A Copula-Based Finite Mixture Sample Selection Model: Application to the Effects of Hearing Problems on Female Wage Earnings

David M. Zimmer[*]
Western Kentucky University

May 4, 2018

Abstract

This paper presents a copula-based sample selection model where the outcome variable follows a finite mixture specification. The estimation approach relies on standard maximum likelihood methods, and thus is relatively easy to implement. A series of Monte Carlo simulation studies vouch for the performance of the proposed model. An empirical application explores the effects of hearing impairments on wage earnings of females. Results find that hearing impairments lead to substantial wage reductions among approximately 24 percent of females characterized by high earnings and high educational attainment. However, among approximately 76 percent of females with lower earnings and lower educational attainment, hearing impairments do not appear to harm wage earnings.

JEL Codes: J14; J30; C34

Keywords: Gaussian; normal; discrete approximation

[*] Department of Economics, Western Kentucky University, Bowling Green, KY 42101; david.zimmer@wku.edu; 270-745-2880

1 Introduction

This paper develops a copula-based finite mixture model that accommodates sample selection, sometimes called “incidental truncation,” in the outcome variable. The proposed specification combines elements of finite mixture setups (Heckman and Singer, 1984; Aitken and Rubin, 1985) with copula-based selection models (Lee, 1983; Smith, 2003). The estimation approach relies on standard maximum likelihood methods, and thus is relatively easy to implement. A series of Monte Carlo simulation studies vouch for the performance of the proposed model. Finally, an empirical application explores the effects of hearing impairments on wage earnings of females, where wage earnings are observed only for females with positive earnings.

Finite mixture models provide an attractive modeling choice when observations in an estimation sample belong to different “groups,” with dependent variables following group-specific distributions. McLachlan and Peel (2004) provide a survey of the statistical literature on those methods. The economics literature provides many applications in labor (Heckman, Robb, and Walker, 1990; Geweke and Keane, 1997), development (Morduch and Stern, 1997; Alfo, Trovato, and Waldmann, 2008), and health care usage (Deb and Trivedi, 1997), among many others.

For example, in the empirical application considered in this paper, wage earning females might fall into, say, two groups. In one of those groups, hearing impairments might correlate with reduced wages, while in the other group, hearing impairments might not affect earnings. Such a pattern would be evident if person-to-person communication for jobs in the second group could effectively operate in written forms.

But as with many econometric tools, the introduction of some complexity, like finite mixtures, often precludes inclusion of other intricacies, although researchers have made some headway blending finite mixtures into larger modeling situations. For example, finite mixture setups have been incorporated into simultaneous equation specifications (Mroz, 1987) and models with endogenous righthand side regressors (Conway and Deb, 2005). Some of those innovations have made their way into commercial statistical software (see the newly-added suite of `fmm` estimators in Stata 15).

But to date, finite mixtures have not been incorporated into selection frameworks. The main reason is that selection models require jointly modeling (1) a selection equation and (2) an outcome equation. And if one of those equations has a finite mixture setup, the joint distribution of both parts, required for maximum likelihood estimation of selection models, becomes unwieldy. This paper sidesteps that obstacle by using copula functions.

The following two sections, in addition to introducing notation used throughout this study, discuss finite mixture models and copula-based selection specifications. The paper then proceeds to combine those ideas into what this paper calls a copula-based finite mixture sample selection model. After presenting that model and testing its performance, the empirical application in this paper finds that hearing impairments lead to substantial wage reductions among approximately 24 percent of females characterized by high earnings and high educational attainment. However, among approximately 76 percent of females with lower earnings and lower educational attainment, hearing impairments do not appear to harm wage earnings. Therefore, the findings potentially offer policymakers important information to better target policies designed to assist people with hearing problems.

2 Basics of Finite Mixture Models

For a data sample $i = 1, ..., n$, let y_i be an outcome variable of interest, and let $\mathbf{x}_i$ be a vector of exogenous explanatory variables. For example, the empirical application explored in this paper considers a sample of females for which the outcome y_i denotes (log) annual wage earnings, and the main explanatory variable of interest included in the vector $\mathbf{x}_i$ is a binary indicator for whether the female has a hearing problem. (Section 6 argues that hearing impairments are relatively exogenous physical conditions.) The aim is to determine the extent, if any, to which hearing problems affect wage earnings.

