A Multivariate Kernel Density Estimator for the Joint Distribution of Bequest Recipients [*]

Jason DeBacker[†]   Richard W. Evans[‡]   Parker Rogers[§]

Kerk L. Phillips[¶]

September 2015

(version 15.12.a)

Abstract

We estimate a flexible joint distribution of bequest recipients in the United States by age and by lifetime income group. We use a multivariate kernel density estimator with a flexible bandwidth parameter. In addition we provide a method with accompanying code to produce the resulting discrete joint distribution over arbitrary age and lifetime income bins.

keywords: multivariate kernel density estimation, bequests, inheritances, survey of consumer finances.

JEL classification: C14, C63, D31, D91, E21

[†]Middle Tennessee State University, Department of Economics and Finance, BAS N306, Murfreesboro, TN 37132, (615) 898-2528, jason.debacker@mtsu.edu.

[‡]Brigham Young University, Department of Economics, 167 FOB, Provo, Utah 84602, (801) 422-8303, revans@byu.edu.

[§]Brigham Young University, Department of Economics, 121B FOB, Provo, Utah 84602, parker.rogers2@gmail.com.

[¶]Brigham Young University, Department of Economics, 166 FOB, Provo, Utah 84602, (801) 422-5928, kerk_phillips@byu.edu.

1   Introduction

The current conventional wisdom among Americans is that inheritances propagate wealth inequality in the United States Wolff (2015). In order to accurately model and predict how inheritances affect wealth inequality, we must better understand which individuals are receiving these wealth transfers. Recent literature suggests that the relationship between age and inheritance reception rates is of the form of an inverted U-shape. More specifically, the rates of inheritance reception among age groups were, 16.2% for the youngest age group, 26.9% for ages 55-64, and 8.3% for ages 75+. Moreover, there is strong positive correlation between income level, and the percentage of households receiving wealth transfers Wolff (2015).

Currently, many dynamic macroeconomic models incorporate how inheritances are distributed using two methods. Each method represents an extreme assumption that is almost assuredly not true.

The first method of inheritance distribution gives all of the inheritances left from a dying individual to individuals who are in their same income group. This method fails to model many altruistic wealth transfers that occur in the economy as well as any distribution of that wealth to other income groups. For example, it fails to simulate individuals who give an altruistic inheritance to poorer family members who belong to a different income group. This hinders the accuracy of the model in predicting the effect of inheritances on wealth inequality, and may lead to incorrect predictions of the effects of certain policy implementations.

The second method of inheritance distribution splits the inheritances from all individuals equally among all income groups. This method assumes that all income groups receive equal inheritances, which does not accurately portray inheritance behavior according to macroeconomic data.

The purpose of this paper is to estimate a function that describes the proportions of inheritances received among different age and income groups. This function can later be incorporated into dynamic macroeconomic models to simulate the effects of inheritances on wealth inequality in the United States. We will use a Multivariate

Kernel Density Estimator (MVKDE) to fit a function to our sample, and we will provide flexibility in our estimator by allowing a desired number of income and age groups. Our sample reflects the behavior of Americans by using data extracted from the Survey of Consumer Finances.

2   Survey of Consumer Finances data

The data we used was pulled from the Survey of Consumer Finances located on The Federal Reserve website. The Survey of Consumer Finances is a triennial survey conducted by The Federal Reserve that began in 1983. Each survey data set contains around 5,400 variables with over 30,000 respondents. Summary variables are included that are adjusted for inflation.of Governers of the Federal Reserve System (2015b).

Another possible data source for our purpose is The Panel Survey of Income Dynamics (PSID). The PSID is a cross-sectional survey conducted by The University of Michigan that began in 1968 and follows 18,000 individuals. You can find data on employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics. Insititute for Social Research (2015)

Here are the variables that we used from The Survey of Consumer Finances when determining our proportion matrix,

X5804 - What the approximate value of the inheritances were at the time the respondent received them.

