such as such as retailing (Lewis and Thomas, 1990), banking (Amel and Rhoades, 1988), pharmaceuticals (Cool and Schendel, 1987), brewing (Tremblay, 1985), and insurance (Fiegenbaum and Thomas, 1990); and
- in varied countries such as Belgium (Houthoofd and Heene, 1997), India (Kumar, 1990), Japan (Nair and Filer, 2002), and Spain (Amel and Rhoades, 1988), aside from the USA (Cool and Schendel, 1987; Fiegenbaum and Thomas, 1990).

The consensus in these research streams seems to be that strategic groups do exist across many types of different industries, and that there is a high degree of convergence in the results across different data types (Ketchen et al., 1997; Nath and Gruca, 1997; Porac and Thomas, 1994). This concept of strategic groups normally refers to the groups of firms in which each firm follows a strategy recipe that is similar to that pursued by other firms in the same strategic group, but different from the strategies followed by firms in other strategic groups (Caves and Porter, 1977; McNamara et al., 2003). Firms within a strategic group resemble one another closely in strategic recipes and, therefore, supposedly compete more fiercely with one another than with firms across strategic groups (Nair and Filer, 2002). There has been a plethora of work done in this area in economics by Newman (1978), Greening (1980), Tremblay (1985), Kumar (1987), Fiegenbaum and Primeaux (1987), Amel and Rhoades (1988), Barney and Hoskisson
Performance groups are defined by Wiggins and Ruefli (1995) as:

 [... a set of firms whose performance levels are statistically indistinguishable from those of other firms in the same performance group but are distinguishable from the performance levels of firms in other performance groups.

Mobility barriers are cited to provide a theoretical resource-based view argument for the existence and stability of strategic groups or performance groups (Caves and Porter, 1977). However, conflicting empirical evidence has been found to support the existence of mobility barriers. Thus, the ex post performance differences between the firms may not necessarily result from competition among relatively close rivals.

Thus, it is imperative to develop a broader concept to better capture the nature of intra-industry competitive heterogeneity, which may serve to resolve some of the limitations in the extant literature. Simply put, we examine both strategies and resulting performance simultaneously to describe intra-industry heterogeneity. The joint distribution of both sets of variables then, by definition, incorporates variation due to each set (strategy and performance), as well as their inter-relationship. Thus, we introduce the concept of latent competitive groups to more fully capture the complete picture of intra-industry competitive heterogeneity in terms of strategy, performance outcomes, and/or interrelationships between strategy and performance.

III. Mathematical derivation and the proposed latent structure methodology

As previously discussed, intra-industry competitive heterogeneity has traditionally been examined rather from either a strategic focus (in the strategic groups literature) or from a performance focus (in the performance groups literature). Do strategies always lead to the same performance measures? Do all firms that employ similar strategies obtain similar outcomes? The methodology that we propose here addresses three critical questions. In a given industry, are all firms employing similar strategies (strategic heterogeneity)? Are all firms equally efficient at implementing those strategies to achieve the associated indicators (impact heterogeneity)? Finally, regardless of the strategies that they use or the means that they use to achieve them, do they obtain similar performances (performance heterogeneity)? There are thus eight possible combinations of answers to these questions (i.e. eight different models) that can potentially characterize the nature of competitive heterogeneity in a given industry.

Consider, for example, the scenario of the presence of impact and performance heterogeneity in the data (answering yes to both the second and third question), but no strategic heterogeneity. This would suggest that firms in the industry use similar strategies but because the firms vary in how effective they are at using those strategies (impact heterogeneity), they obtain different performance levels (performance heterogeneity). This could describe the results found by several authors (Fiegenbaum and Thomas, 1990; Lewis and Thomas, 1990) who found that firms in the same strategic groups obtained different performance. In contrast, when we find that there is the presence of strategic and performance heterogeneity (but no impact heterogeneity), we would expect a correspondence between strategic and performance groups to be obtained (Osborne et al., 2001). Notice how both the “strategic groups” and the “performance groups” concepts each accommodate only one of the three potential
sources of such competitive heterogeneity (and then attempt to answer the remaining
two sources *post hoc* by descriptive analyses of the group members).

Thus, our objective is to conceptualize a modeling framework that will not only
identify the presence of heterogeneity via latent competitive groups, but also
characterize the type(s) of heterogeneity that differentiates those groups in one
integrative framework. It will allow for the isolation of the three types of variability
within/between the derived latent competitive groups (strategic, performance, and
impact) and be able to detect the three above sources simultaneously and separately.
It will also permit probabilistic assignment of each firm to one of the latent competitive
groups obtained. Hard assignments of members to groups (in which they belong to one
and only one group with certainty) should not be assumed as exemplar firms can vary
in how typical they are of one or more latent competitive groups. Finally, the modeling
framework will provide statistical evidence to justify the optimal representation of
intra-industry competitive heterogeneity.

The model that we propose can be estimated with eight variants that correspond to
the possible sources of competitive heterogeneity. Information criteria are used to
provide statistical evidence for optimal model construction. The estimation is done
under the general assumption that both strategic and performance variables come from
a finite mixture of multivariate normal distributions. The proposed model is thus in
line with the family of mixture of multivariate normal models, but it differs from
previous models in that various restrictions on the means and covariances between the
variables allow us to identify precisely the sources of the heterogeneity within an
industry. Keeping these concerns in mind, the next section introduces the mathematical
derivations as the basis for Figure 1 and the three components of intra-industry
heterogeneity, which also provides the foundation for the proposed restricted finite
mixture-based methodology.

A. The restricted finite mixture-based model
As mentioned, the primary objective of our modeling is to derive latent competitive
groups that can embody any or all three components of intra-industry heterogeneity:
strategic, performance, and impact. To see how this arises, we develop the following
notation and definitions.

Let:

\[ i = 1, \ldots, I \text{ firms.} \]

\[ j = 1, \ldots, J \text{ strategic variables.} \]

\[ k = 1, \ldots, K \text{ performance variables.} \]

\[ g = 1, \ldots, G \text{ unknown latent competitive groups.} \]

\[ X = \text{strategic random variables.} \]

\[ Y = \text{performance random variables.} \]

\[ Z = (Y, X) \text{ the concatenation of the two sets of random variables (strategy and performance).} \]
Assume that the joint density \( f(Z) \) is multivariate normal:

\[
\phi(Z) = \frac{1}{(2\pi)^{0.5(J+K)}|\Sigma_z|^{0.5}} \exp \left[ -0.5(\bar{z} - \mu_z)\Sigma_z^{-1}(\bar{z} - \mu_z) \right]
\]

(1)

where:

\[
\mu_z = (\mu_y; \mu_x)
\]

(2)

\[
\Sigma_z = \begin{bmatrix}
\Sigma_{yy} & \Sigma_{yx} \\
\Sigma_{xy} & \Sigma_{xx}
\end{bmatrix}.
\]

(3)

Then, by definition:

\[
\phi(Y, X) = f(Y|X)g(X),
\]

(4)

where:

\[
g(X) = \frac{1}{(2\pi)^{0.5/|\Sigma_{xx}|^{0.5}}} \exp \left[ -0.5(\bar{x} - \mu_x)\Sigma_{xx}^{-1}(\bar{x} - \mu_x) \right]
\]

(5)

and:

\[
f(Y|X) = \frac{1}{(2\pi)^{0.5K}|\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}|^{0.5}} \exp \left[ -0.5\left(\bar{y} - \mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}(\bar{x} - \mu_x)\right)^T \left(\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right)^{-1} \left(\bar{y} - \mu_y - \Sigma_{yx}\Sigma_{xx}^{-1}(\bar{x} - \mu_x)\right) \right]
\]

(6)

which is a conditional multivariate normal distribution with mean vector:

\[
\mu_y = \Sigma_{yx}\Sigma_{xx}^{-1}(\bar{x} - \mu_x)
\]