Let the probability density function (pdf) of y_i be specified as

$$f(y_i|\Omega)$$

where Ω includes all estimable parameters contained in $f(\cdot)$, including regression coefficients attached to $\mathbf{x}_i$ and ancillary parameters such as standard deviations. Those estimable parameters may be estimated by finding their values that maximize the log likelihood function,

$$\sum\nolimits_i \ln f(y_i|\Omega).$$

However, many economic outcomes of interest, especially in micro data, do not adhere to one distribution, but rather a mixture of distributions. In those cases, the pdf of the outcome variable might more accurately be expressed as

$$\sum\nolimits_c \pi_c f_c(y_i|\Omega_c)$$

where $f_c(\cdot)$ denotes each component pdf, and the terms π_c represent the probabilities of belonging to each component, where $\sum_c \pi_c = 1$. Then the estimable parameters in each component and the mixing probabilities may be simultaneously estimated by finding their values that maximize the log likelihood function

$$\sum\nolimits_i \ln \sum\nolimits_c \pi_c f_c(y_i|\Omega_c). \tag{1}$$

As with standard maximum likelihood settings, the likelihood function (1) can be maximized using iterative Newton-type methods, with standard errors obtained using the familiar likelihood-based formulas.

Along with the component-specific parameters Ω_c, the procedure also yields estimates of the mixing probabilities π_c, offering evidence about what proportion of observations belong to each component. However, the procedure does not provide direct information about how those components differ from each other, nor does the procedure indicate the component to which a specific observation i belongs. That information can be approximated, after estimation, by applying a version Bayes' formula,

$$\Pr(\text{Component} = c \mid \mathbf{x}_i) = \frac{\widehat{\pi}_c f_c(y_i|\widehat{\Omega}_c)}{\sum_c \widehat{\pi}_c f_c(y_i|\widehat{\Omega}_c)}, \tag{2}$$

where circumflexes indicate converged parameter estimates obtained by maximum likelihood. Using the calculated probabilities obtained from equation (2), one then may assign data observations to specific components using some classification criteria in order to investigate how those components differ.

3 Copula-Based Selection Models

This section reviews standard selection models and their copula extensions. The section includes sufficient detail to keep this paper relatively self-contained, but readers seeking further exposition should consult Smith (2003) and Zimmer (2018).

3.1 Standard selection model

The standard selection model setup (Heckman, 1979), specifies two equations, both with latent variable outcomes. One equation relates to the outcome of interest, while the other concerns selection. Couched in terms of the empirical application considered below, let y^*_{1i} denotes female i's propensity to work, with the observed selection variable taking the form

$$y_{1i} = \begin{cases} 1 & \text{if } y^*_{1i} > 0 \\ 0 & \text{if } y^*_{1i} \leq 0 \end{cases} .$$

Let y^*_{2i} denote female i's annual (log) wage earnings. Those wage earnings are observed only for females who work, implying that the observed outcome variable takes the form

$$y_{2i} = \begin{cases} y^*_{2i} & \text{if } y^*_{1i} > 0 \\ - & \text{if } y^*_{1i} \leq 0 \end{cases} .$$

Thus, the outcome of interest y_{2i} is observed when $y_{1i} = 1$, but y_{2i} is missing when $y_{1i} = 0$.

The two latent equations comprising the full selection model take the form

$$\begin{aligned} y^*_{1i} &= \mathbf{z}_i^{'}\boldsymbol{\gamma} + \varepsilon_{1i} \\ y^*_{2i} &= \mathbf{x}_i^{'}\boldsymbol{\beta} + \varepsilon_{2i} \end{aligned} \tag{3}$$

with the goal being consistent estimation of the coefficients $\boldsymbol{\beta}$ in the outcome equation. Correlation between the errors $(\varepsilon_1, \varepsilon_2)$ imparts bias on estimates of $\boldsymbol{\beta}$ if the selection problem is ignored. Technically, the coefficients $\boldsymbol{\beta}$ are identified via non-linear distributional assumptions even when the vectors $\mathbf{x}_i$ and $\mathbf{z}_i$ contain the same explanatory variables. For more robust identification, however, many economic applications use "exclusions restrictions," which are variables present in $\mathbf{z}_i$ but excluxed from $\mathbf{x}_i$.

Maximum likelihood estimation (MLE) of the parameters in (3) requires the joint distribution of the latent outcomes (y_{1i}^*, y_{2i}^*). Let the cumulative distribution function (cdf) of that joint distribution be $F(y_{1i}^*, y_{2i}^*)$, with marginal cdfs $F_1(y_{1i}^*)$ and $F_2(y_{2i}^*)$ and corresponding marginal pdfs $f_1(y_{1i}^*)$ and $f_2(y_{2i}^*)$. Then the (unlogged) likelihood function of the selection model takes the form

$$L = \prod\nolimits_0 \Pr(y_{1i}^* \leq 0) \prod\nolimits_1 \Pr(y_{1i} > 0)\ f_{2|1}(y_{2i}|y_{1i}^* > 0)$$

where $f_{2|1}$ is the density function of y_{2i}^* given $y_{1i}^* > 0$. The subscripts on the product operators indicate multiplication over observations for which $y_{1i} = 0$ and $y_{1i} = 1$. The conditional density $f_{2|1}$ is equal to $\frac{1}{\Pr(y_{1i}^*>0)} \frac{\partial}{\partial y_2}(F_2(y_{2i}) - F(0, y_{2i}))$ so that the likelihood function can be re-expressed as

$$L = \prod\nolimits_0 F_1(0) \prod\nolimits_1 \left(f_2(y_2) - \frac{\partial}{\partial y_2} F(0, y_2) \right) \tag{4}$$

where the subscript i has been dropped for notational brevity.

