CrossMark

A dialogue of the deaf in the statistical theater? Adressing structural effects within a geometric data analysis framework

Xavier Bry[1] · Nicolas Robette[2] · Olivier Roueff[3]

Abstract Since their introduction in the late 1960s, the "moderate", and moreover "metrological" and "hypermetrological" uses of regression models quickly became the dominant quantitative approach in the Anglo-Saxon social sciences. This "sociology of the variables" has been the subject of many critical insights, with little impact on its dominance. By contrast, the French situation is quite different, mainly because of the strong association between Pierre Bourdieu's research program and the correspondence analysis methods. In this context, the relationship between geometric data analysis and regression models has turned into a "dialogue of the deaf". Complementarity is sometimes emphasized, correspondence analysis being associated with exploration and description of the data, and regressions being used to explain, reject or confirm assumptions. But regression models may also be used in order to analyze structural effects within a framework of geometrical data analysis, e.g. by visualizing graphically the results of a regression (Rouanet et al. in Math Sci Hum 160:13–46, 2002; Lebaron 2013). We propose a new multi-step approach, "Standardized Factor Analysis", which relies on geometric analysis and uses linear regression in a second stage in order to uncover structural effects in the original space. We illustrate it with data about tastes for cinema in France. We conclude by raising a more general set of questions about causality: social determinisms, even well established, are partial in the sense that they produce their effects only when associated with each other.

Keywords Correspondence analysis · Regression models · Analysis of structural effects · Combination of geometrical and inferential statistics · Standardized factor analysis (SFA)

✉ Olivier Roueff
o.roueff@free.fr

[1] Maître de conférence, I3M, CNRS, Université de Montpellier II, Montpellier, France

[2] Maître de conférence, Printemps, CNRS, Université de Versailles-Saint-Quentin-en-Yvelines, Versailles, France

[3] Chargé de recherche, CRESPPA-CSU, CNRS, Université Paris 8, Saint-Denis, France

It is usual to oppose geometrical data analysis (GDA) with regression models—when GDA is known, considering the wide predominance of regression in the international statistical theater. Some say that the first may be useful for exploring and describing data—but causation would be a matter for regression models only. Others assume that regression models are part of a "sociology of variables" highly questionable, and that only GDA could assess for the complexity of social reality. In this paper, we discuss such assumptions about both statistical frameworks, in order to design a new approach to structural effects, i.e. the effect of some characteristics of a population on the apparent correlation of two or more other characteristics. The article by Louis-André Vallet and Jean-Pierre Caille on the academic achievement of immigrants' children (1995) is a classic example. The authors start from the observation that immigrants' children are less likely to go to high school after 4 years of secondary school. But then they show that with equal family characteristics (occupation of the head of household, parents' degree, etc.), the difference in success disappears: the effect of immigrant origin is nothing but a structural effect, as immigrant parents are more likely to belong to dominated areas of the social space. This way of controlling structural effects—i.e. reasoning "all other properties being equal"—is common with regression models (even it is more often (over)used to eliminate them in order to find the best explanation). But it is far less common with GDA, while GDA is specifically aimed at treating a lot of variables simultaneously (and not only one by one as with regression). We thus discuss the various uses of regression models in social sciences and the issue of its complementarity with GDA. Then we propose a new statistical approach, with a specific tool named Standardized Factor Analysis (SFA), for controlling structural effects within a GDA framework. We illustrate its functioning with an example based on the study of tastes for cinema in France, and discuss shortly some of the epistemological questions raised.

1 Uses of regression

Since their introduction in the late 1960s, the regression models were widely disseminated in the Anglo-Saxon social sciences, and quickly became the dominant quantitative approach, in keeping with (among others) the work of Paul Lazarsfeld. Étienne Ollion has shown that the regressions were only present in about 10 % of articles in the *American Sociological Review* between 1967 and 1970, but that this proportion was close to 70 % in the late 1990s (Ollion 2011).

The "sociology of variables" has been the subject of many critical insights, from general to technical ones, from the ethnomethodologists to the statisticians. In the second part of the 1980s, some American sociologists having some authority in the academic field, as Andrew Abbott (2001), Charles Ragin (1987) and Stanley Lieberson (1985, see also Vallet 2004), attacked the assumptions that underlie the sociology of variables.