(7)

and covariance matrix:

\[
\Sigma_{yy} = \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}
\]

(8)

Note, elements from expression (7) are the regression predicted values of \( Y \) given \( X \), where:

\[
B = \Sigma_{yx}\Sigma_{xx}^{-1}
\]

(9)

are regression coefficients. Then, one can rewrite the joint density function \( \phi(Y, X) \) in expression (4) as:
\[
\phi(Y, X) = \frac{1}{(2\pi)^{0.5(J+K)}|\Sigma_{xx}|^{0.5}|\Sigma_{yy} - \Sigma_{yy}^{-1}\Sigma_{xy}|^{0.5}} \\
\times \exp \left[ -0.5 \left( (y - \mu_y)/(\Sigma_{yy} - \Sigma_{yy}^{-1}\Sigma_{xy}) \right)^{-1} (y - \mu_y) \\
- (x - \mu_x)'\Sigma_{xx}^{-1}\Sigma_{xy} (\Sigma_{yy} - \Sigma_{yy}^{-1}\Sigma_{xy})^{-1} (y - \mu_y) \\
- (y - \mu_y)'(\Sigma_{yy} - \Sigma_{yy}^{-1}\Sigma_{xy})^{-1}\Sigma_{yy}^{-1}(x - \mu_x) \\
+ (x - \mu_x)'\Sigma_{xx}^{-1}(x - \mu_x) \right] = \phi(Z) \\
= \frac{1}{(2\pi)^{0.5(J+K)}|\Sigma_{xx}|^{0.5}|\Sigma_{yy} - B\Sigma_{xy}|^{0.5}} \\
\times \exp \left[ -0.5 \left( (y - \mu_y)/(\Sigma_{yy} - B\Sigma_{xy}) \right)^{-1} (y - \mu_y) \\
- (x - \mu_x)'B(\Sigma_{yy} - B\Sigma_{xy})^{-1}B(x - \mu_x) \\
- (y - \mu_y)'(\Sigma_{yy} - B\Sigma_{xy})^{-1}B(x - \mu_x) \\
+ (x - \mu_x)'B(\Sigma_{yy} - B\Sigma_{xy})^{-1}B(x - \mu_x) \right] 
\]

Note, this decomposition allows one to isolate the various parameters governing the joint distribution of the elements of \( Z = (Y, X) \) including:

- \( \mu_y \) = the means of the performance variables.
- \( \mu_x \) = the means of the strategic variables.
- \( \Sigma_{yy} \) = the variances and covariances underlying the performance variables.
- \( \Sigma_{xx} \) = the variances and covariances underlying the strategic variables.
- \( \Sigma_{xy} = \Sigma_{yx} \) = the covariances between the performance and strategic variables.
- \( B = \Sigma_{yy}^{-1}\Sigma_{xy} \) = the regression coefficients linking strategy to performance.

As to be illustrated shortly, because of the mathematical equivalence between the joint density function and the product of the marginal and conditional distributions presented in expression (4) and expanded in expression (10), one can investigate these estimated parameters to obtain insight into the various components of intra-industry competitive heterogeneity as shown in Figure 1.

Now, to explicitly accommodate heterogeneity, we assume that the metric vector of concatenated strategy and performance variables for firm \( i, z_i \), is distributed as
a finite mixture of $G$ multivariate densities (McLachlan and Peel (2000) on how finite mixture distributions capture heterogeneity vis-à-vis the underlying support mixture distributions):

$$z_i \sim \sum_{g=1}^{G} w_g \phi(Y_i, X_i, \mu^g_x, \mu^g_y, \Sigma^g_x, \Sigma^g_y),$$  \hspace{1cm} (11)

where the $w_g$ are the (unknown) mixing proportions such that:

$$0 < w_g < 1, \hspace{0.5cm} \forall g = 1, \ldots, G \sum_{g=1}^{G} w_g = 1.$$  \hspace{1cm} (12)

Consistent with expression (1), we are now assuming a finite mixture of multivariate normal distributions. For a sample of $I$ independent firms, one can then form the complete likelihood expression:

$$L = \prod_{i=1}^{I} \left[ \sum_{g=1}^{G} w_g (2\pi)^{-\frac{(J+K)}{2}} |\Sigma_z|^{-0.5} \exp \left(-0.5(z - \mu_z)^{-1} (z - \mu_z) \right) \right].$$  \hspace{1cm} (13)

or the corresponding log likelihood:

$$LL = \sum_{i=1}^{I} \ln \left[ \sum_{g=1}^{G} w_g \phi_g (z_i, \mu^g_x, \Sigma^g_x) \right].$$  \hspace{1cm} (14)

Given $X, Y$, and a value of $G$, we need to estimate $w = (w_1, \ldots, w_G)$, $\mu^g_x, \Sigma^g_x$ (actually we use the decomposition in expression (10) to estimate $\mu^g_x, \mu^g_y, \Sigma^g_x, \Sigma^g_y, \Sigma^g_{xy}$ and $B^g$) so as to maximize equation (13) or equation (14) above, subject to the restrictions in equation (12). Note that once estimates of the parameters have been obtained within any iterate, estimates of the posterior probability $p_{ig}$ that firm $i$ comes from latent competitive group $g$ can be calculated for each $\tilde{z}_i$ by Bayes theorem:

$$p_{ig}(z_i|g, -) = \frac{w_g \phi_g (\tilde{z}_i|\mu^g_x, \Sigma^g_x)}{\sum_{l=1}^{G} w_l \phi_l (\tilde{z}_i|\mu^l_x, \Sigma^l_x)}.$$  \hspace{1cm} (15)

Intra-industry competitive heterogeneity is thus captured in this finite mixture-based methodology by these unknown $G$ latent competitive groups. Appendix 1 presents the technical details of the E-M algorithm employed for estimating of the various models’ parameters. Note, while finite mixtures of multivariate normal distributions have been well studied in the multivariate statistics literature (McLachlan and Peel, 2000), the methodological nuance here concerns the imposition of restrictions on subsets of model parameters and estimating mixtures of expression (10). In particular, as shown in Figure 1, $\mu^g_x$ and $\Sigma^g_{xx}$ provide information on strategic heterogeneity. The parameters $\mu^g_y$ and $\Sigma^g_{yy}$ render insight on performance heterogeneity. And, the parameters $\Sigma^g_{xy}$ and $B^g$ render insight on impact heterogeneity. Thus, the mathematical tautology presented in dealing
with the joint distribution of strategy and performance variables forces us to consider
not only the marginal heterogeneity of strategy and performance separately, but also
their interrelationships. The different model options of the algorithm are: strategic
heterogeneity estimated (yes/no), performance heterogeneity estimated (yes/no),
and impact heterogeneity estimated (yes/no). This leads to eight \(2^3\) different models
for us to compare within a specified industry. We now discuss the model selection
heuristics which allow us to determine “optimal model selection” (i.e. which of these eight
model portrayals is “best”) so that we can assess which of the eight models of
intra-industry heterogeneity is present in a given industry application.