The standard selection setup specifies the joint distribution $F(0, y_2)$ as bivariate normal and the marginals $F_1(0)$ and $f_2(y_2)$ as univariate normal. Then, calculating the natural logarithm of (4) and summing over all observations gives the log-likelihood function.

3.2 Forming joint distributions with copulas

Imposing joint normality on $F(y_1^*, y_2^*)$ allows one to write the likelihood in the form of equation (4), but that normality assumption also imposes normality on the marginals. Sklar's Theorem (1973), however, allows for the construction of a multivariate distribution based on *non-normal* marginals. Thus, using the marginal distributions $F_1(y_1^*)$ and $F_2(y_2^*)$ as arguments, a copula function, denoted $C(\cdot)$, allows for a representation of the joint distribution as

$$F(y_1^*, y_2^*) = C(F_1(y_1^*), F_2(y_2^*); \theta)$$

where θ is the copula dependence parameter. The practical benefit of copulas is that, while researchers often know marginal behaviors of individual variables, they have less familiarity with joint relationships. Copulas, therefore, allow researchers to form those higher-dimensional relationships based on the easier-to-formulate univariate marginals.

The most widely-used copula, called the Normal or Gaussian copula, assumes the form

$$F(y_1^*, y_2^*) = \Phi_B(\Phi^{-1}(F_1(y_1^*)), \Phi^{-1}(F_2(y_2^*)); \theta) \tag{5}$$

where Φ_B denotes the bivariate standard normal cdf, and Φ^{-1} represents the quantile function of the standard normal distribution.

3.3 Copula selection models

The normality assumptions commonly imposed in (4) have long been acknowledged as potential sources of misspecification (Goldberger, 1983). A large literature has emerged to relax those assumptions. One strand of that literature, started by Lee (1983) and later generalized by Prieger (2002) and Smith (2003), uses copula functions in place of the most difficult part of (4), the joint distribution $F(0, y_2)$. Using the Normal copula (5) for that joint distribution yields the (unlogged) likelihood function (Smith, 2003),

$$L = \prod\nolimits_0 F_1(0) \prod\nolimits_1 f_2(y_2) \left(1 - \Phi\left(\frac{\Phi^{-1}(F_1(0)) - \theta\Phi^{-1}(F_2(y_2))}{\sqrt{1-\theta^2}}\right)\right).$$

And if the selection equation for y_1 follows a probit specification, as seems natural in many economic applications, then the (log) likelihood function, after reintroducing observation-specific subscripts, takes the form

$$\begin{aligned} \ln L &= (1 - y_{i1})\ln(1 - \Phi(\mathbf{Z}_i'\boldsymbol{\gamma})) + y_{i1}\ln f_2(y_{i2}) \\ &\quad + y_{i1}\ln\left(1 - \Phi\left(\frac{\Phi^{-1}(1 - \Phi(\mathbf{Z}_i'\boldsymbol{\gamma})) - \theta\Phi^{-1}(F_2(y_2))}{\sqrt{1-\theta^2}}\right)\right). \end{aligned} \tag{6}$$

Written in this form, the main modeling decision involves specifying the marginal pdf and cdf of the outcome, $f_2(y_{i2})$ and $F_2(y_{i2})$. If those marginals are normal, then this model is identical to the Heckman-style selection model to be estimated by MLE. But the copula form of selection model allows greater flexibility in that it can accommodate *any* valid form for the marginal distribution of the outcome, including, as outlined in the following section, settings in which the marginal distribution for y_2 follows a finite mixture specification.

The derivation in this subsection uses the Normal copula. To be sure, the statistics literature offers many off-the-shelf copula functions useful for empirical applications, but, unfortunately, most of those have restrictive dependence patterns that make them unappealing for selection problems. For example, the widely-used Clayton, Gumbel, and Joe copulas only permit positive dependence ($\theta > 0$). And although the Farlie-Gumbel-Morgenstern and Ali-Mikhail-Haq copulas permit both positive and negative dependence, their ranges of dependence are relatively limited in the sense that neither permits the full range of Fréchet dependence. *Any* restrictions placed on θ are problematic, because, for selection models, that dependence term captures the direction and magnitude of selection bias, which is often difficult to know *a priori*. In order to remain agnostic about the direction and magnitude of selection bias, this paper suggests using either the Normal or Frank copulas, both of which permit the full range of Fréchet dependence. Owing to its familiarity and ease of coding, this paper opts for the Normal.

4 Copula-Based Finite Mixture Sample Selection Model

As noted in the previous subsection, in contrast to standard selection setups, copula-based selection models allows the marginal distribution of the outcome variable to assume *any* form. This paper exploits that flexibility by letting the outcome variable follow a finite mixture setup.