X5809 - The second reported value of inheritance reception.

X5814 - The third reported value of inheritances reception.

X5805 - The year that they received the first instance of inheritances (corresponding to X5804).

X5810 - The year that they received the second instance of inheritances (corresponding to X5809)

X5815 - The year that they received the third instance of inheritances (corresponding to X5814)

X8022 - age of respondent.

wgt - An SCF summary variable that contains the weights provided by the Federal Reserve to adjust for selection bias in the survey.

networth - An SCF summary variable that contains the self-perceived net worth of the respondent at the time of the survey.

Since the SCF is a triennial survey, we wanted to prevent any overlap of bequest reception between survey years. In order to do this, we used the SCF variables "X5805", "X5810", "X5815" as described above. These variables give the time that each instance of bequest was received, and allow us to make sure that we include only the bequests received in the current, and previous two years of the survey year. For example, the "X5804" instance of bequests was received on the year "X5805" , and if

$$\text{``X5805''} - (\text{current survey year}) < 3$$

then we would include that particular instance of bequests.

To gather the amount of bequests received, conditional upon both age and ability types (networth), we separated the bequests reported by respondents within three years of the survey, by separating the bequests received by certain age and ability types into separate bins. After, we divided the amount in each age-income bin by the total amount of bequests reported within the three year interval of the survey for all respondents. This resulted in a proportion matrix, which tells us what proportion of bequests each age-income category received. Upon examining the following resulting graph of proportions over the years 2001-2013, we proceeded to apply a different variable provided by the SCF called "weights".

**Figure 1: Distribution of Inheritances Before Weights Applied**
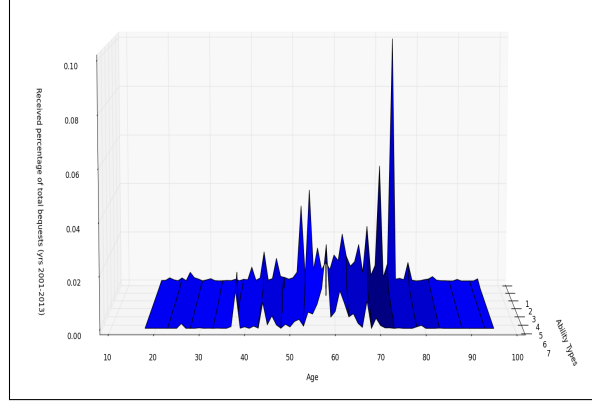
In order to reduce selection bias, the SCF provides a Summary Variable called "weights". The weights provided are meant to "compensate for unequal probabilities of selection in the original design and for unit nonresponse (failure to obtain an interview)" of Governers of the Federal Reserve System (2015a). The distribution before these weights were applied is seen in Figure 1. We applied these weights to all our variables by multiplying the weights by each bequest value.

Lastly, we adjusted the nominal values for bequests to account for inflation between the used survey years of 1998-2013. We used the St. Louis Federal Reserve's CPI for all urban consumers: all items of St. Louis (2015) . We then set 2013 as the base year and calculated the percentage inflation factors for all the years down to 1988. The inflation factors, which will be multiplied by the bequest amounts that correspond to each year, are given as follows,

1987	1988	1989	1990	1991	1992	1993
1.51229385	1.492299173	1.467973318	1.439144582	1.415496948	1.397721517	1.379834479
1994	1995	1996	1997	1998	1999	2000
1.363737434	1.345889029	1.326679888	1.310939123	1.300280732	1.284934882	1.260857994
2001	2002	2003	2004	2005	2006	2007
1.240039148	1.227912707	1.210171616	1.189103802	1.161807505	1.134803101	1.109966432
2008	2009	2010	2011	2012	2013	
1.076012397	1.078969961	1.063898833	1.034477726	1.014431538	1	

The final results showing the distribution of bequests received over the years 1998-2013, after applying our weights and inflation factors, are as follows,