First, this "general linear reality", in the words of Abbott, while seeking to explain a feature by causal factors, hypothesizes that these causal factors act independently of each other (what Ragin calls "net effects thinking"). They have a specific effect and, therefore, do not conjugate but add up.[1] Then one could understand the risk of being affected by

---

[1] It is technically possible to add interaction effects between explanatory variables in a regression model but besides that it is a relatively uncommon practice, the multiplication of combinations quickly becomes unworkable. An interesting approach is developed by Judea Pearl in the framework of analytical sociology, based on structural causality (Pearl 2009).

poverty by separating the effect of the social origin from the one of the educational attainment: being without a degree would have the same consequences for the child of a blue collar and for the child of the CEO of a big company.

Similarly, these causal factors, i.e. social characteristics of individuals, have a unique meaning (they are univocal) and apply uniformly at all points of the social space, regardless of the context. For example, an indicator of cultural practices such as going out to the movie theater once a month has the same meaning in Paris or in rural areas, for the generations born after WWII as for those of the 70's economic crisis, etc.[2]

In addition, cases (individuals, countries, etc.) are—also—supposed to be independent of each other: what happens to a case has no influence on what happens to others. And yet, by focusing—for different countries—on the impact of the outbreak of war on fertility during WWII, Lieberson shows that Norway's entry into the war currently affected fertility behavior, but that fertility has also changed in Sweden which, nevertheless, remained neutral: it is what he calls a contamination effect (or diffusion effect). Here, the causal factor may have an effect even where it is not directly present. It is more generally the issue of social interactions that is excluded by the assumption of case independence.

Moreover, one often reasons "all things being equal". But the social sciences can never reproduce the experimental framework of the natural sciences (control of population, etc.).[3] There are still properties that are not observed, because we do not have the information or because they are not reducible to a simple indicator: the control procedure is still incomplete.

Finally, the regression models generally assume reversibility of causal processes: if a factor has a given result, its disappearance will result in the return to the initial situation. Thus one ignores that certain social phenomena are hardly reversible, due for example to ratchet effects.[4]

Ultimately, these various attacks have had little impact on the dominance of the approach by the variables in the Anglo-Saxon social sciences. Moreover, the dominant paradigm counter-attacked by offering ever more complex statistical models. These technical refinements are expected to raise some objections, but never really call into question the epistemological foundations of the approach.

What then are the practical uses of regression and their theoretical assumptions, which often remain unthought? Broadly speaking, we can distinguish three types of uses of regression models. The first, a "moderate" use, consists in controlling the structure of the population under study, i.e. to reason "all social properties being equal". The article by Louis-André Vallet and Jean-Pierre Caille on the academic achievement of immigrants' children (1995) is a classic example. The authors start from the observation that immigrants' children are less likely to go to high school after 4 years of secondary school. But then they show that with equal family characteristics (occupation of the head of household, parents' degree, etc.), the difference in success disappears: there is no specific effect of immigrant origin. It is in this first type of use that one can locate Durkheim's *Suicide*, even if his "concomitant variations" do not yet rely upon the actual regression techniques.

The second use, which could be described as "metrological" (Desrosières 2001), consists in using regression in order to rank the importance of different causal factors.

---

[2] There is already criticism of this assumption in Halbwachs' *The causes of suicide* (2002 [1930]), about the relationship between religion and urbanization.

[3] This argument is also very present in the epistemological reflections of Passeron (1991).

[4] We cite here some of the main criticisms, leaving aside—among others—those of Abbott about temporality issues (2001).

Attempts, for example, measure whether it is social origin or educational level (ascription or achievement) that has the greatest influence on tastes and lifestyles; or whether cultural norms or social environment explain the integration (or non-integration) of immigrants.[5]

The third use, that we call "hyper-metrological", is more directly related to policy assessment procedures. As Etienne Pénissat ([2011](#)) shows in the case of employment policies, it is a matter of measuring as precisely as possible the effectiveness, i.e. the "pure" effect, of return to work policies, by reproducing the approach of experimental sciences: statistical tools are then involved in redefining the action of the bureaucratic field.[6]

2 Geometric data analysis and regression: what complementarity?

In France, the situation is quite different from the Anglo-Saxon countries. Firstly because quantitative approaches in general are less common. Regarding more specifically the regression techniques, one sees some diffuse appropriations from the 80s, including studies on stratification and social mobility from large surveys, but it is still far from a mass phenomenon.