Using notation introduced earlier in this paper, let the pdf and cdf of the outcome take the forms

$$f_2(y_{i2}) = \sum\nolimits_c \pi_c f_c(y_{i2}|\Omega_c) \tag{7}$$

$$F_2(y_{i2}) = \sum\nolimits_c \pi_c F_c(y_{i2}|\Omega_c). \tag{8}$$

In sum, the copula-based finite mixture sample selection model substitutes (7) and (8) into the log likelihood function given in (6).

The final modeling decision involves the number and the forms of component distributions in the finite mixture part of the model. In practice, finite mixtures with two components often have appealing economic interpretations, and two-component models lessen the convergence and overfitting issues common in mixtures with more components (Deb and Trivedi, 1997). Thus, the applications below use two components. Furthermore, for many microeconometric applications, it seems plausible to let the component distributions come from the same family, but with component-specific parameters Ω_c. Because the outcome in the empirical application presented below, log annual earnings, appears approximately normal, estimates below use normal distributions for both components. Thus, the pdf and cdf are expressed as

$$f_2(y_{i2}) = \pi f_1(y_{i2}|\mathbf{x}_i'\boldsymbol{\beta}_1, \sigma_1) + (1-\pi) f_2(y_{i2}|\mathbf{x}_i'\boldsymbol{\beta}_2, \sigma_2)$$

$$F_2(y_{i2}) = \pi F_1(y_{i2}|\mathbf{x}_i'\boldsymbol{\beta}_1, \sigma_1) + (1-\pi) F_2(y_{i2}|\mathbf{x}_i'\boldsymbol{\beta}_2, \sigma_2)$$

where f and F are the pdf and cdf of the normal distribution with component-specific means $\mathbf{x}_i'\boldsymbol{\beta}_c$ and standard deviations σ_c. Then the full set of estimable parameters is $(\boldsymbol{\gamma}, \boldsymbol{\beta}_1, \sigma_1, \boldsymbol{\beta}_2, \sigma_2, \theta, \pi)$, where $\boldsymbol{\gamma}$ represents the vector of slopes attached to the regressors $\mathbf{z}_i$ in the selection equation, and θ is the copula dependence term that captures the direction and magnitude of selection.

5 Simulation Study

This section conducts a set of Monte Carlo experiments designed to gauge the performance of the proposed copula-based finite mixture sample selection model. The experiments specify a selection equation given by

$$y_{1i} = \mathbf{1}(\gamma_0 + \gamma_0 x_i + \gamma_2 z_i + \varepsilon_{1i} > 0)$$

where x and z are drawn from (independent) standard normal distributions, and remain fixed throughout each replication of each experiment. The notation $\mathbf{1}()$ indicates that the selection variable y_{1i} equals 1 if the condition in parentheses is true, and 0 otherwise.

The outcome equation follows a finite mixture setup

$$\text{component 1: } y_{2i} = \beta_{10} + \beta_{11} x_i + \varepsilon_{2i}$$

$$\text{component 2: } y_{2i} = \beta_{20} + \beta_{21} x_i + \varepsilon_{2i}.$$

In all experiments, the probability that an observation belongs to component 1 is 0.60, implying a 0.40 probability of belonging to component 2. The selection problem arises from (i) the disturbances being jointly distributed as

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right),$$

and from (ii) y_{2i} being set to missing when y_{1i} equals 0.

The experiments set the sample size to 10,000 and use 500 replications. For each replication, the disturbance terms are drawn from the aforementioned bivariate normal distribution, which produces new realizations for y_{1i} and y_{2i} for each replication. Using those simulated data, the copula-based finite mixture sample selection model is estimated as specified in the previous section. A similar finite mixture model that ignores the selection problem also is estimated for sake of comparison.

Table 1 shows means and standard deviations of the 500 replications, along with "true" values of the parameters. Recognizing that the main focus of this paper is the coefficients attached to explanatory variables in the outcome part of the model (β_{11} and β_{21}), the model that ignores selection appears to badly misestimate those parameters. The copula-based finite mixture sample selection model, by contrast, appears to more accurately estimate those values. In fact, the selection model appears to satisfactorily estimate most parameters, the exception being the copula dependence parameter at the bottom of the table. However, for the dependence parameter, the "true" value reflects the overall correlation between the error terms ε_{1i} and ε_{2i}, with the mixture part of the model then splitting the second error term into two distributions, each with different values for the first two moments, and each with different proportional contributions to the overall mixture. Thus, it becomes difficult to compare the estimated dependence parameter to its "true" value.