B. Model selection tests
There are a number of different issues involving which latent structure model variant
best describes the structure in the input data. First, how does one go about selecting the
“best” value of \(G\) (the number of latent competitive groups)? The standard likelihood
ratio test of \(H_0: G = G + 1\) latent competitive group does not apply here as the
regularity conditions of this test are not satisfied given that the null hypothesis
corresponds to a boundary of the parameter space for the alternative hypothesis.
Second, how do we go about and select which combination of the three components of
intra-industry is optimal? For such questions, researchers have utilized various
information criteria for model selection in such finite mixture specifications. Such
measures attempt to balance the increase in fit obtained when a larger number of
parameters are estimated (i.e. when there are more latent competitive groups or when
additional types of intra-industry heterogeneity are estimated) with the need for a
parsimonious model that does not estimate unnecessary parameters. Wedel and
Kamakura (2000) provided a general form for such information measures (termed the
general information criterion (GIC)):

\[
GIC = -2 \ln L + Pd,
\]  

where \(P\) is the number of parameters estimated and \(d\) is some specified constant.
This \(d\) constant imposes a penalty on the likelihood, which reflects the increase in fit
(more parameters yield a higher likelihood) against the number of parameters
estimated. The constant \(d\) thus attempts to penalize models that have many
parameters, which do not significantly increase the likelihood. The classical Akaike
(1974) information criterion, designated as Akaike information criterion (AIC), arises
when \(d = 2\). Two other information criteria exist which penalize the likelihood more
heavily for additional parameters to be estimated. The Bayesian information criterion
(BIC) (Schwartz, 1978), BIC, occurs when \(d = \ln n\). The consistent AIC, CAIC
(Bozdogan, 1987), is formed in expression (16) above when \(d = \ln (N - 1)\). In all these
variants of expression (16) discussed above, one selects the model solution with the
lowest GIC measure (see Burnham and Anderson (2002) for a discussion of other model
selection heuristics). Note that both BIC and CAIC impose an additional sample size
penalty on the likelihood, and tend to be more conservative than AIC in favoring more
parsimonious models (i.e. model solutions with fewer groups); they also tend to result
in similar model selections. It has been argued that when we have reasons to believe
that the true model is included in the set, BIC is preferable due to consistency
properties (Kuha, 1994), and outperforms AIC (McQuarrie and Tsai, 1998). Yang and
Yang (2007) found that AIC obtains a decrease in average accuracy rates as sample
sizes increase. Thus, we will be emphasizing use of the BIC and CAIC criteria. Note, a variety of Monte Carlo simulations were conducted to test the performance of the proposed methodology. In particular, a number of synthetically created data sets were created whose structure was known as a number of data, model, error, etc. factors were altered in order to test the performance of the estimation algorithm and model selection tests in recovering the true structure. In all cases tested, the proposed methodology performed extremely well in converging to the true solution in terms of identifying the sources of heterogeneity, the number and sizes of latent competitive groups, and model parameters.

IV. Application: US public banks
A. The data
To illustrate the proposed latent competitive groups framework and latent structure methodology, we collected archival data from the COMPUSTAT Banks Database for the year 2006. The banking industry has been studied by many strategy researchers (Amel and Rhoades, 1988; McNamara et al., 2003; DeSarbo and Grewal, 2008). Furthermore, characterizing heterogeneity in the banking industry is nontrivial because the industry represents a turbulent environment with fuzzy boundaries (Fiegenbaum and Thomas, 1993). Consistent with the vast literature on strategic groups, we collect archival data pertaining to various performance and strategic variables (Cool and Schendel, 1987; Fiegenbaum and Thomas, 1990; DeSarbo and Grewal, 2008). Although researchers frequently use archival data, others also have employed perceptual cognitive measures (which can easily be incorporated in the proposed methodology). In an examination of alternative methods to assess strategic groups, Nath and Gruca (1997) compare archival methods with perceptual-cognitive data and find convergence (Ketchen et al., 1997). Thus, archival data on strategy and performance variables seem appropriate for our illustrative purposes in this research (DeSarbo and Grewal, 2008).

The COMPUSTAT Banks Database for the year 2006 consists of data on 749 banks, but we obtain complete data after imputation for 649 of them. The data did not exhibit severe patterns of missing data across the selected variables in Z. Missing values were imputed using the Bayesian multiple-imputation approach recommended by Schafer (1997) and Schafer and Graham (2002). Note, we use the total usable sample of N = 649 banks to summarize the pattern of competitive heterogeneity present in the entire industry of US public banks, as opposed to discerning inferences about aspects of competition which may be geographically constrained as noted in DeSarbo and Grewal (2008). Although, as noted by Petersen and Rajan (2002), banks since the 1990s are much more likely to lend over long geographical distances than they had earlier as the restrictions on geographical expansion became looser over time, and banks began to operate branches across state lines since, 1995. The advent of internet-based shopping for lower interest rates also aided in interstate banking trends.

B. Measures
We examined the major areas of strategy and performance dimensions for the banking industry through investigating a variety of financial ratios as obtained from an extensive literature search across multiple academic disciplines and in congruence with the recent framework presented by DeSarbo and Grewal (2008) for public banks.
Existing research seems to focus on a diverse set of strategic variables such as indicators of financial (Fiegenbaum and Thomas, 1990), product (Cool and Schendel, 1987), and marketing (Frazier and Howell, 1983) strategies. Consistent with research in finance (Brealey and Myers, 1988) and in the strategic groups area using financial variables (Baird et al., 1988) concerning banks (McNamara et al., 2003), we employ five different constructs of strategic variables as input into our procedure:

1. liquidity ratios;
2. leverage ratios;
3. product portfolio of loans;
4. product portfolio of deposits; and
5. asset quality ratios.

The selection of these measures is consistent with the strategies that banks use to manage the variety of risks (Park and Peristiani, 2007).

A listing of these strategic constructs and specific COMPUSTAT measures/ratios and computation in COMPUSTAT can be found in Appendix 2. Note, the selection of these specific items under each of the constructs was guided by mention in the literature, amounts of missing data in COMPUSTAT, collinearity, etc. Here, as an illustration of the methodology, we selected the current ratio (CR) of assets to liabilities to capture firm liquidity where the CR provides the defensive cash and near-cash resources for firms to meet claims for payment (Miller, 1997). We selected the times interest earned (TIE) as an indicator of leverage where the lower the TIE, the less debt and more equity the bank uses to finance its assets, i.e. the bigger the bank’s equity cushion (Brealey and Myers, 1988). To assess the product portfolio of loans, we used the ratio of gross loans to total investment securities (“LIS”; Rose, 1999). For the product portfolio of deposits, we used total investment securities as a ratio of total worldwide deposits (“ISD”; Rose, 1999). Finally, to reflect asset quality, we selected the noncurrent loans to loans (NCLL) ratio to secure the loan loss allowance to be adequate to cover potential loan losses (“NCLL”; Shaffer and Hoover, 2008; Allen and Rai, 1996).

In addition to these strategic variables, the literature supports three major bank performance constructs:

1. profitability;
2. financial market value ratios; and
3. efficiency ratios.

A listing of these performance constructs and specific COMPUSTAT measures/ratios and computation in COMPUSTAT can be found in Appendix 2 as well. Here, too, the selection of these specific items under each of the constructs was guided by mention in the literature, amounts of missing data in COMPUSTAT, collinearity, etc. Specifically, we use Tobin’s Q to assess the firms’ market value (Brealey and Myers, 1988). This ratio signals the intangible value of the firm and captures future earnings potential, in addition to current earnings (Tobin, 1969; Wernerfelt and Montgomery, 1988). To operationalize Tobin’s Q, we use the approximation detailed by Chung and Pruitt (1994) which is often used in empirical research (Bharadwaj et al., 1999)[3]. To assess bank efficiency, we use technical efficiency (TE) (Brealey and Myers, 1988).
Finally, we use net profit margin (NPM) to reflect profitability. Note, all three groups of variables selected for analysis here involving strategic and performance variables are consistent with DeSarbo and Grewal (2008) who conducted strategic group analysis on banks with 2001 COMPUSTAT data.

As an illustration and direct input to this new latent competitive group methodology, we use the five strategic variables \( (X) \) and the three performance variables \( (Y) \) as described above as input \( (Z) \) into the proposed methodology. All eight variables were column standardized to preclude scaling effects. Note that our latent competitive group conceptualization and restricted finite mixture-based latent structure scheme for estimation is sufficiently general to accommodate any specification of input variables in any specified industry.