Moreover, this French situation is probably linked to the trajectory of Pierre Bourdieu, particularly the strong association between his research program and the correspondence analysis methods (Rouanet et al. [2000](#); Lebaron [2010](#)). In 1966, in a chapter of the book *Le partage des bénéfices* (*The share of profits*) entitled "La fin du malthusianisme" ("The end of Malthusianism"), Pierre Bourdieu and Alain Darbel ([1966](#)) propose a mathematical equation of fertility by income level, based on econometric models of consumption, and thus provide some explanation of the "Baby Boom" France was then experiencing. But the modeling of fertility by income is immediately followed by a violent attack against the sociology of variables, and the willingness to separate the respective influences of social determinants of fertility, when it is rather the "structure of the factors' constellation" that must be described, and how this structure is refracted in a given social group.

A few years later, in 1976, Bourdieu—temporarily—formalized its field theory, new statistical tools are available (in particular through the work of Benzécri), and he publishes with Monique de Saint- Martin "L'anatomie du goût" ("The anatomy of taste"), his first application of correspondence analysis to his theoretical model (Bourdieu and de Saint-Martin [1976](#)). This application will be followed by many others, up to *Les structures sociales de l'économie* (*The social structure of the economy*) in 2000, and Bourdieu repeatedly highlighted the affinity between correspondence analysis—or geometric data analysis—and his relational approach to the social world (Lebaron [2010](#)).

In this context, the relationship between geometric data analysis and regression models has been under discussion, which may turn into a "dialogue of the deaf" (Desrosières [2003](#)). Rather than opposing approaches, complementarity is sometimes emphasized (Des Nétumières [1997](#)): the division of statistical labor associates correspondence analysis with exploration and description of the data when the regressions are used to explain, reject or confirm assumptions. But this dichotomy is somewhat simplistic: well used,

---

[5] One may ask if we would not understand better some scientific debates by analyzing these kinds of competitions on the "size" and "robustness" of effects as a form of male competition.

[6] This is the same logic that explains the recent success, within state organizations or NGOs, of the "randomized evaluation" promoted by Esther Duflo ([2010](#)) to test the effectiveness of development assistance programs.

correspondence analysis "is not the purely descriptive method that those who oppose it to regression analysis want to see" (Bourdieu 2000). More generally, "statistics explain nothing, but give possible elements of explanation" (Lebart et al. 2000). The idea of complementarity is not acceptable in these terms and must be reformulated: Rouanet and Lebaron (2006) propose to "seek a synthesis between data analysis and regression by integrating regression into geometric data analysis, and then giving meaning to the sociology of variables, seen in the context of the social space." In other words, it is a matter of keeping the social group or individual as the main actor of the "statistical theater" (Desrosières 2003), and therefore using regression (in the first of its uses identified above) without adopting the "language of variables" (Desrosières 2001).

In the remainder of this article, we describe a method of geometric data analysis, building on the work of Rouanet et al. (2002), which proposes "not to eliminate the structural effects"—running the risk famously highlighted by Simiand[7]—"but to analyze them."

3 The space of tastes for cinema

We focus here on the space of tastes for cinema, constructed from the survey about French cultural practices conducted in 2008 by the Research Department of the French Ministry of Culture (Donnat 2009)[8]. The variables used to build this space describe actual attendance, positive preferences in the abstract (declarative taste) and negative preferences in the abstract (declarative distaste):

- kinds of favorite movies, from a list of movie genres (comic (comedies), action, historical and biopics, police (crime) and thrillers, adventure, comedy-dramas, animation, horror, *auteur*, documentaries, westerns, erotic, sci-fi, love (romance), musicals);
- kinds of movies never watched, from the same list;
- movies already seen from a list (*Les Bronzés*, *Pirates of the Caribeans*, *Camping*, *Star Wars*, *Brice de Nice*, *Shrek*, *Matrix*, *Titanic*, *The Lord of the Rings*, *Astérix et Obélix*, *Les Visiteurs*, *The Lives of Others (Das Leben der Anderen)*, *Brokeback Mountain*);
- among these films, those that are particularly liked;
- among these films, those that one does not like or does not want to see;
- frequency of cinema attendance.