For the experiments in Table 1, approximately 20 percent of observations have missing values for y_{2i}. Does the model's performance suffer when more observations are missing? To explore that possibility, Table 2 shows a similar experiment where γ_0 is set to 0.50, rendering approximately 33 percent of observations missing. Continuing that trend, Table 3 then sets γ_0 equal to 0.00, meaning approximately 50 percent of observations are missing. Those tables show that, as the percentage of missing observations increases, the performance of the model that ignores selection becomes progressively worse, highlighted by the parameters of interest (β_{11} and β_{21})

moving farther from their true values. The performance of the selection model, however, does not appear to suffer. It should be noted that, as the percentage of missing observations increases, Newton's method seems to require more iterations. Nonetheless, the selection model always manages to achieve convergence on each replication, regardless of the number of missing observations.

Table 4 returns to the original true values from Table 1, but it removes the variable z from the selection equation. Eliminating that variable means that identification comes solely from the nonlinear functional forms built into the model. The selection model appears to perform satisfactorily, but that should not be interpreted as an endorsement of models that lack exclusion restrictions. The satisfactory performance in Table 4 could be a consequence of the true values selected for the experiment, but even so, selection models without exclusion restrictions often become difficult to interpret through the lens of economic theory. Good practice still suggests finding appropriate exclusion restrictions.

6 Empirical Application

This section seeks to estimate the effects of hearing impairments on annual wage earnings among females. The presence of a hearing impairment should be relatively exogenous with respect to earnings, because most causes of hearing impairments stem from sources outside a person's control, such as non-contagious infections, birth defects, genetics, and aging (Koffler, Ushakov, and Avraham, 2015). However, if unmeasured individual-specific factors that lead a person into employment also correlate with earnings potential, then estimators that ignore such selectivity might misrepresent the true link between hearing problems and earnings.

6.1 Data

Data come from the 2012, 2013, 2014, 2015, 2016, and 2017 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). The estimation sample focuses on females ages 25-64 who do *not* report themselves as self employed. The final estimation sample includes $n = 298{,}064$ observations.

The outcome variable is annual wage earnings, converted to 2017 dollars using the Consumer Price Index. The main explanatory variable of interest is a dummy indicator for whether the person reports any difficulty hearing. Table 5 reports sample means, partitioned by hearing status. Females with hearing problems appear to have lower wages, and females with hearing problems also are less likely to have any wages. That pattern of labor market participation opens the possibility of selection bias if one's goal is to assess the impacts of hearing impairments on wage earnings.

The remainder of the table shows means for control variables. Of particular note, the bottom two rows show the number of children under 5 and the number of children 5 or older present in the household. Following past studies that estimate selection models of female wages (Mroz, 1987), the selection models estimated here use those two variables as exclusion restrictions, meaning they appear as explanatory variables in the selection equation but not in the outcome wage equation.

6.2 Estimates

Table 6 reports estimates that ignore the selection problem, focusing only on females with positive wage earnings. The left panel, which reports OLS results, shows that hearing impairments correlate with an approximate 25 percent reduction in wage earnings. The right panel, using a finite mixture setup, provides evidence that the sample consists of two components, with 21 percent of females belonging to the first component, and the remaining 79 percent coming from the second component. Although hearing impairments associate with reduced wage earnings in both components, the magnitude appears much larger among the 21 percent of females in the first component. That first component shows a 66 percent reduction in wage versus a 7.5 percent reduction in the second component.

Table 7 shows estimates from a Heckman-style selection model, estimated by maximum likelihood. The selection term (ρ) is negative, large in magnitude, and precisely estimated. The implication is that unmeasured attributes that correlate with increased labor market participation also tend to associate with reduced wages. Accounting for that selection yields a *positive* link between hearing problems and wage earnings. Also of note is that the two exclusions restrictions – number and children under 5 and number of children 5 or older – both appear to significantly and nontrivially reduce the likelihood of positive earnings in the selection equation.

Finally, Table 8 presents results for the copula-based finite mixture sample selection model proposed in this paper. The copula dependence parameter (θ), analogous to the term ρ in the previous paragraph, finds negative, large, and precisely estimated evidence of selection. The finite mixture part of the model estimates that 24 percent of females belong to the first component, which is similar to the nonselection estimate reported in Table 6. The finite mixture part of the model also finds that, among the first component, hearing problems lead to a 34 percent reduction in wages. For the other 76 percent of females who belong to the second component, hearing problems correlate with an approximate 9.8 percent *increase* in wages. Thus, to the extent that hearing problems reduce wage earnings among females, that harm, while large in magnitude, appears to be concentrated among approximately 24 percent of females.

6.3 Characteristics of the components

Finite mixture models separate samples into components, but they do not inform upon *which* observations come from which component. But, after estimation, finite mixture models do allow one to calculate the *probability* that an observation belongs to a particular component, using the expression given in equation (2). Using that formula and the estimated values presented in the finite mixture part of the model in Table 8, observations with calculated probabilities of belonging to the first component larger than 0.50 are assigned to the first component, and subjects with probabilities less than 0.50 are assigned to the second component. (The distribution of calculated probabilities is heavily bimodal, with little mass at 0.50. Thus, different cutoff values do not alter the main findings.)