C. Model selection

The purpose of the following analyses is to identify whether groups of banks exhibit different choices in the strategy that they pursue, the resulting performance that they obtain, and the extent to which the chosen strategies impact the obtained performance. To address this issue, the finite mixture model proposed above was first applied for estimating \( g = 1, \ldots, G \) latent competitive groups for the completely saturated case (all three types of heterogeneity are assumed in the model for now). The results are presented in Table I. As the BIC and CAIC statistics are both minimum at \( G = 5 \), we select \( G = 5 \) as the appropriate number of latent competitive groups. Note that \( G = 1 \) represents the model in which there is no intra-industry heterogeneity. This \( G = 1 \) aggregate model solution is clearly rejected by all the model selection heuristics examined in Table I. These results leave little doubt that there is some form of heterogeneity (be it strategic, performance, or impact) in the data.

Once the number of latent competitive groups is identified, we now need to determine the nature of the intra-industry heterogeneity. The eight different models were tested, each varying whether each type of heterogeneity should be estimated at the group level: strategic (yes/no), performance (yes/no), and impact (yes/no). The results of the estimation are presented in Table II. We find that the saturated model (with all three types of intra-industry heterogeneity present) is the one that fits the data best with the lowest BIC and CAIC results. Thus, the \( G = 5 \) saturated finite mixture model appears to best recreate the intra-industry heterogeneity present in this particular application, suggesting that there is objective statistical evidence for strategic, performance, and impact heterogeneity present in the dataset. Note, there are

<table>
<thead>
<tr>
<th>( G )</th>
<th>LNL</th>
<th>PAR</th>
<th>BIC</th>
<th>CAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-6,255.3</td>
<td>44</td>
<td>12,795.6</td>
<td>12,839.6</td>
</tr>
<tr>
<td>2</td>
<td>-5,091.4</td>
<td>89</td>
<td>10,759.0</td>
<td>10,848.0</td>
</tr>
<tr>
<td>3</td>
<td>-4,372.7</td>
<td>134</td>
<td>9,613.2</td>
<td>9,747.2</td>
</tr>
<tr>
<td>4</td>
<td>-4,035.8</td>
<td>179</td>
<td>9,230.7</td>
<td>9,409.7</td>
</tr>
<tr>
<td>5</td>
<td>-3,828.7</td>
<td>224</td>
<td>9,108.0</td>
<td>9,332.0</td>
</tr>
<tr>
<td>6</td>
<td>-3,741.8</td>
<td>269</td>
<td>9,225.6</td>
<td>9,494.6</td>
</tr>
<tr>
<td>7</td>
<td>-3,661.0</td>
<td>314</td>
<td>9,355.3</td>
<td>9,669.3</td>
</tr>
<tr>
<td>8</td>
<td>-3,544.6</td>
<td>359</td>
<td>9,413.9</td>
<td>9,772.9</td>
</tr>
</tbody>
</table>

**Table I.** Selection of the number of latent competitive groups
two important aspects illustrated in Table II that are additionally worth stressing. One, Model 5 which represents the strategic groups only formulation of competitive heterogeneity, is clearly rejected in this application and dominated statistically by the saturated model. Two, the same result holds for Model 3 which represents the performance groups only formulation, and is also rejected and dominated by the saturated latent competitive groups formulation. Thus, for this particular application, the two more traditional conceptual frameworks for representing intra-industry competitive heterogeneity (strategic and performance groups) provide very incomplete portrayals of such variation in this particular application.

The five extracted latent competitive groups were then examined for their composition and sizes. One does not wish to see evidence of outliers manifested by derived latent competitive groups with sparse membership (e.g. less than ten or so banks). The mixing proportions of latent competitive group membership \( w_1 = 0.226, w_2 = 0.225, w_3 = 0.145, w_4 = 0.340, \) and \( w_5 = 0.065 \) seem well dispersed (respective sample sizes per derived group were \( n_1 = 147, n_2 = 144, n_3 = 92, n_4 = 224, \) and \( n_5 = 42 \)). To further assess the degree of separation amongst the latent competitive groups, we calculated an entropy-based measure of the fuzziness of membership (Ramaswamy et al., 1993). The relative entropy measure is a number ranging between zero and one, where numbers close to zero indicate poor separation between the centroids of the conditional distributions for the derived latent competitive groups in the mixture distribution formulation. We see an entropy of 0.923, suggesting that the \( G = 5 \) latent competitive groups were very well separated. This is also indicative of the fact that the estimated posterior probabilities of latent competitive group membership displayed substantial variation and were close to that obtained by hard partitions (i.e. the estimated \( p_{kg} \) were close to zero or one).

V. Empirical results
A. Performance heterogeneity
Figure 2 shows the means for the various standardized performance measures across the derived five latent competitive groups. Here, zero reflects average performance given the variable standardization performed prior to the analyses. Latent competitive Group 4 achieves the highest profitability as represented by NPM, but at the same time it suffers from the lowest market value (Tobin’s Q). In comparison, latent competitive Group 3 obtains the highest market value, yet is the worst performer among all the groups in

<table>
<thead>
<tr>
<th>Model</th>
<th>LNL</th>
<th>PAR</th>
<th>BIC</th>
<th>CAIC</th>
<th>Heterogeneity included?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-6,255.3</td>
<td>44</td>
<td>12,795.5</td>
<td>12,839.5</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>-5,152.4</td>
<td>184</td>
<td>11,593.4</td>
<td>11,792.4</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>-6,627.5</td>
<td>55</td>
<td>13,611.1</td>
<td>13,666.1</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>-5,101.2</td>
<td>199</td>
<td>11,393.9</td>
<td>11,577.9</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>-5,745.1</td>
<td>65</td>
<td>11,911.1</td>
<td>11,976.1</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>-4,068.1</td>
<td>209</td>
<td>9,489.6</td>
<td>9,688.6</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>-5,463.1</td>
<td>80</td>
<td>11,444.2</td>
<td>11,524.2</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>-3,828.7</td>
<td>224</td>
<td>9,107.9</td>
<td>9,331.9</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table II. Model selection: intra-industry heterogeneity sources
terms of profitability. Also, both latent competitive Groups 3 and 4 have average TE when compared to the other the groups. Latent competitive Group 5 has the highest TE, but it achieves lower profitability and market value. Similarly, latent competitive Group 2 obtains decent TE, but is behind in market value as well as in profitability. Finally, latent competitive Group 1 has average profitability and market value, along with much higher technical inefficiency. For labeling, we hereafter refer to latent competitive Group 1 as the “under-performing” group containing banks such as First State Bancorporation, Atlantic Bancorp, Inc., etc. We refer latent competitive Group 2 as the “mediocre-performing” group containing such member banks as First Charter Corp., and Mercantile Bank Corp., etc. Latent competitive Group 3 is labeled as “market-driven” group, which involves banks such as M&T Bank Corp., Fifth Third Bancorp., etc. Latent competitive Group 4 is called “profit-driven” group, which includes the banks such as Omega Financial Corp., Citizens First Bank, etc. Finally, our latent competitive Group 5 is the “efficiency-driven” group, which consists of banks such as Sovereign Bancorp., Northwest Bancorp., etc. Considering that performance outcomes might be obtained by:

- different strategies adopted by different latent competitive groups, as well as; and
- the differential bank effectiveness of such strategies across these latent competitive groups, we turn our attention to each of these groups’ strategic orientation, and their impact on the performance dimensions.

B. Strategic heterogeneity

As shown in Figure 3, the banks in the market-driven group pursue high-asset liquidity as exemplified by the highest CR in our study. In particular, most of the banks in the market-driven group emphasize a high proportion of total investment securities over total worldwide deposits (ISD), and a low ratio of gross loans over total investment securities (LIS). Banks in the market-driven group use more debt than equity to finance their assets (e.g. the lowest TIE). More interestingly, banks in market-driven group have the highest NCLL, which may indicate problems in the long-term if their borrowers encounter financial problems of their own. In sum, it seems that market-driven banks are trading off levels of equity capital to absorb the fluctuations or variability of cash flows for high-asset liquidity and perceived value in
the eyes of the investors. In other words, compared to the other latent competitive groups, banks in market-driven group are the best candidates to deal with liquidity risks which represent the risks like a sudden and unexpected increase in liability, where withdrawals may require the commercial bank to liquidate assets in a very short period of time and at low prices (Acharya and Pedersen, 2005).