A Multiple Correspondence Analysis (MCA) is performed on these variables, focusing on the active population (N = 3008). The first axis of the MCA may be interpreted as the dimension of the practice intensity. On the left side of the graph (Fig. 1), one finds all the movies already seen, twelve movies particularly liked among the thirteen proposed (only *Titanic* is on the right) and the highest frequencies of cinema attendance (from five to eleven times a year, at least once a month).

---

[7] «This method [the elimination of structural effects] leads to the study and the comparison between the behaviors of a reindeer in the Sahara and of a camel at the North Pole» (cited in Maurice Halbwachs, «La statistique en sociologie», *in* Centre internationale de synthèse, *La statistique. Ses applications. Les problèmes qu'elles soulèvent*, Paris, PUF, pp. 113–160).

[8] For a study of tastes for cinema in France, see Duval (2011).

Fig. 1 The space of tastes for cinema (axes 1–2; «classical» MCA). The asymmetry of axes' scales has no statistical signification; it is only aimed at facilitating the reading

Reading: "Kinds of favorite movies" variables are prefixed by "fav", "kinds of movies never watched" by "never", "movies already seen" by "seen", "movies liked" by "like", "movies disliked" by "disl" and frequency of cinema attendance by "freq". "1" suffix means the presence of a category while "0" means its absence.

But it is above all the fact of having seen or not the proposed movies which is the principle of the first dimension of the space of cinematographic tastes: the thirteen variables about movies already seen contribute 55 % of the construction of the first axis, the variables about the movies particularly liked 26 %, and the frequency variable 2 %.

We then introduce several social characteristics as supplementary variables—here limited to sex, age and level of education, as we only aim at illustrating the approach. This dimension of intensity of practice is closely associated with age, sex and level of education of respondents (Fig. 2): age categories are almost perfectly aligned with the axis and "explain" 16 % of the variance among individuals;[9] men are importantly opposed to women (8 % of the variance) and college graduates to non-degree (3 % of the variance). The women, the individuals who are more than 50 years, and those without a degree are on the side of a low intensity of cinema practice.

---

[9] We use "eta²" indicator here, which is similar to R^2 in linear regression analysis (Le Roux and Rouanet 2004).
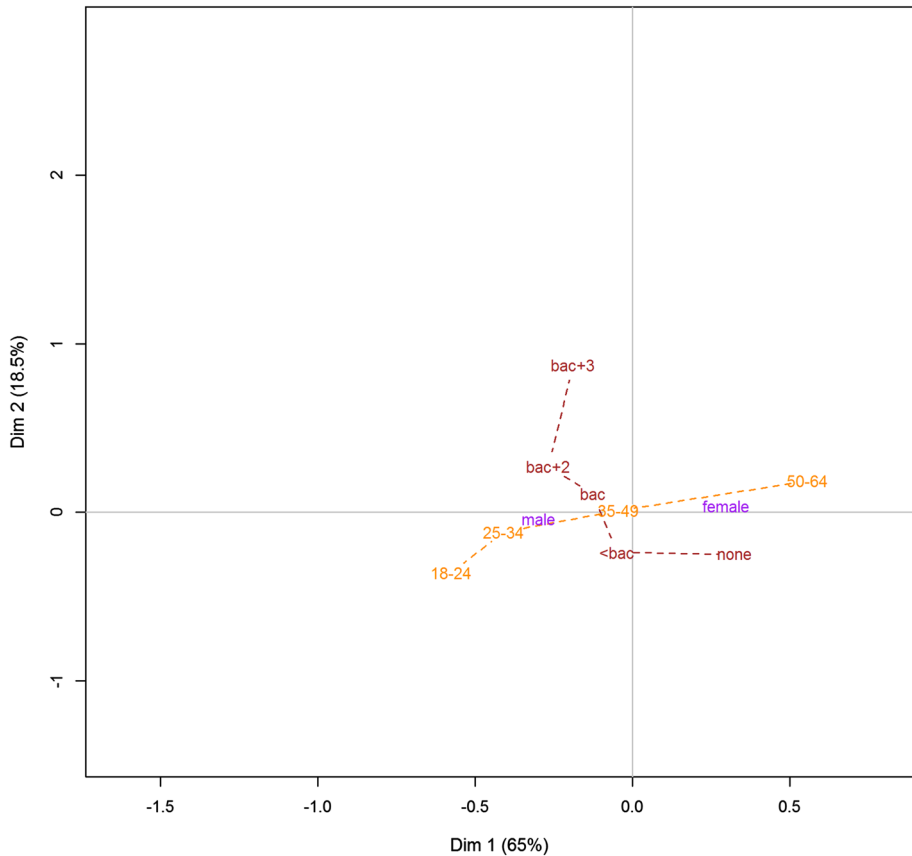
Fig. 2 Main structuring factors of the space of tastes for cinema (axes 1–2; «classical» MCA)

