WORKING PAPER 14-17

Modelling unobserved heterogeneity in distribution

Finite mixtures of the Johnson family of distributions

August 2017

Peter Willemé, pw@plan.be

Avenue des Arts 47-49 – Kunstlaan 47-49
1000 Brussels

E-mail: contact@plan.be
http://www.plan.be

Federal Planning Bureau

The Federal Planning Bureau (FPB) is a public agency that carries out, in support of political decision-making, forecasts and studies on economic, social-economic and environmental policy issues and examines their integration into a context of sustainable development. It shares its expertise with the government, parliament, social partners, national and international institutions.

The FPB adopts an approach characterised by independence, transparency and the pursuit of the general interest. It uses high-quality data, scientific methods and empirical validation of analyses. The FPB publishes the results of its studies and, in this way, contributes to the democratic debate.

The Federal Planning Bureau is EMAS-certified and was awarded the Ecodynamic enterprise label (three stars) for its environmental policy.

url: http://www.plan.be
e-mail: contact@plan.be

Publications

Recurrent publications:
Outlooks

Planning Papers (latest publication):
The Planning Papers aim to diffuse the FPB's analysis and research activities.

115 Les charges administratives en Belgique pour l'année 2014 /
De administratieve lasten in België voor het jaar 2014
Chantal Kegels, Dirk Verwerft - February 2016

Working Papers (latest publication):

13-17 Accuracy assessment of the FPB medium-term outlooks - An update
Igor Lebrun - September 2017

Responsible publisher: Philippe Donnay
Legal Deposit: D/2017/7433/35

Federal Planning Bureau
Avenue des Arts - Kunstlaan 47-49, 1000 Brussels
phone: +32-2-5077311
fax: +32-2-5077373
e-mail: contact@plan.be
http://www.plan.be

Modelling unobserved heterogeneity in distribution

Finite mixtures of the Johnson family of distributions

August 2017

Peter Willemé, pw@plan.be

Abstract - This paper proposes a new model to account for unobserved heterogeneity in empirical modelling. The model extends the well-known Finite Mixture (or Latent Class) Model by using the Johnson family of distributions for the component densities. Due to the great variety of distributional shapes that can be assumed by the Johnson family, the method does not impose the usual a priori assumptions regarding the type of densities that are mixed. As a consequence, the component densities are allowed to vary over a wide range of shapes as measured by the skewness and kurtosis parameters. By lifting the assumptions regarding the shape of the component densities, the method generalizes the Finite Mixture Model to modelling unobserved heterogeneity in distribution. The method has been implemented in R. This paper outlines the algorithm to estimate the parameters of a mixture of Johnson distributions and provides a proof of principle that the method is feasible and a potential improvement over current latent class modelling practice. The method has not yet been tested for a mixture of regression models, which will be an obvious next step in turning it into a practical research tool.

Jel Classification - C10, C13, C46, C51, C52
Keywords - Unobserved heterogeneity, Finite Mixture Models, Johnson family of distributions.

Acknowledgements - The author wishes to thank Benoît Laine for his comments on an earlier version of this paper.

Table of contents

List of tables

List of graphs

Executive summary

With the increasing availability of micro data in many fields of applied research, finite mixture models (FMM) are becoming increasingly popular as a tool to model unobserved heterogeneity between subjects. FMMs (also known as Latent Class Models, LCM) are based on the assumption that the observations in a sample derive from an (unknown) number of heterogeneous subgroups or classes, and allow for the estimation of the parameters by subgroup. They have been used in economics to analyze health care utilization and expenditures, labour supply, productivity analysis, and market segmentation, among other topics. The models are also used extensively in applied research areas such as biology, psychology, biostatistics etc. The unobserved heterogeneity modeled with FFMs usually pertains to the mean of the distribution, although the variance has also been modelled (sometimes implicitly, as in the case of the gamma distribution). The current practice in applied economic research amounts to choosing a distributional form (normal, lognormal, gamma, Poisson, etc.) for the components, usually based on a priori considerations regarding the support and the shape of the population distribution.

A drawback of this approach is that it places a priori restrictions on the nature of the unobserved heterogeneity in at least two ways. First, the choice of the distribution is in general somewhat arbitrary and as a rule not tested against a more general (unrestricted) alternative. Second, while the ‘true’ number of latent classes is in principle unknown, it is also routinely assumed that the mixed distributions are of one kind. That is, the mixed components only differ from each other in terms of the parameters of the chosen distribution but not in terms of the probability density functions themselves.

This paper addresses these problems by lifting some of these implicit assumptions. This is achieved by postulating a flexible form for the component distributions. Several such flexible forms have been proposed and studied long ago, such as the Pearson and Johnson families, among others. Both families share the property that they can assume a wide variety of shapes depending on the value of their four parameters. In fact, most commonly used distributions are special cases of both families. The paper outlines an algorithm that can be used to estimate the parameters of a mixture of Johnson distributions and provides a proof of principle that the method is feasible and a potential improvement over current latent class modelling practice.