In contrast to this approach, the banks in profit-driven group seek a high level of equity capital in its assets. Such banks enjoy the highest TIE ratio, which implies that adequate levels of equity capital equip these banks for fluctuations that might occur. They also minimize noncurrent loans in their portfolio to avoid potential loss from the loans. Banks in the profit-driven group do not pay as much attention to liquidity risk as do the banks in market-driven group. Thus, they are content with mediocre levels of CR, product loan ratio, and product deposit ratio. In short, the banks in the profit-driven group are more capable to cope with insolvency risks which refer to the risk that a commercial bank may not have enough capital to offset a sudden decline in the value of its assets relative to its liabilities, than the banks in other latent competitive groups (Kim and Santomero, 1998).

The banks in the efficiency-driven group pursue higher quality loans. Specifically, these firms manage to have the highest LIS ratio and the lowest ISD ratio. This can be explained by the fact that the revenue generated from investment securities is generally low compared to that from loans, thus minimizing the amount of held investment in securities and using the loans to generate the larger flow of revenue income. A potential problem with this approach is that loans are also, for most banks, the least liquid asset items and a major source of credit and liquidity risk. Not surprisingly, members of the efficiency-driven group acknowledge the “dark side” of loans and manage to have the

Figure 3.
Strategic heterogeneity
lowest noncurrent loans over total loan ratio. In this way, these banks make the most efforts to secure the quality of their loans compared to what the banks in other latent competitive groups do. Given the discussion thus far, banks in the efficiency-driven group appear as the best candidates to deal with credit risks which refer to the risk that promised cash flows from loans and securities held by commercial banks may not be paid in full (Jarrow and Turnbull, 1995).

For the banks in both mediocre-performing and under-performing groups, it is not surprising to find that they do not stand apart in any of the strategies they employ compared to market-driven, profit-driven, or efficient-driven groups. Moreover, little difference in the strategies (with the exception of perhaps ISD) have been observed between the banks in mediocre-performing and under-performing groups. This affirms the theory of competitive heterogeneity that reasons that other than sticky resources (or strategies in our study) should be accounted for concerning performance heterogeneity. Consequently, we argue that it is also important to examine impact heterogeneity to investigate how similar strategies may lead to different performances.

C. Impact heterogeneity

Thus far, we have focused on high and low levels of performance and strategy variables across the five derived latent competitive groups. Table III presents the estimated $B^g$ coefficients per performance variable by derived latent competitive group, which aptly illustrate that the heterogeneous impacts of strategy on performance play an important role in the formation of intra-industry heterogeneity. There are a number of interesting findings here in light of our discussion thus far with respect to Figures 2 and 3. First, as mentioned earlier, we mentioned that the banks in the mediocre-performing and the under-performing groups did not differ that much from each other in terms of mean

<table>
<thead>
<tr>
<th>Performance Variable</th>
<th>Under-performing</th>
<th>Mediocre-performing</th>
<th>Market-driven</th>
<th>Profit-driven</th>
<th>Efficiency-driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net profit margin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR</td>
<td>0.046</td>
<td>-0.085</td>
<td>-0.064 *</td>
<td>-0.009</td>
<td>-0.262 **</td>
</tr>
<tr>
<td>TIE</td>
<td>0.760 **</td>
<td>0.723 **</td>
<td>1.073 **</td>
<td>0.806 **</td>
<td>0.856 **</td>
</tr>
<tr>
<td>ISD</td>
<td>0.342</td>
<td>1.586 **</td>
<td>0.207 **</td>
<td>1.180 **</td>
<td>1.018</td>
</tr>
<tr>
<td>LIS</td>
<td>2.682</td>
<td>1.441 **</td>
<td>2.578 **</td>
<td>3.193 **</td>
<td>0.116</td>
</tr>
<tr>
<td>NCLL</td>
<td>-0.041</td>
<td>-0.130</td>
<td>-0.165 **</td>
<td>-0.175 *</td>
<td>0.238</td>
</tr>
<tr>
<td>Tobin’s Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR</td>
<td>0.118</td>
<td>0.190 **</td>
<td>0.305 **</td>
<td>0.159 **</td>
<td>0.271 **</td>
</tr>
<tr>
<td>TIE</td>
<td>0.320 **</td>
<td>0.268 **</td>
<td>0.168</td>
<td>0.376 **</td>
<td>0.212 **</td>
</tr>
<tr>
<td>ISD</td>
<td>1.272 **</td>
<td>4.043 **</td>
<td>0.530 **</td>
<td>3.385 **</td>
<td>-1.078</td>
</tr>
<tr>
<td>LIS</td>
<td>15.610 **</td>
<td>3.250 **</td>
<td>5.737 **</td>
<td>11.620 **</td>
<td>-0.074</td>
</tr>
<tr>
<td>NCLL</td>
<td>0.363 **</td>
<td>0.445 **</td>
<td>0.230 **</td>
<td>0.335 **</td>
<td>1.058 **</td>
</tr>
<tr>
<td>Technical efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR</td>
<td>-0.904</td>
<td>-0.011</td>
<td>0.042</td>
<td>0.096 **</td>
<td>0.221 **</td>
</tr>
<tr>
<td>TIE</td>
<td>-0.002</td>
<td>-0.052</td>
<td>-0.385 **</td>
<td>-0.021</td>
<td>-0.004</td>
</tr>
<tr>
<td>ISD</td>
<td>1.545 **</td>
<td>4.919 **</td>
<td>0.442 **</td>
<td>2.699 **</td>
<td>-0.050</td>
</tr>
<tr>
<td>LIS</td>
<td>20.183 **</td>
<td>4.705 **</td>
<td>11.200 **</td>
<td>10.714 **</td>
<td>-0.034</td>
</tr>
<tr>
<td>NCLL</td>
<td>0.128 *</td>
<td>0.246 **</td>
<td>0.007</td>
<td>0.248 **</td>
<td>1.093 **</td>
</tr>
</tbody>
</table>

Table III. Estimates of impact heterogeneity

Note: Significance at: * $p < 0.05$ and ** $p < 0.01$ levels
strategies or performance outcomes. However, when we examine Table III carefully, we can identify dramatic differences in impact heterogeneity between these two groups. For example, CR, which is used to equate firms to deal with liquidity risks has no impact on the market value for the firms in the under-performing group, but has significant positive impact ($0.190, p < 0.01$) on the market value for the firms in mediocre-performing groups. Across the three performance variables, the impact of LIS is substantially greater for the under-performing group as compared to the mediocre performing group. Indeed, improving LIS will have greatest impact for the under-performing group for increasing all three performance measures, whereas increasing ISD and LIS will most benefit the mediocre-performing group across all three performance measures.

Second, from Figure 2, we see that banks in profit-driven group seem to possess below average market loan portfolios (LIS). However, Table III shows that LIS is one of the most effective means to increase all three performance results for banks in this profit-driven group. This group also is characterized by below average product portfolio of deposits (ISD), which also significantly impacts all three performance measures. An argument can therefore be made that this profit-driven group is engaging in very sub-optimal strategies.

Third, a similar description can also be made for banks in the market-driven group. Figure 2 shows that banks in market-driven group seem to possess below average market loan portfolios (LIS). And, Table III shows that LIS is one of the most effective means to increase all three performance results for banks in this market-driven group too. Unlike the previous group, however, this group also is characterized by the highest product portfolio of deposits (ISD) which also significantly impacts/improves all three performance measures. Improving the loan portfolio appears to reap the most benefits for this group.