Figure 2: Distribution of Inheritances 1998-2013
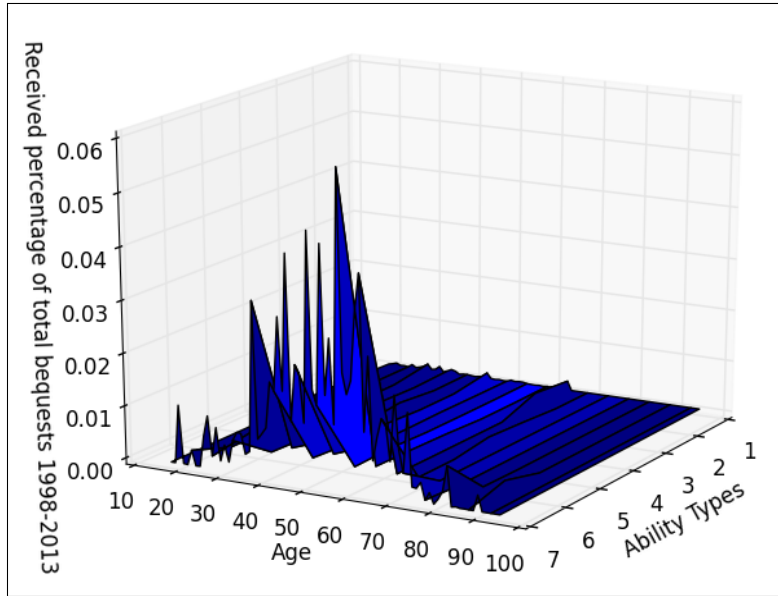
where the ability types one through seven are different net worth groups categorized using the following net worth amounts:

1: Under $15,000,

2: $15,000 - $24,999,

3: $25,000 - $49,999,

4: $50,000 - $74,999,

5: $75,000 - $99,999,

6: $100,000 - $250,000,

7: $250,000 or more

Notice the peak of inheritance reception at the age group 55-65 years and ability type 7. This suggests that most inheritances fall on middle-aged, wealthy individuals.

3   Estimated Joint Distribution of Bequests

After gathering the data on the proportion of inheritances, we needed to fit a functional form to our data set in order to have a smooth representation of our sample and a robust way of altering both age groups and income groups according to different user inputs. For example, if the user wanted only four ability-type cohorts and twenty age cohorts, then we need a way to create new proportion bins according to these specified conditions. If we were to use only our sparse non-linear data set, it would be difficult to compose the desired proportions for each age-ability category. However, a functional form allows us to impose any conditions on the bequest distribution that the user desires. This is why we chose to fit a Multivariate Kernel Density Estimator to our data set.

3.1   Multivariate Kernel Density Estimator (MVKDE)

Kernel Density Estimation is a non-parametric way to approximate a density function when the parametric form of the sampling density is unknown. This proved to be useful in our approximation of the underlying density function of our data in Figure 2. We used a Multivariate Kernel Density Estimator to estimate the sample density as there were no clear parametric density forms that matched our data set. We put forward both the univariate and the multivariate Kernel Density Estimators, and display our Python code used to execute this estimation.

The Univariate Kernel Density Estimator is defined as follows:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$$

where $K_h(x) = K(x/h)/x$ is the kernel function of choice, h is the bandwidth, and n is the number of sample points Scott (2015). Much like the histogram, a Kernel Density Estimator, creates functional values that correspond to a chosen kernel function at each sample point. Moreover, upon summing up each kernel function corresponding to its respective sample point, this sum of kernels is evenly scaled by dividing the functional values by the total number of sample points and the bandwidth selection.

The multivariate case of kernel estimation is given by the following,

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{x}_i)$$

In our case, we used a Gaussian Kernel , which yields the following equation for this general Multivariate Kernel Density Estimator for $\mathbf{x} \in \mathbb{R}^d$

$$\hat{f}(\mathbf{x}) = \frac{1}{n2\pi^{d/2}|\Sigma|^{1/2}}\sum_{i=1}^{n} \exp\left[-\frac{1}{2}(x - x_i)^T \Sigma^{-1}(x - x_i)\right]$$

David W. Scott (2015)

In order to implement our Multivariate Kernel Density Estimator we used the following python code,

```python
import numpy as np
from matplotlib import pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from scipy import stats
from scipy.stats import norm
from scipy.stats import kde

def MVKDE(s, e, filename, plot = False):
    '''
    Generates a Multivariate Kernel Density Estimator and returns a matrix
    representing a probability distribution according to given age categories,
    and ability type categories.
```

```
13
14    Inputs:
15        s                    = scalar, the number of age groups in the model.
16
17        e                    = scalar, the number of ability type groups in the model.
18
19        filename             = string, the file name of the .txt document that contains
20                                the original proportion matrix created by SCFExtract.py.
21
22        plot                 = boolean, whether or not you want a plot of the probability
23                                distribution generated by your given age and ability type
24                                groups.
25
26    Functions called:
27      kde.gaussian_kde       = scipy function that generates a Kernel
28                                Density Estimator from given data
29
30    Objects in function:
31        proportion_matrix    = [78, 7], array containing the proportion (0 < x < 1) of
32                                the total bequests that each age−income category
      receives.
33                                Derived in SCFExtract.py
34
35      age_probs              = [78,], array containing the frequency, or how many times,
36                                that random drawn numbers fell into the 78 different
      age bins
37
38      income_probs           = [7,], array containing the frequency, or how many times,
39                                that random drawn numbers fell into the 7 different
      ability
40                                type bins
41
42      age_frequency          = [70000,], array containing repeated age values for each
43                                age number at the frequency given by the age_probs
      vector
44
45      xmesh                  = complex number, the number of age values that will be
46                                evaluated in the Kernel Density Estimator.
47
48      freq_mat               = [70000, 2], array containing age_frequency and
49                                income_frequency stacked
50
51      density                = object, class given by scipy.stats.gaussian_kde.
52                                The Multivariate Kernel Density Estimator for the given
      data set.
```

```
53
54     age_min , age_max        = scalars , the minimum and maximum age values and minimum
55        income_min , income_max   and maximum income values
56
57
58     agei                = [s , e] , array containing the age values to be evaluated in
59                                   the Kernel Estimator ( ranging from 18−90)
60
61     incomei             = [s , e] , array containing the income values to be evaluated
62                                   in the Kernel Estimator ( ranging from 1−7)
63
64     coords                  = [2 , s∗e] , array containing the raveled values of agei
65                                   and incomei stacked
66
67     estimator               = [s , e] , array containing the new proportion values for
68                                   s age groups and e ability type groups that are
       evaluated
69                                   using the Multivariate Kernel Density Estimator
70
71     estimator_scaled        = [s , e] , array containing the new proportion values for
72                                   s age groups and e ability type groups that are
       evaluated using
73                                   the Multivariate Kernel Density Estimator , but scaled
       so that
74                                   the sum of the array is equal to one.
75
76     Returns : estimator_scaled
77     '''
78
79     proportion_matrix = np.loadtxt ( filename , delimiter = ',')
80     proportion_matrix_income = np.sum( proportion_matrix , axis = 0)
81     proportion_matrix_age = np.sum( proportion_matrix , axis = 1)
82     age_probs = np.random.multinomial (70000 , proportion_matrix_age )
83     income_probs = np.random.multinomial (70000 , proportion_matrix_income )
84     age_frequency = np.array ([])
85     income_frequency = np.array ([])
86     age_mesh = complex( str ( s)+'j')
87     income_mesh = complex( str ( e)+'j')
88
89     j = 18
90     '''creating a distribution of age values '''
91     for i in age_probs :
92        listit = np.ones ( i )
93        listit ∗= j
94        age_frequency = np.append( age_frequency , listit )
```

9

```
95        j+=1

96

97     k = 1
98     '''creating a distribution of ability type values'''
99     for i in income_probs:
100        listit2 = np.ones(i)
101        listit2 *= k
102        income_frequency = np.append(income_frequency, listit2)
103        k+=1

104

105    freq_mat = np.vstack((age_frequency, income_frequency)).T
106    density = kde.gaussian_kde(freq_mat.T, bw_method=0.3)
107    age_min, income_min = freq_mat.min(axis=0)
108    age_max, income_max = freq_mat.max(axis=0)
109    agei, incomei = np.mgrid[age_min:age_max:age_mesh, income_min:income_max:
          income_mesh]
110    coords = np.vstack([item.ravel() for item in [agei, incomei]])
111    estimator = density(coords).reshape(agei.shape)
112    estimator_scaled = estimator/float(np.sum(estimator))
113    if plot == True:
114        fig = plt.figure()
115        ax = fig.gca(projection='3d')
116        ax.plot_surface(agei,incomei, estimator_scaled, rstride=5)
117        ax.set_xlabel("Age")
118        ax.set_ylabel("Ability Types")
119        ax.set_zlabel("Received percentage of total bequest")
120        plt.show()
121    return estimator_scaled

122

123 newproportion = MVKDE(80,7,'fileout.txt', True)

124
```

Listing 1: MVKDE.py

We used the kernel function from the imported scipy module for our estimation. This kernel function works much like a histogram, therefore we changed the nature of our data from a matrix of proportions to vectors filled with the frequencies that different age and ability type groups occurred. This transformation occurs on lines 80-103.

Using the multinomial function from numpy, we loaded the function with our proportion vectors for age and ability type. We obtained these vectors by summing up our proportion matrix over both the age and then the ability type axes. This resulted

in two vectors that summed to one, which provide the proportions that each age group and ability type group received. The multinomial function then uses those proportions as probabilities and takes random draws according to those given proportions. We specified the function to take 70,000 random draws with the given probabilities that correspond to the age and ability proportions. For our age probability vector, this creates a new vector where each index corresponds to an age group in ascending order (18-95), where each index contains the number of times that a random draw resulted in that particular age. For our ability probability vector, multinomial creates a similar vector in ascending order (1-7), with the number of draws that fell into into each ability type group/index.

After creating these vectors we created two new vectors that transform the data set from proportions to age numbers and ability type numbers. The frequency that each age and ability type number occurs is determined by how many times the random draw from our multinomial function resulted in each age and ability type respectively. For example, if our "age_probs" vector contains 1,876 at index 0, then this means that age 18 was drawn 1,876 times. This means that we would put 1,876 instances of "18" in our new vector called "age_frequency". This is done similarly with our "income_probs" and "income_frequency" vectors.

These new vectors now correspond to the required data format for scipy's kernel function. We then create the appropriate mesh grid to be entered into the "density" estimator function. Finally we scale the estimate to sum to one, and graph the resulting estimator.

3.2   Distribution of bequests

Our results for a Multivariate Kernel Density Estimator are illustrated in the following graph,