Table 9 presents sample means partitioned by assigned components. The most important finding is that females who likely belong to the first component have significantly higher wage earnings, and also higher levels of educational attainment, compared to their counterparts assigned to the second component. Recalling that the first component consists of the approximately 24 percent of females for whom hearing problems lead to large reductions in earnings, the implication is that hearing problems have the largest proportional effects among high earning, and highly educated, females. Hearing impairments do not appear to harm wage earnings among lower earning, and less educated, females.

7 Conclusion and Practical Considerations

This paper presents a copula-based finite mixture sample selection model. The model is relatively easy to code, as it relies upon standard maximum likelihood methods and the optimizers often employed with those methods. An empirical application considers the effects of hearing impairments on female wage earnings. Results find that hearing impairments lead to substantial wage reductions among approximately 24 percent of females characterized by high earnings and high educational attainment. However, among approximately 76 percent of females with lower earnings and lower educational attainment, hearing impairments do not appear to harm wage earnings.

Finite mixture models, in general, and the selection versions proposed here, in particular, warrant a few notes of caution. First, although finite mixture specifications can handle more than the two-component scenarios presented here, the Newton-type optimizers tend to encounter more flat spots and areas of nonconcavity on the likelihood surface. Furthermore, adding more components risks overfitting, as some components might capture a small proportion of observations. Thus, unless a researcher has strong reasons for pursuing more components, two-component setups have intuitive and practical appeal (Deb and Trivedi, 1997).

Even in two-component settings, finite mixture models might exhibit local optima on the likelihood surface. Changing starting values is a useful, albeit informal, method for checking whether estimates adhere to the global optimum. Finally, finite mixture specifications, especially the selection-based version presented in this paper, seem to perform better with larger datasets. Trying to identify components and selection effects in small data settings proved onerous in preliminary explorations. That is the main reason the sample sizes used in this paper are relatively large.

References

Aitken, M. and D. Rubin (1985). "Estimation and Hypothesis Testing in Finite Mixture Models." *Journal of the Royal Statistical Society, B*, 47, 67-75.

Alfo, M., Trovado, G., and R. Waldmann (2008). "Testing for Country Heterogeneity in Growth Models using a Finite Mixture Approach." *Journal of Applied Econometrics*, 23, 487-514.

Blackwell, D., Lucas, J., and T. Clarke (2014). "Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2012." *Vital and Health Statistics*, 260, 1-161.

Conway, K. and P. Deb (2005). "Is Prenatal Care Really Ineffective? Or, is the 'Devil' in the Distribution?" *Journal of Health Economics*, 24, 489-513.

Deb, P. and P. Trivedi (1997). "Demand for Medical Care by the Elderly in the United States: A Finite Mixture Approach." *Journal of Applied Econometrics*, 12, 313-336.

Geweke, J. and M. Keane (1997). "Mixture of Normals Probit Models." Federal Reserve Bank of Minneapolis, Staff Report 237.

Goldberger, A. (1983). "Abnormal Selection Bias." In S. Karlin, T. Amemiya, and L. Goodman, eds., *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press

Heckman, J. (1979). "Sample Selection as a Specification Error." *Econometrica*, 47, 153-161.

Heckman, J. and B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data." *Econometrica*, 52, 271-320.

Heckman, J. (1990). "Varieties of Selection Bias." *American Economic Review*, 80, 313-318.

Heckman, J., Robb, R., and M. Walker (1990). "Testing the Mixture of Exponentials Hypothesis and Estimating the Mixing Distribution by the Method of Moments." *Journal of the American Statistical Association*, 85, 582-589.

Koffler, T., Ushakov, K., and K. Avraham (2015). "Genetics of Hearing Loss – Syndromic." *Otolaryngologic Clinics of North America*, 48, 1041-1061.

Lee, L.-F. (1983). "Generalized Econometric Models with Selectivity." *Econometrica*, 51, 507-512.

McLachlan, G. and D. Peel (2004). *Finite Mixture Models.* John Wiley: New York.

Morduch, J. and H. Stern (1997). "Using Mixture Models to Detect Sex Bias in Health Outcomes in Bangladesh." *Journal of Econometrics*, 77, 259-276.

Mroz, T. (1987). "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." *Econometrica*, 55, 765-799.

Prieger, J. (2002). "A Flexible Parametric Selection Model for Non-Normal Data with Application to Health Care Usage." *Journal of Applied Econometrics*, 17, 367-392.

Sklar, A. (1973). "Random Variables, Joint Distributions, and Copulas." *Kybernetica*, 9, 449–460.

Smith, M. (2003). "Modelling Sample Selection Using Archimedean Copulas." *Econometrics Journal*, 6, 99-123.

Zimmer, D. (2018) "The Effect of Hearing Impairments on Wage Earnings: Evidence from a Copula-Based Spline Selection Model." Working Paper, Western Kentucky University.

Table 1: Means (standard deviations) from experiment 1, approximately 20 percent missing outcome observations, (n = 10,000; replications = 500)

		true	Ignoring selection	Selection model
Component 1	β_{10}	10.0	10.12	10.03
			(0.02)	(0.03)
	β_{11}	0.25	0.34	0.28
			(0.05)	(0.06)
	σ_1	1.0	0.95	0.98
			(0.02)	(0.03)
Compontent 2	β_{20}	10.0	10.23	10.05
			(0.04)	(0.04)
	β_{21}	-0.50	-0.42	-0.51
			(0.09)	(0.09)
	σ_2	1.0	0.95	0.97
			(0.03)	(0.04)
Probability of component 1	π	0.60	0.63	0.60
			(0.08)	(0.09)
Selection equation	γ_0	1.0	–	1.00
				(0.02)
	γ_1	-0.60	–	-0.60
				(0.02)
	γ_2	-0.10	–	-0.10
				(0.02)
Copula dependence	θ	0.50	–	0.38
				(0.10)

Table 2: Means (standard deviations) from experiment 2, approximately 33 percent missing outcome observations, (n = 10,000; replications = 500)

		true	Ignoring selection	Selection model
Component 1	β_{10}	10.0	10.23	10.06
			(0.03)	(0.06)
	β_{11}	0.25	0.37	0.29
			(0.05)	(0.07)
	σ_1	1.0	0.93	0.97
			(0.02)	(0.04)
Compontent 2	β_{20}	10.0	10.33	10.09
			(0.04)	(0.10)
	β_{21}	-0.50	-0.39	-0.50
			(0.09)	(0.10)
	σ_2	1.0	0.93	0.96
			(0.04)	(0.04)
Probability of component 1	π	0.60	0.63	0.61
			(0.08)	(0.09)
Selection equation	γ_0	0.50	–	0.50
				(0.01)
	γ_1	-0.60	–	-0.60
				(0.02)
	γ_2	-0.10	–	-0.10
				(0.01)
Copula dependence	θ	0.50	–	0.36
				(0.11)

Table 3: Means (standard deviations) from experiment 3, approximately 50 percent missing outcome observations, (n = 10,000; replications = 500)

		true	Ignoring selection	Selection model
Component 1	β_{10}	10.0	10.37	10.12
			(0.03)	(0.11)
	β_{11}	0.25	0.40	0.30
			(0.05)	(0.08)
	σ_1	1.0	0.92	0.97
			(0.02)	(0.05)
Compontent 2	β_{20}	10.0	10.46	10.16
			(0.05)	(0.14)
	β_{21}	-0.50	-0.36	-0.49
			(0.10)	(0.12)
	σ_2	1.0	0.92	0.95
			(0.04)	(0.05)
Probability of component 1	π	0.60	0.63	0.63
			(0.08)	(0.08)
Selection equation	γ_0	0.00		0.00
				(0.01)
	γ_1	-0.60	–	-0.60
				(0.01)
	γ_2	-0.10	–	-0.10
				(0.01)
Copula dependence	θ	0.50	–	0.33
				(0.13)

Table 4: Means (standard deviations) from experiment 4, approximately 20 percent missing outcome observations, no exclusion restriction, (n = 10,000; replications = 500)

		true	Ignoring selection	Selection model
Component 1	β_{10}	10.0	10.12	10.03
			(0.02)	(0.04)
	β_{11}	0.25	0.34	0.28
			(0.05)	(0.06)
	σ_1	1.0	0.95	0.98
			(0.02)	(0.03)
Compontent 2	β_{20}	10.0	10.23	10.07
			(0.04)	(0.08)
	β_{21}	−0.50	−0.42	−0.50
			(0.09)	(0.10)
	σ_2	1.0	0.95	0.97
			(0.03)	(0.04)
Probability of component 1	π	0.60	0.63	0.61
			(0.08)	(0.09)
Selection equation	γ_0	1.0	–	1.00
				(0.02)
	γ_1	−0.60	–	−0.60
				(0.02)
Copula dependence	θ	0.50	–	0.35
				(0.13)

Table 5: Sample means

	Hearing problem	No hearing problem
	n = 3,523	n = 294,541
Wage (if any)	34,636	42,870*
Any wage?	0.47	0.71*
Black	0.11	0.13*
Hispanic	0.14	0.18*
Married	0.44	0.60*
Metropolitan residence	0.74	0.82*
Age	49.9	43.7*
Less than high school (omitted)	—	—
Highest education is high school	0.01	0.01
Highest education is some college	0.31	0.29*
Highest education is college	0.21	0.35*
Number of children in household under 5	0.09	0.21*
Number of children in household 5 or older	0.63	0.97*

* indicates that "No hearing problem" mean differs from "Hearing problem" mean at $p < .05$

Table 6: Log wage regressions where wage > 0, no correction for selection

	OLS	Finite mixture	
		Comp 1	Comp 2
Hearing problem	**−0.253***	**−0.664***	**−0.075***
	(0.023)	**(0.090)**	**(0.017)**
Black	−0.061*	−0.028	−0.086*
	(0.006)	(0.026)	(0.004)
Hispanic	−0.136*	−0.035	−0.166*
	(0.006)	(0.024)	(0.004)
Married	0.002	−0.103*	0.035*
	(0.004)	(0.018)	(0.003)
Metropolitcan residence	0.178*	0.237*	0.177*
	(0.005)	(0.022)	(0.004)
Age	0.068*	0.112*	0.053*
	(0.002)	(0.007)	(0.001)
Age squared	−0.001*	−0.001*	−0.001*
	(0.000)	(0.000)	(0.000)
Education - high school	−0.186*	−0.136	−0.175*
	(0.022)	(0.086)	(0.016)
Education - some college	0.263*	0.212*	0.276*
	(0.005)	(0.022)	(0.004)
Education - college	0.771*	0.847*	0.748*
	(0.005)	(0.021)	(0.004)
Constant	8.180*	6.250*	8.737*
	(0.035)	(0.139)	(0.025)
Probability of component 1		0.206*	
		(0.002)	

Standard errors in parentheses, * $p < .05$

Table 7 – Selection model

	wage equation	selection equation
Hearing problem	**0.134***	– 0.428*
	(**0.024**)	(0.021)
Black	– 0.030*	– 0.073*
	(0.007)	(0.007)
Hispanic	– 0.068*	– 0.116*
	(0.007)	(0.006)
Married	0.113	– 0.109
	(0.005)	(0.005)
Metropolitcan residence	0.198*	– 0.016*
	(0.006)	(0.006)
Age	0.013*	0.077*
	(0.002)	(0.002)
Age squared	0.000	– 0.001*
	(0.000)	(0.000)
Education - high school	– 0.060*	– 0.106*
	(0.024)	(0.021)
Education - some college	0.037*	0.325*
	(0.006)	(0.006)
Education - college	0.429*	0.510*
	(0.006)	(0.006)
Number of children under 5	–	– 0.143*
		(0.004)
Number of children 5 or older	–	– 0.048*
		(0.002)
Constant	9.760*	– 0.910*
	(0.039)	(0.040)
ρ	– 0.928*	
	(0.001)	

Standard errors in parentheses, * p < .05

Table 8 – Copula-based finite mixture sample selection model

	wage equation		selection equation
	Comp 1	Comp 2	
Hearing problem	$\mathbf{-0.340}^{*}$	$\mathbf{0.098}^{*}$	-0.537^{*}
	(0.085)	**(0.019)**	(0.022)
Black	0.042	-0.075^{*}	-0.065^{*}
	(0.027)	(0.005)	(0.008)
Hispanic	0.052^{*}	-0.144^{*}	-0.115^{*}
	(0.024)	(0.005)	(0.007)
Married	0.027	0.081^{*}	-0.125^{*}
	(0.019)	(0.004)	(0.005)
Metropolitcan residence	0.242^{*}	0.183^{*}	-0.030^{*}
	(0.023)	(0.004)	(0.006)
Age	0.066^{*}	0.029^{*}	0.092^{*}
	(0.007)	(0.001)	(0.002)
Age squared	-0.001^{*}	-0.0002^{*}	-0.001^{*}
	(0.000)	(0.000)	(0.000)
Education - high school	-0.013	-0.138^{*}	-0.142^{*}
	(0.085)	(0.018)	(0.022)
Education - some college	-0.050^{*}	0.189^{*}	0.369^{*}
	(0.022)	(0.005)	(0.006)
Education - college	0.454^{*}	0.607^{*}	0.597^{*}
	(0.022)	(0.005)	(0.006)
Number of children under 5	–		-0.235^{*}
			(0.005)
Number of children 5 or older	–		-0.072^{*}
			(0.002)
Constant	8.026^{*}	9.403^{*}	-1.063^{*}
	(0.144)	(0.030)	(0.043)
σ	1.756^{*}	0.566^{*}	
	(0.011)	(0.002)	
Probability of component 1	0.240^{*}		
	(0.003)		
θ	-0.711^{*}		
	(0.007)		

Standard errors in parentheses, * $p < .05$

Table 9 – Component-specific means,
based on calculated component probabilities from estimates reported in Table 8

	Comp 1	Comp 2
Wage	45,721	6,281*
Black	0.13	0.13
Hispanic	0.16	0.21*
Married	0.58	0.62*
Metropolitcan residence	0.82	0.81*
Age	43.0	45.0*
Education - high school	0.01	0.02*
Education - some college	0.30	0.27*
Education - college	0.40	0.26*

* indicates that “Component 2” mean differs from “Component 1” mean at $p < .05$