The method has been tested using data generated from different distributions, chosen to cover a wide range of combinations of skewness and kurtosis. The first results are encouraging. The method generally converges in about the same number of iterations as standard models that mix normal or gamma distributions. More importantly, when the data are generated from mixed distributions that differ substantially from the standard assumptions (identical component densities and ‘regular’ skewness and kurtosis values), the mixture of Johnson distributions generally fits the data better than the standard models.

The method has not yet been tested for a mixture of regression models, which will be an obvious next step in turning it into a practical research tool.

Synthèse

Grâce à la disponibilité croissante de microdonnées dans de nombreux domaines de la recherche appliquée, les modèles de mélange fini (finite mixture models ou FMM) deviennent un outil de plus en plus populaire pour modéliser l'hétérogénéité non observée entre sujets. Les FMM, également appelés modèles de classe latente (latent class models ou LCM) partent de l'hypothèse que les observations d'un échantillon proviennent d'un nombre (inconnu) de sous-groupes ou classes hétérogènes et permettent d'estimer les paramètres par sous-groupe. Ils ont été utilisés dans le domaine économique pour analyser notamment l'utilisation et les dépenses de soins de santé, l'offre de travail, la productivité et la segmentation de marché. Les modèles sont également abondamment utilisés dans d'autres domaines de la recherche appliquée comme la biologie, la psychologie, la biostatistique, etc. L'hétérogénéité non observée modélisée à l'aide des FMM porte habituellement sur la moyenne de la distribution, même si la variance a également été modélisée (parfois de manière implicite, comme dans le cas de la distribution gamma). La pratique actuelle en recherche économique appliquée revient à choisir une forme de distribution (normale, log-normale, gamma, Poisson, etc.) pour les composants, généralement sur la base de considérations a priori relatives à l'étendue et à la forme de la distribution de population.

Un inconvénient de cette approche est qu'elle impose de deux manières au moins des restrictions a priori quant à la nature de l'hétérogénéité non observée. Tout d'abord, le choix de la distribution est généralement assez arbitraire ; elle n'est habituellement pas confrontée à une alternative plus générale (moins restrictive). Deuxièmement, alors que le nombre « réel » de classes latentes est en principe inconnu, on suppose systématiquement que les composants suivent la même distribution. En d'autres termes, on suppose que les composants mixtes ne diffèrent entre eux qu'en ce qui concerne les paramètres de la distribution choisie, mais pas en ce qui concerne la distribution elles-même.

Cette étude aborde ces problèmes en assouplissant certaines de ces hypothèses implicites. Elle se base sur une forme flexible pour les distributions de composants. Plusieurs formes flexibles ont été proposées et étudiées par le passé, dont les familles Pearson et Johnson. Ces familles ont ceci en commun qu'elles peuvent adopter des formes très diverses en fonction de la valeur de leurs quatre paramètres. En réalité, la plupart des distributions utilisées couramment sont des cas spéciaux de ces deux familles. L'étude décrit un algorithme pouvant être utilisé pour estimer les paramètres d'un mélange de distributions Johnson et donne une preuve de principe que la méthode fonctionne et constitue une possible amélioration par rapport à la pratique courante pour les modèles de classe latente.

La méthode a été testée sur des données générées à partir de différentes distributions choisies pour couvrir un large éventail de combinaisons d'asymétrie et d'aplatissement. Les premiers résultats sont encourageants. La méthode converge pratiquement aussi vite que les méthodes standard qui mélangent des distributions normales ou gamma. Plus important encore, lorsque les données sont générées à partir de distributions mixtes qui diffèrent sensiblement des hypothèses standard (distributions de composants identiques et valeurs 'régulières' d'asymétrie et d'aplatissement), le mélange de distributions Johnson donne généralement de meilleurs résultats (qualité de l'ajustement) que les modèles standard.

La méthode n'a pas encore été testée pour un mélange de modèles de régression. C'est naturellement l'étape qu'il faudra franchir pour en faire un instrument de recherche pratique.

Synthese

Met de toenemende beschikbaarheid van microdata voor empirisch onderzoek worden 'finite mixture' modellen (FMM) een steeds populairder werktuig om niet-geobserveerde heterogeniteit tussen subjecten te modelleren. FMMs (die ook 'latent class' modellen worden genoemd, afgekort LCM) gaan uit van de veronderstelling dat de observaties in een steekproef afkomstig zijn uit een (onbekend) aantal heterogene subgroepen of klassen, en laten toe de parameters van deze subgroepen te schatten. Ze zijn in diverse domeinen van het economisch onderzoek gebruikt, onder meer voor de analyse van het gebruik en de uitgaven van medische zorg, het arbeidsaanbod, productiviteit, en marktsegmentatie. De modellen worden ook frequent toegepast in andere onderzoeksgebieden zoals biologie, psychologie, biostatistiek en zo verder. De niet-geobserveerde heterogeniteit die wordt gemodelleerd met FMMs heeft gewoonlijk betrekking op de verwachtingswaarde van de verdeling, hoewel soms ook de variantie wordt gemodelleerd (soms impliciet, zoals in het geval van de gamma-verdeling). De huidige praktijk in toegepast economisch onderzoek komt neer op het kiezen van een kansverdeling (normaal, lognormaal, gamma, Poisson, enz.) voor de componenten gebaseerd op a priori overwegingen betreffende het domein en de vorm van de populatieverdeling.