Fourth, although banks that are members of the other latent competitive groups greatly benefit from increases in adopting loan portfolios to improve their efficiency, the members of the efficiency-driven group seem to be characterized by their superior liquidity capabilities. It is also interesting to note how members of the efficiency-driven group could improve their efficiency even more by improving their noncurrent loans ratio. Perhaps, their best strategy would thus be to reduce their liquidity and focus on their noncurrent loans, simultaneously improving their efficiency, market value, and profitability.

Thus, we see different paths (impact heterogeneity) to increased performance for each of the various latent competitive groups, as well as potential inefficiencies with respect to resource allocation across the various strategic variables. The methodology aptly contrasts levels of strategic investment (in Figure 3) with the impact of those strategic investments (Table III) across the three performance measures.

D. Empirical comparison with traditional methodology
Cluster analysis has been the primary methodology utilized to estimate and identify strategic and performance groups in the strategy literature (DeSarbo and Grewal, 2008). How would a more traditional clustering approach involving the popular K-MEANS cluster analysis (MacQueen, 1967) perform in comparison with the proposed new methodology in modeling intra-industry competitive heterogeneity? Conceptually, K-MEANS would only accommodate first moment heterogeneity involving the means.
of the input strategic and performance variables (not the covariances). Additionally, it cannot explicitly account for impact heterogeneity.

For a statistical comparison, we performed a K-MEANS analysis for \( G = 5 \) clusters using \( Z = [X; Y] \) as input. This produced a log likelihood value of \(-7,025.55\) with \( \text{BIC} = 14,427.52 \) and \( \text{CAIC} = 14,471.52 \). As can be seen from a comparison of the various model selection heuristics from the eight different models fit in Table II, the K-MEANS solution is far worse than any of these eight solutions; in fact, it is statistically dominated by all of them!

There are many plausible explanations for this. One, K-MEANS is restricted to estimate partitions and thus cannot account for the fuzziness of individual membership of banks to the different groups. Two, as mentioned above, K-MEANS cannot accommodate impact heterogeneity or second moment heterogeneity (involving covariances). Third, the error sums-of-squares loss function optimized in K-MEANS is different from the log likelihood expression involved in the proposed approach. Finally, K-MEANS is a deterministic procedure whereas the proposed restricted finite mixture approach is stochastic.

VI. Discussion
A. Summary
In this paper, we provide both a conceptual and mathematical treatment of intra-industry competitive heterogeneity. In particular, we have made the case for examining both strategy and performance variables simultaneously in assessing competitive heterogeneity in an industry. And, when dealing with the joint distribution of both sets of variables, we demonstrate mathematically that three different components of intra-industry heterogeneity can potentially exist in any given industry. This research shows that strategic, performance, and/or impact heterogeneity can occur simultaneously in a given industry and can be successfully represented through our proposed finite mixture-based methodology in deriving latent competitive groups. We introduce the concept of latent competitive groups to illustrate the nature of intra-industry heterogeneity from a perspective that reconciles the conflicts in the extant literature regarding the convergence of strategic and performance heterogeneity (McNamara et al., 2003; Wiggins and Ruefli, 1995). Furthermore, the proposed methodology is able to generate individual level estimates by individual firm to explore this intra-industry competitive heterogeneity (Appendix 1). Finally, this new methodology can be utilized with any set of input variables (not just those specified in the application with the banks).

The proposed finite mixture-based model that we proposed permits the identification of such latent competitive groups and has been shown to bring to light all three types of intra-industry heterogeneity (strategy, performance, and impact) in an application to a sample of 649 commercial public banks in the USA. The results indicated the presence of five latent competitive groups that were then tied to interpretations of their strategy, their performance, and the impact their strategy has on their subsequent performance. The model selection criteria indicated that the intra-industry heterogeneity in this market should be explained in terms of all three components: strategic, performance, and impact heterogeneity. We then illustrated how these three types of heterogeneity are manifested in the data. We have first found that banks differ in the performance results they obtain from their different strategies. We found that a small sample of banks was driven by results in efficiency, but that more banks were driven by results in market
value or profitability. Additionally, we found that a large chunk of the banks seemed to be rather average on most performance indicators. More interesting results appeared when we investigated how these various results were achieved through the (in)efficient use of various strategies.

With differences not only in the use of strategies but also on their impact on various aspects of performance, it becomes clear that the industry displays sources of heterogeneity from how banks choose different strategies and how they manage to leverage them (successfully or not) to obtain performance results. (Interestingly, to further profile the derived latent competitive groups, we assessed the scale of the operations of the banks and examined the number of employees, total assets, gross loans, total worldwide deposits, total interest expenses, net income, and total borrowing by derived latent competitive group, but found no statistical differences between these groups with respect to these firmographic variables.)

B. Limitations
Banks are unique enterprises whose operations are largely affected by state, federal, and/or internal financial regulations, and thus certain strategies adopted by the banks might reflect the pressure of such regulations rather than managerial decisions. For example, US commercial banks are prohibited from making loans exceeding 10 percent of their own equity capital funds to any one company or borrower.

Second, we only selected one variable to capture each construct/dimension of strategy and performance as an illustration of the proposed methodology. A major part of this limitation was due to the incomplete data available from COMPUSTAT concerning many other financial variables/ratios. Although we acknowledge omitted variables might constrain the analysis, we do not expect such a limitation to significantly alter the identification and the implications of latent competitive groups. In addition, as DeSarbo and Grewal (2008) mention, many variables such as advertising budgets, media vehicles, promotion mix allocation, etc. are extremely rare for the researcher to gain access to. However, these data limitations do not constrain our methodological contributions.

Third, we did not explicitly model intra-industry heterogeneity dynamics either by using lagged strategic and performance variables (which can be accommodated in the presented methodology), or utilizing time series of data (which would require modifications in the likelihood function). The primary contribution of this paper is methodological and we utilize the data presented as an illustration of the methodology (as opposed to presenting externally valid conclusions about the entire banking industry).

C. Future research
First, our research calls for future studies on understanding the dynamics of how intra-industry competitive heterogeneity is formed. Hambrick et al. (2005), from a managerial discretion perspective, acknowledge an increase in intra-industry heterogeneity over the latter decades of the twentieth century. However, identifying the temporal changes of the strategic, performance, and impact heterogeneities remains available for further investigation. Second, more Monte Carlo testing of the proposed methodology is desired to examine the effects of such things as outliers, model mis-specification, non-normality, etc. Finally, we hope that our study inspires scholars to apply our finite mixture methodology to identifying competitive latent
groups in other industries and other contexts when multiple sources of heterogeneity might be present. For example, additional external validity can be obtained by perhaps augmenting the COMPUSTAT data with FDIC bank data, which contains observations for over 8,000 banks.

Notes
1. Caves (1984) refers to this joint research in economics and business as an interchange that has been “fruitful” and goes on to suggest that “more fruit remains for the picking.”
2. Overall, 85 percent of cases had no missing data and less than 3 percent of the cells were missing (in Z).
3. We specify, \( Q = \frac{(\text{MVE} + \text{PS} + \text{DEBT})}{\text{TA}} \), where \( Q \) = Tobin’s Q; \( \text{MVE} = (\text{closing price of share at end of the financial year}) \times (\text{number of common shares outstanding}) \); \( \text{PS} \) = liquidating value of the firm’s outstanding preferred stock; \( \text{DEBT} = (\text{current liabilities} - \text{current assets}) + (\text{book value of inventories}) + (\text{long-term debt}) \); and \( \text{TA} \) = book value of total assets.
4. A complete listing of the \( N = 649 \) banks by derived latent competitive group can be obtained from the authors.