Reading: Educational categories (in red) are built according to the French "baccalauréat", which is more or less equivalent to a high school degree. "<bac" means that one left school before baccalauréat, "bac + 2" means two years after baccalauréat, "bac + 3" three years or more after baccalauréat.

The second axis seems to show a contrast between the taste for French mainstream comedies (*Brice de Nice*, *Camping*, *Astérix et Obélix*, *Les Bronzés*, *Les Visiteurs*) and the taste for foreign *auteur* movies (*The Lives of Others*, *Brokeback Mountain*). This dimension ranks tastes in terms of legitimacy, with US action or science fiction movies in an intermediate position. Moreover, we note that the variables of "tastes" (especially liked movies), expressing more directly the distinctive principle of cultural legitimacy, contribute more to the construction of the second axis (40 %) than the variables of "practice" (films already seen, 21 %). In the same logic, the variables of "distaste" make a significant appearance (10 % against 5 % for the first axis). Unsurprisingly, there is a clear homology between the distribution of cultural capital (in its academic form) and the tastes legitimacy: the level of education explains 14 % of the variance of the cloud of individuals on the second axis. In contrast, the weight of age and sex as structuring factors is negligible here.

We focus now on the taste for the different movies available in the space of cinema tastes, particularly in the dimension of the practice intensity (Fig. 3).
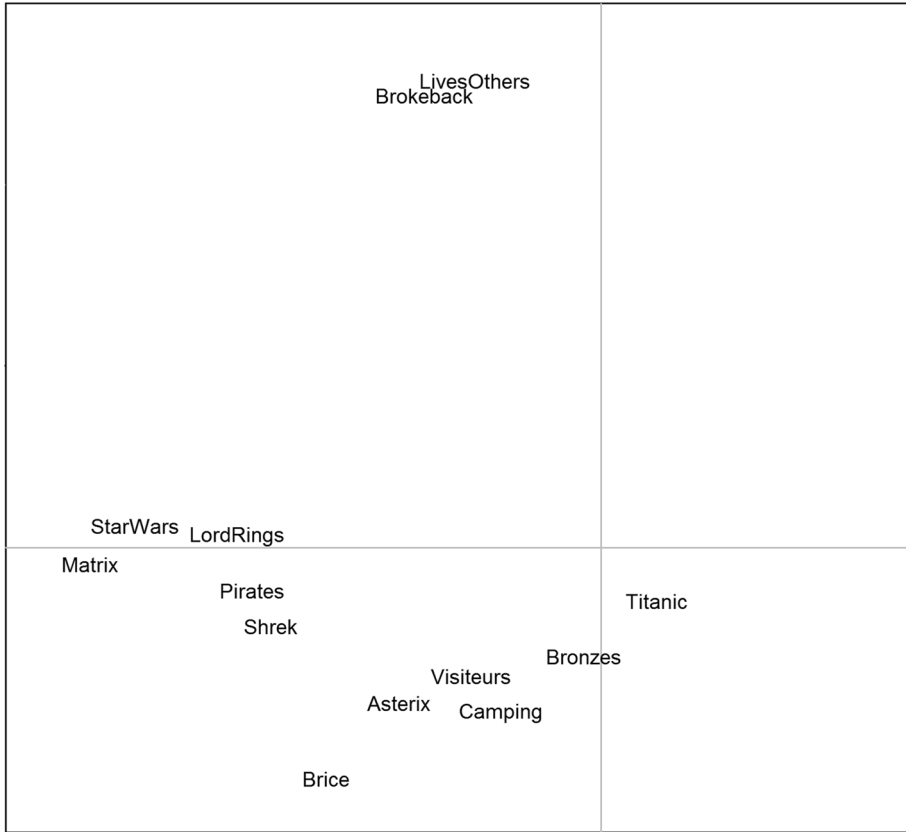
Fig. 3 The movies «especially liked» in the space of tastes for cinema (axes 1–2; «classical» MCA)

US action and science fiction movies form a relatively homogeneous group at the extreme left of the graph, while the French comedies (top) and *auteur* movies (bottom) are further right. This suggests that the taste for action and science fiction movies concerns the most regular cinema lovers, while other films are more appreciated by individuals with less intense practices. However, we also know (French Ministry of Culture, DEPS 2008) that it is women over 30 years old who have the most highbrow tastes in cinema (as in other cultural fields). But it is also this group that attends the least cinema, as was shown above. It is therefore possible that the observed association between the taste for art films and a moderate intensity of practice is an effect of the age and sex structure of the population of highbrow cinema lovers.

4 Studying structural effects: the Standardized Factor Analysis

In order to make sure, one should be able to control for the structure of sex and age. In the framework of the geometric data analysis, this would mean forcing the axes of the MCA to be independent from (i.e. orthogonal to) individuals' properties of sex and age, that is to say, performing a MCA "all things (of age and sex) being equal." To do this, we propose a

multi-step approach, which we call "Standardized Factor Analysis" (SFA), by analogy with the standardization methods in demography (Léridon and Toulemon 1997; Deauvieau 2011).[10]

- First, we start from the "classic" MCA previously performed to build a new data table (called C) with the coordinates of the individuals on each axis of this MCA. There is thus a table Individuals x Coordinates, with i rows and p columns (i is the number of individuals and p is the number of principal components of the MCA).[11]
- Then, p linear regressions are performed, i.e. one by column of Table C. The coordinates of the individuals of the MCA (i.e. the variables of C) are used one after the other as the dependent variable, with sex and age as systematically independent variables.
- The residuals are retained for each of the regressions: one has then a new table (called R) with i rows and p columns. The quantitative variables of R correspond to the coordinates of individuals in the initial MCA "net" of the sex and age structure.
- A Principal Component Analysis (PCA) is then performed on table R, which provides a new cloud of individuals, i.e. a new space.
- The variables that were used to build the original MCA are then projected on this space as supplementary variables: what appears is the space of tastes for cinema "net" of the sex and age structure of the population.[12]

In our application, we start by the "classic" MCA of the tastes for cinema. We choose to retain the first five dimensions as they comprehend most part of the interpretable information in our dataset. Then we perform five linear regressions: each of them uses the individual coordinates as dependent variable and sex and age as independent variables. For each of these regressions, we calculate the residuals: we get five quantitative variables. Afterwards, these five variables are used as input in a Principal Component Analysis (PCA). Finally, we project the variables of tastes for cinema as supplementary variables on the plot produced with PCA results.

When applying a standardized factor analysis (by sex and age) to our data, the result is clear (Fig. 4). There are still three groups of relatively similar films, but the French comedies have slightly moved to the right of the first axis. Moreover, the two *auteur* movies are now at the extreme left of the graph, at the same level of intensity as US action and science fiction movies. The comparison of the two spaces, one with structural effects and one without, thus shows that highbrow amateurs' practices are as intense as those of action-movie lovers when sex and age are equivalent. Or, to put it differently, the observed

---

[10] In demography, one of the most widely used indicators to study mortality is the crude death rate. When comparing Sweden and Mexico, for example, we see that the crude death rate in Sweden is higher than that of Mexico, which may seem counterintuitive in that health conditions are better and life expectancy is higher in Sweden. This is due to the fact that the crude death rate is influenced by the age structure of the population under study: Sweden has an older population than Mexico, and the elderly have higher death rates than other categories of age, which leads to a relatively high crude death rate. Demographers thus use statistical techniques of "standardization" to neutralize the effect of the age structure, i.e. to calculate the mortality indicators that are net of this structural effect. More generally, SFA is applied to MCA in this article but it could apply to any other factor analysis technique, such as principal component analysis (PCA).

[11] At this stage, one can choose to retain only the first principal components, those that contain the most information.

[12] To the extent that the PCA is not performed directly on the variables of the original MCA, one can't obtain the contributions of these variables. However, all the other tools for the interpretation of supplementary variables may be used ($\cos^2$, v test, eta^2).
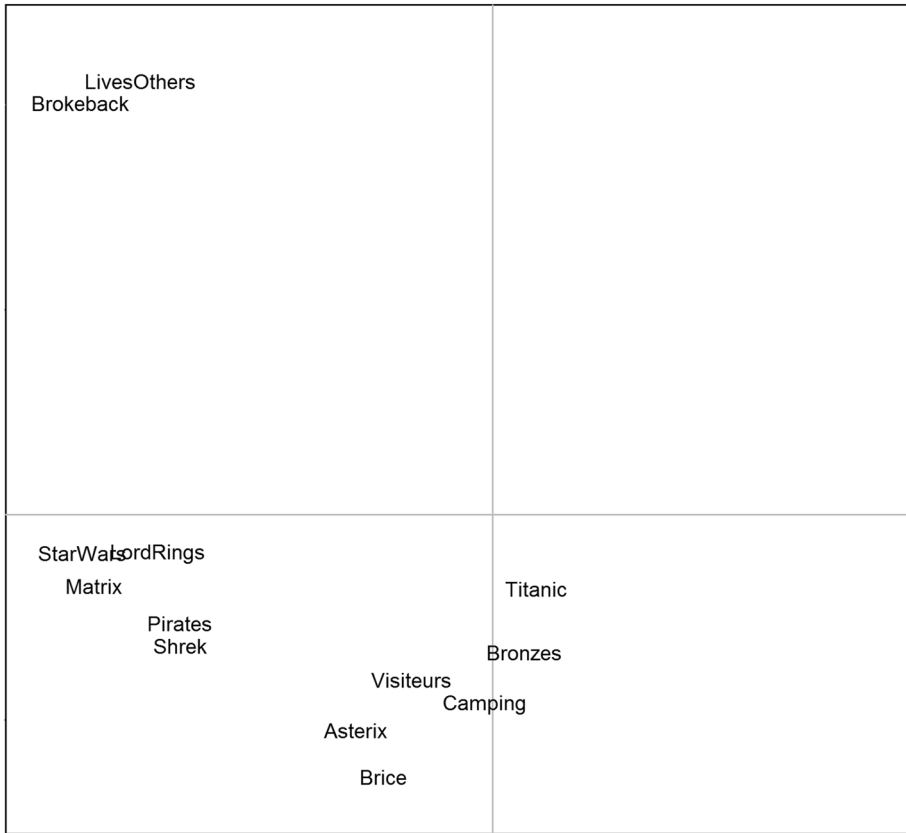
Fig. 4 The movies «especially liked» in the space of tastes for cinema (axes 1–2; Standardized Factor Analysis)

association between the taste for art films and a relatively limited level of practice is the result of the sex and age structure of the population.

5 Conclusion

Analyzing structural effects within a framework of geometric data analysis may be done by graphically visualizing the results of a regression (Rouanet et al. 2002; Lebaron 2013). It means building the multidimensional space formed by the independent variables of the model and then projecting the dependent variables on this space: one can thus compare the overall and partial effects of independent variables. We propose a complementary line of research, "SFA" (for standardized factor analysis), which relies on geometric data analysis—MCA here—and uses linear regression in a second stage in order to uncover structural effects in the original space. The main common point of both approaches is that they are intended to compare the results with and without structural effects rather than eliminate them, without any random model or a priori probabilistic framework (Lebaron 2013).

SFA also raises a more general set of questions. The many applications we have attempted (and we can't recount in detail within this article) all point to the same phenomenon: when structural effects are neutralized, the initial multidimensional space remains almost the same in its general pattern. Some categories may move as *auteur* movies in the example shown here, but the overall structure of the cloud of categories and the interpretation that can be given of its different factors withstand the neutralization, including when several highly "explanatory" variables are controlled (such as social class, sex, age, and educational level). This probably illustrates a more general epistemological assumption: social determinisms, even well established, are probabilistic and not mechanical, and they are partial in the sense that they produce their effects only when associated with each other. Empirically observed correlations between social properties, e.g. relatively consistent taste patterns, can be explained by various social determinants, but these determinants never fully "exhaust" social reality. Massive structural effects, such as reversals in the case of academic achievement among immigrants' children studied by Vallet and Caille (1995), are actually quite rare. But this does not diminish the need to analyze these structural effects—for themselves, or during the process of exploring a dataset.