Figure 3: MVKDE of Distribution of Inheritances 1998-2013
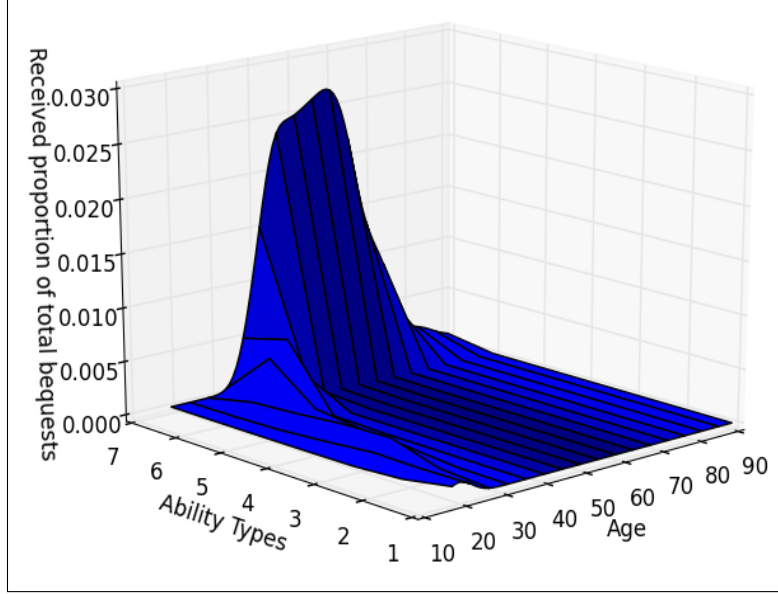
Note that the peak of bequest reception still lies in the age group 55-65 years and ability type 7 like Figure 2. Our estimated distribution is ultimately defined by a matrix of proportions whose dimensions are determined by the user. In this particular case, we specified 78 age groups (ages 18-95) and 7 ability type groups. Any desired amount of age groups and ability type groups can be given, and the resulting matrix will match these amounts. This flexibility allows our distribution to change and alter according to the different specifications in macroeconomic models.

Also, a noteworthy characteristic of the MVKDE the use of a bandwidth parameter. This parameter allows for the graph to be more or less smooth. Smoothness corresponds to higher values of the "bw_method" in the "kde.gaussian_kde" function, and represents a distribution that is less determined by the data. The lower the bandwidth selection, the rougher the graph is and the more jagged the distribution; however, the lower bandwidths more accurately portray the behavior of the sample set.

4 Conclusion

Our estimation of a bequest distribution allows for a data-driven allocation of inheritances among different age and ability type groups. This allows for more heterogeneity among agents within macroeconomic models, and can ultimately lead to more accurate results. Previous methods of bequest allocation in dynamic macroeconomic models are most assuredly not true. The assumption that inheritances are distributed equally among different income groups, or that they are distributed only to individuals who belong to the same income group as those giving the inheritances, is not true according to our sample. By using a more accurate distribution of inheritances we can better predict the effects that inheritance taxes have on the distribution of wealth.

By modeling our distribution using a MVKDE, we are able to have flexible age and ability categories. If a macroeconomic model is defined by 3-period lived agents, then one can input the need for three distinct age categories, and the resulting estimated proportion matrix will be returned with three rows representing the three age groups. Similarly, if a model defines five ability type groups in the economy, our estimator can return a new matrix of proportions with five columns of ability types. This flexibility will allow this estimator to be used in a wide variety of macroeconomic models.

Even though we were able to find data in The Survey of Consumer Finances regarding individuals who receive bequests by both age and income, we were unable to find data that links these individuals with those who are giving bequests. It would be useful to know which individuals, by age and income categories, are giving bequests so that we can also accurately model this component of bequest distribution.

References

of Governers of the Federal Reserve System, Board, "Files: Survey of Consumer Finances Codebook," 2015.

_ , "Research Resources: Survey of Consumer Finances About," 2015.

of Michigin Insititute for Social Research, University, "PSID The Panel Study of Income Dynamics - PSID - is the longest running longitudinal household survey in the world.," 2015.

of St. Louis, Federal Reserve Bank, "FRED Economic Data Index for All Urban Consumers: All Items," 2015.

Scott, David W., *Kernel Density Estimators, in Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley and Sons, Inc., 2015.

Scott, Stephan R. Sain David W., *9-Multidimensional Density Estimation*, John Wiley and Sons, Inc., 2015.

Wolff, Edward N., *Inheriting Wealth in America: Future Boom or Bust?*, Oxford University Press, 2015.