References


Further reading

Appendix 1. The E-M estimation algorithm employed for deriving latent competitive groups and intra-industry heterogeneity
We introduce non-observed data, $h_{ig}$, indicating 1 when bank $i$ belongs to latent competitive group $g$, 0 otherwise. We assume that the $h_{ig}$ are i.i.d. multinomial:

$$f(h_{1g}, \ldots, h_{lg}) = \prod_{i=1}^{I} w_{ig}^{h_{ig}}.$$  \hspace{1cm} (A1)

It is also assumed that $z_{i}$ given $(h_{1g}, \ldots, h_{lg})$ are conditionally independent and that $z_{i}$ given $(h_{ig}, \ldots, h_{lg})$ has density:

$$\phi_{g}(z_{i}|h_{ig}) = \prod_{i=1}^{I} \phi_{g}(z_{i} | \mu_{g}^{z}, \Sigma_{g}^{z})^{h_{ig}}.$$  \hspace{1cm} (A2)

Using equations (13) and (A2), we can form the log-likelihood:

$$LL(z|h, \ldots) = \sum_{g=1}^{G} \sum_{i=1}^{I} h_{ig} \left( \ln w_{ig} + \ln \phi_{g}(z_{i} | \mu_{g}^{z}, \Sigma_{g}^{z}) \right).$$  \hspace{1cm} (A3)

In the E-step, the log-likelihood is replaced by its expectation, based on estimates of $\mu_{g}^{z}, \Sigma_{g}^{z}$. In the M-step, the expectation of the log-likelihood is maximized with respect to $\mu_{g}^{z}, \Sigma_{g}^{z}$ to obtain new estimates.

**E-step**
To get the expectation of equation (A2), we replace $h_{ig}$ by their expected value $E(h_{ig} | \mu_{g}^{z}, \Sigma_{g}^{z})$.

The conditional distribution of $z_{i}$ given $\mu_{g}^{z}, \Sigma_{g}^{z}$:

$$E(h_{ig} | z_{i}, \ldots) = \frac{w_{ig} \phi_{g}(z_{i} | \mu_{g}^{z}, \Sigma_{g}^{z})}{\prod_{i=1}^{I} w_{ig} \phi_{g}(z_{i} | \mu_{g}^{z}, \Sigma_{g}^{z})}.$$  \hspace{1cm} (A4)
which is equal to the posterior \( p_\theta \) as defined in equation (15). Estimates of \( \tilde{p}_{ig}, \tilde{\theta}_{ig} \), are obtained by evaluating equation (A4) at current estimates of \( \hat{\mu}_i, \hat{\Sigma}_i \). Estimates of \( w_g, \hat{w}_g \), are obtained by:

\[
\hat{w}_g = \frac{\sum_{I=1}^{I} \tilde{p}_{ig}}{I}.
\]

\textit{M-step}

To maximize the expectation of the log-likelihood in equation (A3) with respect to \( (\mu_i^g, \Sigma_i^g, \Sigma_{ii}^g, \Sigma_{iy}^g, \Sigma_{xy}^g) \), \( h_{ig} \) are replaced in equation (A3) by \( \tilde{p}_{ig} \), their current expectations. Setting the first partial derivative to zero with respect to \( \mu_i^g \), we obtain:

\[
\frac{\partial (LL)}{\partial \mu_i^g} \propto \sum_{i=1}^{I} \tilde{p}_{ig} \left( (\Sigma_i^g)^{-1} + (\Sigma_i^g)^{-1} \right) \mu_i^g - \left( (\Sigma_i^g)^{-1} + (\Sigma_i^g)^{-1} \right) \epsilon_i = 0.
\] (A5)

Solving for \( \mu_i^g \), we get that:

\[
\mu_i^g = \frac{\left( (\Sigma_i^g)^{-1} + (\Sigma_i^g)^{-1} \right) \sum_{i=1}^{I} \tilde{p}_{ig} \left( (\Sigma_i^g)^{-1} + (\Sigma_i^g)^{-1} \right) \epsilon_i}{\sum_{i=1}^{I} \tilde{p}_{ig}}.
\] (A6)

Additionally, if both vectors of means are not at the latent competitive group level, we obtain:

\[
\mu_i = \left( \sum_{g=1}^{G} \sum_{i=1}^{I} \tilde{p}_{ig} (\Sigma_i^g)^{-1} + (\Sigma_i^g)^{-1} \right) \left( \sum_{i=1}^{I} \tilde{p}_{ig} (\Sigma_i^g)^{-1} + (\Sigma_i^g)^{-1} \right) \epsilon_i.\] (A7)

If we wish to solve for the means of the strategic variables at the latent competitive group level and the means of the performance variables not at the latent competitive group level, we can set first partial derive with respect to \( \mu_i^g \):

\[
\frac{\partial (LL)}{\partial \mu_i^g} \propto \partial \left( \sum_{i=1}^{I} \tilde{p}_{ig} \left( (\Sigma_i - \mu_i^g) (\Sigma_{xx}^{-1} g - \Sigma_{xy}^{-1} g) (\Sigma_{yx}^{-1} g - \Sigma_{yy}^{-1} g) \right)^{-1} \left( \epsilon_i - \mu_i^g \right) \right.
\]

\[
- \left( \epsilon_i - \mu_i^g \right) (\Sigma_i^g - \Sigma_{xy}^{-1} g) (\Sigma_{yx}^{-1} g) \left( \Sigma_{xx}^{-1} g - \Sigma_{xy}^{-1} g \right) \left( \epsilon_i - \mu_i^g \right)
\]

\[
+ \left( \epsilon_i - \mu_i^g \right) (\Sigma_i^g - \Sigma_{xy}^{-1} g) (\Sigma_{yx}^{-1} g) \left( \Sigma_{xx}^{-1} g - \Sigma_{xy}^{-1} g \right) \left( \epsilon_i - \mu_i^g \right)
\]

\[
\times + \left( \epsilon_i - \mu_i^g \right) (\Sigma_i^{-1} g) \left( \mu_i - \mu_i^g \right) \right) / \partial \mu_i^g.
\] (A8)

Now, let:

\[
\delta_i = \Sigma_{xx}^{-1} g \Sigma_{xy}^{-1} g \left( \Sigma_{yy}^{-1} g - \Sigma_{yx}^{-1} g \Sigma_{xy}^{-1} g \right) \left( \epsilon_i - \mu_i^g \right),
\]

\[
\epsilon_i = \left( \epsilon_i - \mu_i^g \right) \left( \Sigma_{xy}^{-1} g \left( \Sigma_{xx}^{-1} g - \Sigma_{yx}^{-1} g \Sigma_{xy}^{-1} g \right) \right) \left( \Sigma_{yy}^{-1} g - \Sigma_{yx}^{-1} g \Sigma_{xy}^{-1} g \right) \left( \epsilon_i - \mu_i^g \right)
\]

and:

\[
F_i = \Sigma_{xx}^{-1} g \Sigma_{yx}^{-1} g \Sigma_{yy}^{-1} g \left( \Sigma_{xx}^{-1} g - \Sigma_{yx}^{-1} g \Sigma_{xy}^{-1} g \right) \left( \Sigma_{yy}^{-1} g - \Sigma_{yx}^{-1} g \Sigma_{xy}^{-1} g \right) \left( \epsilon_i - \mu_i^g \right).
\]
Then, setting the partial derivative equal to zero, we obtain:

\[
\frac{\partial (LL)}{\partial \mu^g_y} \approx \min_{i=1}^l \hat{p}_g \left[ \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' \right) \mu^g_y - \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' \right) \xi_i \right]
\]

\[+ (E + E')\mu^g_y - (E + E')\xi_i - d_i - \xi'_i \right] = 0.
\]

Solving for \( \mu^g_y \), we can obtain:

\[
\sum_{i=1}^l \hat{p}_g \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' + E + E' \right) \mu^g_y
\]

\[= \sum_{i=1}^l \hat{p}_g \left[ \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' + E + E' \right) \xi_i - d_i - \xi'_i \right],
\]

and:

\[
\mu^g_y = \frac{\sum_{i=1}^l \hat{p}_g \left[ \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' + E + E' \right) \xi_i - d_i - \xi'_i \right]}{\sum_{i=1}^l \hat{p}_g}
\]