References

Abbott, A.: Time Matters. On Theory and Method. The University of Chicago Press, Chicago (2001)

Bourdieu, P.: Les Structures Sociales de L'économie. Seuil, Paris (2000)

Bourdieu, P., Darbel, A.: La fin d'un malthusianisme. In: Darras (eds) Le Partage des Bénéfices, pp.135–154. Minuit, Paris (1966)

Bourdieu, P., de Saint-Martin, M., .: Anatomie du goût. Actes Recherche Sci. Soc. **2**(5), 2–81 (1976)

Deauvieau, J.: Est-il possible et souhaitable de traduire sous forme de probabilités un coefficient logit ? Réponse aux remarques formulées par Marion Selz à propos de mon article paru dans le BMS en 2010. Bull. Méthodol. Sociol. **112**(1), 32–42 (2011)

Des Nétumières, F.: Méthodes de régression et analyse factorielle. Hist. Mesure **12**(3–4), 272–297 (1997)

Desrosières, A.: Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative. Genèses **43**, 112–127 (2001)

Desrosières, A: Bourdieu et les statisticiens. Une rencontre improbable et ses deux héritages. In: P. Encrevé, P., Lagrave, R.M. (eds) Travailler avec Bourdieu. Flammarion, Paris (2003)

Donnat, O.: Les pratiques culturelles des Français à l'ère numérique. Enquête 2008. La Découverte, Paris (2009)

Duflot, E.: Le Développement humain. Lutter contre la pauvreté (I). Le Seuil/République des idées, Paris (2010)

Duval, J.: L'offre et les goûts cinématographiques en France. Sociologie **2**(1), 1–18 (2011)

Halbwachs, M.: Les causes du suicide. PUF, Paris (2002 [1930])

Le Roux, B., Rouanet, H.: Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis. Springer, New York (2004)

Lebaron, F.: L'analyse géométrique des données dans un programme de recherche sociologique : le cas de la sociologie de Bourdieu. Modulad **42**, 102–109 (2010)

Lebaron, F.: La régression peut-elle faire progresser? Approches et usages critiques des modèles de régression. Working Paper for: Sociologie quantitative et sociologie de la quantification. PRINTEMPS, University of Versailles-Saint-Quentin-en-Yvelines (22 March 2013)

Lebart, L., Morineau, A., Piron, M.: Statistique Exploratoire Multidimensionnelle. Dunod, Paris (2000)

Léridon, H., Toulemon, L.: Démographie.Approche Statistique et Dynamique des Populations. Economica, Paris (1997)

Lieberson, S.: Making It Count. The Improvement of Social Research and Theory. University of California Press, Berkeley (1985)

Ministère de la Culture et de la Communication: Les résultats complets de l'enquête 2008. http://www.pratiquesculturelles.culture.gouv.fr/08resultat.php (2008). Accessed 20 Sept 2014

Ollion, E.: De la sociologie en Amérique. Éléments pour une sociologie de la sociologie étasunienne contemporaine. Sociologie **2**(3), 277–294 (2011)

Passeron, J.-C.: Le raisonnement sociologique, L'espace non-poppérien du raisonnement naturel. Nathan, Paris (1991)

Pearl, J.: Causal inferences in statistics: an overview. Stat. Surv. **3**, 96–146 (2009)

Pénissat, E.: Quantifier l'effet "pur" de l'action publique: entre luttes scientifiques et redéfinition des politiques d'emploi en France. Sociol. Soc. **43**(2), 223–247 (2011)

Ragin, C.: The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies. University of California Press, Berkeley (1987)

Rouanet, H., Ackermann, W., Le Roux, B.: The geometric analysis of questionnaires: the lesson of Bourdieu's La distinction. Bull. Méthodol. Sociol. **65**(1), 5–18 (2000)

Rouanet H., Lebaron F.: La preuve statistique: regard critique sur la régression. Working Paper presented at: Qu'est-ce que Faire preuve?, University of Amiens (2006)

Rouanet, H., Lebaron, F., Hay, V.L., Ackermann, W., Le Roux, B.: Régression et analyse géométrique des données: réflexions et suggestions. Math. Sci. Hum. **160**, 13–46 (2002)

Vallet, L.-A.: A propos d'un ouvrage peu connu dans la sociologie française: making it count. The improvement of social research and theory, de Stanley Lieberson. Revue Eur. Sci. Soc, **XLII**(129), 341–348 (2004)

Vallet, L.-A., Caille, J.-P.: Les carrières scolaires au collège des élèves étrangers ou issus de l'immigration. Éduc. Form. **40**, 5–14 (1995)