Een nadeel van deze benadering is dat ze a priori beperkingen oplegt over de aard van de niet-geobserveerde heterogeniteit, op tenminste twee manieren. Ten eerste is de keuze van de verdeling over het algemeen nogal arbitrair en wordt ze gewoonlijk niet getest tegen meer algemene (minder restrictieve) alternatieven. Ten tweede wordt, los van de vraag hoeveel latente klassen er in werkelijk bestaan, stelselmatig verondersteld dat de componenten dezelfde verdeling volgen. Met andere woorden, men veronderstelt dat de componenten enkel van elkaar verschillen wat de parameters van de verdeling betreft, maar niet in termen van de verdeling zelf.

Deze paper poogt deze problemen aan te pakken door sommige van de gebruikelijke veronderstellingen te versoepelen. Dit wordt bereikt uit te gaan van een flexibele vorm voor de component-verdelingen. Diverse zulke flexibele vormen zijn in het verleden bestudeerd, waaronder de zogenaamde Pearson en Johnson families van verdelingen. Deze families delen de eigenschap dat ze, afhankelijk van de waarden van hun parameters, heel diverse vormen kunnen aannemen. Bovendien zijn de verdelingen die courant worden gebruikt speciale gevallen van beide families. De paper beschrijft een algoritme waarmee de parameters van de Johnson familie van verdelingen kunnen worden geschat en levert een 'proof of principle' dat de methode werkt en een potentiële verbetering is ten opzichte van de huidige gangbare praktijk voor latente klassen-modellen.

De methode werd getest op gegevens gegenereerd met verschillende verdelingen die een groot bereik van scheefheid en kurtosis omvatten. De eerste resultaten zijn bemoedigend. De methode convergeert ongeveer even snel als de standaardmethoden met een mix van normale of gamma-verdelingen. Nog belangrijker, wanneer de data werden gegenereerd met verdelingen die ver afwijken van de standaardveronderstellingen (identieke componentverdelingen en 'reguliere' scheefheid en kurtosis), geeft de nieuwe methode in het algemeen betere resultaten ('goodness-of-fit') dan de standaardmodellen.

De methode werd nog niet getest in de context van een mix van regressiemodellen. Dit is de voor de hand liggende volgende stap om ze bruikbaar te maken als een praktisch onderzoeksinstrument.

1. Introduction

With the increasing availability of micro data in many fields of applied research, finite mixture models (FMMs) are becoming increasingly popular as a tool to model unobserved heterogeneity between subjects (McLachlan and Peel 2000). FMMs (also known as Latent Class Models, LCMs) have been used in economics to analyze health care utilization and expenditures (see e.g. Deb and Holmes 2000, Jiménez-Martín, Lebeaga, and Martínez-Granado 2002) , labour supply (Pacifico 2012), productivity analysis (De Vries and Koetter 2011), ordered choice (Greene et al. 2008), market segmentation (Wedel and Kamakura 2000) and many other topics. The models are also used extensively in applied research areas such as biology, psychology, biostatistics etc. The unobserved heterogeneity modeled with FFMs usually pertains to the mean of the distribution, although the variance has also been modelled (sometimes implicitly, as in the case of the gamma distribution). The current practice in applied economic research amounts to choosing a distributional form (normal, lognormal, gamma, Poisson, etc.) usually based on a priori considerations regarding the support and the shape of the population distribution.

A drawback of this approach is that it places a priori restrictions on the nature of the unobserved heterogeneity in at least two ways. First, the choice of the distribution is in general somewhat arbitrary and as a rule not tested against a more general (unrestricted) alternative. Second, while the 'true' number of latent classes is in principle unknown, it is also routinely assumed that the mixed distributions are of one kind. That is, the mixed components only differ from each other in terms of the parameters of the chosen distribution but not in terms of the probability density functions themselves.

This paper addresses these problems by lifting some of these implicit assumptions. This is achieved by postulating a flexible form for the component distributions. Several such flexible forms have been proposed and studied long ago, such as the Pearson and Johnson families, among others (see N. L. Johnson 1949, Johnson, Kotz, and Balakrishnan 1994, Kendall and Stuart 1969). Both families share the property that they can assume a wide variety of shapes depending on the value of their four parameters. In fact, most commonly used continuous distributions are special cases of both families. The Pearson family, for instance, contains the normal, lognormal, t, gamma, beta and F distributions as special cases. This paper outlines an algorithm that can be used to estimate the parameters of a mixture of Johnson distributions and provides a proof of principle that the method is feasible and a potential improvement over current latent class modelling practice.

The remainder of the paper gives a brief overview of the characteristics of the Johnson family in section 2, followed by an outline of the estimation algorithm (section 3). Section 4 provides some Monte Carlo experiments demonstrating the feasibility of the procedure and comparing the results with two commonly used alternatives. Section 5 concludes and discusses the next research steps.

2. The Johnson family of distributions

The Johnson family of distributions (N. L. Johnson 1949) is obtained by a set of transformations of a standard normal variable. Each of the four types corresponds with one particular element of the set, the general form of which is:

$$Z = \gamma + \delta f\left(\frac{X - \xi}{\lambda}\right) \quad (1)$$

$$Z = N[0, 1]$$

The transformations, which are shown here as transformations to the standard normal variable *Z*, are defined by the particular specification of the function *f*(.), as follows:

$$Type\ SN: f(.) = (.) \quad (2)$$

$$Type\ SL: f(.) = \ln(.) \quad (3)$$

$$Type\ SB: f(.) = \ln\left(\frac{X - \xi}{\xi + \lambda - X}\right) \quad (4)$$

$$Type\ SU: f(.) = \sinh^{-1}\left(\frac{X - \xi}{\lambda}\right) \quad (5)$$

Type SN is simply the normal distribution (transformed only by a location and scale parameter), while the SL transformation produces the lognormal. Types SB ('bounded') and SU ('unbounded') refer to the support of the transformed variable, which is $[\xi, \xi+\lambda]$ for the former and $(-\infty, +\infty)$ for the latter. The distributions can be characterized (apart from the location and scale parameters) by the skewness (β_1) and kurtosis (β_2) values they can accommodate. In fact, in the (β_1, β_2) plane, the SB and SU distributions occupy a region of values separated by the SL curve (which is a limiting distribution of both). The normal distribution in turn is a limiting case of the lognormal at the ($\beta_1 = 0$, $\beta_2 = 3$) point.

The wide range of (β_1, β_2) values covered by the Johnson family implies a great variety of distributional shapes, making it an ideal choice to model heterogeneity in distribution without the need to impose strong a priori assumptions. The same objective could be achieved using the Pearson family, which produces very similar distributions for a given choice of (β_1, β_2) (Kendall and Stuart 1969). The choice for the Johnson family in this paper was made for convenience: the software we used (R) currently does not provide a weighted maximum likelihood estimator for the Pearson family, while it does for the Johnson family.

3. An algorithm to combine Johnson distributions in a finite mixture model

Latent class models are composed of two elements: a number (K) of component distributions and a 'mixing distribution' that weighs each component. The mixing distribution can be viewed as a multinomial distribution over K possible outcomes. The combined distribution can be formalized as (we follow the notation of Leisch 2004):

$$h(y|x,\psi) = \sum_{k=1}^{K} \pi_k f(y|x,\theta_k) \tag{6}$$

$$\pi_k \geq 0, \qquad \sum_{k=1}^{K} \pi_k = 1$$

where

y = a dependent variable with conditional density h;
x = a vector of independent variables;
π_k = the prior probability of component k;
θ_k = the component-specific parameter vector of the density f; and
$\psi = (\pi_1, \ldots, \pi_K, \theta_k, \ldots, \theta_K)$ is the vector of all model parameters.

Since class membership is unknown, the π_{ik} (the probability that observation i is a member of component k) values have to be estimated together with the parameters of the component distributions. This is done using the 'Expectation-Maximization' (EM) algorithm (Dempster, Laird, and Rubin 1977), which iterates over two steps until convergence. In the E-step, the class membership probabilities are estimated taking the parameters of the component distributions as given. In the M-step, the parameters of the component distributions are estimated for given (current) values of the group membership probabilities. This step is essentially a weighted maximum likelihood estimate, with the group membership probabilities as weights. The two steps are repeated until convergence.

The EM estimation procedure needs to be adapted slightly in the case of a mixture of (potentially) different component distributions, since the probability density functions, which are the elements of the likelihood function, differ between components and are unknown. We propose to add a 'selection' step to the EM algorithm, which selects the component distribution based on the current value of the weights. This can be achieved by estimating the parameters of the Johnson SU and SB densities[1] and selecting the type of distribution which fits the observations best (based on the value of the likelihood function or the Akaike or Bayesian Information Criteria). Since the weighted ML parameters must be estimated in the 'selection' step, the solution with the maximum likelihood immediately produces the M-step result. Consequently, the 'selection' step automatically produces the M-step.

---

[1] There should be no need to estimate the SL and SN subtypes since they are limiting forms of SU and SB. As a consequence, SL and SN can be approximated arbitrarily closely by either of these distributions.

It should be noted that the procedure described above can, in principle, be used to combine any number of different component distributions. These distributions need not be restricted to the Johnson (or any other) family. The advantage of using the Johnson family is that, due to its flexible form, a great variety of distributions can be emulated without having to try them all [2]. Whichever combination of distributions is used, the proposed generalization modifies equation (6) to:

$$h(y|x,\psi) = \sum_{k=1}^{K} \pi_k f_k(y|x,\theta_k) \tag{7}$$

$$\pi_k \geq 0, \qquad \sum_{k=1}^{K} \pi_k = 1$$

A general problem in statistical modelling pertains to the identifiability of the model parameters. This is particularly relevant in the case of mixture models because of their great flexibility. Early work in this area (Teicher 1967) provided a condition for 'strict identifiability', but this condition turned out to be too strict and was later relaxed (see Allman, Matias, and Rhodes 2009). In fact, the general identifiability of mixtures of many different distribution families has been established in the meantime, such as for the X^2, Pareto and power function distributions (Chandra 1977), the exponential (Henna 1994), log-gamma and inverse log-gamma distributions (Atienza, Garcia-Heras, and Muñoz-Pichardo 2006) and even mixtures of components of different types such as the lognormal, gamma and Weibull distributions (Atienza, Garcia-Heras, and Muñoz-Pichardo 2006). Since most of these distributions are special cases of the Johnson family, one might conjecture that mixtures of Johnson distributions are also identifiable. This conjecture, however, remains to be investigated.

An issue related to identifiability is the problem of overparameterization, which is again a serious concern in the case of mixture models. Simply put, any data set can be fit arbitrarily closely given enough free parameters in the model. This is the rationale for imposing a penalty for additional parameters in goodness-of-fit statistics such as the AIC and BIC. In the case of the Johnson distribution, only two additional parameters are introduced relative to standard models such as the normal or gamma FFMs, so the risk of over-fitting does not appear to be very great relative to the gain in flexibility.

[2] In fact the 'flexmix' package does allow the user to combine different distributions, but these have to be specified when the function is called.

4. A Monte Carlo experiment

The FMM with Johnson distributions has been implemented using the 'flexmix' package in R. This package allows users to define their own M-step, which we have extended with a 'selection' step as explained above. The Johnson distribution has been implemented in several R packages, but only R´s 'SuppDist' package provides a weighted maximum likelihood routine (other R packages only offer method of moments estimators of the parameters). The current version of the package only implements the Johnson SU and SB types but this should not be a matter of great concern, as explained in footnote 1.

With the Johnson_M-step defined for the 'flexmix' package, all the necessary R code was available to put the 'Johnson Finite Mixture Model' (J-FFM) to the test. In what follows we have restricted ourselves to estimating the parameters of various mixtures of distributions, without trying to estimate mixtures of regression models (so there are no independent variables x in equation (7)). This will be the obvious next step in the application of the model. For now, we just provide a proof of principle that the method is feasible and has the potential to outperform the current practice of mixing a number of a priori chosen and identical distributions.

To test the model, we have generated data from five known distributions and tested whether the Johnson FFM is capable of reproducing the shape of the mixed distributions and how the model fit compares to a standard mixture of normal and gamma distributions. Draws of 100 pseudo random numbers with different shapes of three population component distributions have been generated and estimated using the Johnson FFM and normal/gamma FFMs [3]. The population parameters were chosen such that the skewness (β_1) and kurtosis (β_2) parameters cover a suitably wide range of values in the (β_1, β_2) plane that contains commonly used distributions such as normal, t, gamma, lognormal, beta and F distributions. Five sets of (β_1, β_2) were selected (see Table 1). The random numbers were drawn from Pearson distributions (whose types correspond to the common distributions mentioned earlier).

Table 1 Skewness and kurtosis values of the component distributions

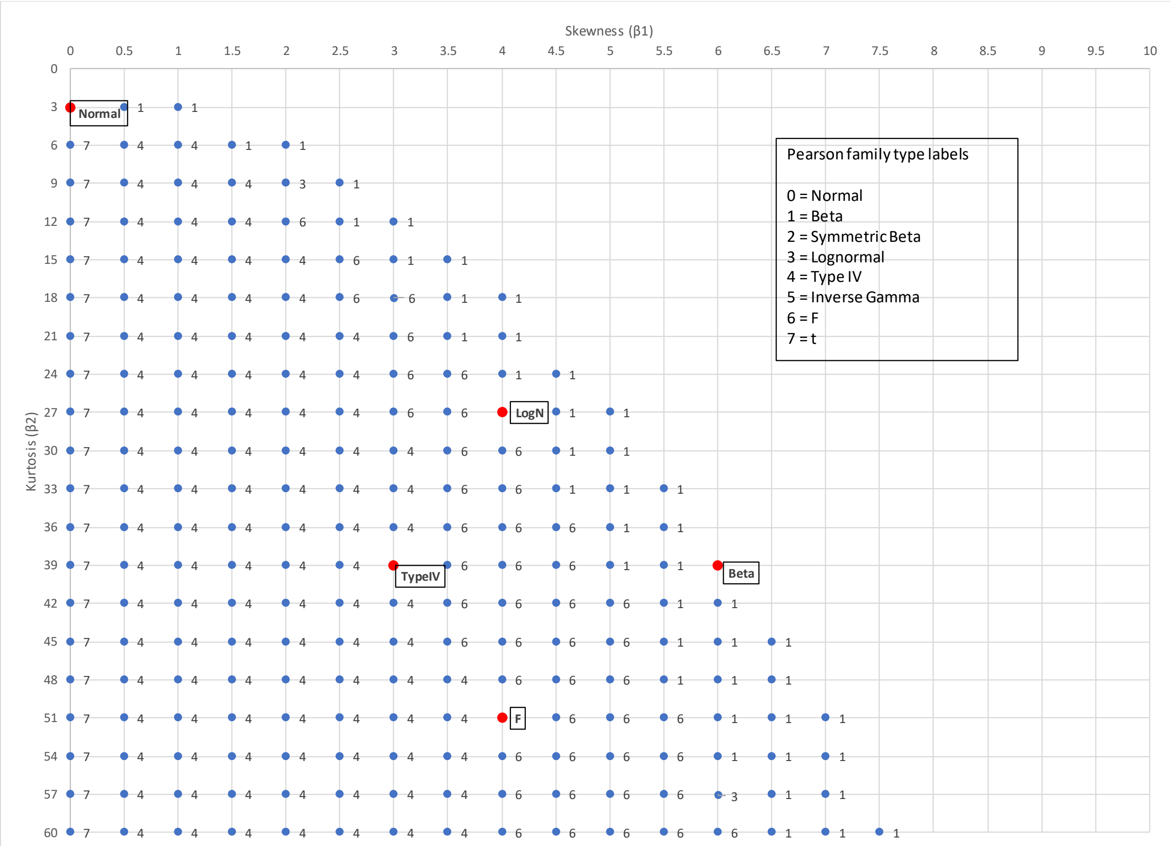
	β_1	β_2
Normal	0	3
F	4	51
Pearson Type IV	3	39
Beta	6	39
Lognormal	4	27

Figure 1 illustrates the Pearson types in a (β_1, β_2) grid and highlights the five selected distributions used in the experiment. The five skewness/kurtosis pairs for each of the three component distributions result in 35 different data generating processes (DGPs). The capacity for each FMM to estimate the DGP is evaluated by the Bayesian Information Criterion (BIC), a measure based on the (negative) log-likelihood with a penalty for the number of parameters used in the estimation process (increasing the number of parameters improves the fit, so the correction is needed to avoid overfitting, see section 3). This is

[3] The estimation process may not converge for every model and every random draw, and converged solutions may differ between draws since the likelihood function may converge to different local maxima. Convergence may also depend on the sample size. This needs to be investigated further.

particularly relevant in the current experiment, as the Johnson SU and SB distributions have four parameters, whereas the normal and gamma distributions only have two.

Figure 1 Pearson distributions in the (β_1, β_2) plane



The estimation results are presented in Table 2. Each row of the table represents a different DGP, as explained above. The first line is the benchmark of a mixture of three normal distributions, while subsequent lines show combinations of the five selected component distributions. The columns on the left side of the table show the input parameters of the DGPs, and the columns on the right show the BIC value for the fitted models (three normal, three gamma and three Johnson distributions respectively). It should be obvious that the three Johnson distributions need not be of the same type, since this is precisely the feature that provides the flexibility for which the method is proposed.

Table 2 Goodness of fit for selected DGPs estimated with mixtures of normal, gamma and Johnson distributions

Data Generating Process						Mixture model		
Component 1		Component 2		Component 3		3 normal	3 Gamma	3 Johnson
β_1	β_2	β_1	β_2	β_1	β_2	BIC	BIC	BIC
0	3	0	3	0	3	1664.0	1660.8	1694.1
0	3	0	3	4	51	1599.3	1597.0	1591.2
0	3	0	3	3	39	1677.9	1677.5	1669.6
0	3	0	3	6	39	1711.4	1704.2	1417.1
0	3	0	3	4	27	1703.5	1700.9	1578.9
0	3	4	51	4	51	1656.0	1553.7	1481.8
0	3	4	51	3	39	1652.5	1645.2	1599.1
0	3	4	51	6	39	1715.8	1711.7	1391.2
0	3	4	51	4	27	1576.4	1569.7	1478.9
0	3	3	39	3	39	1652.5	1646.6	1628.8
0	3	3	39	6	39	1720.2	1717.4	1490.2
0	3	3	39	4	27	1643.8	1632.9	1584.0
0	3	6	39	6	39	1726.0	1723.5	1735.2
0	3	6	39	4	27	1721.0	1722.0	1450.5
0	3	4	27	4	27	1493.3	1433.0	1391.5
4	51	4	51	4	51	1569.5	1673.8	1542.8
4	51	4	51	3	39	1652.9	1503.1	1457.7
4	51	4	51	6	39	1706.9	1695.9	1162.9
4	51	4	51	4	27	1462.2	1462.7	1502.0
4	51	3	39	3	39	1685.1	1680.7	1553.4
4	51	3	39	6	39	1750.9	1742.7	1421.5
4	51	3	39	4	27	1702.8	1693.4	1494.7
4	51	6	39	6	39	NC	4931.4	3259.3
4	51	6	39	4	27	1705.7	1698.0	1227.9
4	51	4	27	4	27	1700.6	1691.4	1283.1
3	39	3	39	3	39	1674.1	1671.2	1572.9
3	39	3	39	6	39	1684.8	1676.4	1397.8
3	39	3	39	4	27	1517.2	1514.8	1552.5
3	39	6	39	6	39	NC	NC	1050.3
3	39	6	39	4	27	NC	NC	1265.0
3	39	4	27	4	27	1597.9	1433.0	1355.6
6	39	6	39	6	39	3308.0	3286.3	1293.4
6	39	6	39	4	27	2693.5	2684.3	2790.4
6	39	4	27	4	27	1554.4	1550.3	1029.2
4	27	4	27	4	27	1700.6	1692.3	1173.3

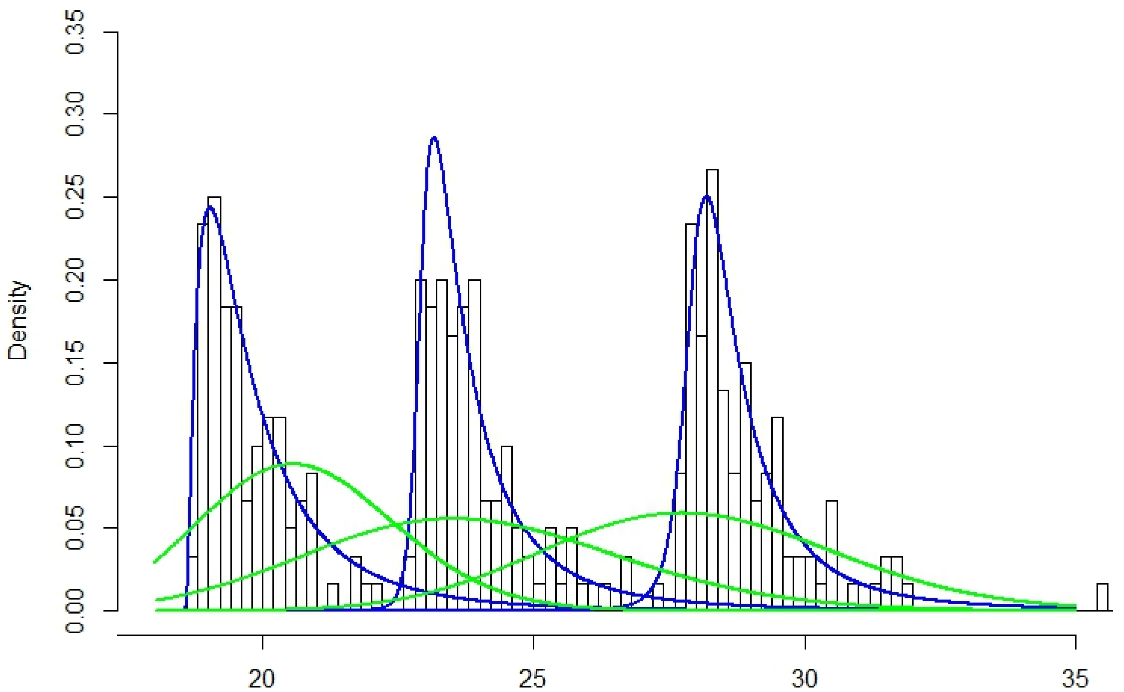
BIC: Bayesian Information Criterion. Lower values indicate better model fit.
NC: No convergence of the estimation procedure.

The results in Table 2 prove the feasibility of the method. Most DGPs can be estimated, although the estimation process may be slow as a consequence of convergence problems. This is not surprising, since the selection step involves the estimation of two types of Johnson distribution and the algorithm may alternate between local optima of the likelihood function as a result of jumps between the Johnson types.

This is especially likely when the true parameters are close to the lognormal distribution, which can be approximated by the neighbouring SU and SB distributions.

Comparing the fit of the FMMs, it is no surprise that the Johnson mixture is not better than either the mix of normal or gamma distributions when the DGP is in fact based on these distributions. However, for DGPs that are a mix of different distributions that cannot be well approximated by either normals or gammas, the fit of the Johnson FMM is usually (but not always) at least as good and often substantially better than the standard models. This is illustrated in Figure 2 for the case of three F distributions (shaded in Table 2).

Figure 2 Data histogram and fitted gamma (green) and Johnson (blue) distributions
DGP: 3x ($\beta_1 = 4$, $\beta_2 = 51$)



The observation that the Johnson FMM only outperforms standard mixture models when the DGP is in fact not normal, gamma or another standard distribution (a fact that is obviously unknown to the researcher) naturally leads to the idea to consider the standard FMMs as restrictions on a more general model (Johnson). Since many common distributions are special cases of the Johnson and Pearson families, a model fitted with any of these distributions may be compared against the 'general' or 'unrestricted' Johnson FMM using the likelihood ratio (LR) test. While the test statistic is not asymptotically Chi-square distributed in the case of non-nested hypotheses (as is the case here), it may still be used as a model selection criterion. In the example discussed above LR = 278.4, which would lead to strongly reject the restriction imposed by the gamma FMM against the more general Johnson model (the 'χ^2' - value would be evaluated with six ((4 - 2) x 3) degrees of freedom, since each gamma component has two parameters while each Johnson has four.

5. Conclusion

This paper introduces a method that extends the finite mixture model (FMM) to account for unobserved heterogeneity in distribution. FFMs have become popular in economics and many other fields of applied research to model heterogeneity in subject characteristics that are not observed in the data set and so cannot be controlled for. The model is typically used to model micro data where the units of observation are persons, firms or other individuals that differ systematically from each other in unknown ways in such a way that they effectively belong to groups or sub-populations. The current standard practice in FMM modelling consists of selecting the number of groups (usually labelled 'components') and of choosing a probability distribution for the component densities. The number of groups is usually determined by trial and error or based on a goodness-of-fit criterion, while the choice of the component distributions is typically based on a priori considerations regarding the support and the anticipated shape of the population distribution. A drawback of this approach is that it places a priori restrictions on the nature of the unobserved heterogeneity in at least two ways. First, the choice of the distribution is in general somewhat arbitrary and as a rule not tested against a more general (unrestricted) alternative. Second, while the 'true' number of latent classes is in principle unknown, it is also routinely assumed that the mixed distributions are of one kind. That is, the mixed components only differ from each other in terms of the parameters of the chosen distribution but not in terms of the probability density functions themselves.

The method proposed in this paper addresses these problems by lifting some of these implicit assumptions. This is achieved by postulating a flexible form for the component distributions. Several such flexible forms have been proposed and studied long ago, such as the Pearson and Johnson families, among others. Both families share the property that they can assume a wide variety of shapes depending on the value of their four parameters. In fact, most commonly used distributions are special cases of both families. The paper outlines an algorithm that can be used to estimate the parameters of a mixture of Johnson distributions and provides a proof of principle that the method is feasible and a potential improvement over current latent class modelling practice.

The method has been tested using data generated from different distributions, chosen to cover a wide range of combinations of skewness and kurtosis. The first results are encouraging. The method generally converges in about the same number of iterations as standard models that mix normal or gamma distributions. More importantly, when the data are generated from mixed distributions that differ substantially from the standard assumptions (identical component densities and 'regular' skewness and kurtosis values), the mixture of Johnson distributions generally fits the data better than the standard models.

Several steps still need to be taken to turn the method into a fully operational research tool. First, the model needs to be tested for mixtures of regression models, which are the main application in applied economics research and many other fields. Second, convergence problems related to the numerical optimization routines remain to be addressed. Third, the model needs to be tested with real-world data sets, ideally data that have been analysed with the standard FMM approach and with the results published in the academic literature. Finally, if the method passes all these tests, it should be made available to the research community as (part of) a statistical package.

References

Allman, Elizabeth S., Catherine Matias, and John A. Rhodes. 2009. 'Identifiability of Parameters in Latent Structure Models with Many Observed Variables'. *The Annals of Statistics* 37 (6A):3099–3132. https://doi.org/10.1214/09-AOS689.

Atienza, N., J. Garcia-Heras, and J. M. Muñoz-Pichardo. 2006. 'A New Condition for Identifiability of Finite Mixture Distributions'. *Metrika* 63 (2):215–21. https://doi.org/10.1007/s00184-005-0013-z.

Chandra, Satish. 1977. 'On the Mixtures of Probability Distributions'. *Scandinavian Journal of Statistics* 4 (3):105–12.

De Vries, Gaaitzen J., and Michael Koetter. 2011. 'ICT Adoption and Heterogeneity in Production Technologies: Evidence for Chilean Retailers*: ICT Adoption and Heterogeneity'. *Oxford Bulletin of Economics and Statistics* 73 (4):539–55. https://doi.org/10.1111/j.1468-0084.2010.00622.x.

Deb, Partha, and Ann M. Holmes. 2000. 'Estimates of Use and Costs of Behavioural Health Care: A Comparison of Standard and Finite Mixture Models'. *Health Economics* 9 (6):475–89. https://doi.org/10.1002/1099-1050(200009)9:6<475::AID-HEC544>3.0.CO;2-H.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. 'Maximum Likelihood from Incomplete Data via the EM Algorithm'. *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1):1–38.

Greene, William H., N. Harris, Mark, Bruce Hollingworth, and Pushkar Maitra. 2008. 'A Bivariate Latent Class Correlated Generalized Ordered Probit Model with an Application to Modeling Observed Obesity Levels'. SSRN Scholarly Paper ID 1281910. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=1281910.

Henna, Jogi. 1994. 'Examples of Identifiable Mixture'. *Journal of the Japan Statistical Society* 24 (2):193–200. https://doi.org/10.11329/jjss1970.24.193.

Jiménez-Martín, Sergi, José M. Lebeaga, and Maite Martínez-Granado. 2002. 'Latent Class versus Two-Part Models in the Demand for Physician Services Across the European Union'. In *Econometric Analysis of Health Data*, edited by Andrew M. Jones and Owen O'Donnell, 101–16. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/0470846313.ch7.

Johnson, N. L. 1949. 'Systems of Frequency Curves Generated by Methods of Translation'. *Biometrika* 36 (1/2):149. https://doi.org/10.2307/2332539.

Johnson, Norman Lloyd, Samuel Kotz, and N. Balakrishnan. 1994. *Continuous Univariate Distributions*. 2nd ed. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Kendall, Maurice G., and Alan Stuart. 1969. *The Advanced Theory of Statistics*. 3rd ed. London: Griffin.

Leisch, Friedrich. 2004. 'FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in *R*'. *Journal of Statistical Software* 11 (8). https://doi.org/10.18637/jss.v011.i08.

McLachlan, Geoffrey J., and David Peel. 2000. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Applied Probability and Statistics Section. New York: Wiley.

Pacifico, Daniele. 2012. 'On the Role of Unobserved Preference Heterogeneity in Discrete Choice Models of Labour Supply'. *Empirical Economics*, August. https://doi.org/10.1007/s00181-012-0637-6.

Teicher, Henry. 1967. 'Identifiability of Mixtures of Product Measures'. *The Annals of Mathematical Statistics* 38 (4):1300–1302.

Wedel, Michel, and Wagner A. Kamakura. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. 2nd ed. Boston: Kluwer Academic.