If we are interested in \( \mu_x \) instead of \( \mu^g_y \), i.e. one believes that the means of the strategic variables do not differ per latent competitive group, a similar formulation to equation (A11) follows:

\[
\mu_x = \left( \sum_{g=1}^G \sum_{i=1}^l \hat{p}_g \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' + E + E' \right) \right)^{-1}
\]

\[\times \sum_{g=1}^G \sum_{i=1}^l \hat{p}_g \left[ \left( \sum_{xx}^{-1(g)} + \left( \sum_{xx}^{-1(g)} \right)' + E + E' \right) \xi_i - d_i - \xi'_i \right]
\]

If we wish to solve for the means of the strategic variables at the latent competitive group level and the means of the performance variables not at the latent competitive group level, we can set first partial derivate with respect to \( \mu^g_x \):

\[
\frac{\partial (LL)}{\partial \mu^g_x} \approx \min_{i=1}^l \hat{p}_g \left[ \left( \sum_{xy}^{-1(g)} + \left( \sum_{xy}^{-1(g)} \right)' \right) \right] \mu^g_x - \left( \sum_{xy}^{-1(g)} + \left( \sum_{xy}^{-1(g)} \right)' \right) \xi_i \right]
\]

\[+ \left( \sum_{xy}^{-1(g)} + \left( \sum_{xy}^{-1(g)} \right)' \right) \xi_i - d_i - \xi'_i \right] / \partial \mu^g_x
\]

Now, let:

\[
A = \left( \sum_{xy} - \sum_{xy} \sum_{xx}^{-1(g)} \sum_{xy} \right)^{-1},
\]

\[
b_i = \left( \sum_{xy} - \sum_{xy} \sum_{xx}^{-1(g)} \sum_{xy} \right)^{-1},
\]
and:
\[ \xi = \left( \Sigma_{xy}^{g} - \Sigma_{xy}^{g} \Sigma_{xx}^{-1(g)} \Sigma_{yx}^{g} \right)^{-1} \Sigma_{xy}^{g} \Sigma_{xx}^{-1(g)} \left( \bar{x}_i - \mu_x^g \right). \]

Then, setting the partial derivative equal to zero:
\[
\frac{\partial(\mathrm{LL})}{\partial \mu_y^g} = \sum_{i=1}^{l} \hat{p}_{ig} \left[ (A + A') \mu_y^g - (A + A') y_i - b_i' - \xi \right] = 0. \tag{A14}
\]
Solving for \( \mu_y^g \), we obtain:
\[
\sum_{i=1}^{l} \hat{p}_{ig} (A + A') \mu_y^g = \sum_{i=1}^{l} \hat{p}_{ig} [ (A + A') y_i - b_i' - \xi ] \tag{A15}
\]
and:
\[
\mu_y^g = \frac{(A + A')^{-1} \sum_{i=1}^{l} \hat{p}_{ig} [ (A + A') y_i - b_i' - \xi ]}{\sum_{i=1}^{l} \hat{p}_{ig}}. \tag{A16}
\]

If we are interested in \( \mu_y \) instead of \( \mu_y^g \), a similar formulation to equation (A16) follows:
\[
\mu_y = \left( \sum_{g=1}^{G} \sum_{i=1}^{l} \hat{p}_{ig} (A + A') \right)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{l} \hat{p}_{ig} [ (A + A') y_i - b_i' - \xi ] \tag{A17}
\]

With regards to the covariance matrices, the case where both strategic and performance variables are estimated at the latent competitive group level has already been established (He et al., 2006; McLachlan and Krishnan, 1997):
\[
\Sigma_y^g = \frac{\sum_{h=1}^{l} \hat{p}_{ig} (z_i - \mu_y^g) (z_i - \mu_y^g)'}{\sum_{i=1}^{l} \hat{p}_{ig}}, \tag{A18}
\]
and the case in which neither are estimated at the latent competitive group level is:
\[
\Sigma_y = \frac{\sum_{g=1}^{G} \sum_{i=1}^{l} \hat{p}_{ig} (z_i - \mu_y^g) (z_i - \mu_y^g)'}{I}. \tag{A19}
\]

Because the computation of the covariance matrix among the strategic variables does not involve the performance variables, the latent competitive group level covariance matrix for the strategic variables \( \Sigma_{xx} \) follows from equation (A15) as:
\[
\Sigma_{xx}^g = \frac{\sum_{h=1}^{l} \hat{p}_{ig} (x_i - \mu_x^g) (x_i - \mu_x^g)'}{\sum_{i=1}^{l} \hat{p}_{ig}} \tag{A20}
\]
and from simplifications for the covariance matrix not at the latent competitive group level it follows that:
\[
\Sigma_{xx} = \frac{\sum_{g=1}^{G} \sum_{i=1}^{l} \hat{p}_{ig} (x_i - \mu_x^g) (x_i - \mu_x^g)'}{I}. \tag{A21}
\]
A similar argument can be made for the performance variables. This leads to:

\[
\Sigma_{yy}^{g} = \sum_{i=1}^{I} \hat{\mu}_{yi} \left( y_i - \mu_{yi} \right) \left( y_i - \mu_{yi} \right)'
\]

when at the latent competitive group level, and:

\[
\Sigma_{yy} = \sum_{g=1}^{G} \sum_{i=1}^{I} \hat{\mu}_{yi} \left( y_i - \mu_{yi} \right) \left( y_i - \mu_{yi} \right)'
\]

when not at the latent competitive group level. Finally, the covariance between strategic and performance variables can be obtained using:

\[
\Sigma_{xy}^{g} = \sum_{i=1}^{I} \hat{\mu}_{yi} \left( x_i - \mu_{xi} \right) \left( y_i - \mu_{yi} \right)'
\]

when at the latent competitive group level, otherwise:

\[
\Sigma_{xy} = \sum_{g=1}^{G} \sum_{i=1}^{I} \hat{\mu}_{yi} \left( x_i - \mu_{xi} \right) \left( y_i - \mu_{yi} \right)'
\]

We can then employ variations of expression (9) to obtain either aggregate or latent competitive group level regression coefficients to explore the effects of strategy on performance. Individual level parameter estimates are also obtainable in such mixture models as convex combinations of the estimated conditional distribution centroid parameters where the weights are the estimated posterior probabilities of latent competitive group membership. Note, mixtures of multivariate normal distributions are typically identified (see McLachlan and Peel (2000), for conditions), although spurious maximizers have been encountered in the heteroscedastic cases where separate covariance matrices are estimated by group necessitating multiple runs per value of \( G \) in such cases or imposing constraints on the covariance matrices.

**Appendix 2. Strategic and performance ratios utilized in the illustrative analysis**

**Performance**

Profits: \( \text{NPM} = \frac{\text{net income}}{\text{total current operating revenue}} \)

Market value: Tobin’s \( \text{Q} = \frac{\text{market value of equity} + \text{book value of liabilities}}{\text{book value of assets}} \).

Efficiency: \( \text{TE} = \frac{\text{loans} + \text{treasury bonds}}{\text{deposits} + \text{fixed assets} + \text{labor}} \).

**Strategy**

Asset liquidity ratios: \( \text{CR} = \frac{\text{current assets}}{\text{current liabilities}} \).

Capital adequacy ratio: \( \text{TIE} = \frac{\text{earnings before interest and taxes} + \text{depreciation}}{\text{interest}} \).

Product ratios – deposit: \( \text{ISD} = \frac{\text{total investment securities}}{\text{total worldwide deposits}} \).
Product ratios – loans: \( LIS = \frac{\text{gross loans}}{\text{total investment securities}} \).

Asset quality ratios: \( NCLL = \frac{\text{noncurrent loans}}{\text{loans}} \).

Corresponding author
Wayne S. DeSarbo can be contacted at: wsd6@psu.edu

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints