Appendices for

"The Evolution of Uncertainty in Labor Earnings in the U.S. Economy"

Flavio Cunha and James Heckman

September 30, 2007

Contents

List of Tables

List of Figures

Web Supplement I

Description of the Data

We use white males age 22 to 36 from National Longitudinal Survey of the Youth - 1979 (NLSY/1979) and National Longitudinal Sample of the Young Men - 1966 (NLS/1966)[1]. In the original NLSY/1979 sample there are 3379 white males. We impose the following sample restrictions in this order. From the original population, we drop the 1153 individuals who are in an oversample of poor people, who are in the military sample, or who die in the survey period. We discard 212 individuals because parental education is missing. We restrict the NLSY/1979 sample to white males with a high school or college degree. We define high school graduates as individuals who have a high school degree or have completed 12 grades and do not report ever attending college. We define participation in college as having a college degree or having completed at least 16 years in school. These rules produce a sample of 1,360 individuals.

Tuition at age 17 is average tuition in colleges in the county of residence at 17. If there is no college in the county, average tuition in the state is used instead. For details on the construction of this variable see Cameron and Heckman (2001).

In 1980, NLSY/1979 respondents were administered a battery of ten achievement tests referred to as the Armed Forces Vocational Aptitude Battery (*ASVAB*) (see Cawley, Conneely, Heckman, and Vytlacil, 1997, for a complete description of these tests). The math and verbal components of the *ASVAB* can be aggregated into the Armed Forces Qualification Test (AFQT) scores.[2] Many studies have used the overall AFQT score as a regressor, arguing that this is a measure of scholastic ability. We argue that AFQT is an imperfect proxy for scholastic ability because of measurement error. Accounting for measurement error in test scores has a substantial effect on our estimates. We also avoid a potential aggregation bias by using the cognitive components of the *ASVAB* as a separate measure of ability. In our empirical work, we use five components of the ASVAB test battery: arithmetic reasoning, word knowledge, paragraph comprehension, math knowledge and coding speed.

We use annual labor earnings in our analysis. We extract this variable from the NLSY/1979 reported annual earnings from wages and salary. Earnings (in thousands of dollars) are discounted to 2000 using the Consumer Price Index reported by the Bureau of Labor Statistics. Missing values for this variable occur for two reasons: first, because respondents do not report earnings for wages/salary; and second, because the NLSY becomes a biannual survey after 1994 and this prevents us from observing respondents when they reach certain ages. For example, because the NLSY/1979 was not conducted in 1995, we do not observe individuals born in 1964 when they are 31 years old.

In the original NLS/1966 sample there are 3734 white males who are born between 1942 and 1952. We drop 316 individuals who die or are dropped from the sample in the survey period. We also drop 627

[1]For a description of the NLSY/1979 see Miller (2004). For a description of the NLS/1966 see the questionaries and codebook supplement available at http://www.nlsinfo.org/ordering/display_db.php3

[2]Implemented in 1950, the AFQT score is used by the army to screen applicants.

individuals missing information on parental education. We only keep the white males who are high-school or college graduates, according to the definition given for NLSY/1979. This results in a sample size of 1872 individuals from NLS/66 .

The construction of variables in both data sets is comparable except for the construction of region of residence at age 14 (reported in NLSY/1979, but not in NLS/1966) and the tests of aptitude. To generate Southern residence at age 14 for the NLS sample, we proceed as follows. First, everyone who is born in 1952 and lives in the South in 1966 is assigned South14 = 1 (others are assigned the value 0). These are the only people we can establish residence in the South region at age 14 without making any assumptions. Second, we look at the location of the last high school that the respondents attended. Third, we look for people who live in the South in 1966 and report that they were born in the same Census division as they live in 1966. We assign South14 = 1 if South1966 = 1 and the person was born in the same division that he now lives. If a person does not live in the South in 1966 and was born in the same division that he lives then he is assigned South14 = 0. The assumption is made that a person who lives in the same division as she was born in did not grow up in a different region of the country. Fourth, for all remaining respondents we assign the value South14 = 0 (respectively South14 = 1) if South1966 = 0 (respectively South1966 = 1) and the year of birth is 1948 or later. Summary statistics for both samples are presented in Tables I-1 and I-2.

In the NLS/1966 there are many different achievement tests. We use the two most commonly reported ones: the OTIS/BETA/GAMMA and the California Test of Mental Maturity (CTMM). One problem in the NLS/1966 sample is that there are no respondents for whom we observe scores on two (or more) achievement tests. That is, for each respondent we observe at most one test score. Test scores are standardized on a common scale. We supplement the information from these test scores by using other proxies for cognitive achievement. These are the tests on ”Knowledge of the World of Work”. There are three different tests. The first is a question regarding occupation: the respondent is asked about the duties of a given profession, say draftsman. For this specific example, there are three possible answers: (a) makes scale drawings of products or equipment for engineering or manufacturing purposes, (b) mixes and serves drinks in a bar or tavern, (c) pushes or pulls a cart in a factory or warehouse. The second test asks the respondent to provide the level of education associated with each occupation in the first test. The third test is an earnings comparison test. Specifically, it asks the respondent who he/she believes makes more in a year, given two distinct occupations. In Table I-3 we show that even after controlling for parental education, number of siblings, family income in 1966, urban residence at age 14, Southern residence at age 14, and dummies for year of birth, the ”Knowledge of the World of Work” test scores are correlated with the cognitive test scores. The correlation with OTIS/BETA/GAMMA and CTMM is stronger for the occupation and education tests than for the earnings-comparison test.

Web Data Appendix Table I-1

Summary Statistics – NLSY/79

Earnings in Ten Thousand Dollars CPI-adjusted (base year 2000)

Variable	High School Sample			College Sample		
	Observations	Mean	Standard Error	Observations	Mean	Standard Error
Mother's Education	873	3.8522	1.2087	487	5.1417	1.5324
Father's Education	873	3.8958	1.5112	487	5.7207	1.8526
Family Income in 1979	873	1.9871	1.0928	487	2.6511	1.5744
Number of Siblings	873	3.1123	1.9527	487	2.4682	1.6785
South Residence at age 14	873	0.2314	0.4220	487	0.2423	0.4289
Urban Residence at age 14	873	0.6850	0.4648	487	0.8029	0.3982
Local Tuition at 4-year college[2]	873	0.2207	0.0851	487	0.2075	0.0772
Enrolled at School at ASVAB Test Date	845	0.4414	0.4969	481	0.9252	0.2634
Age at ASVAB Test Date	873	19.4479	2.1957	487	19.3265	2.2375
Highest Grade Completed at ASVAB Test Date	845	11.0686	1.2534	481	12.0395	2.1939
ASVAB - Arithmetic Reasoning[3]	822	0.3557	0.8994	472	1.3030	0.7218
ASVAB - Word Knowledge[3]	822	0.2894	0.7704	472	0.9775	0.4106
ASVAB - Paragraph Composition[3]	822	0.1663	0.8438	472	0.8867	0.4275
ASVAB - Coding Speed[3]	822	0.0037	0.8187	472	0.6195	0.7469
ASVAB - Math Knowledge[3]	822	0.0625	0.8197	472	1.4034	0.7034

[1]The sample consists of white males born between 1957 and 1964 who are high school or college graduates

[2]In ten thousand dollars. The tuition figures are inflation-adjusted using the CPI. The base year is 2000.

[3]Not available for PSID respondents.

Web Data Appendix Table I-2

Summary Statistics – Pooled NLS/1966 and PSID[1]

Variable	High School			College		
	Observations	Mean	Standard Error	Observations	Mean	Standard Error
Mother's Education	1089	3.2617	1.1681	783	4.4074	1.5988
Father's Education	1089	3.0248	1.3781	783	4.5045	1.8962
Family Income in 1966	1089	6.9275	1.8853	783	8.1507	1.7534
Number of Siblings	1089	3.0781	2.1715	783	2.3078	1.7609
South Residence at age 14	1085	0.3005	0.4587	781	0.2676	0.4430
Urban Residence at age 14	1089	0.6630	0.4729	783	0.7995	0.4006
Local Tuition at 4-year college[2]	1089	0.1592	0.0189	783	0.1578	0.0192
Otis/Beta/Gamma Test[3]	226	-0.4235	0.7907	167	0.7500	0.7525
California Test of Mental Maturity[3]	164	-0.3170	0.8868	98	0.6843	0.7958
Work Knowledge, Occupations[3]	1070	0.0766	0.8111	776	0.6366	0.7632
Work Knowledge, Education[3]	1075	0.0815	0.8365	778	0.5993	0.7556
Work Knowledge, Earnings Comparison[3]	1082	0.0058	0.9422	781	0.3833	0.9342

[1]The sample consists of white males born between 1942 and 1952 who are high school or college graduates

[2]In ten thousand dollars. The tuition figures are inflation-adjusted using the CPI. The base year is 2000.

[3]Not available for PSID respondents.

Web Data Appendix Table I-3
Raw Correlation of Test Scores from NLS/1966[1]

	Otis/Beta/Gamma Test	California Test of Mental Maturity	Work Knowledge, Occupations	Work Knowledge, Education and Occupation	Work Knowledge, Earnings Comparison
Otis/Beta/Gamma Test	1.0000	N/A[2]	0.4064	0.4149	0.0752
California Test of Mental Maturity	N/A[2]	1.0000	0.2719	0.2586	0.0848
Work Knowledge, Occupations	0.4064	0.2719	1.0000	0.9358	0.4244
Work Knowledge, Education	0.4149	0.2586	0.9358	1.0000	0.0782
Work Knowledge, Earnings Comparison	0.0752	0.0848	0.4244	0.0782	1.0000

[1]We control for mother's and father's education, family income in 1966, south residency at age 14, urban residency at age 14, and birth dummies.
[2]Individuals report either Otis/Beta/Gamma or the California Test of Mental Maturity, but not both.

Web Supplement II

Properties of the Model

Web Data Appendix Table II-1
Normalizations on Factor Loadings
NLSY/79 and NLS/66

	High School Earnings Equations			College Earnings Equations		
Age	Loading on Factor 1	Loading on Factor 2	Loading on Factor 3	Loading on Factor 1	Loading on Factor 2	Loading on Factor 3
22		1.00	0.00			0.00
23			0.00			0.00
24			1.00			
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						

[1]The empty cells correspond to factor loadings that are estimated, not normalized.

Web Data Appendix Table II-2

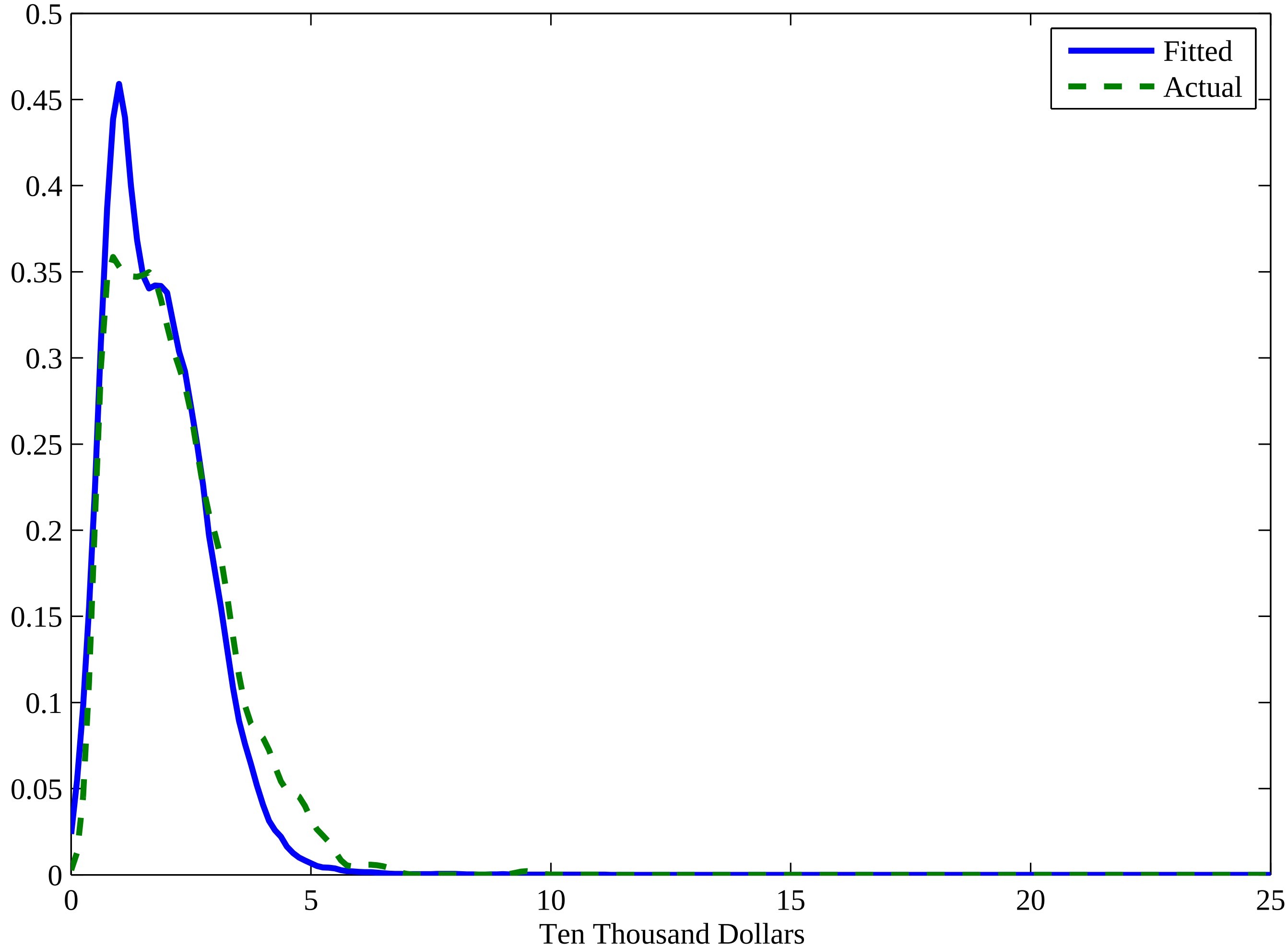
χ^2 Goodness of Fit Test*

	NLSY/66 - White Males				NLSY/1979 - White Males			
	High School		College		High School		College	
Period	χ^2 statistic	Critical Value	χ^2 statistic	Critical Value	χ^2 statistic	Critical Value	χ^2 statistic	Critical Value
22	20.1081	28.8693	11.3674	15.5073	57.6139	37.6525	17.0958	16.9190
23	31.2115	31.4104	18.6951	21.0261	47.3941	38.8851	36.2178	21.0261
24	45.3552	35.1725	29.9623	26.2962	39.5717	40.1133	71.0567	22.3620
25	29.8860	33.9244	19.8444	26.2962	33.4427	38.8851	50.3284	24.9958
26	25.9980	31.4104	17.9291	26.2962	39.9348	41.3371	24.4273	24.9958
27	47.5875	31.4104	22.8861	26.2962	30.9588	41.3371	23.2935	24.9958
28	19.7162	27.5871	16.2158	23.6848	28.1225	40.1133	23.2251	24.9958
29	33.3359	26.2962	60.8435	23.6848	23.7535	40.1133	24.5953	24.9958
30	21.4356	26.2962	24.7761	23.6848	33.6092	37.6525	12.6124	22.3620
31	20.6412	22.3620	32.1826	21.0261	19.7494	37.6525	18.7615	21.0261
32	18.7421	19.6751	16.7482	16.9190	12.2899	31.4104	11.9100	19.6751
33	14.6120	18.3070	15.9341	16.9190	22.5333	31.4104	14.3782	18.3070
34	9.1826	12.5916	12.0096	14.0671	22.1509	27.5871	11.6416	16.9190
35	8.2744	9.4877	4.3654	9.4877	14.3759	27.5871	32.9993	15.5073
36	3.2788	7.8147	0.7022	7.8147	12.2162	22.3620	6.1417	12.5916

* 95% Confidence, equiprobable bins with aprox. 25 people per bin. A χ^2 statistic lower than the critical value indicates a "good" fit.

Web Data Appendix Figure II-1.1

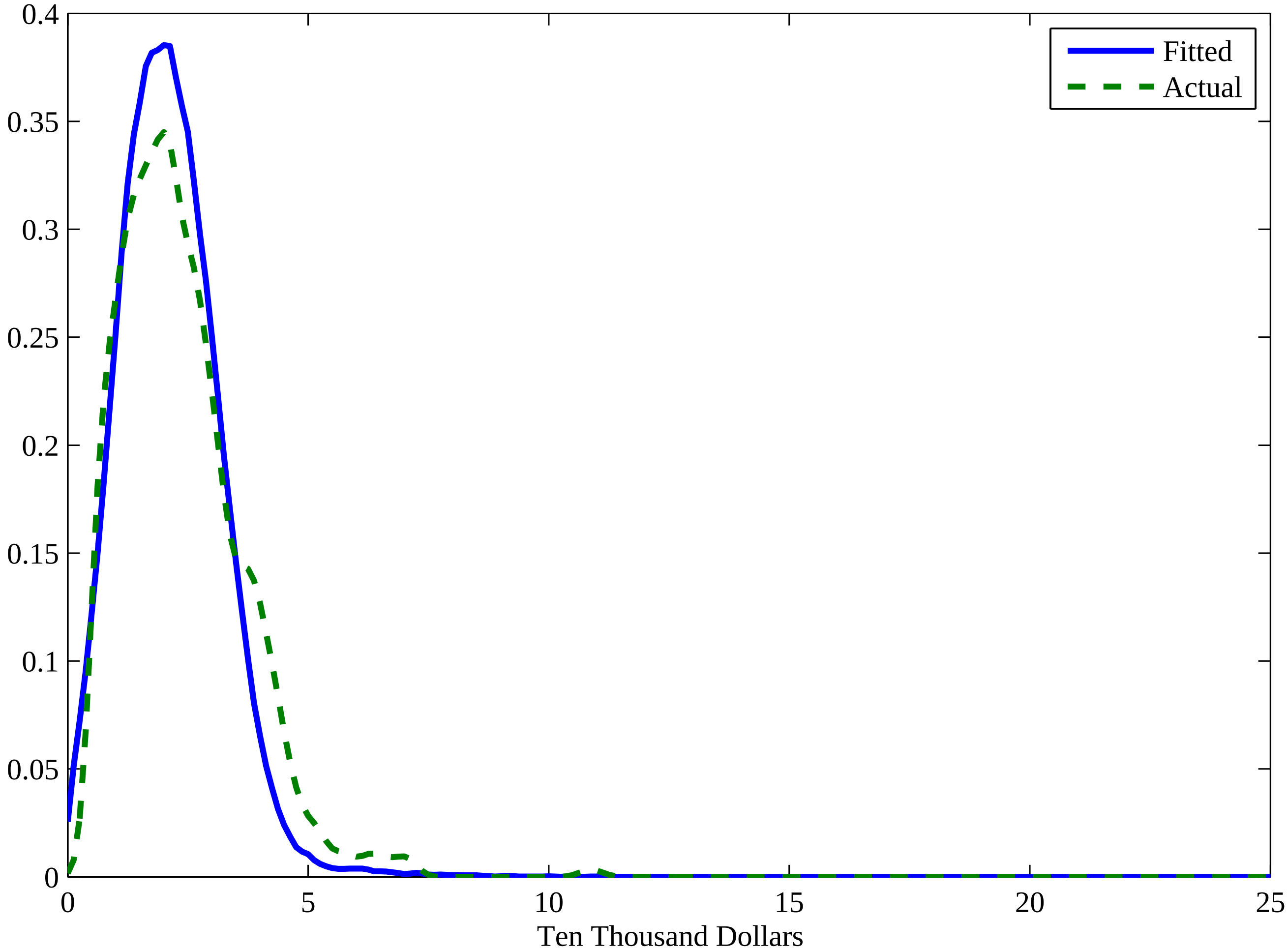
Densities of earnings at age 22 Overall Sample NLSY/79



Let Y denote earnings at age 22 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

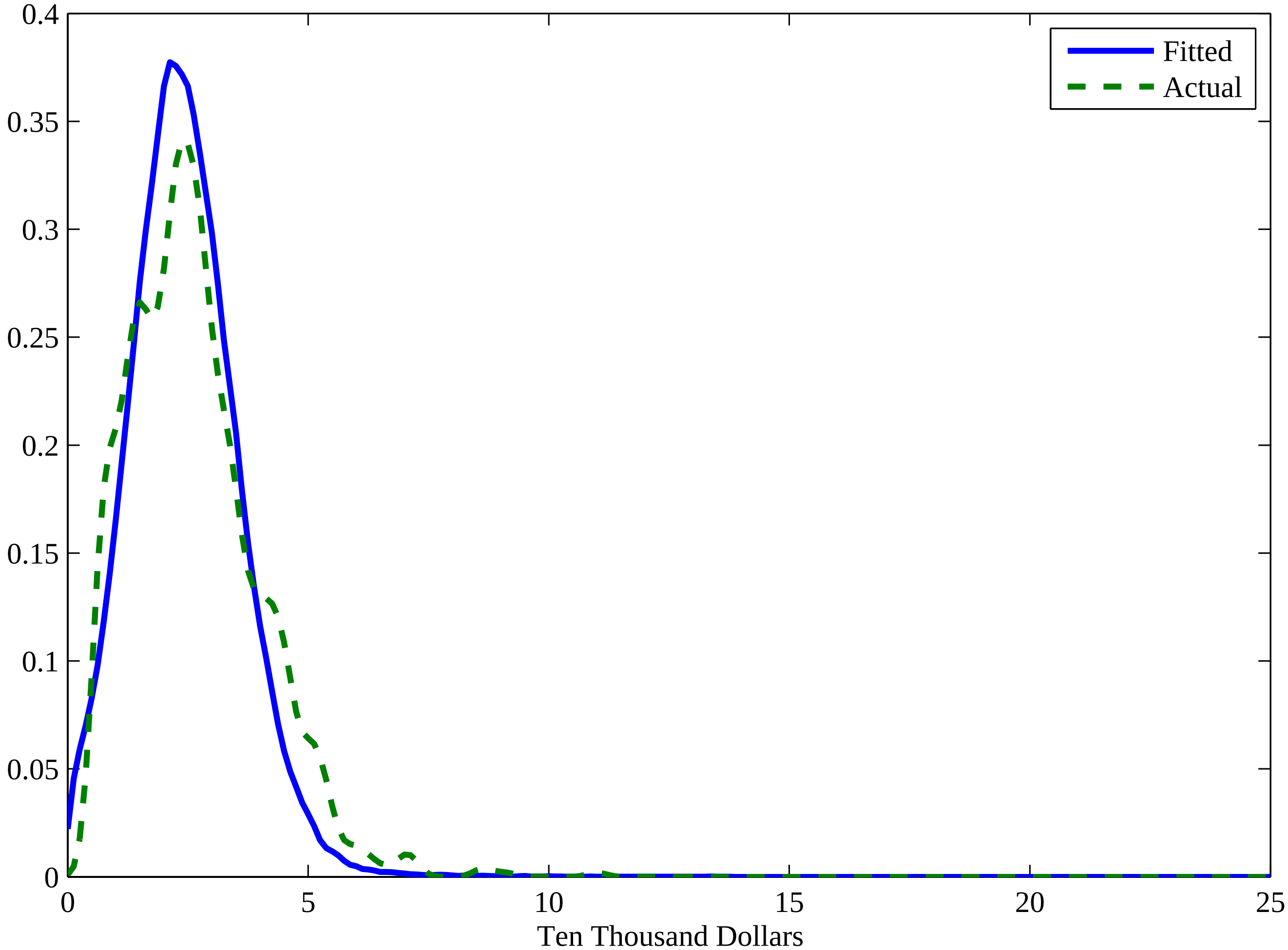
Web Data Appendix Figure II-1.2

Densities of earnings at age 23 Overall Sample NLSY/79



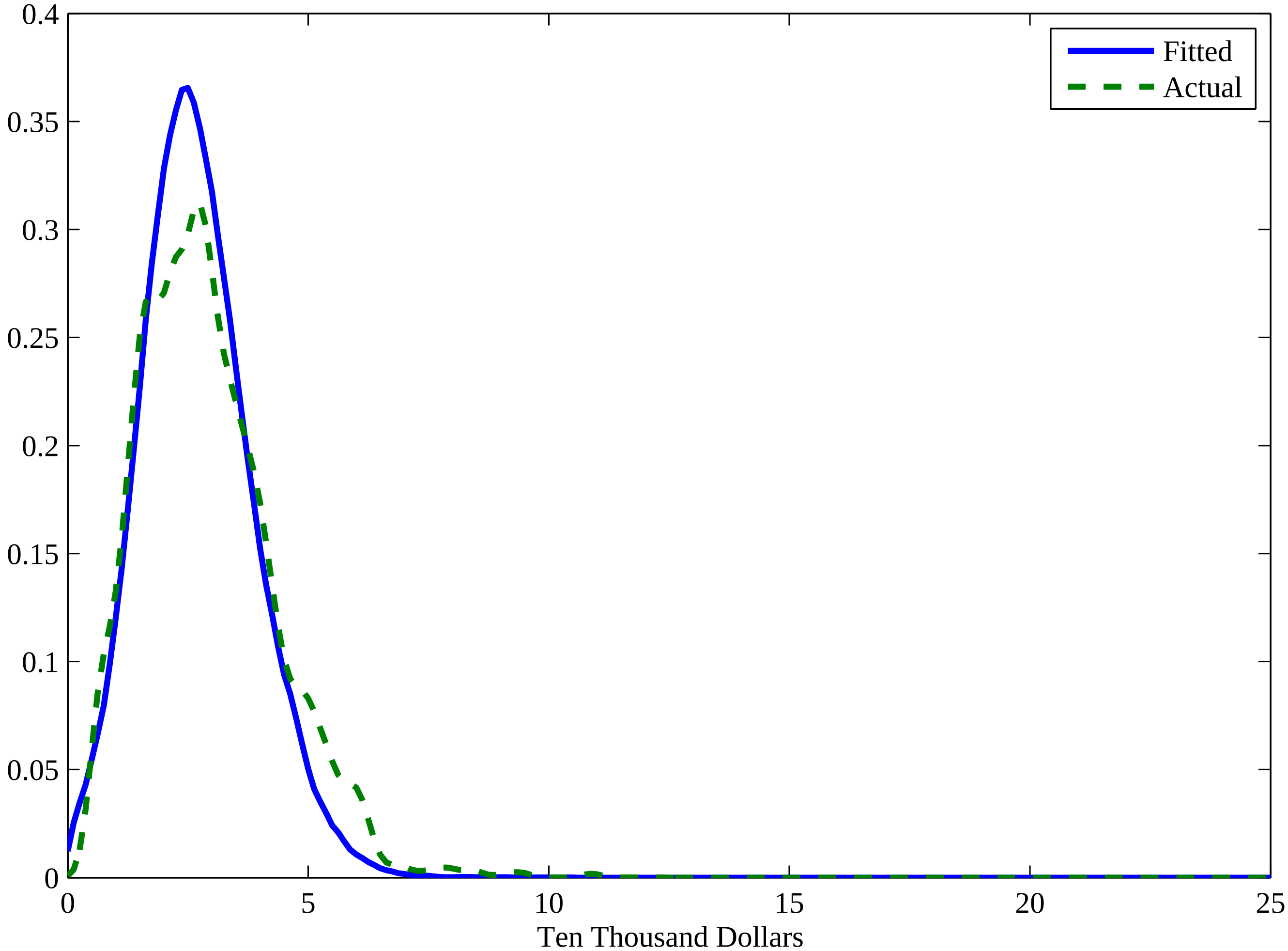
Let Y denote earnings at age 23 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.3
Densities of earnings at age 24 Overall Sample NLSY/79



Let Y denote earnings at age 24 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

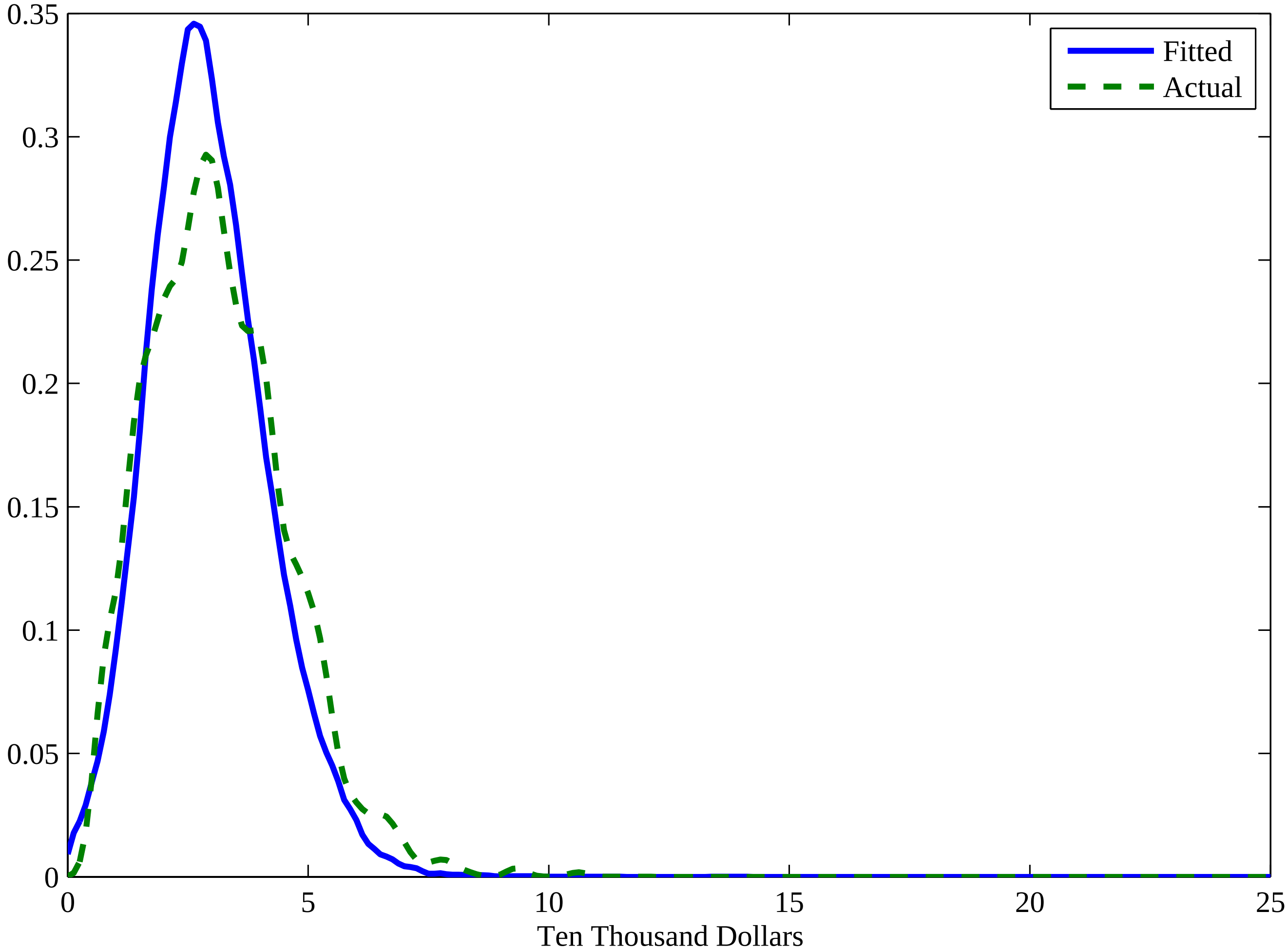
Web Data Appendix Figure II-1.4
Densities of earnings at age 25 Overall Sample NLSY/79



Let Y denote earnings at age 25 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.5

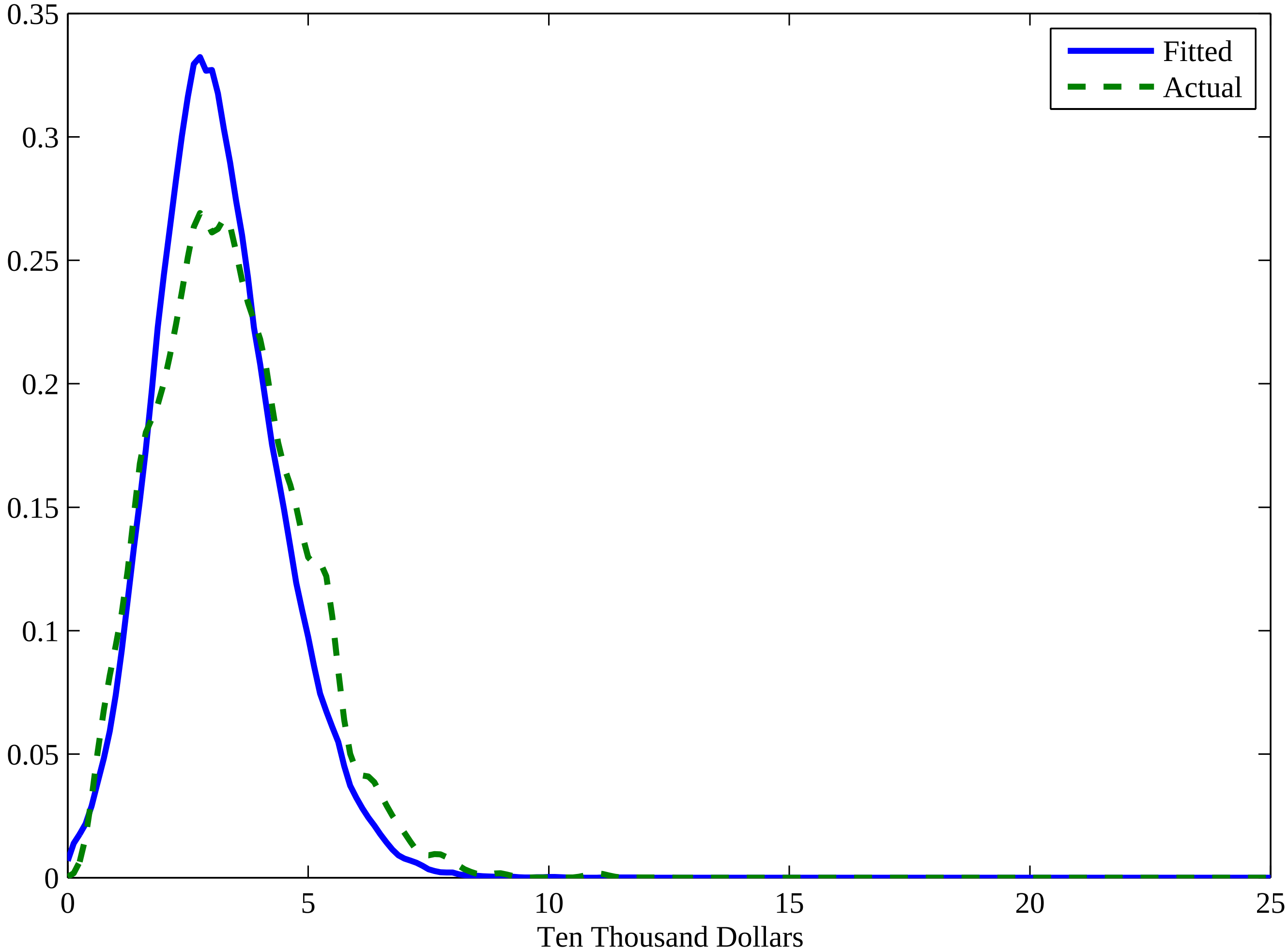
Densities of earnings at age 26 Overall Sample NLSY/79



Let Y denote earnings at age 26 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.6

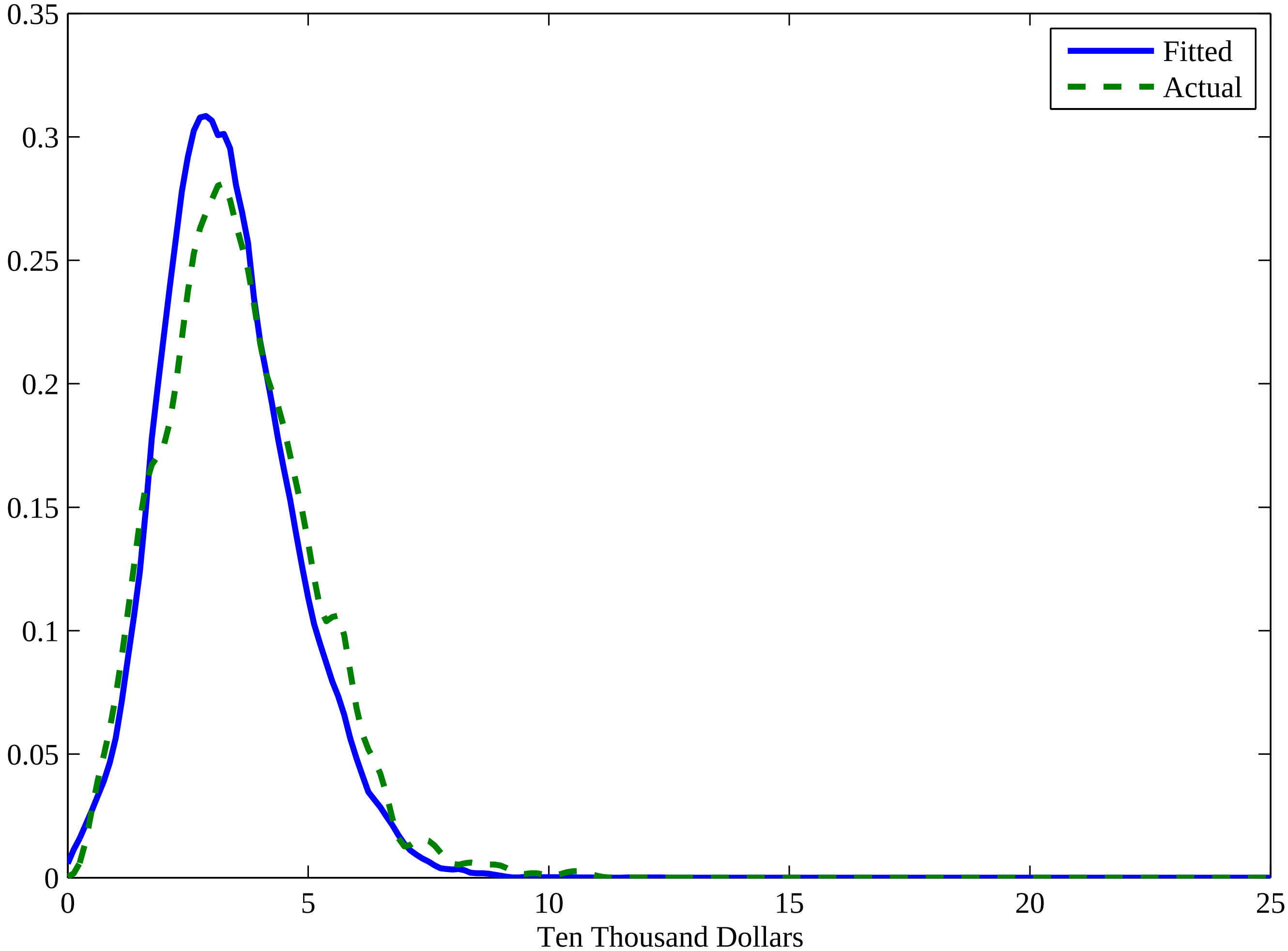
Densities of earnings at age 27 Overall Sample NLSY/79



Let Y denote earnings at age 27 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.7

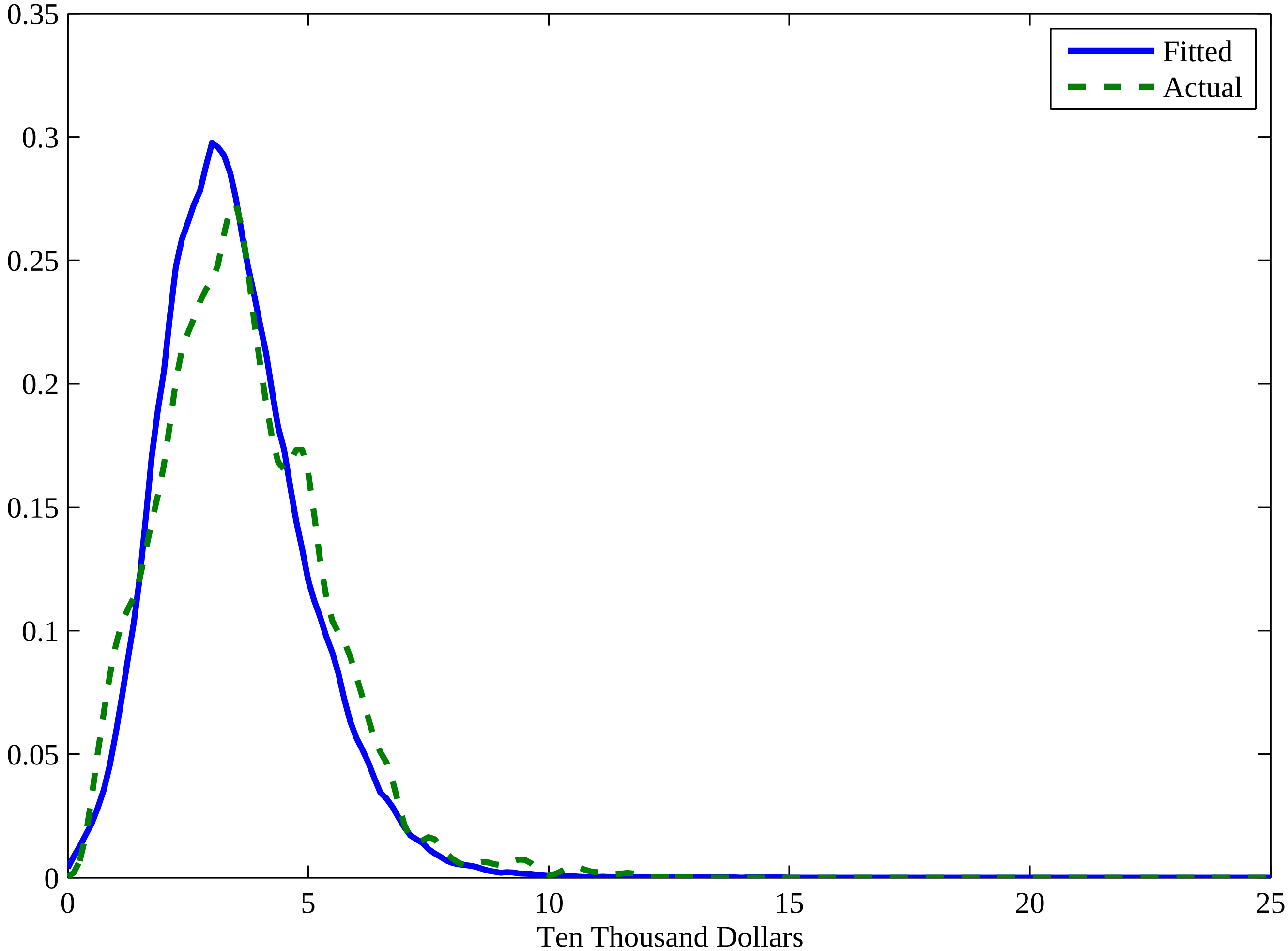
Densities of earnings at age 28 Overall Sample NLSY/79



Let Y denote earnings at age 28 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

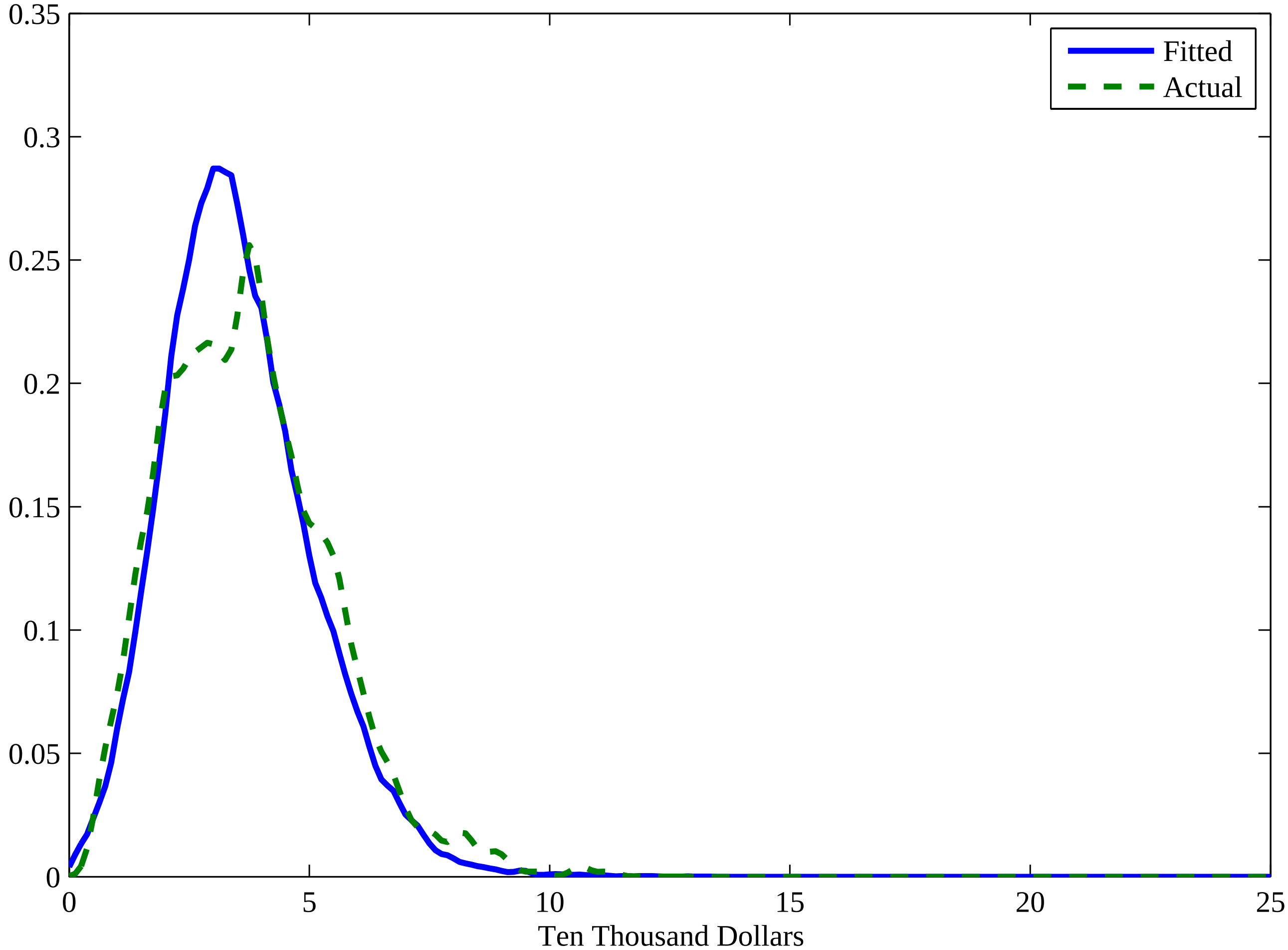
Web Data Appendix Figure II-1.8

Densities of earnings at age 29 Overall Sample NLSY/79



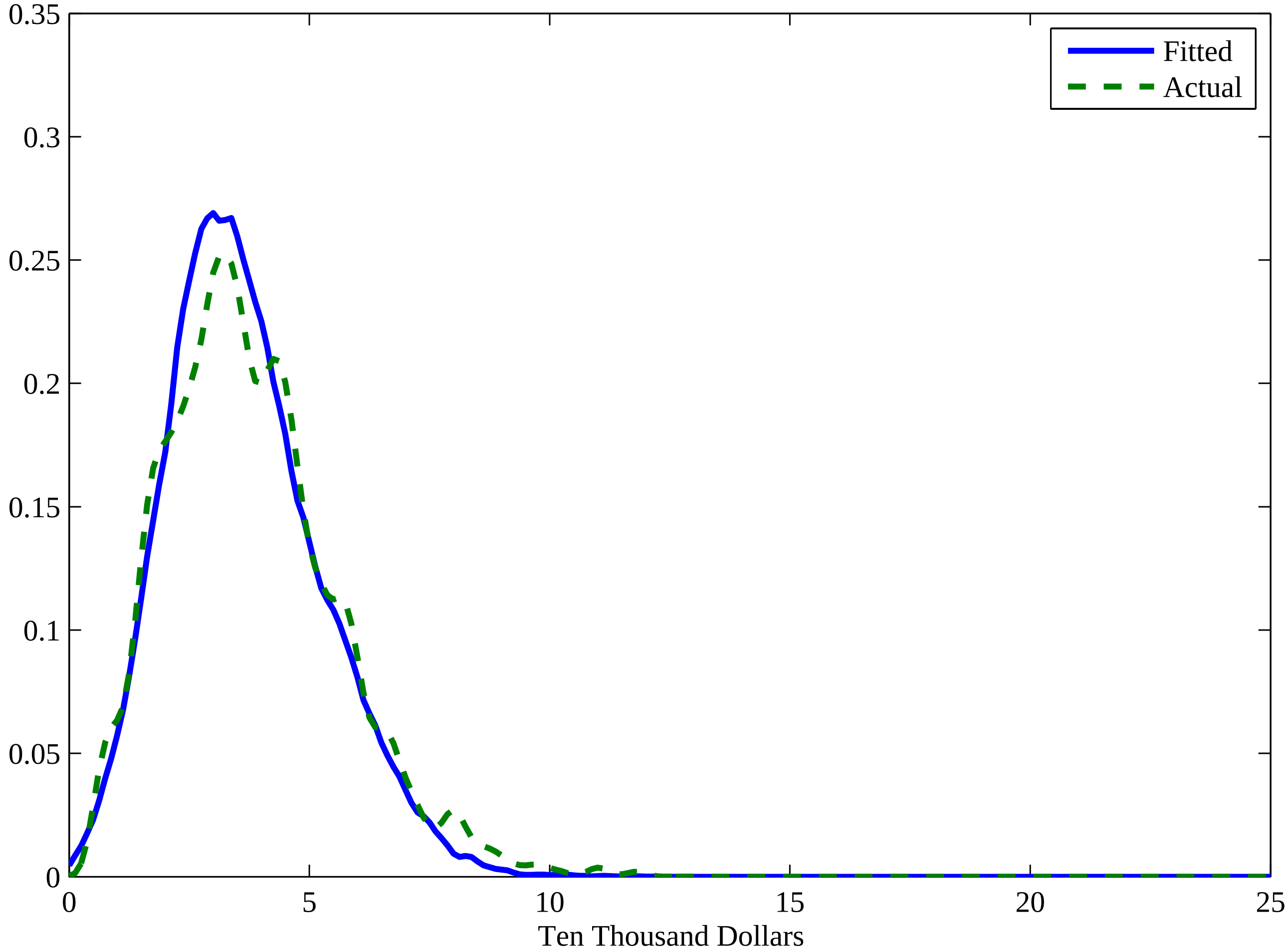
Let Y denote earnings at age 29 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.9
Densities of earnings at age 30 Overall Sample NLSY/79



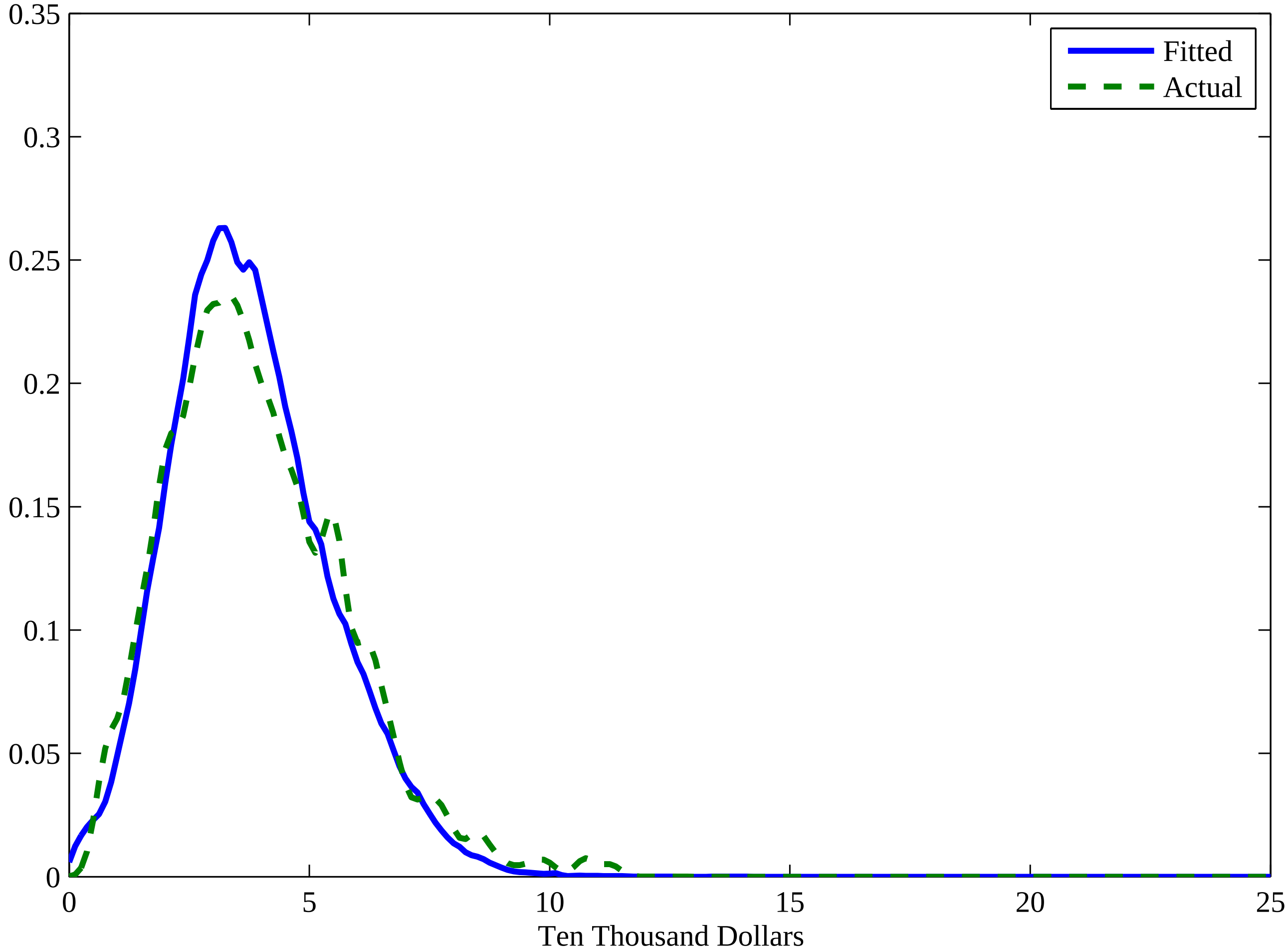
Let Y denote earnings at age 30 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.10
Densities of earnings at age 31 Overall Sample NLSY/79



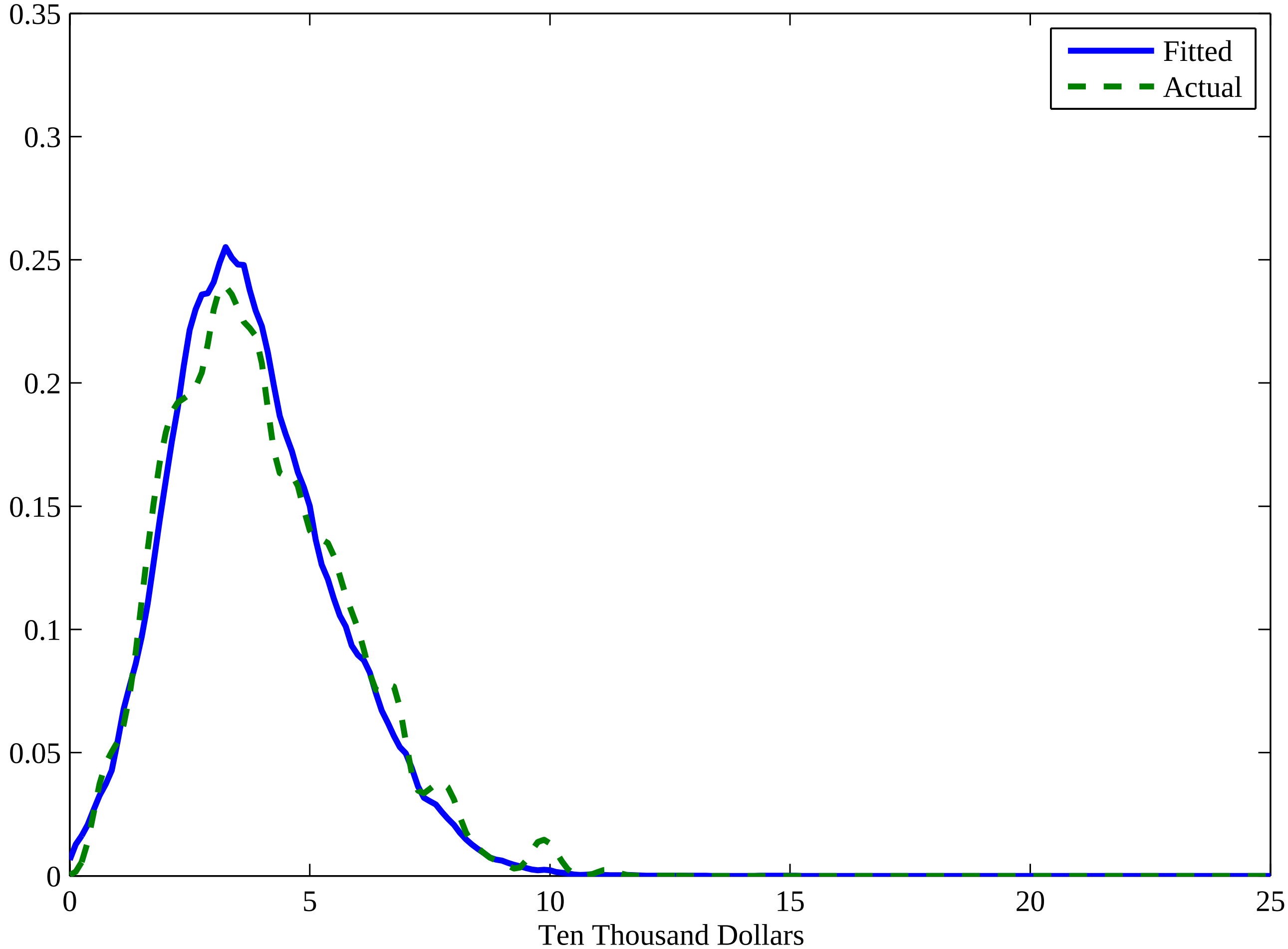
Let Y denote earnings at age 31 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.11
Densities of earnings at age 32 Overall Sample NLSY/79



Let Y denote earnings at age 32 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

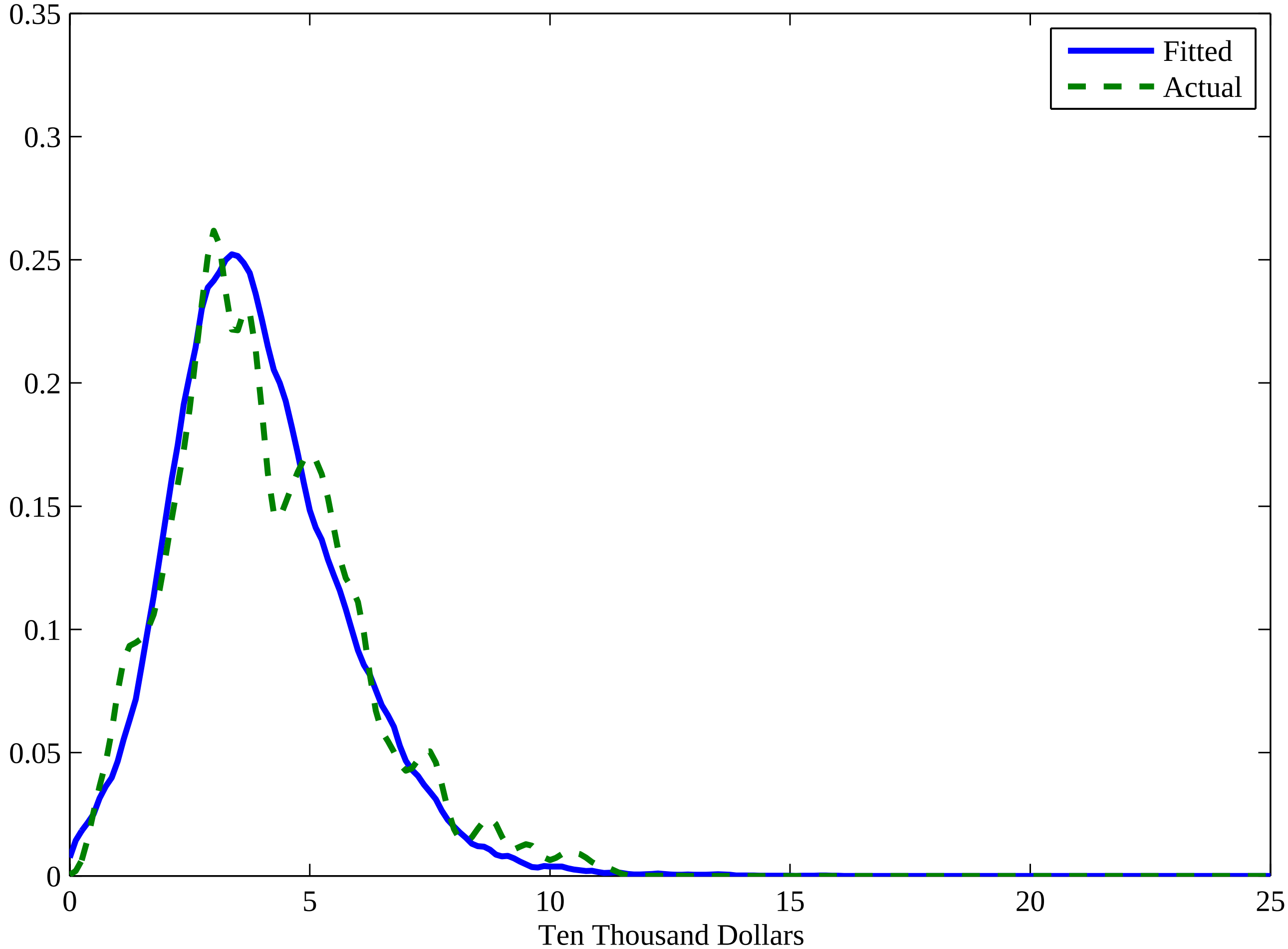
Web Data Appendix Figure II-1.12
Densities of earnings at age 33 Overall Sample NLSY/79



Let Y denote earnings at age 33 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.13

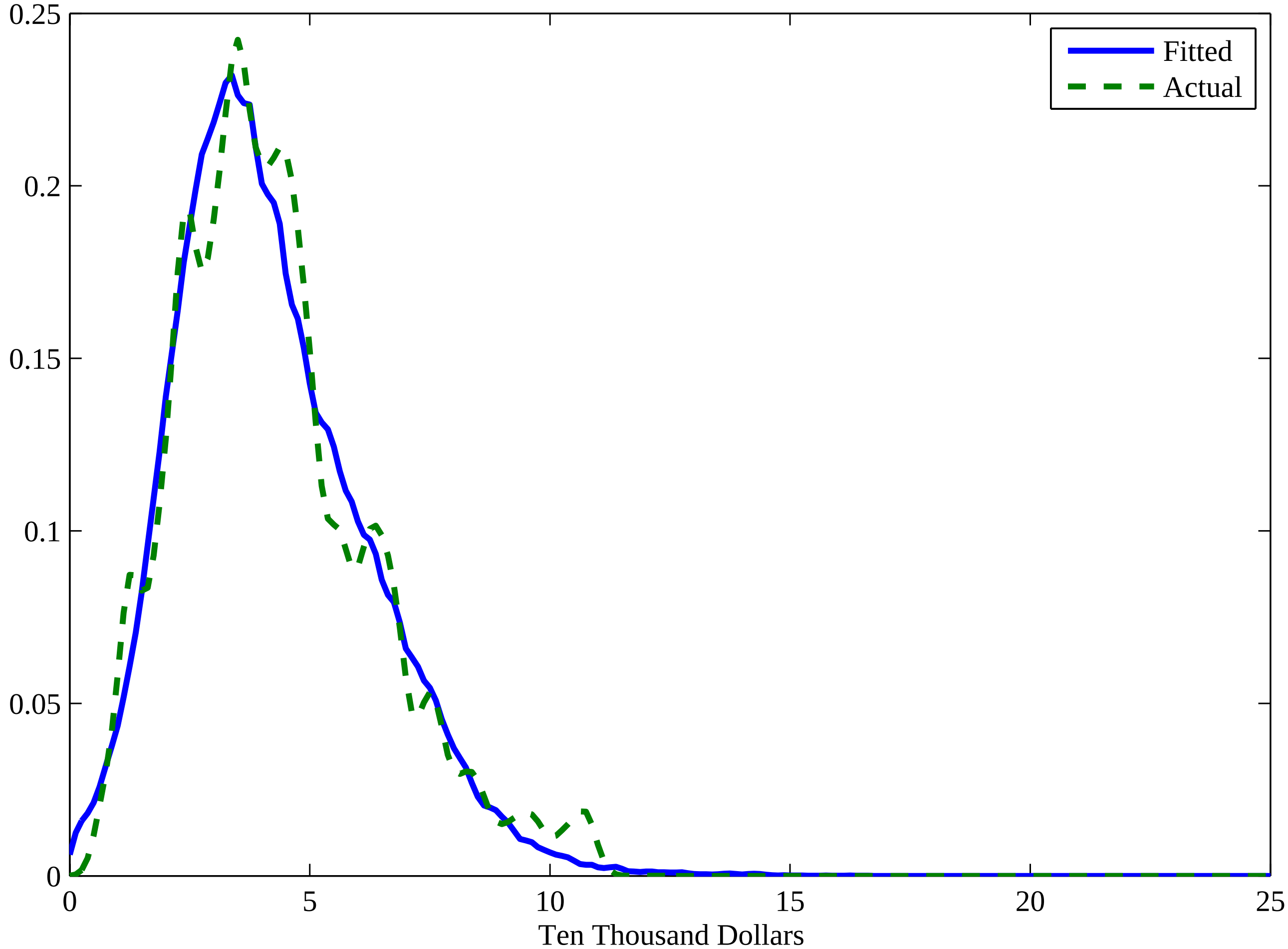
Densities of earnings at age 34 Overall Sample NLSY/79



Let Y denote earnings at age 34 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

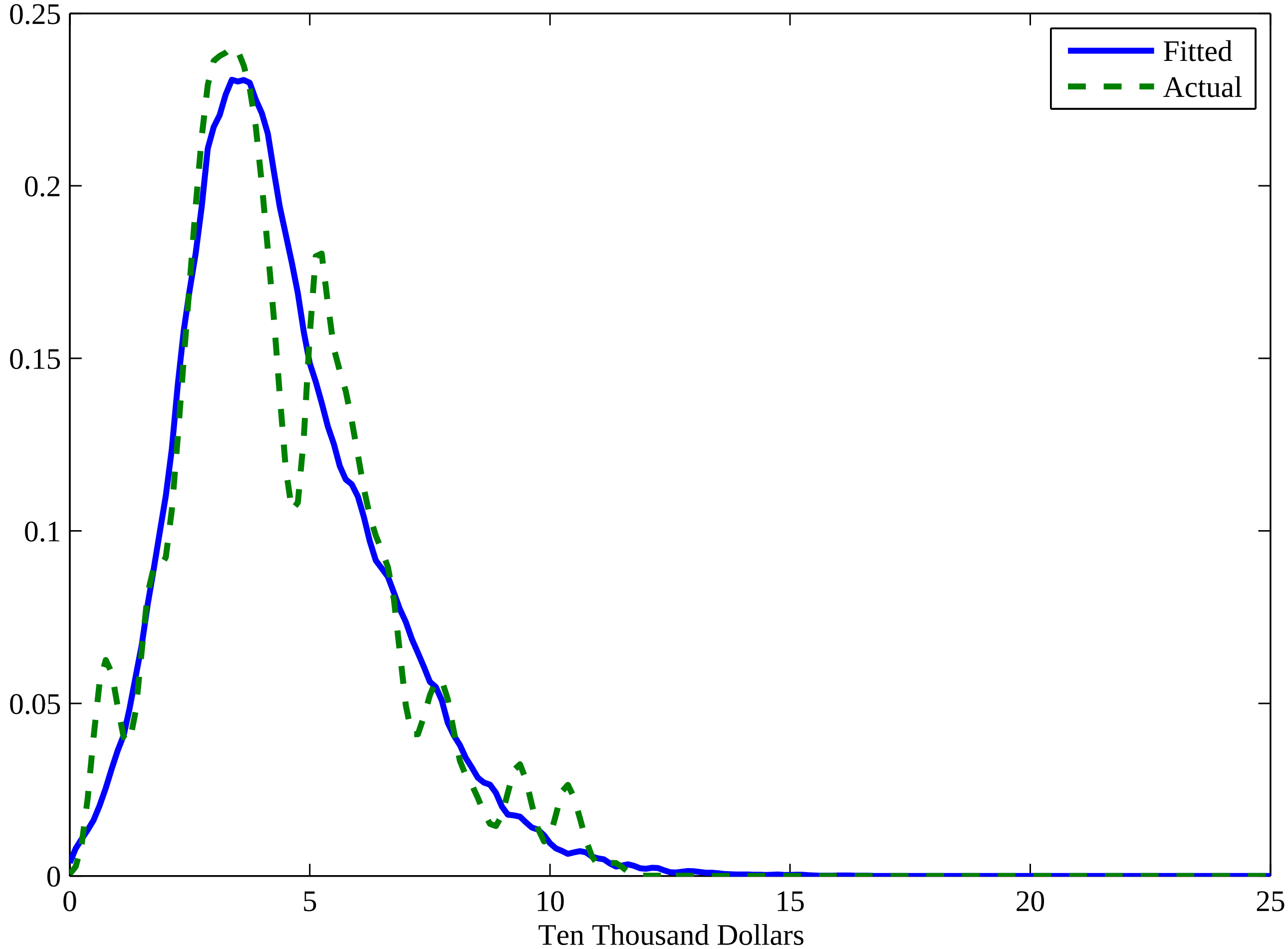
Web Data Appendix Figure II-1.14

Densities of earnings at age 35 Overall Sample NLSY/79



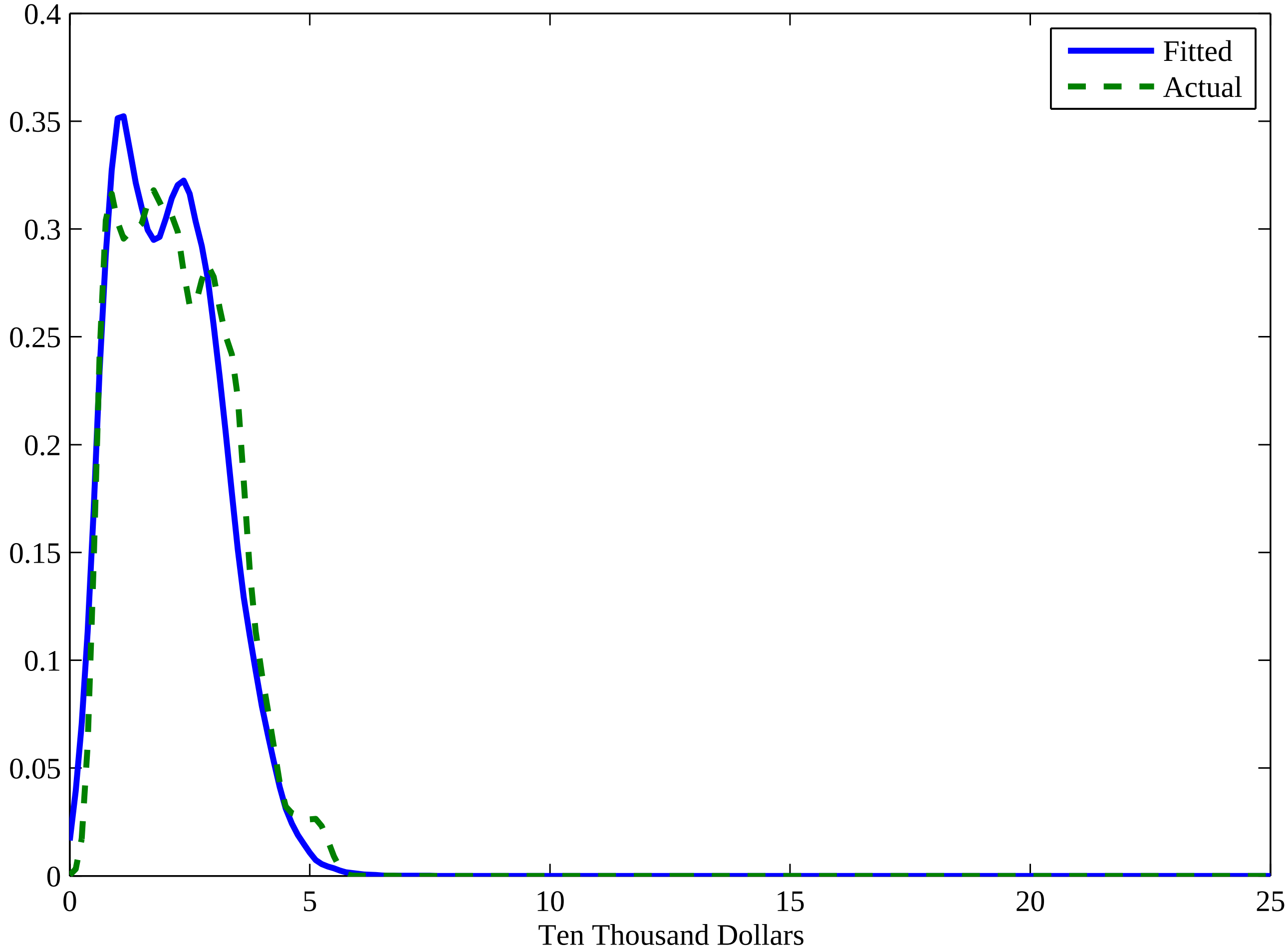
Let Y denote earnings at age 35 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.15
Densities of earnings at age 36 Overall Sample NLSY/79



Let Y denote earnings at age 36 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

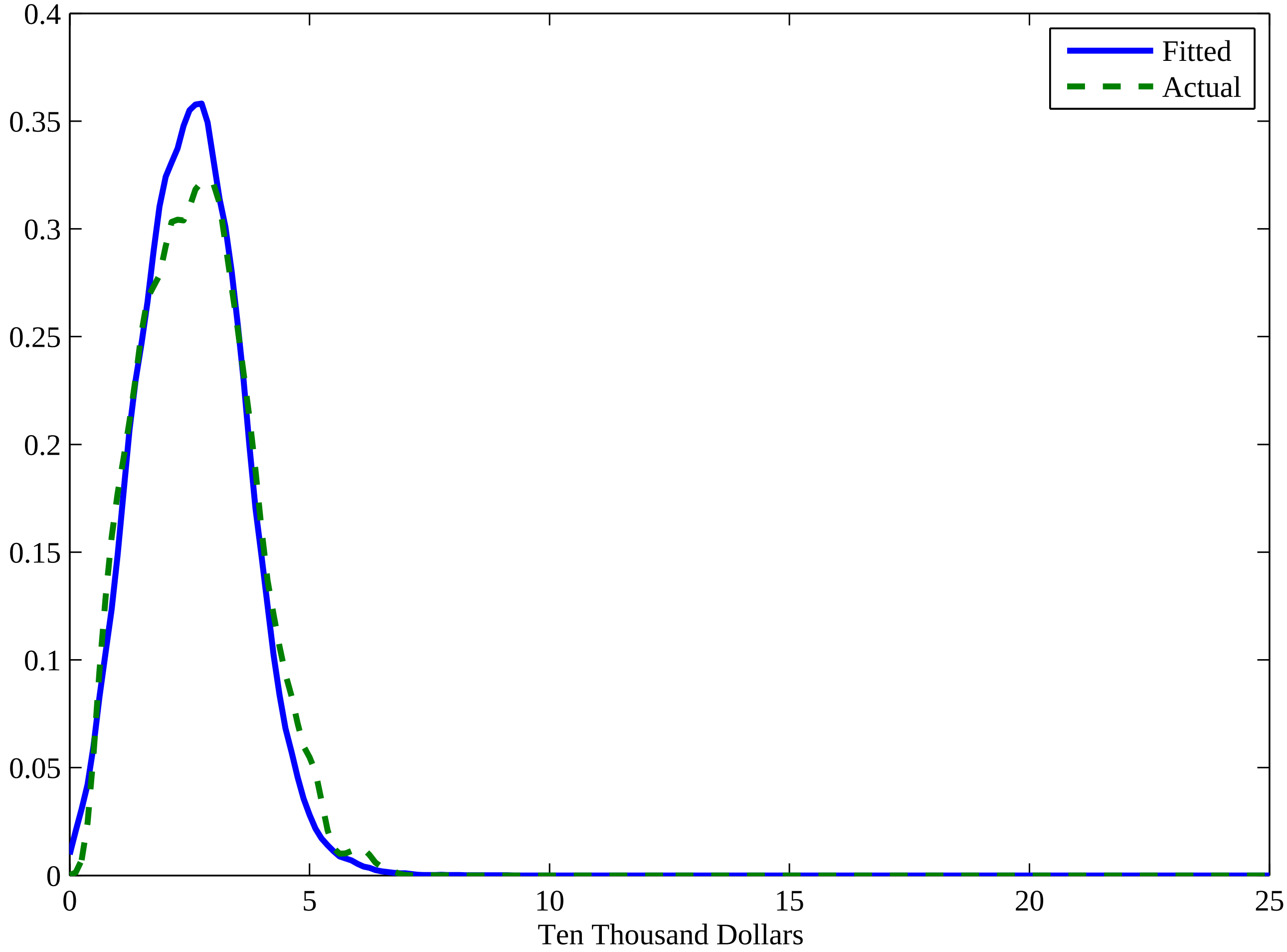
Web Data Appendix Figure II-1.16
Densities of earnings at age 22 Overall Sample NLS/66



Let Y denote earnings at age 22 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

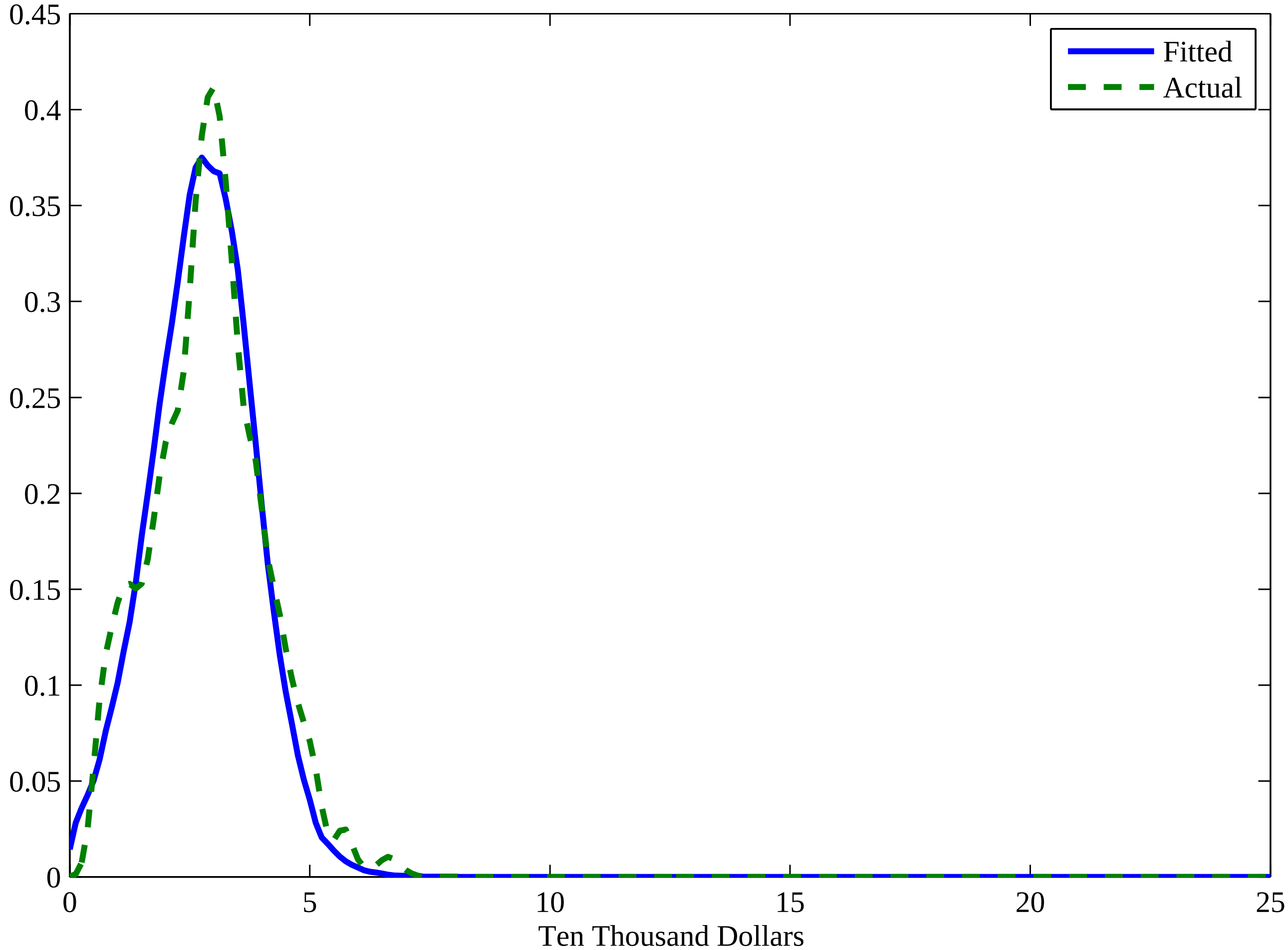
Web Data Appendix Figure II-1.17

Densities of earnings at age 23 Overall Sample NLS/66



Let Y denote earnings at age 23 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

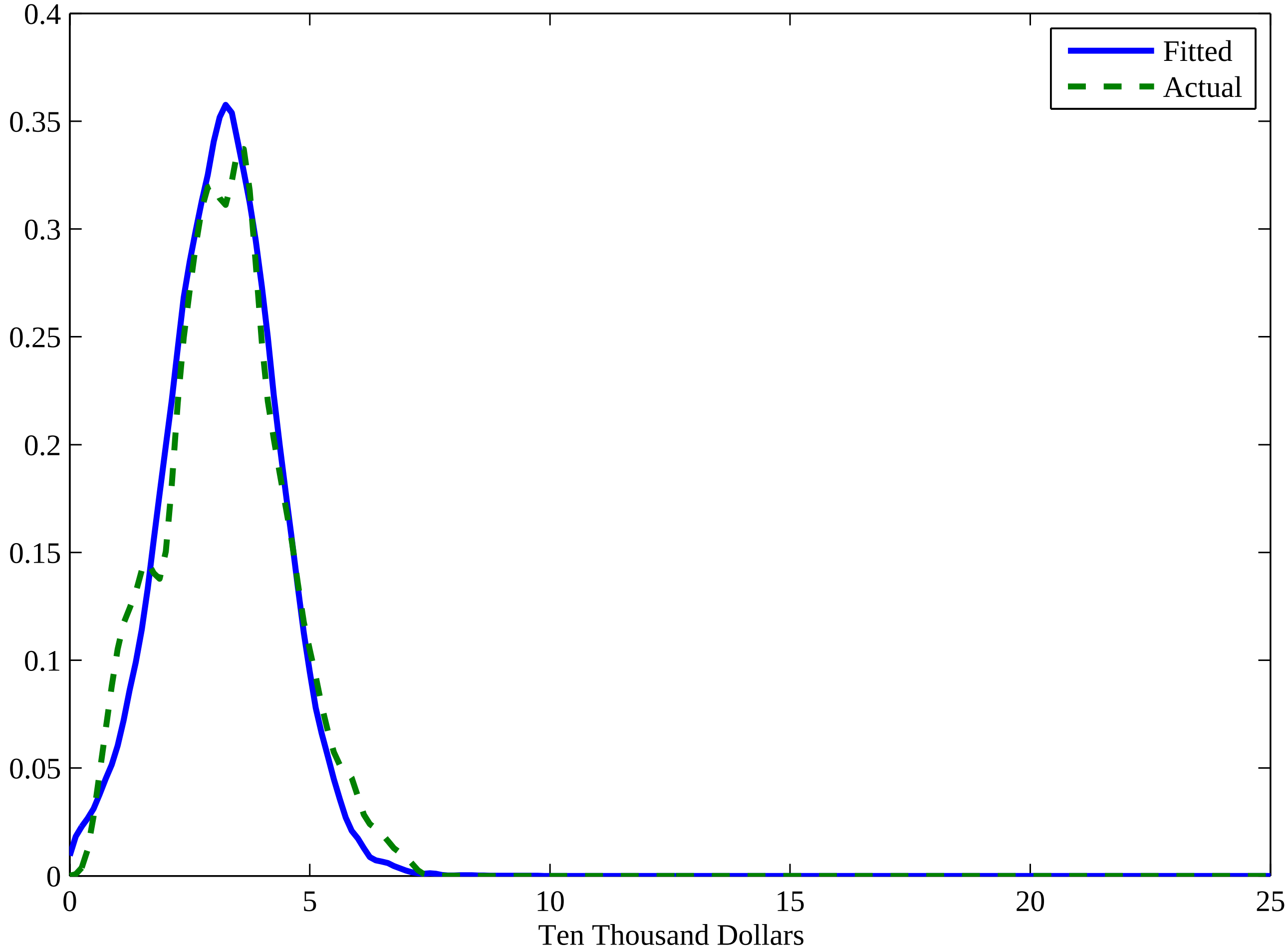
Web Data Appendix Figure II-1.18
Densities of earnings at age 24 Overall Sample NLS/66



Let Y denote earnings at age 24 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.19

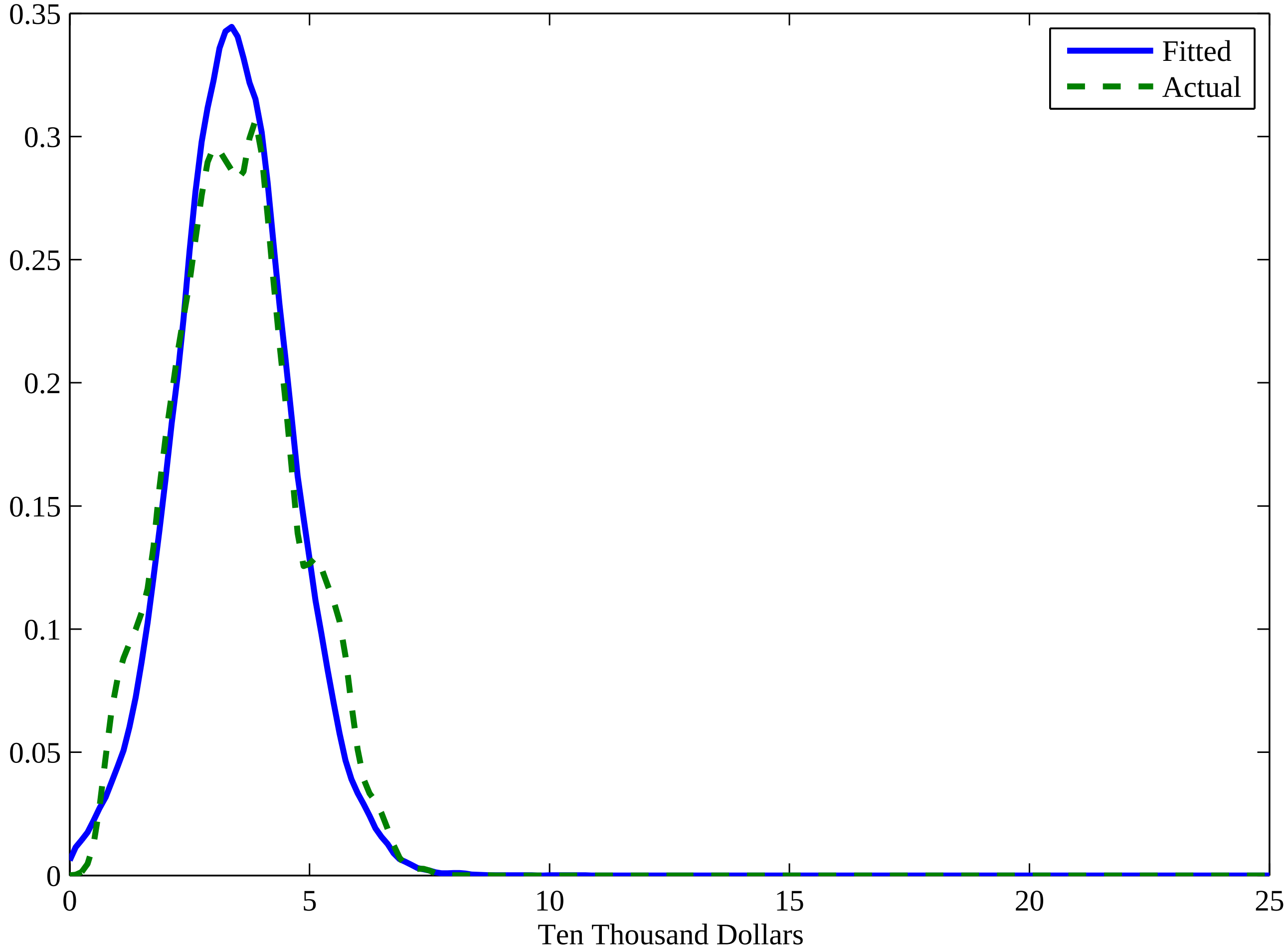
Densities of earnings at age 25 Overall Sample NLS/66



Let Y denote earnings at age 25 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.20

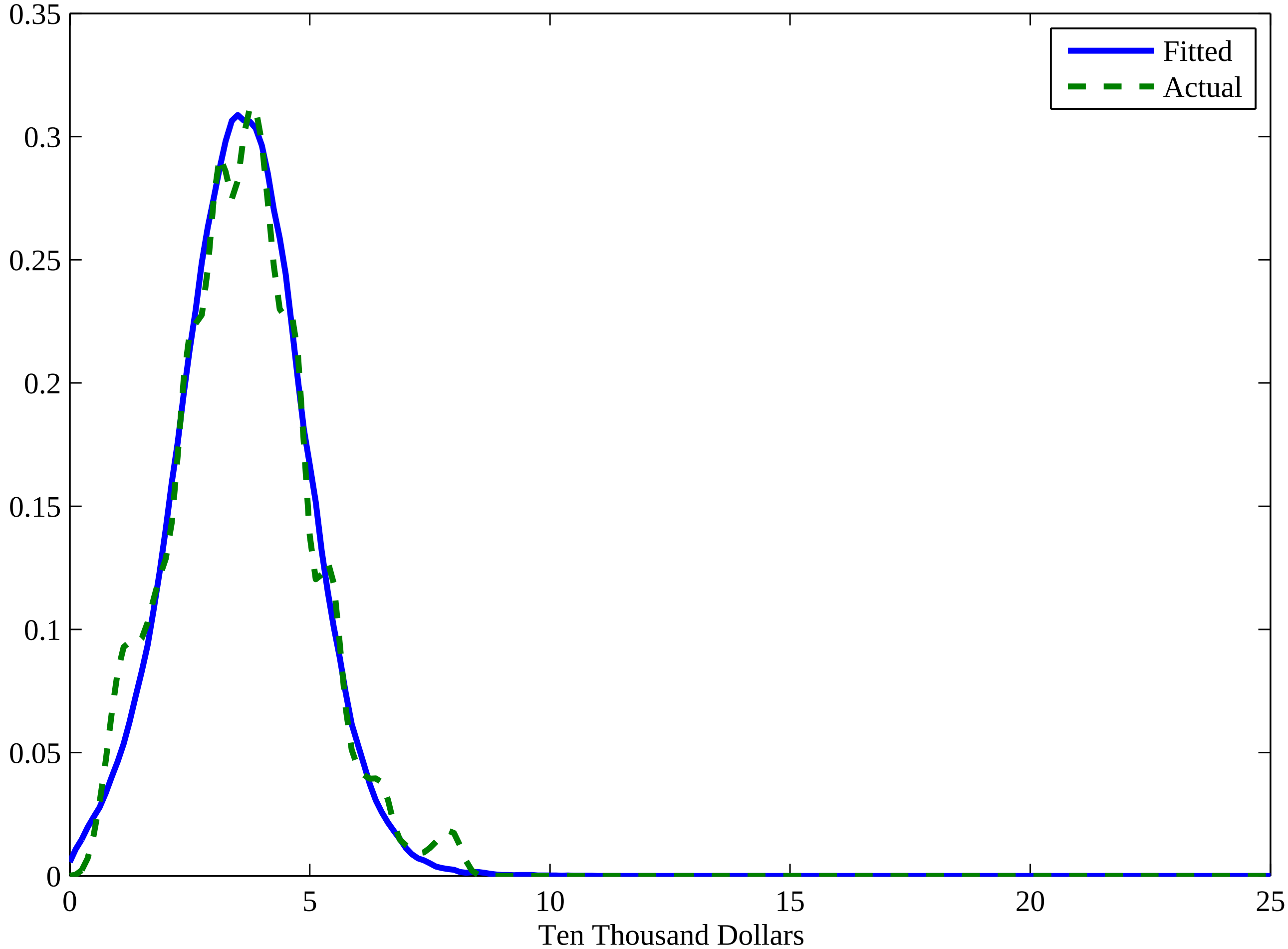
Densities of earnings at age 26 Overall Sample NLS/66



Let Y denote earnings at age 26 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.21

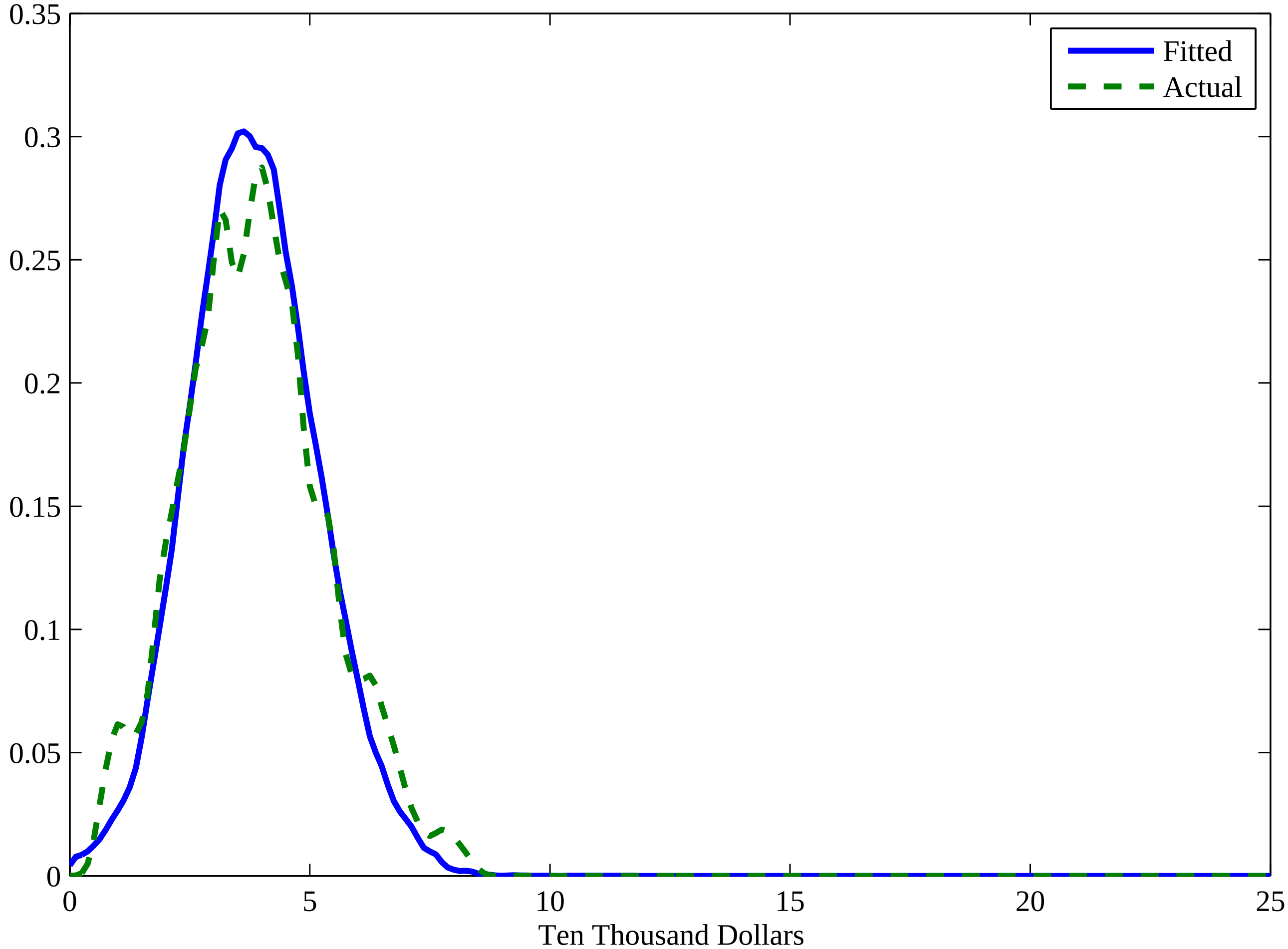
Densities of earnings at age 27 Overall Sample NLS/66



Let Y denote earnings at age 27 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.22

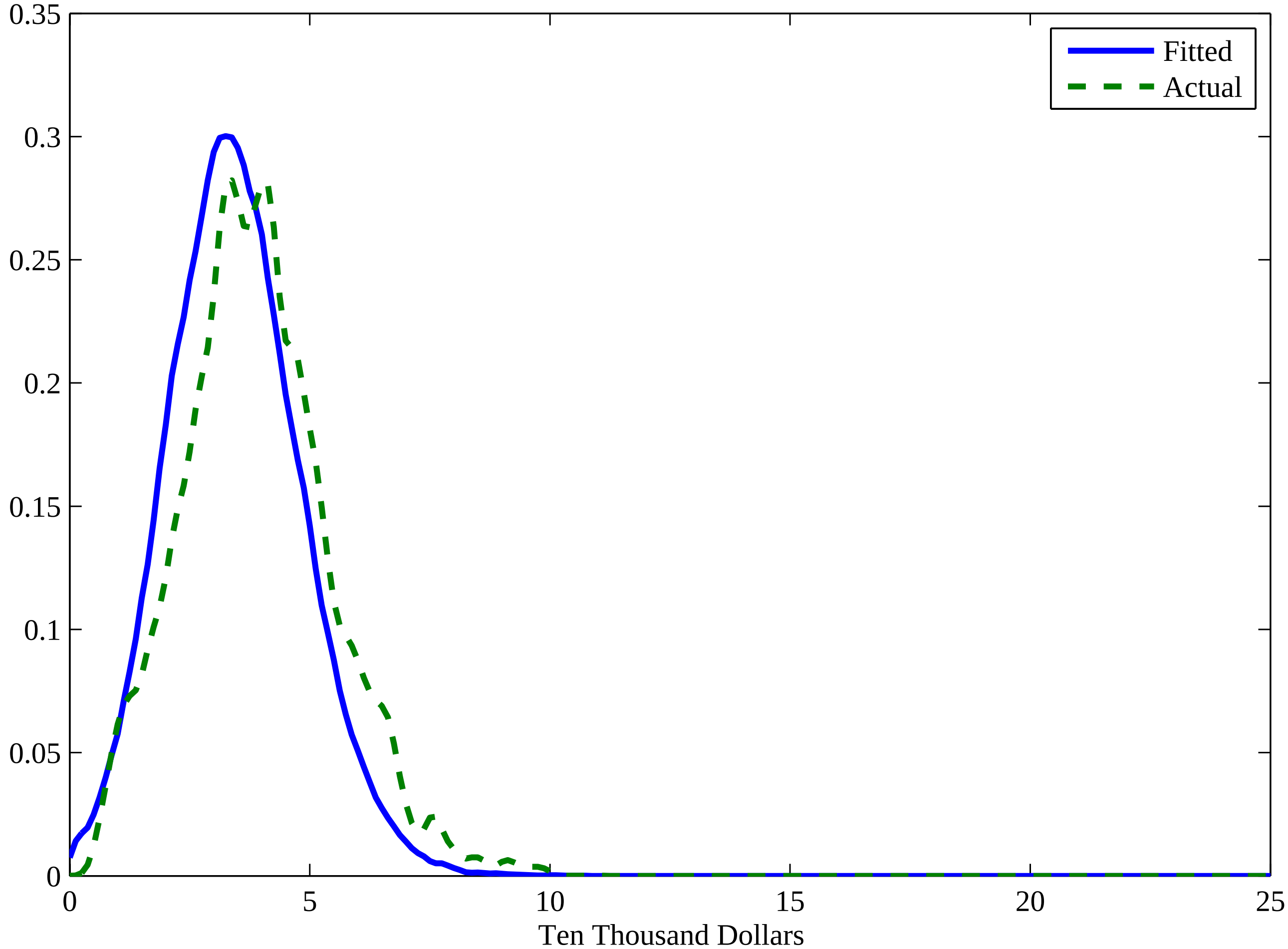
Densities of earnings at age 28 Overall Sample NLS/66



Let Y denote earnings at age 28 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

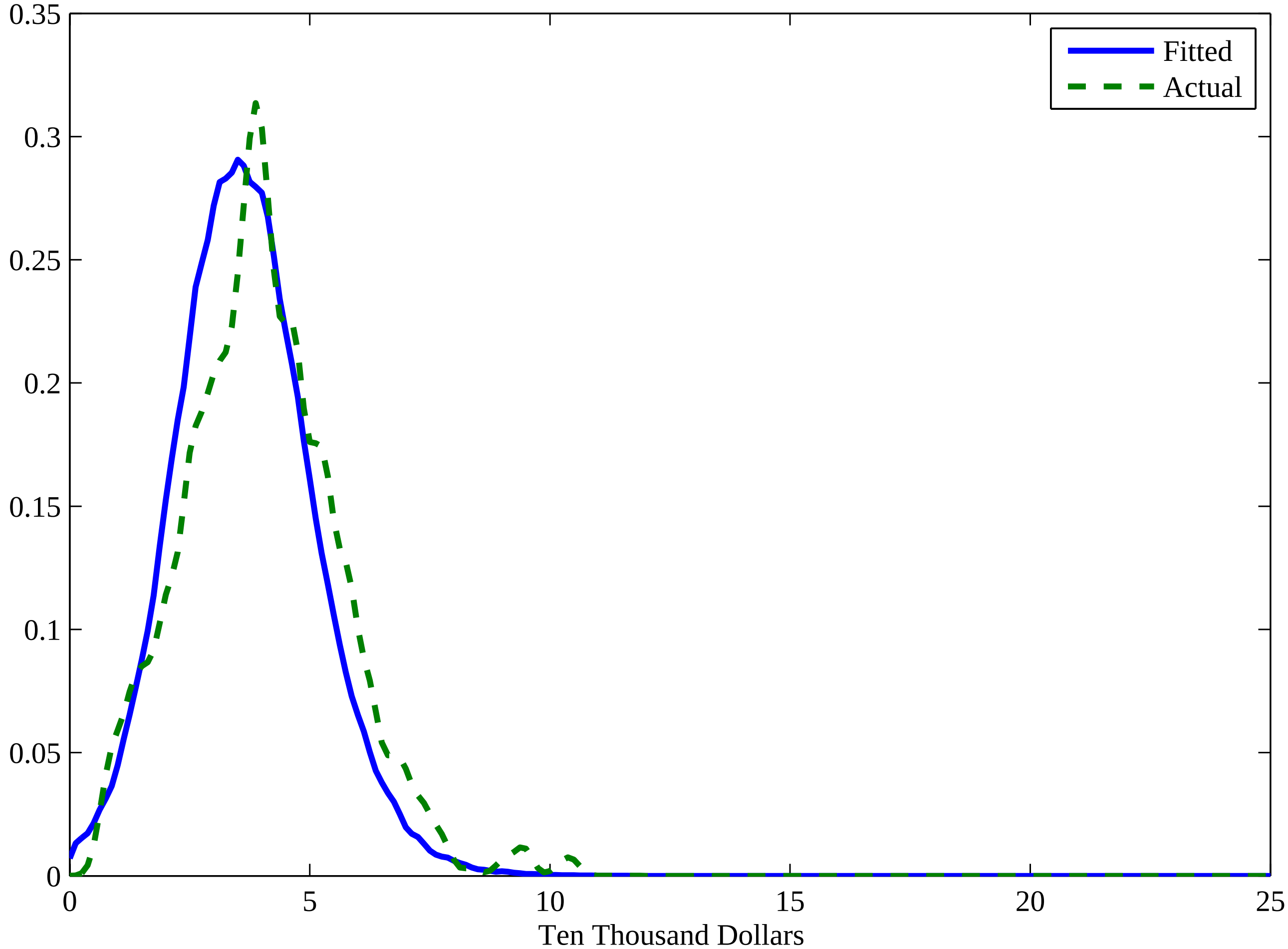
Web Data Appendix Figure II-1.23

Densities of earnings at age 29 Overall Sample NLS/66



Let Y denote earnings at age 29 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

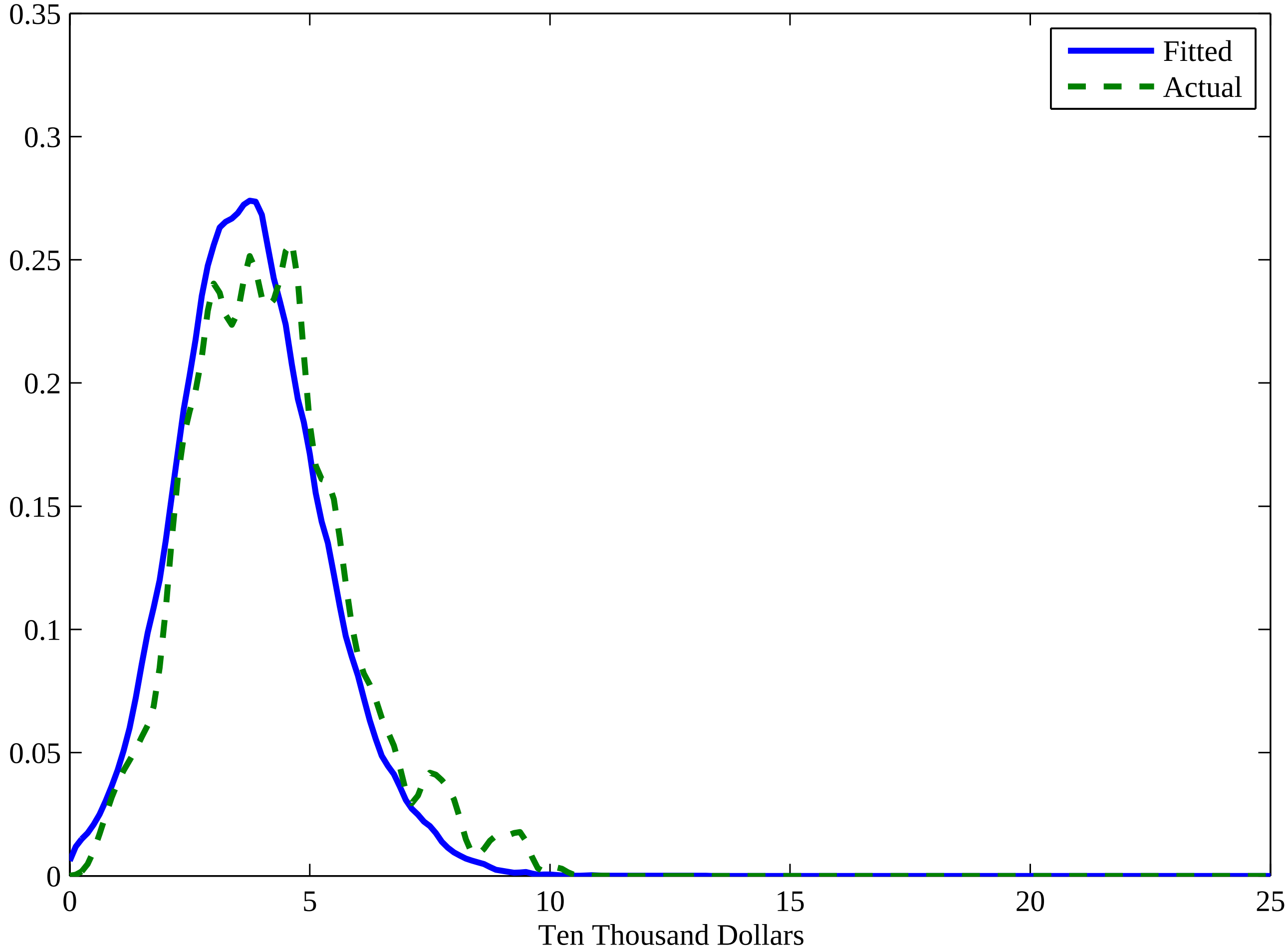
Web Data Appendix Figure II-1.24
Densities of earnings at age 30 Overall Sample NLS/66



Let Y denote earnings at age 30 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

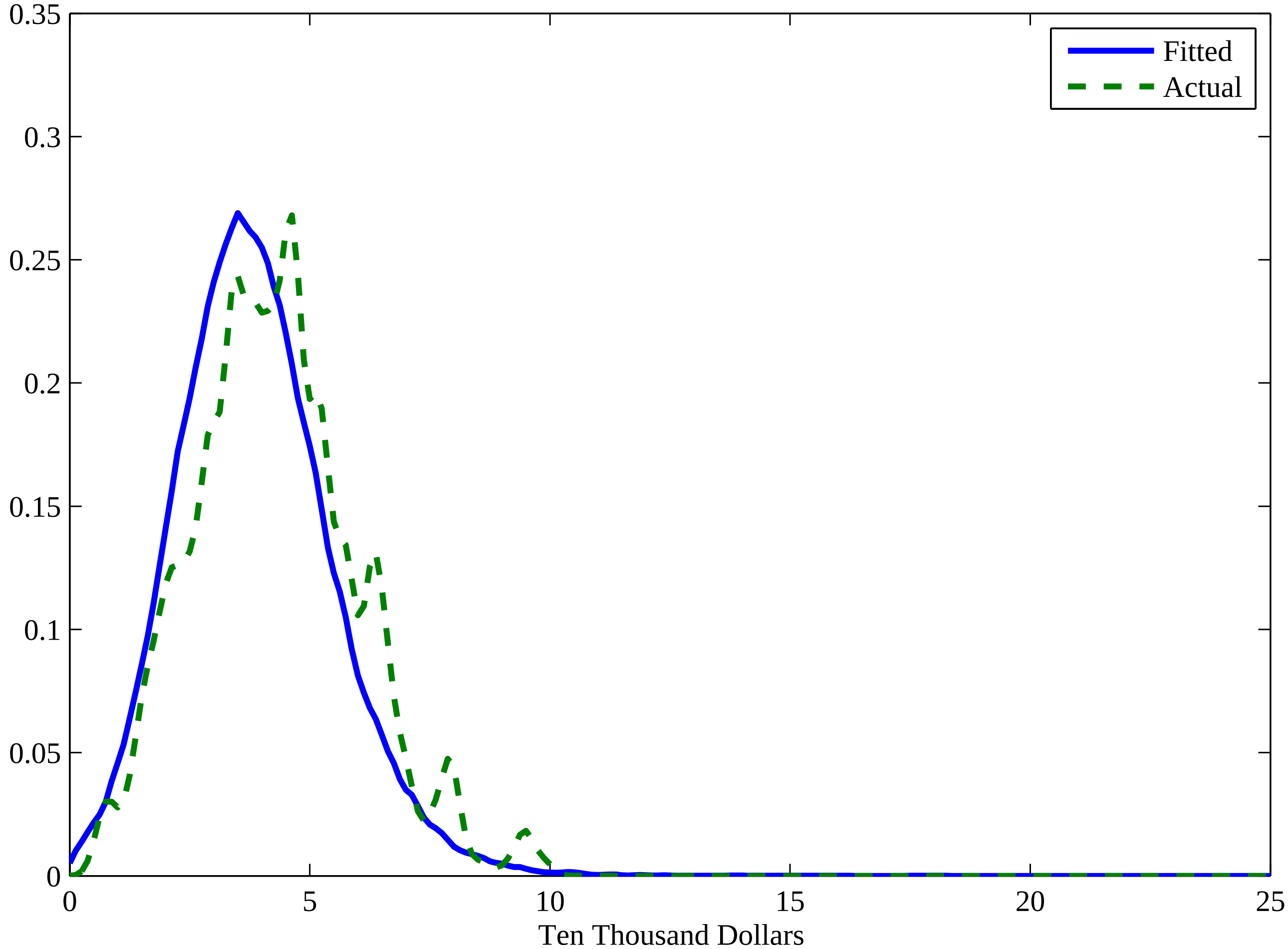
Web Data Appendix Figure II-1.25

Densities of earnings at age 31 Overall Sample NLS/66



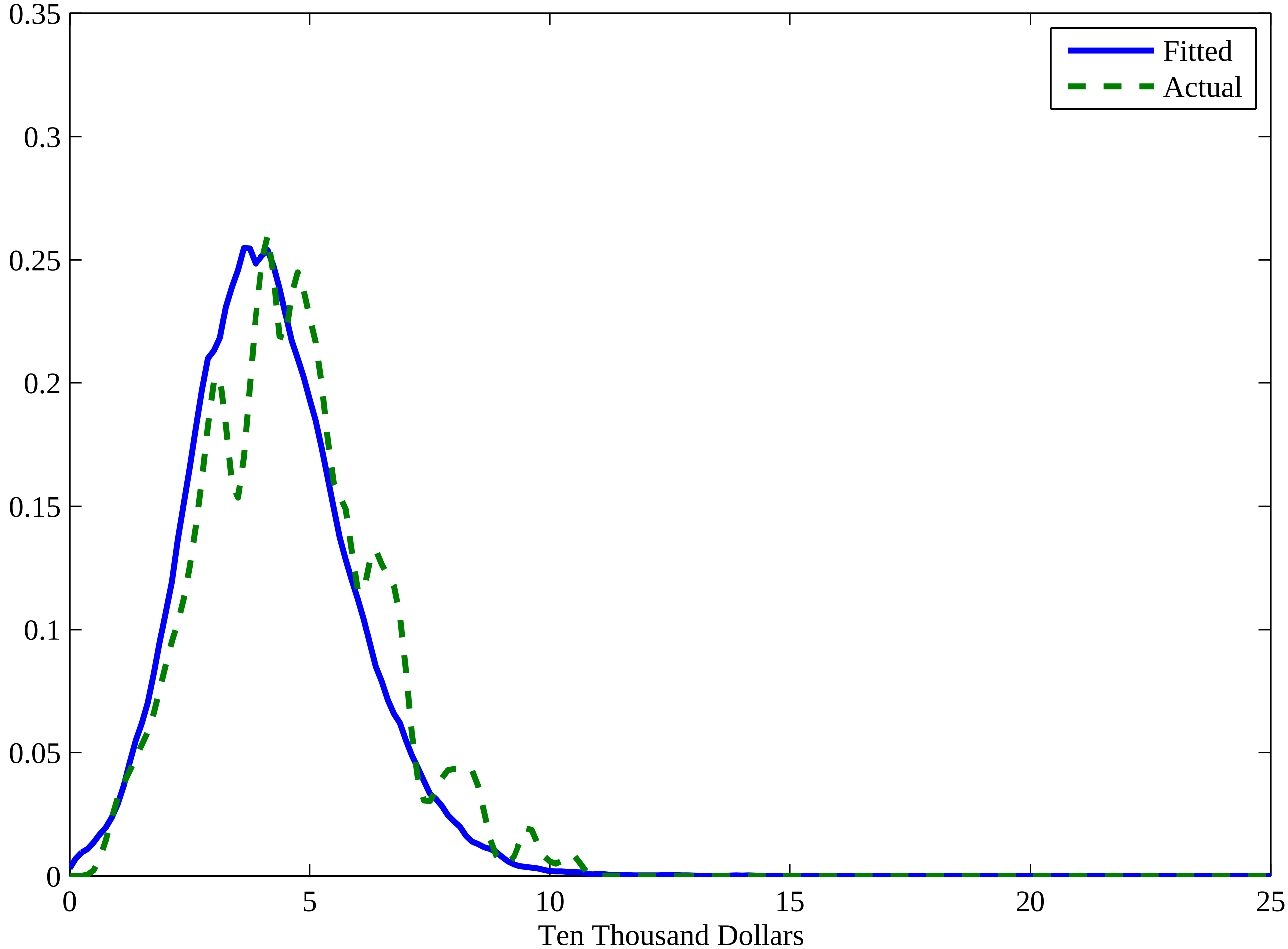
Let Y denote earnings at age 31 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.26
Densities of earnings at age 32 Overall Sample NLS/66



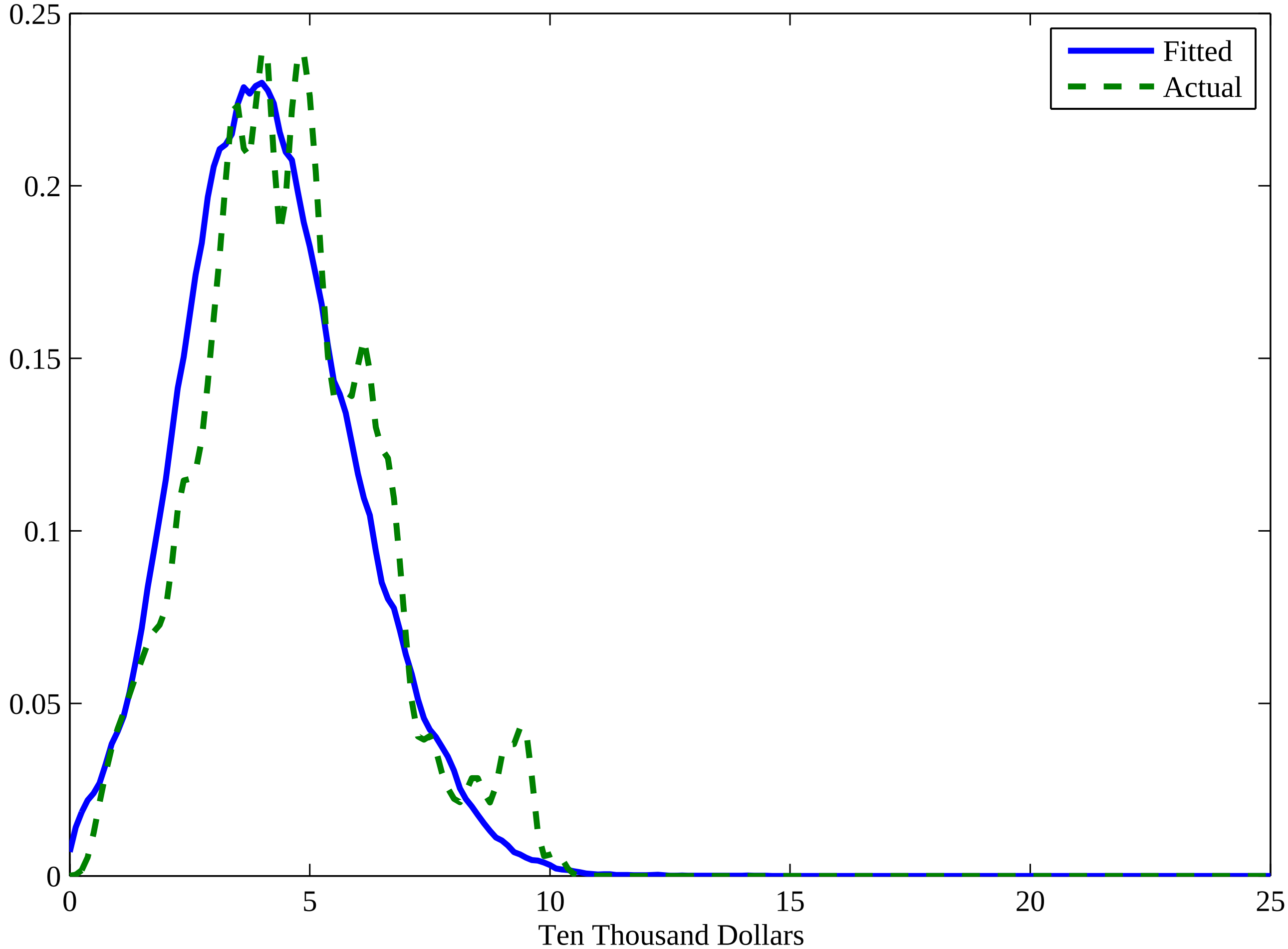
Let Y denote earnings at age 32 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.27
Densities of earnings at age 33 Overall Sample NLS/66



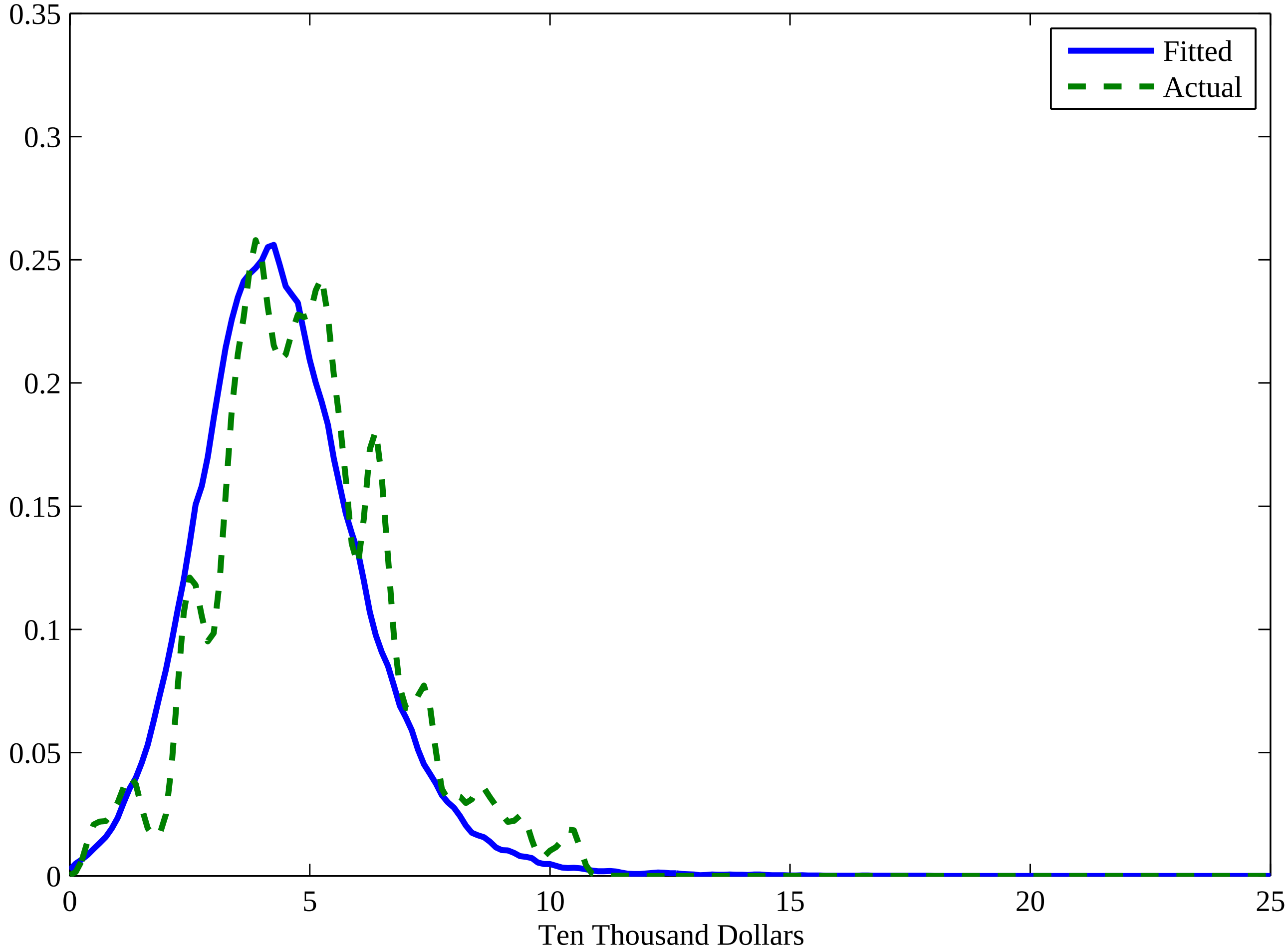
Let Y denote earnings at age 33 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-1.28
Densities of earnings at age 34 Overall Sample NLS/66



Let Y denote earnings at age 34 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

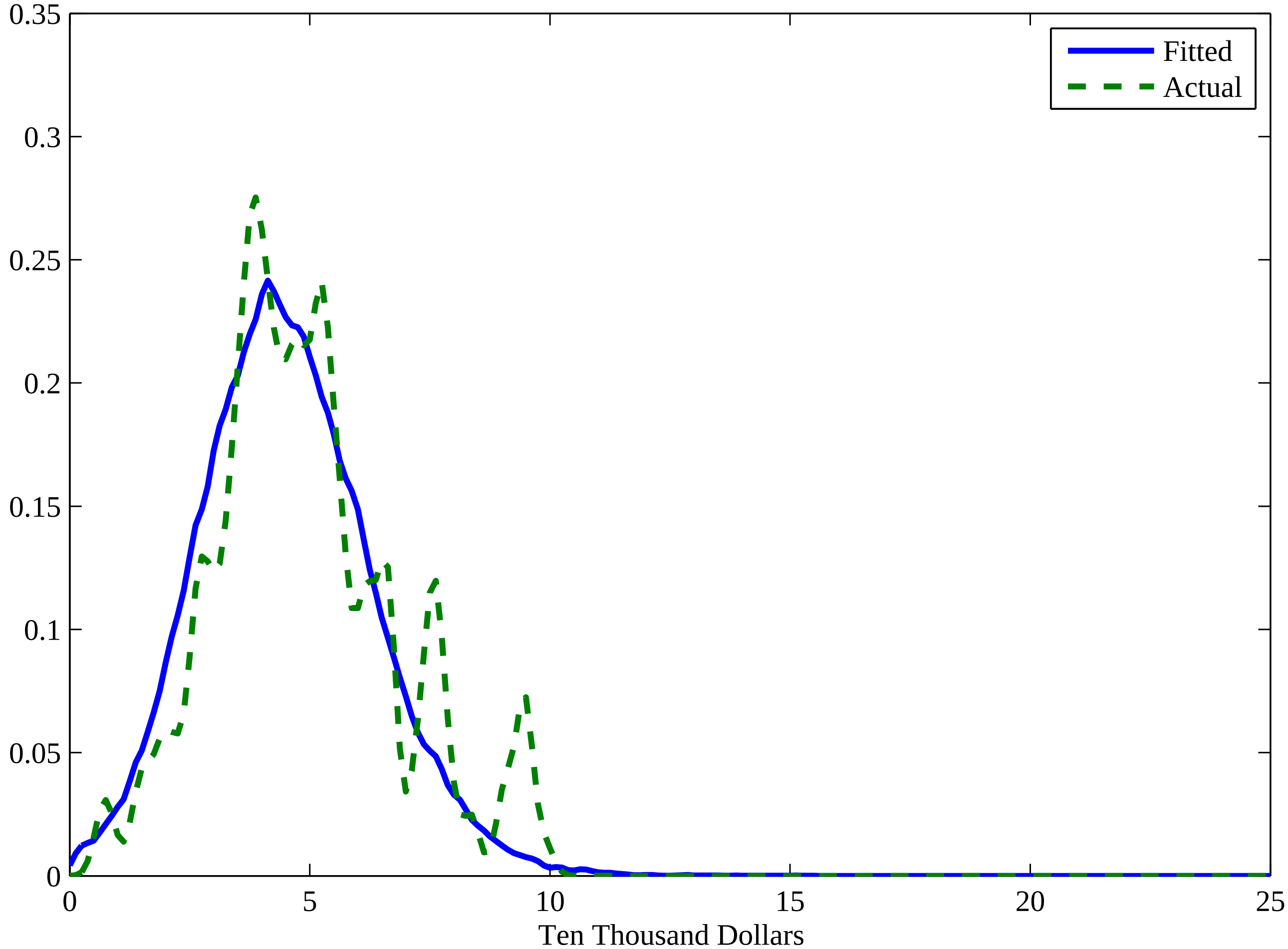
Web Data Appendix Figure II-1.29
Densities of earnings at age 35 Overall Sample NLS/66



Let Y denote earnings at age 35 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

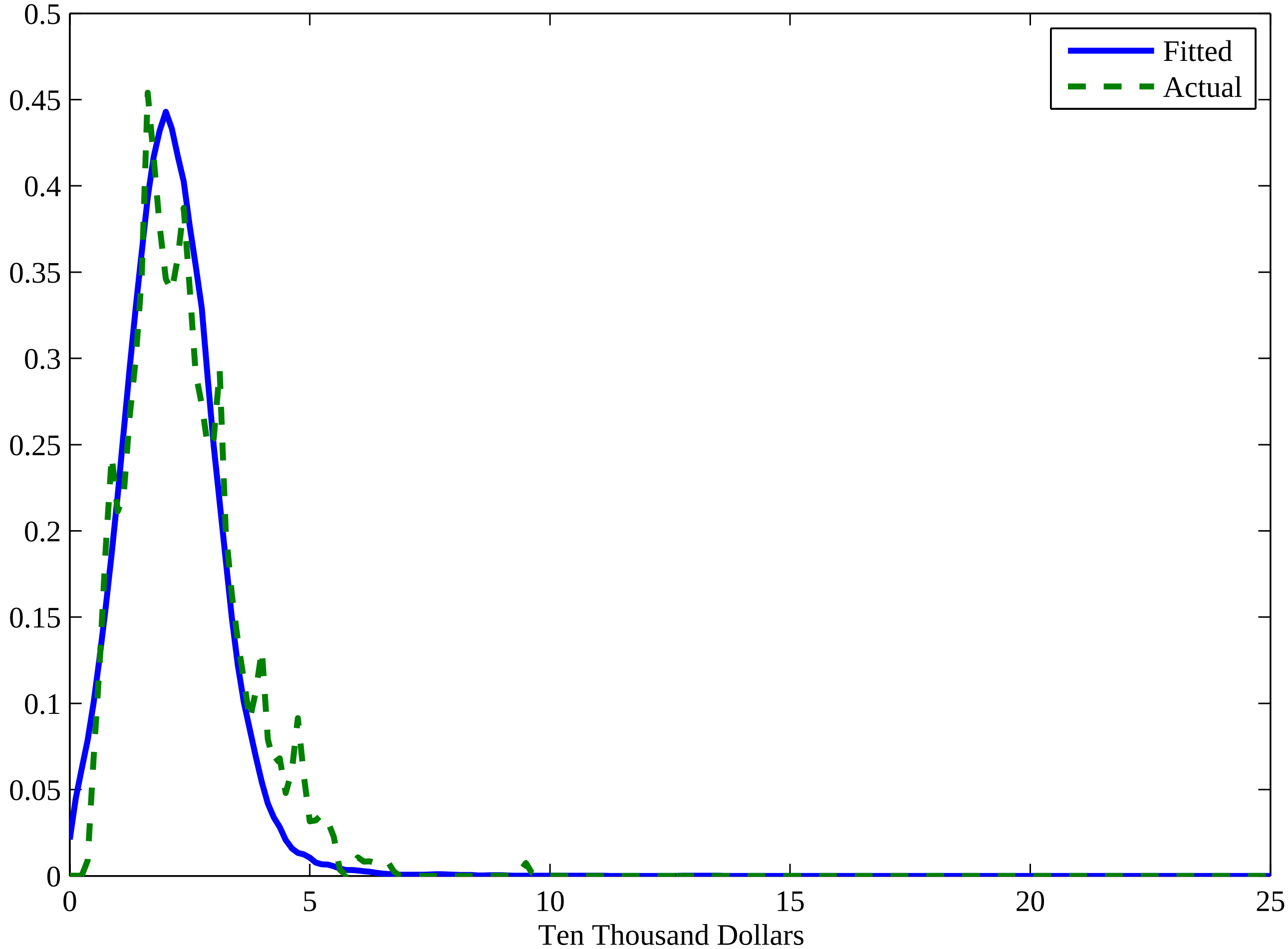
Web Data Appendix Figure II-1.30

Densities of earnings at age 36 Overall Sample NLS/66



Let Y denote earnings at age 36 in the overall sample.
Here we plot the density functions f(y) generated from the data (the solid curve), against that predicted by the model (the dashed line).

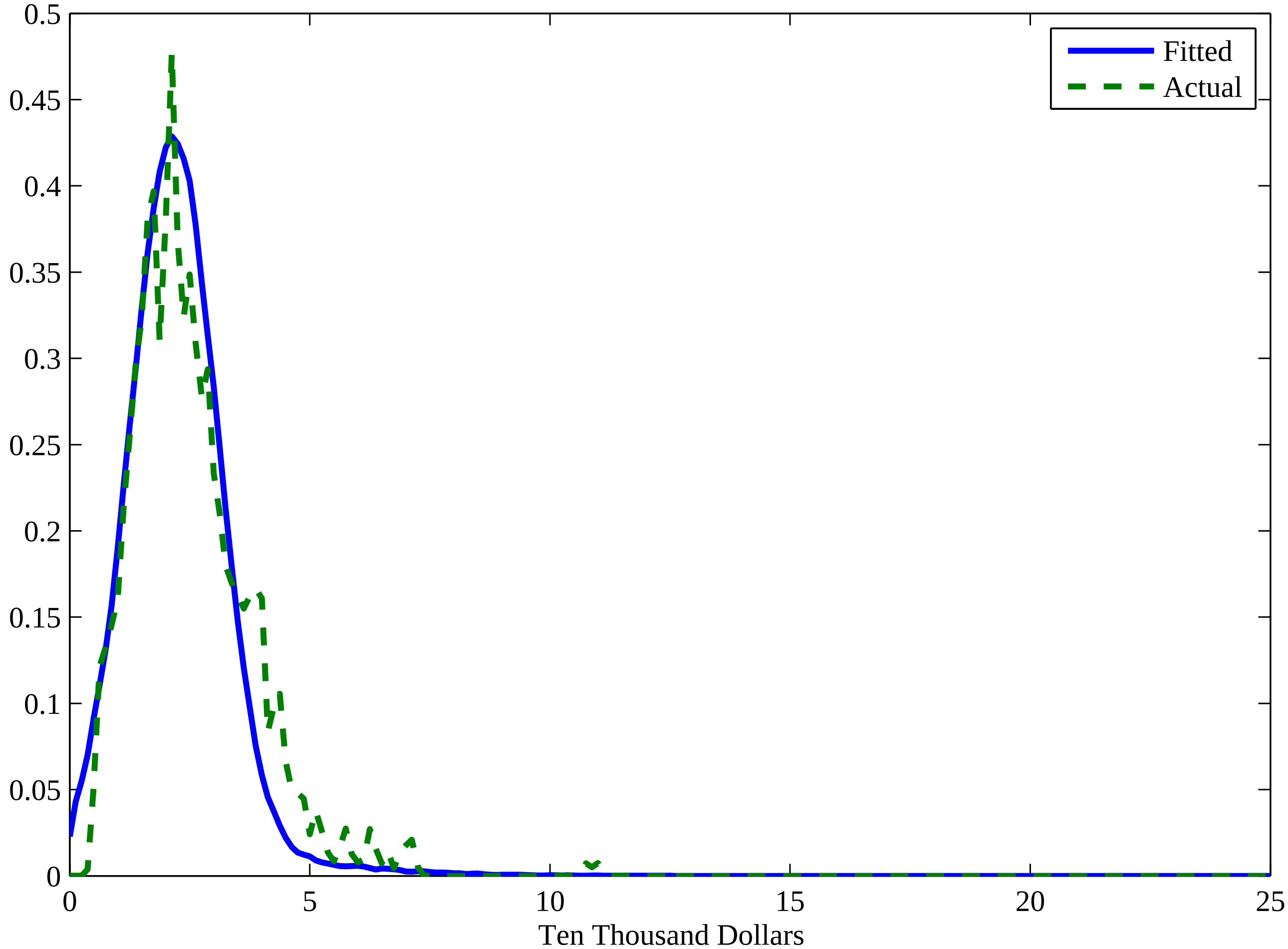
Web Data Appendix Figure II-2.1
Densities of earnings at age 22 High School Sample NLSY/79



Let Y_0 denote earnings at age 22.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

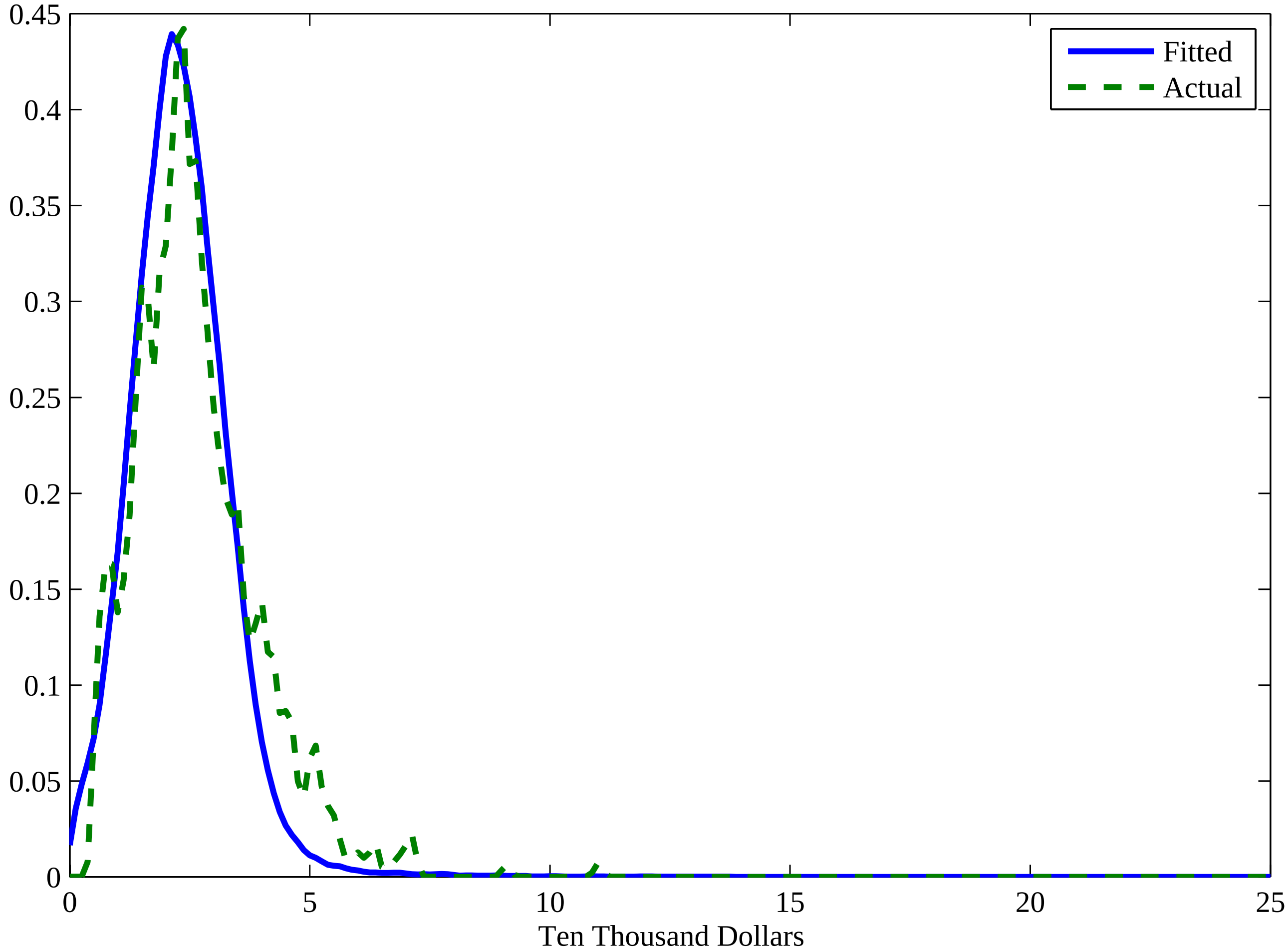
Web Data Appendix Figure II-2.2

Densities of earnings at age 23 High School Sample NLSY/79



Let Y_0 denote earnings at age 23.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

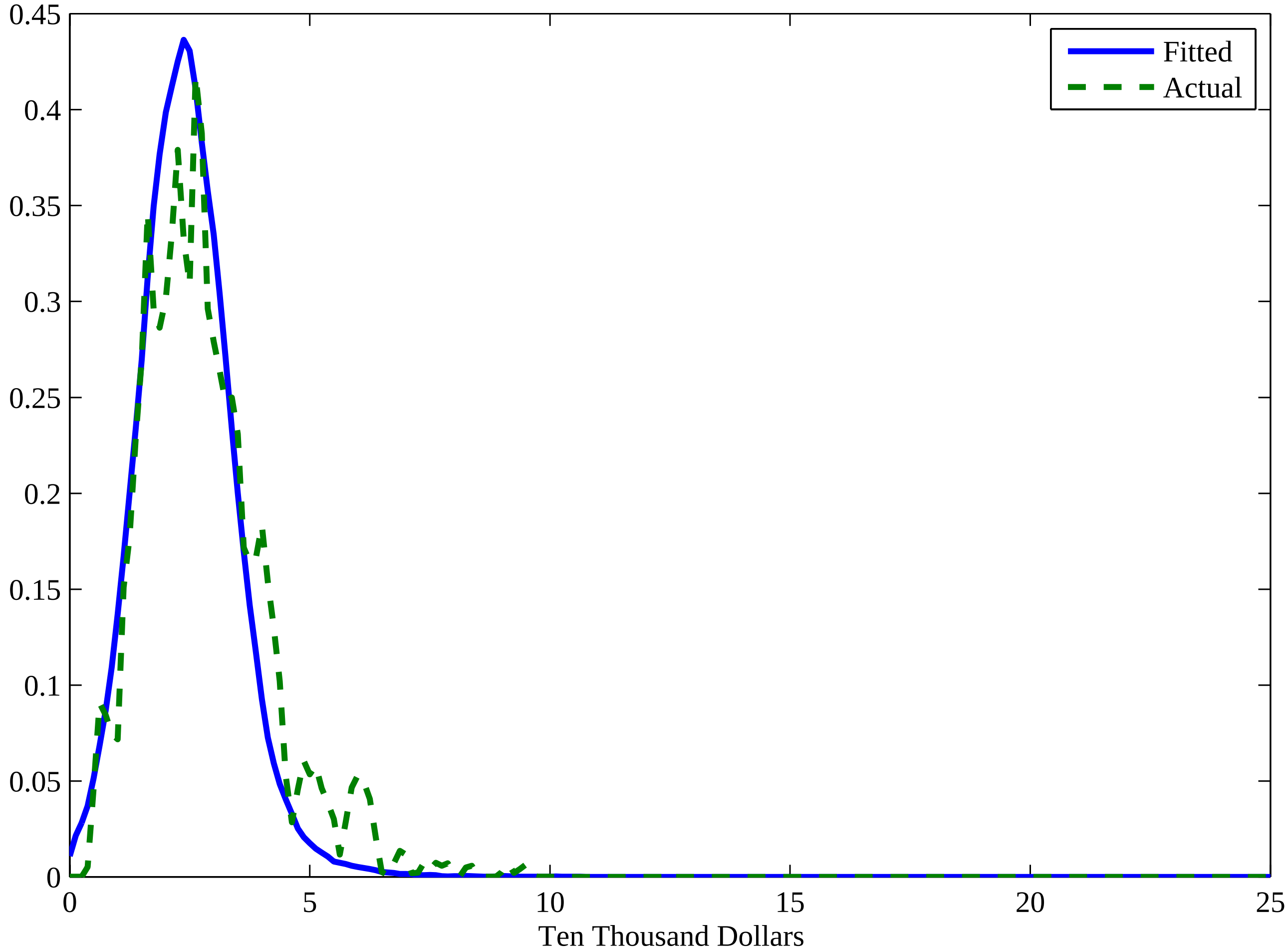
Web Data Appendix Figure II-2.3
Densities of earnings at age 24 High School Sample NLSY/79



Let Y_0 denote earnings at age 24.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.4

Densities of earnings at age 25 High School Sample NLSY/79

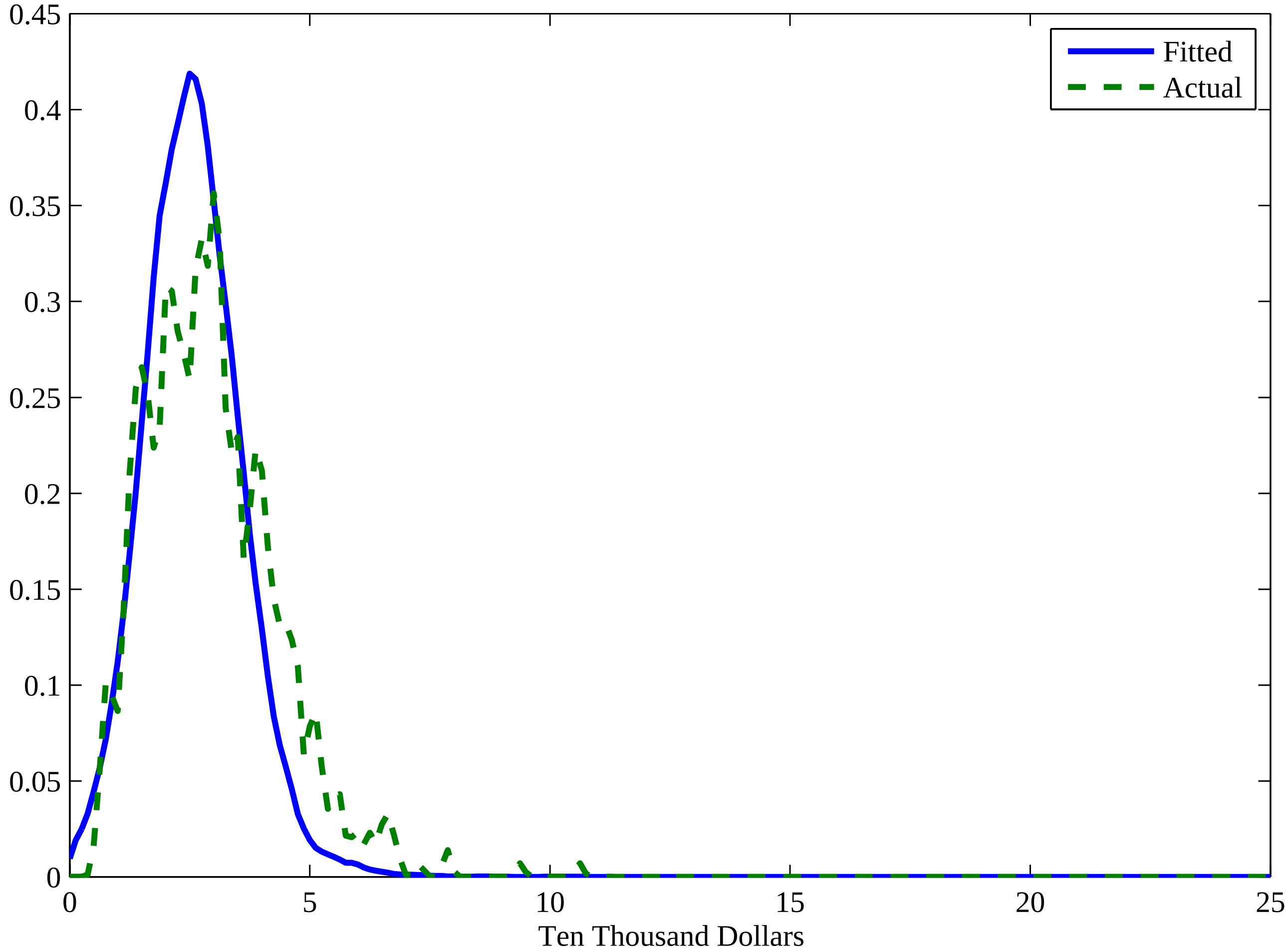


Let Y_0 denote earnings at age 25.

Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

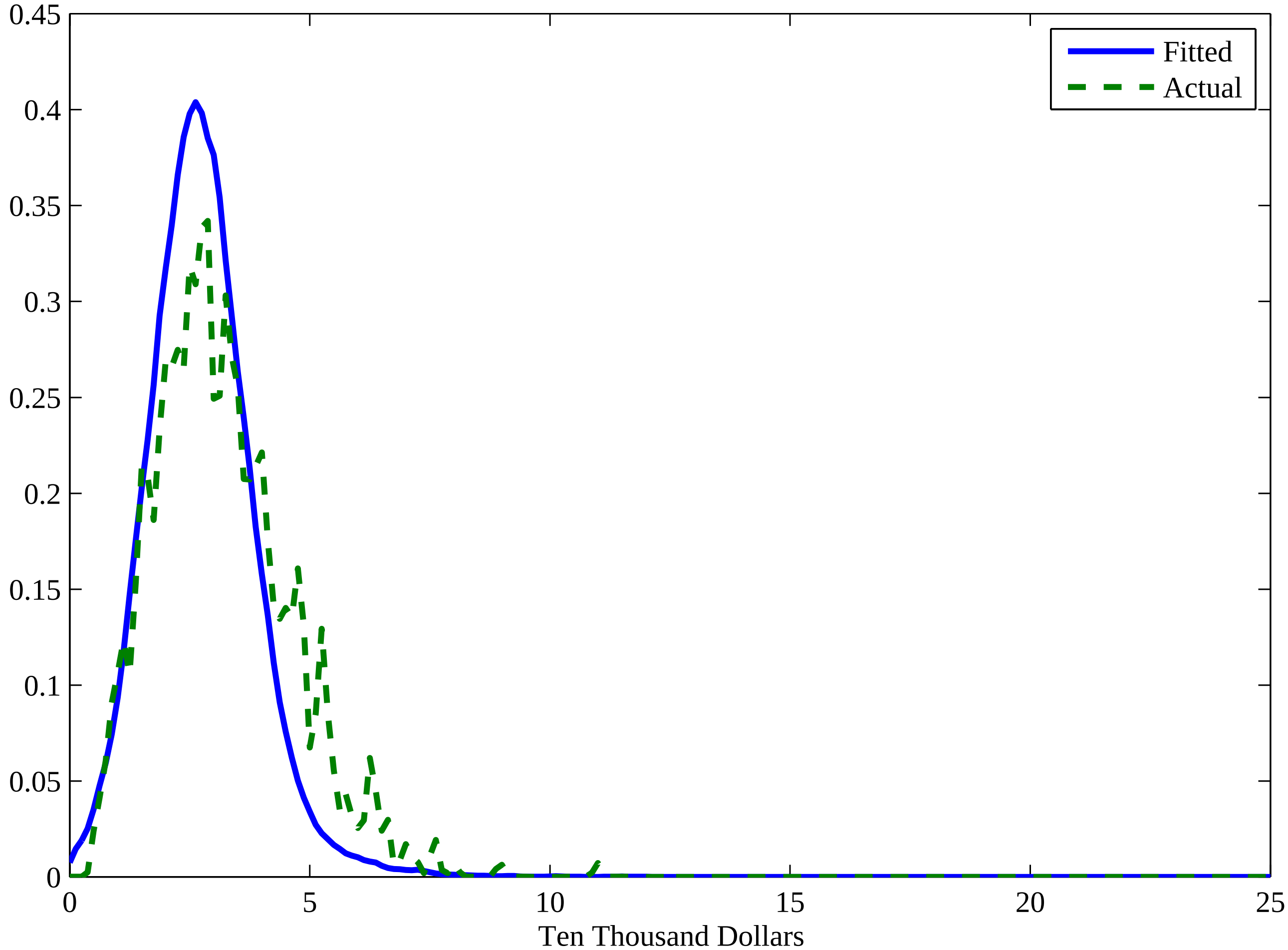
Web Data Appendix Figure II-2.5

Densities of earnings at age 26 High School Sample NLSY/79



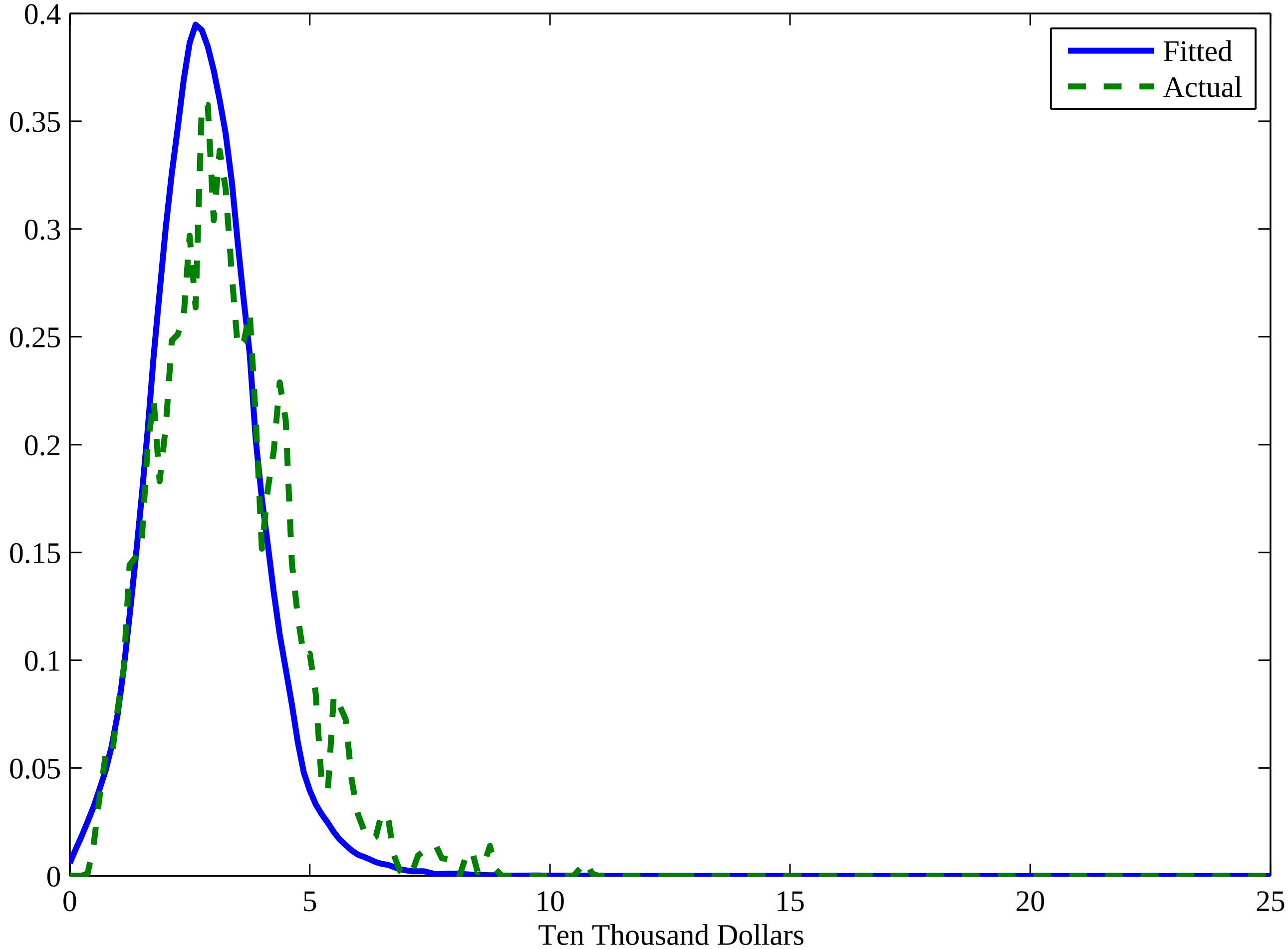
Let Y_0 denote earnings at age 26.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.6
Densities of earnings at age 27 High School Sample NLSY/79



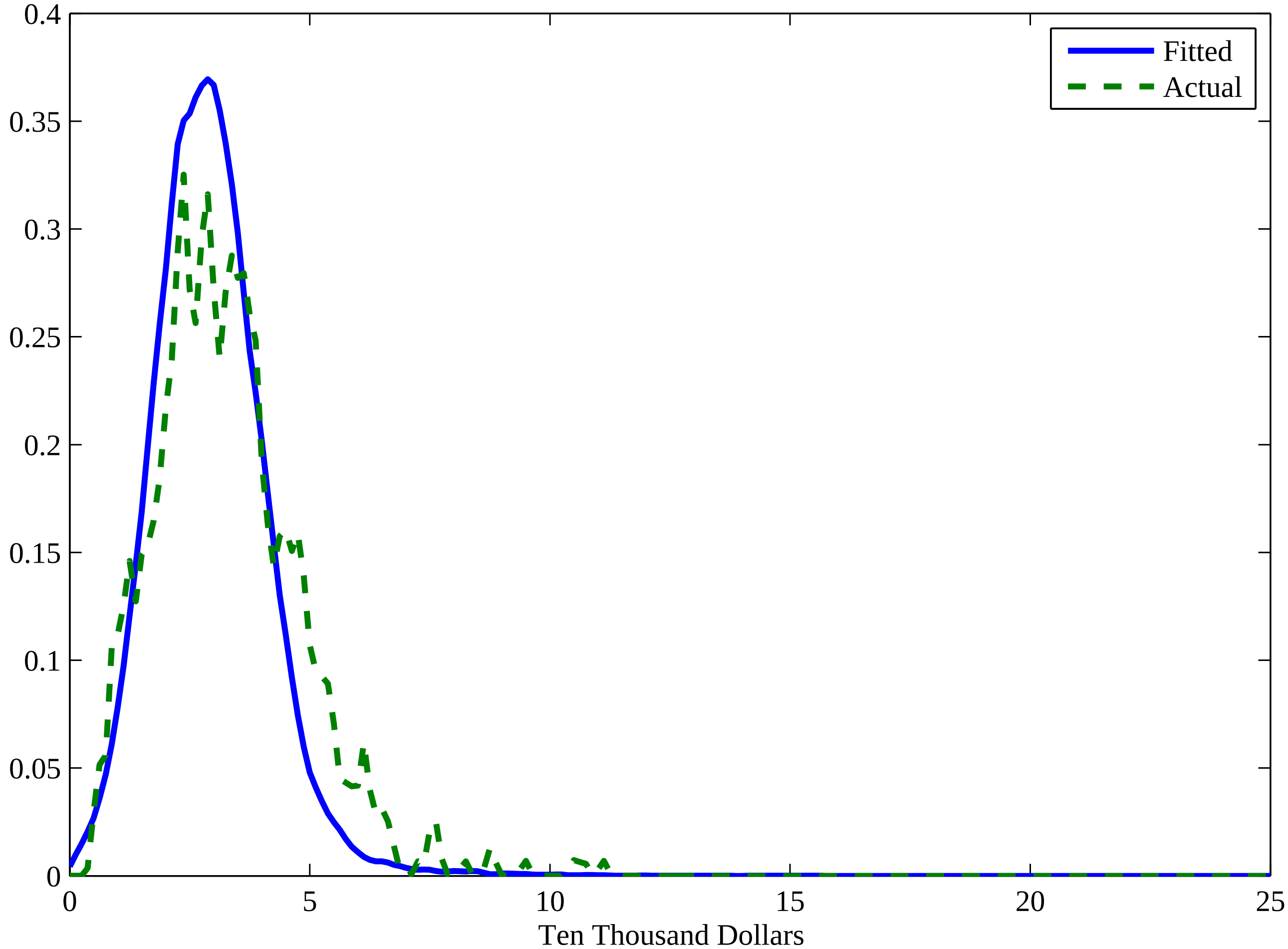
Let Y_0 denote earnings at age 27.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.7
Densities of earnings at age 28 High School Sample NLSY/79



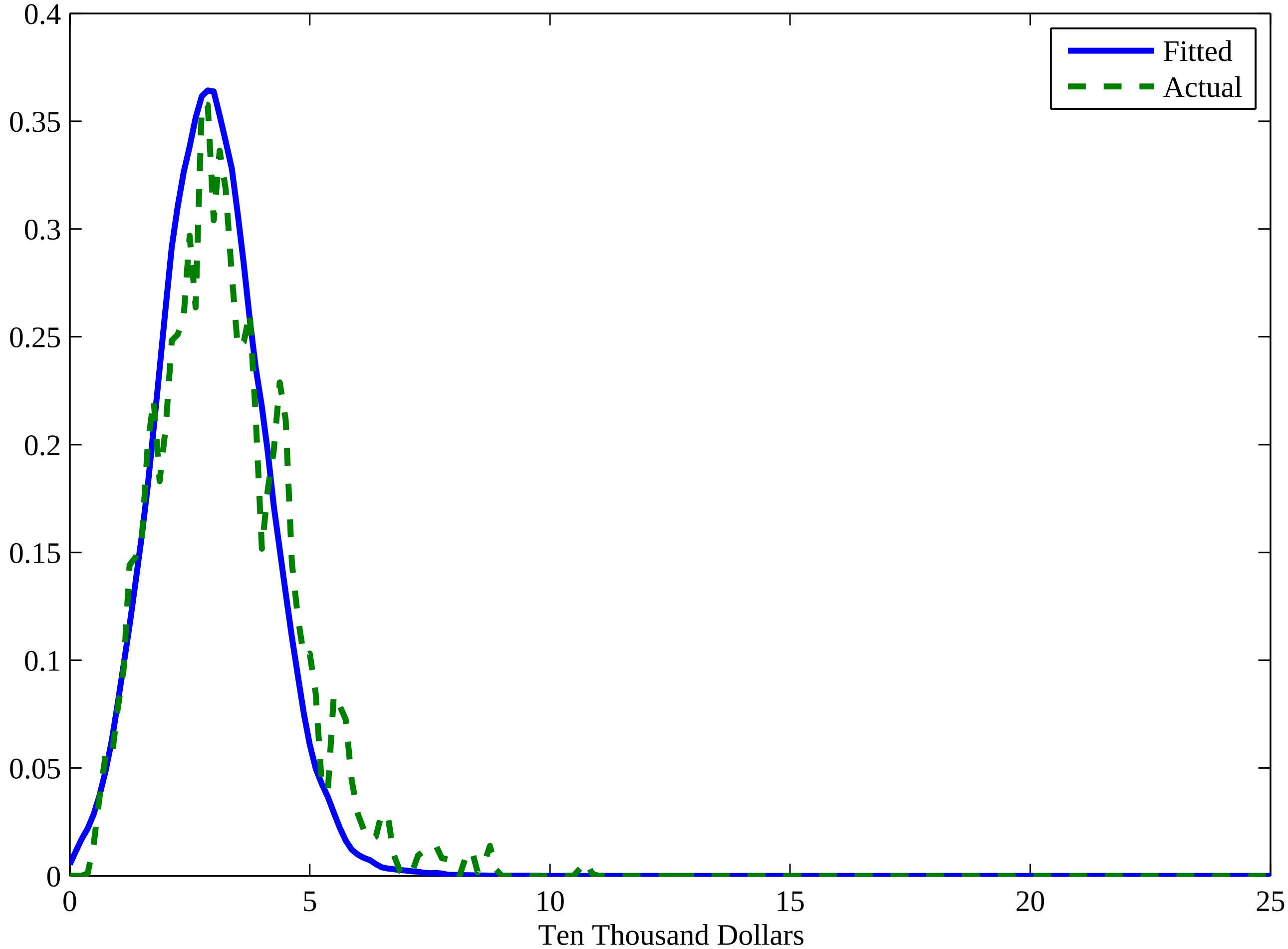
Let Y_0 denote earnings at age 28.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.8
Densities of earnings at age 29 High School Sample NLSY/79



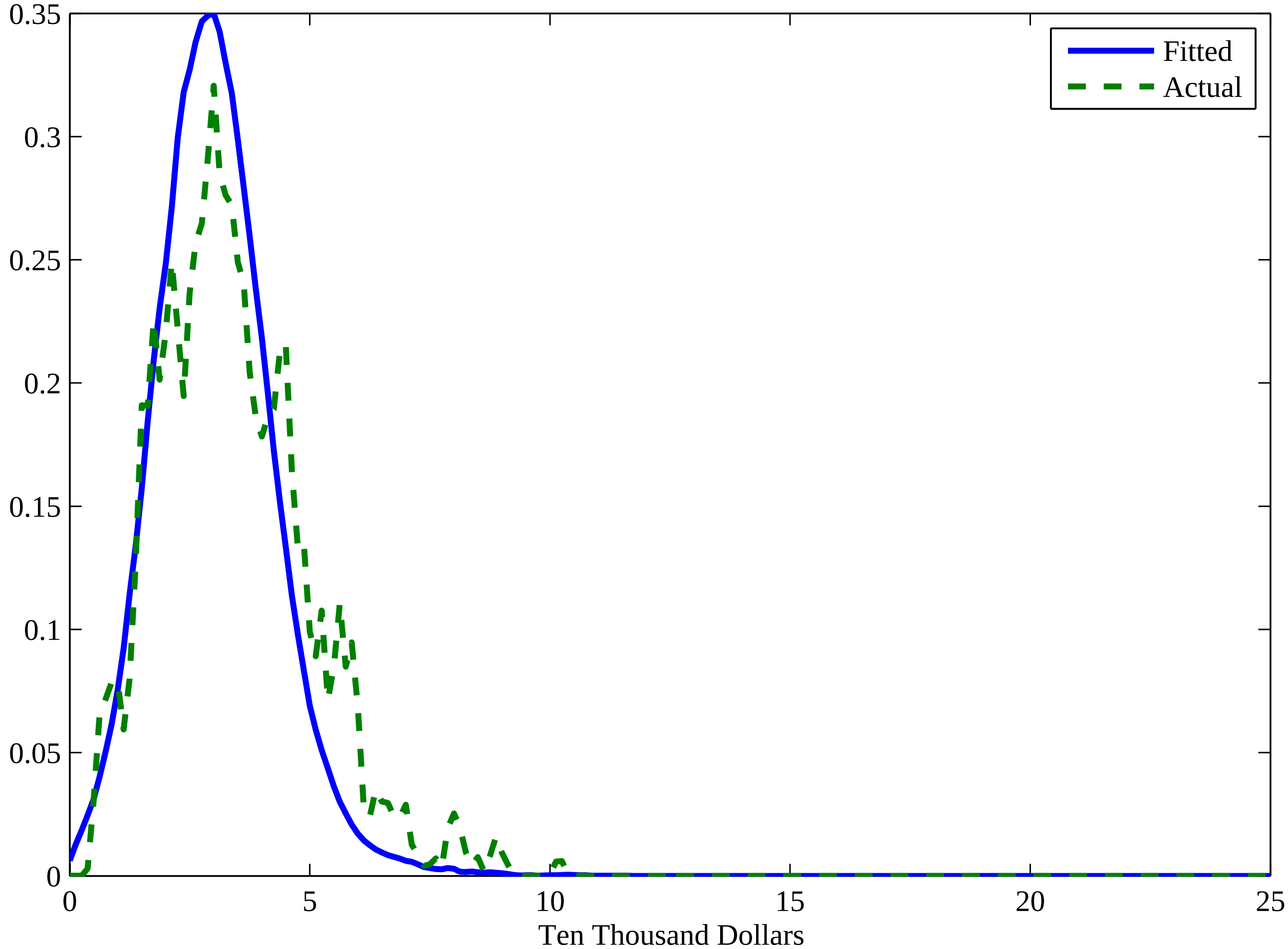
Let Y_0 denote earnings at age 29.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.9
Densities of earnings at age 30 High School Sample NLSY/79



Let Y_0 denote earnings at age 30.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

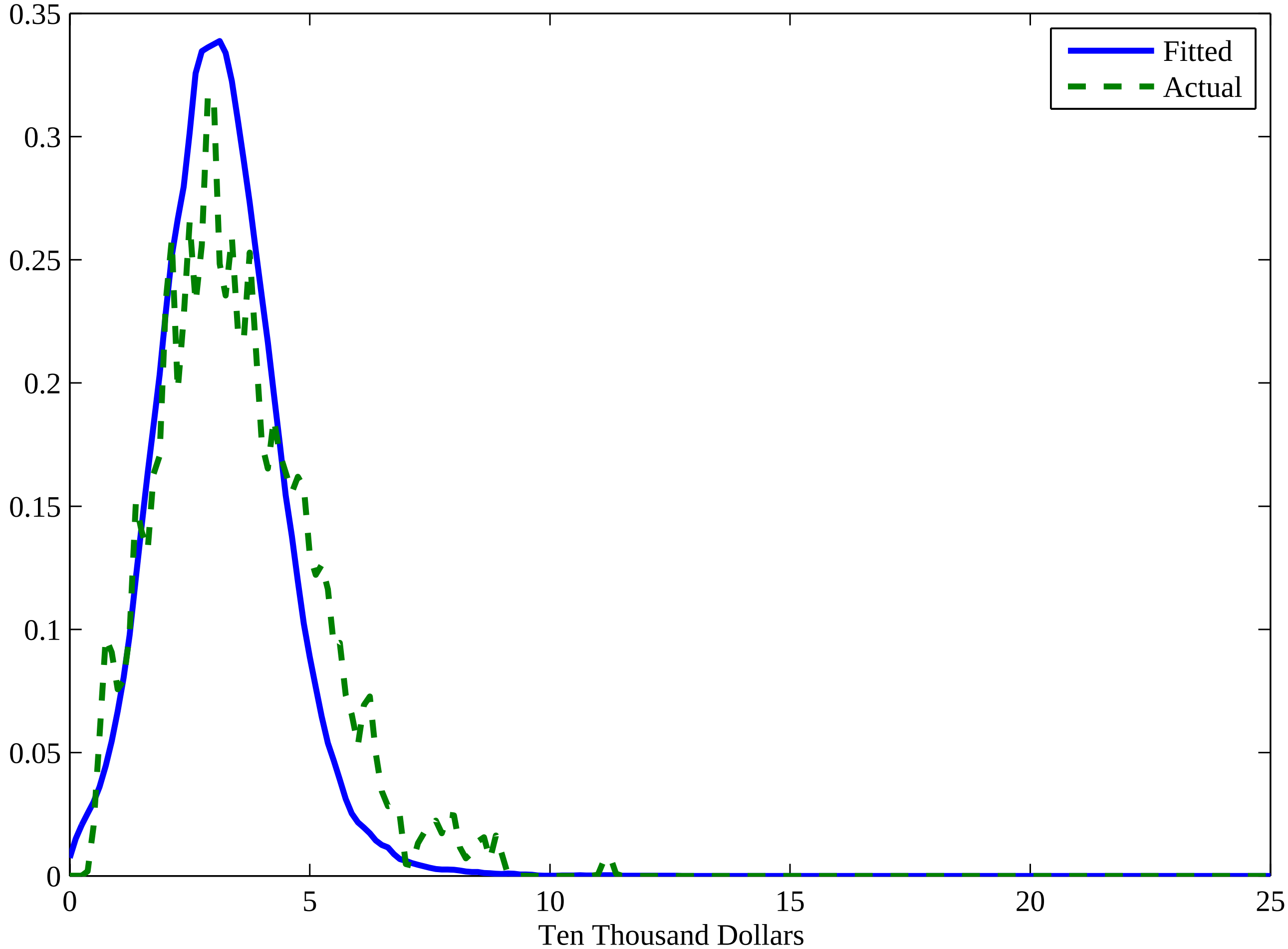
Web Data Appendix Figure II-2.10
Densities of earnings at age 31 High School Sample NLSY/79



Let Y_0 denote earnings at age 31.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.11

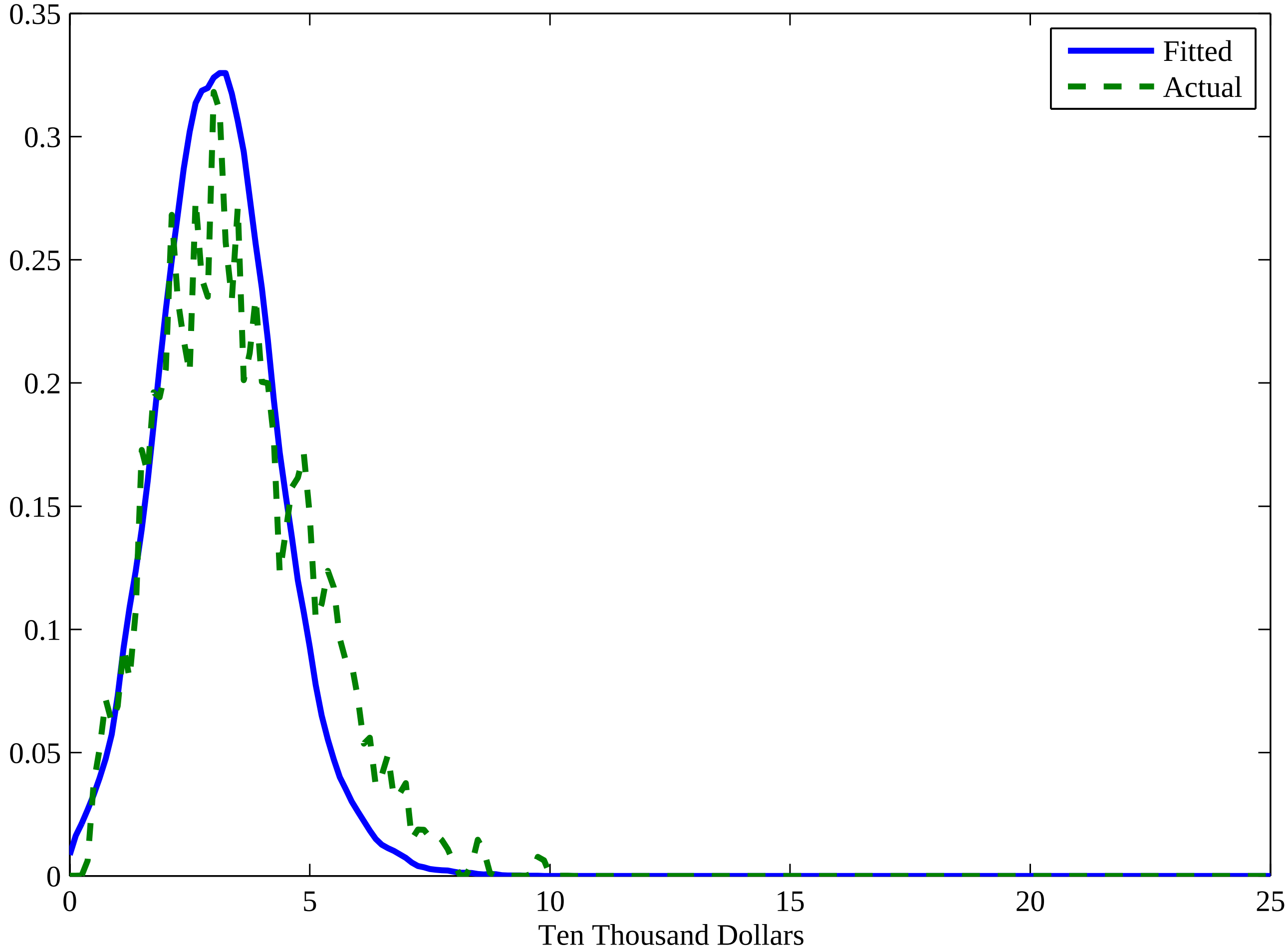
Densities of earnings at age 32 High School Sample NLSY/79



Let Y_0 denote earnings at age 32.

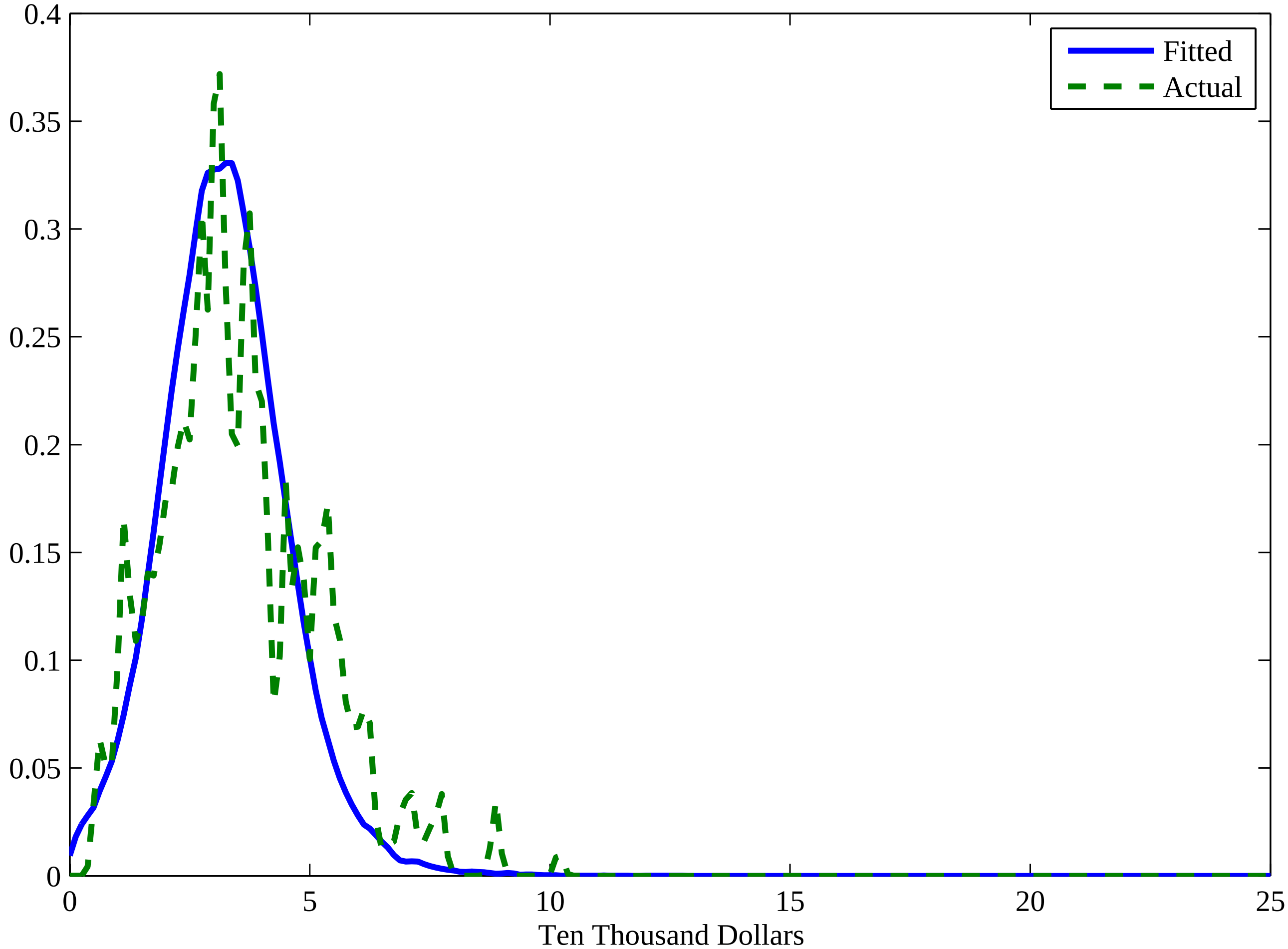
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.12
Densities of earnings at age 33 High School Sample NLSY/79



Let Y_0 denote earnings at age 33.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

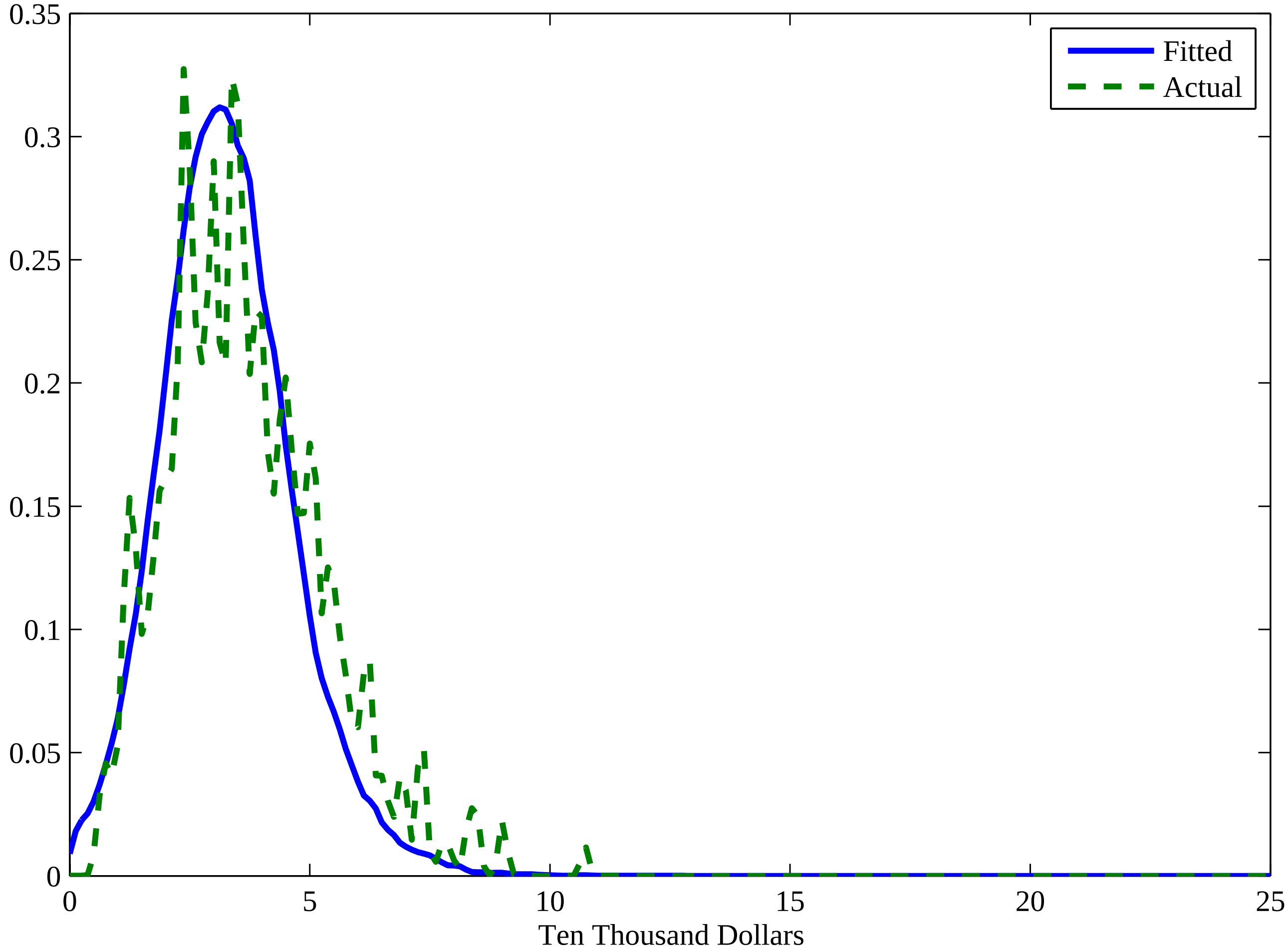
Web Data Appendix Figure II-2.13
Densities of earnings at age 34 High School Sample NLSY/79



Let Y_0 denote earnings at age 34.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.14

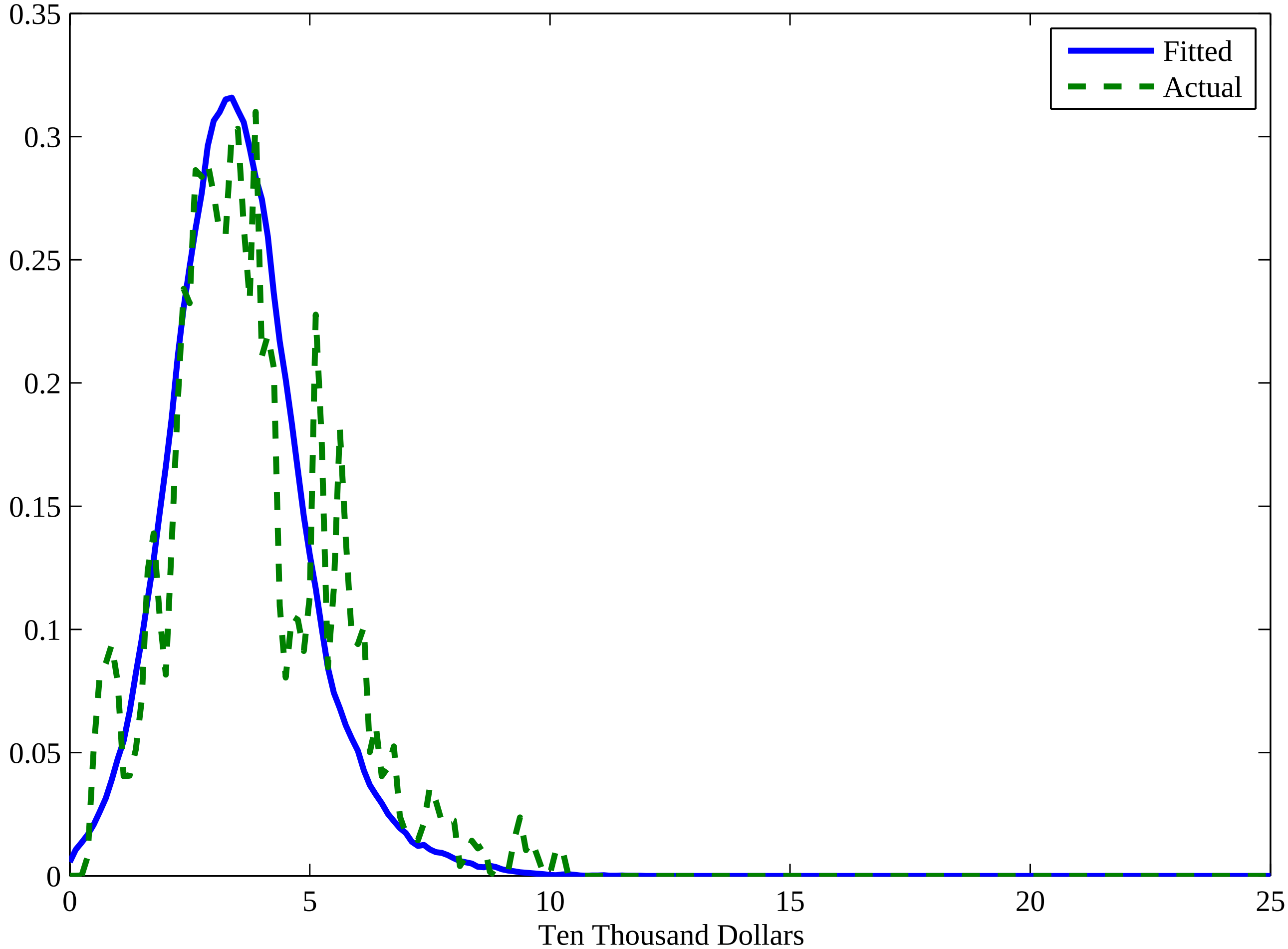
Densities of earnings at age 35 High School Sample NLSY/79



Let Y_0 denote earnings at age 35.

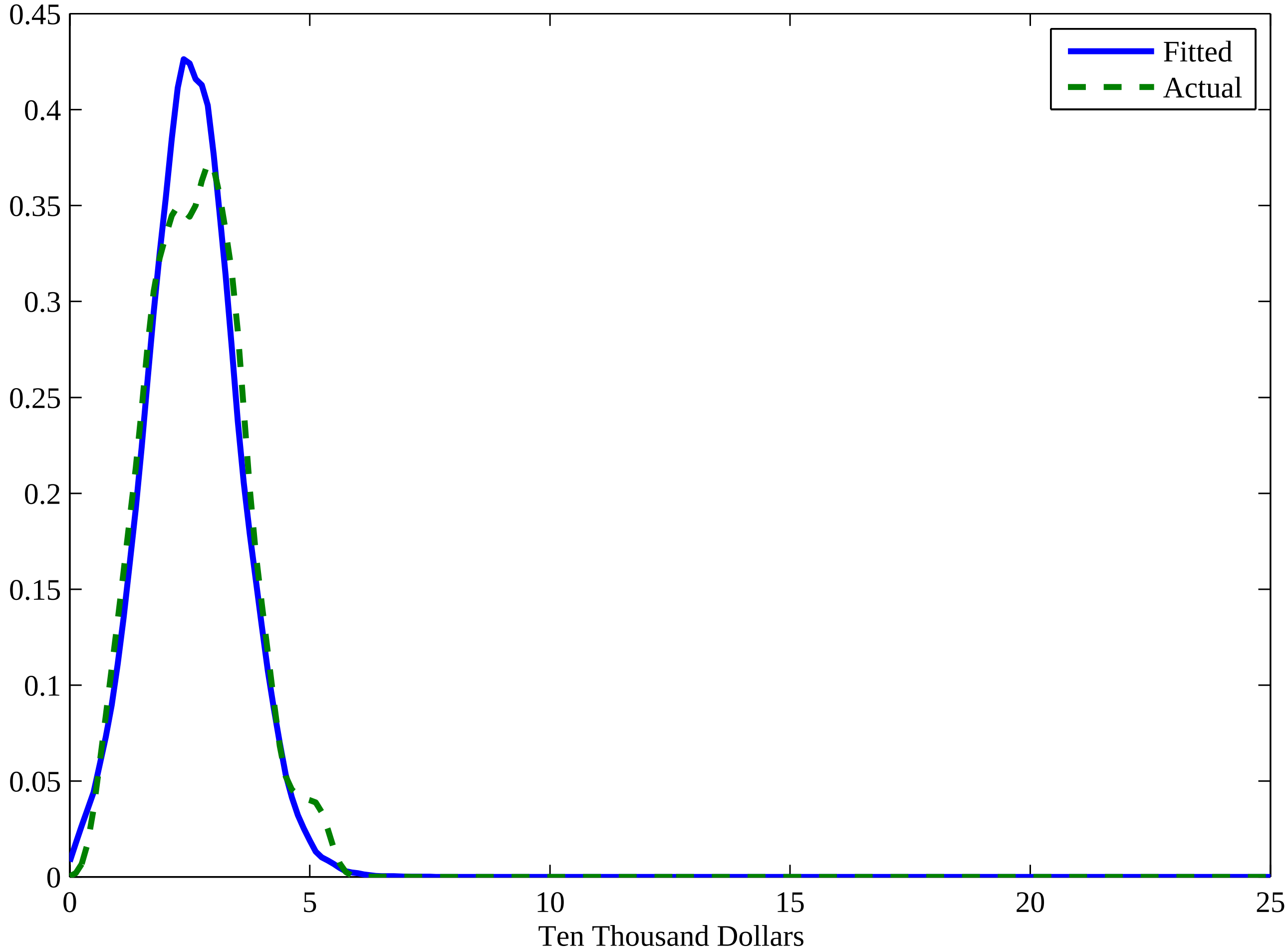
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.15
Densities of earnings at age 36 High School Sample NLSY/79



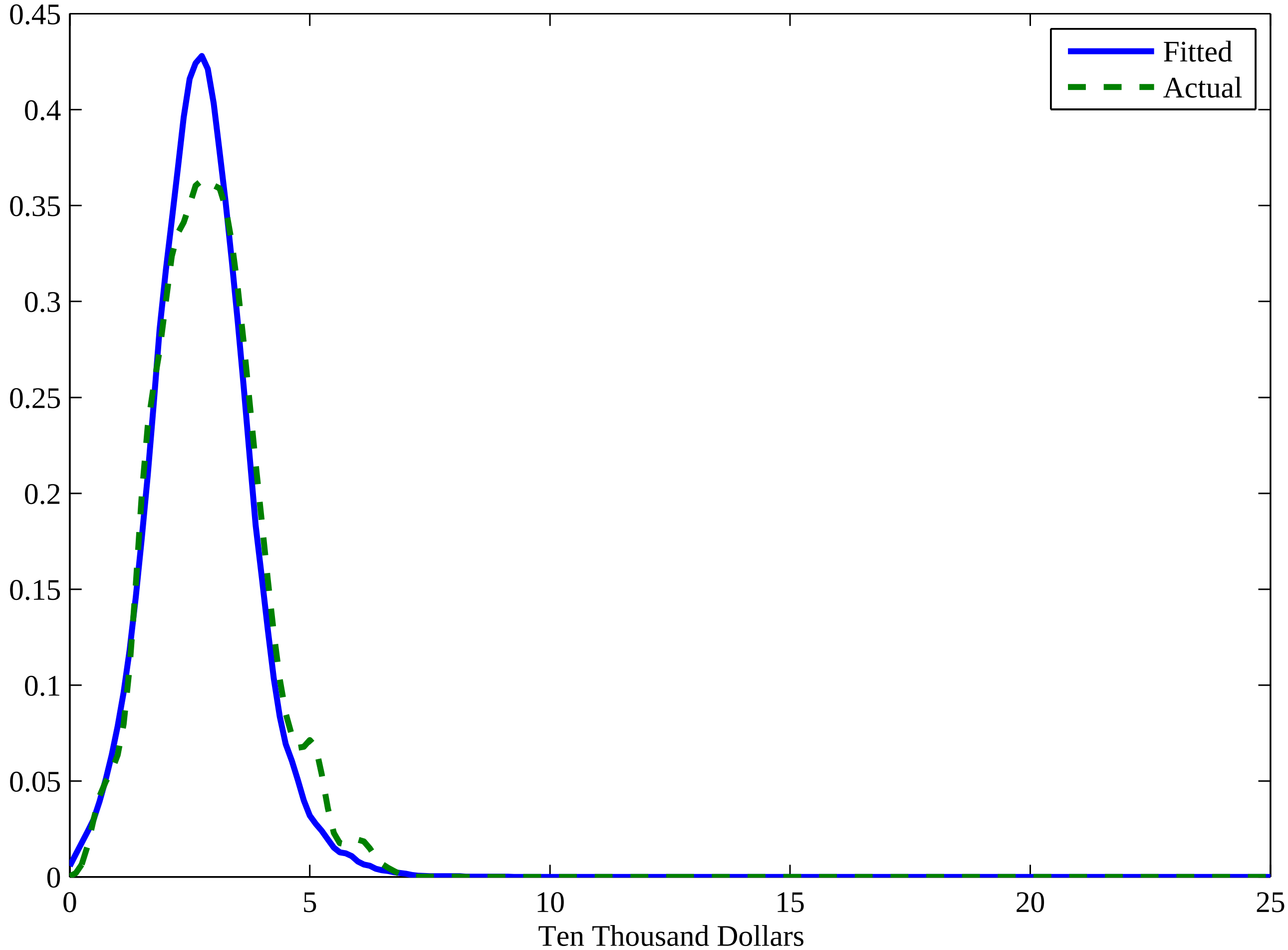
Let Y_0 denote earnings at age 36.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.16
Densities of earnings at age 22 High School Sample NLS/66



Let Y_0 denote earnings at age 22.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

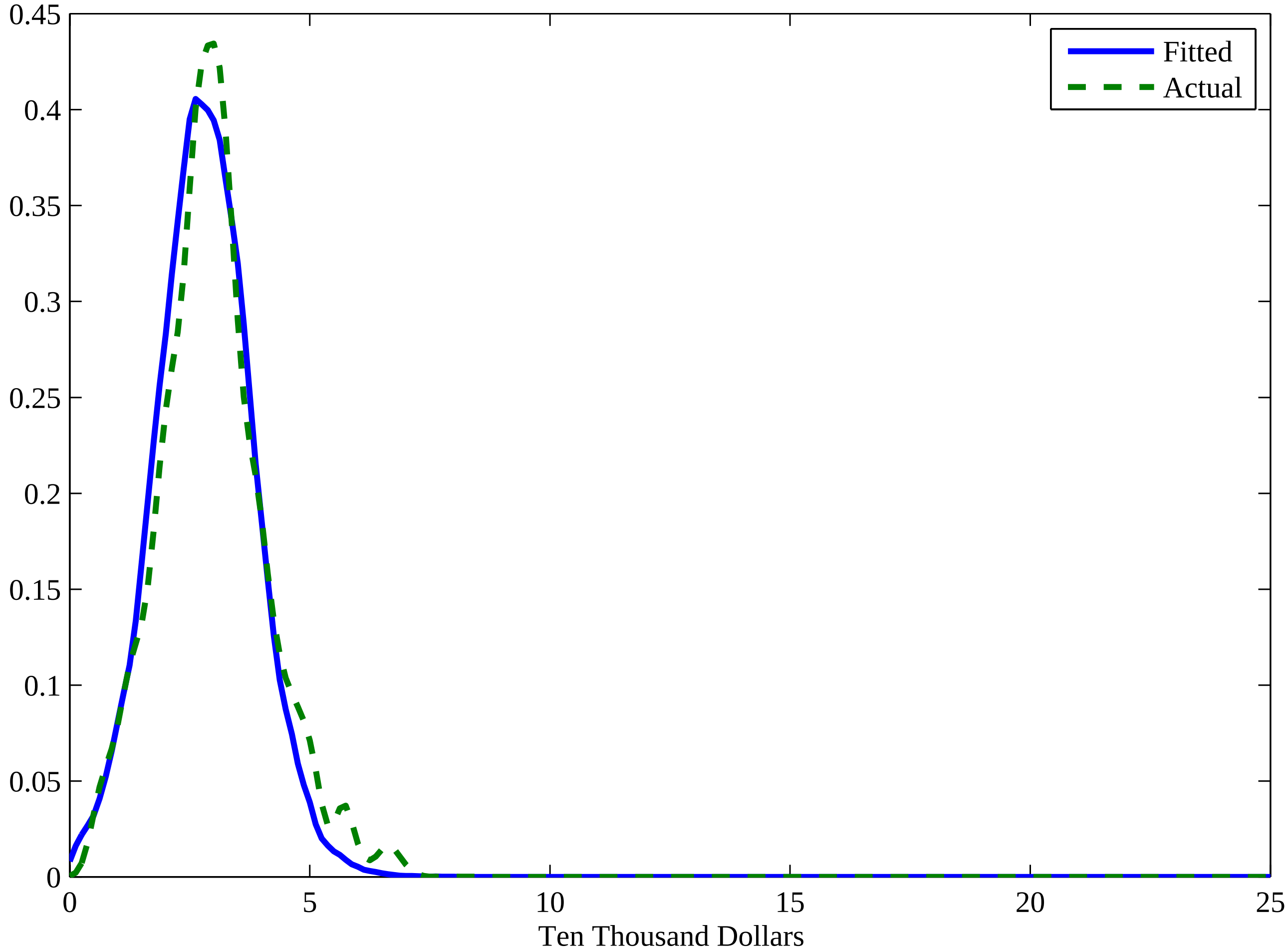
Web Data Appendix Figure II-2.17
Densities of earnings at age 23 High School Sample NLS/66



Let Y_0 denote earnings at age 23.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.18

Densities of earnings at age 24 High School Sample NLS/66

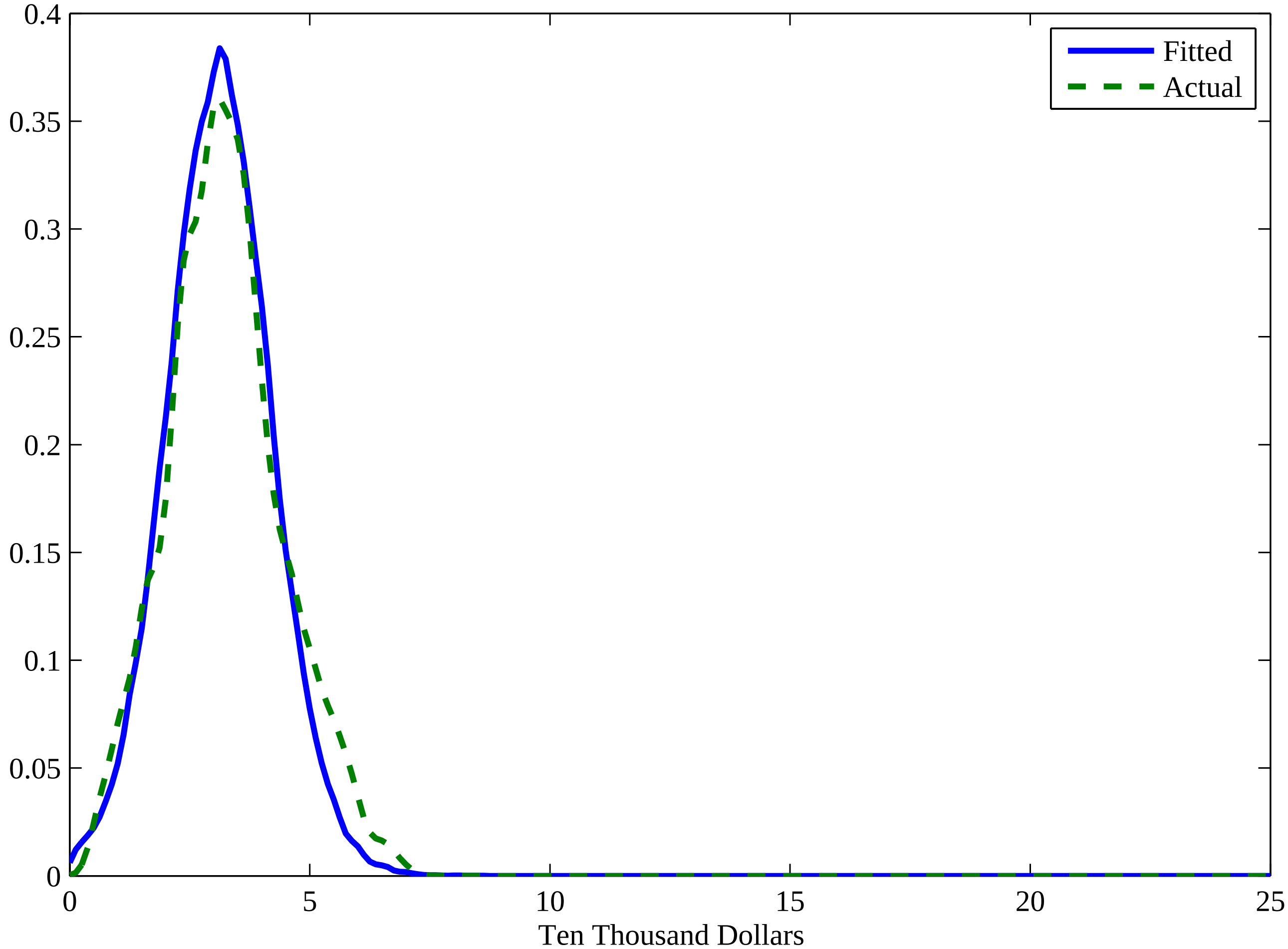


Let Y_0 denote earnings at age 24.

Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.19

Densities of earnings at age 25 High School Sample NLS/66

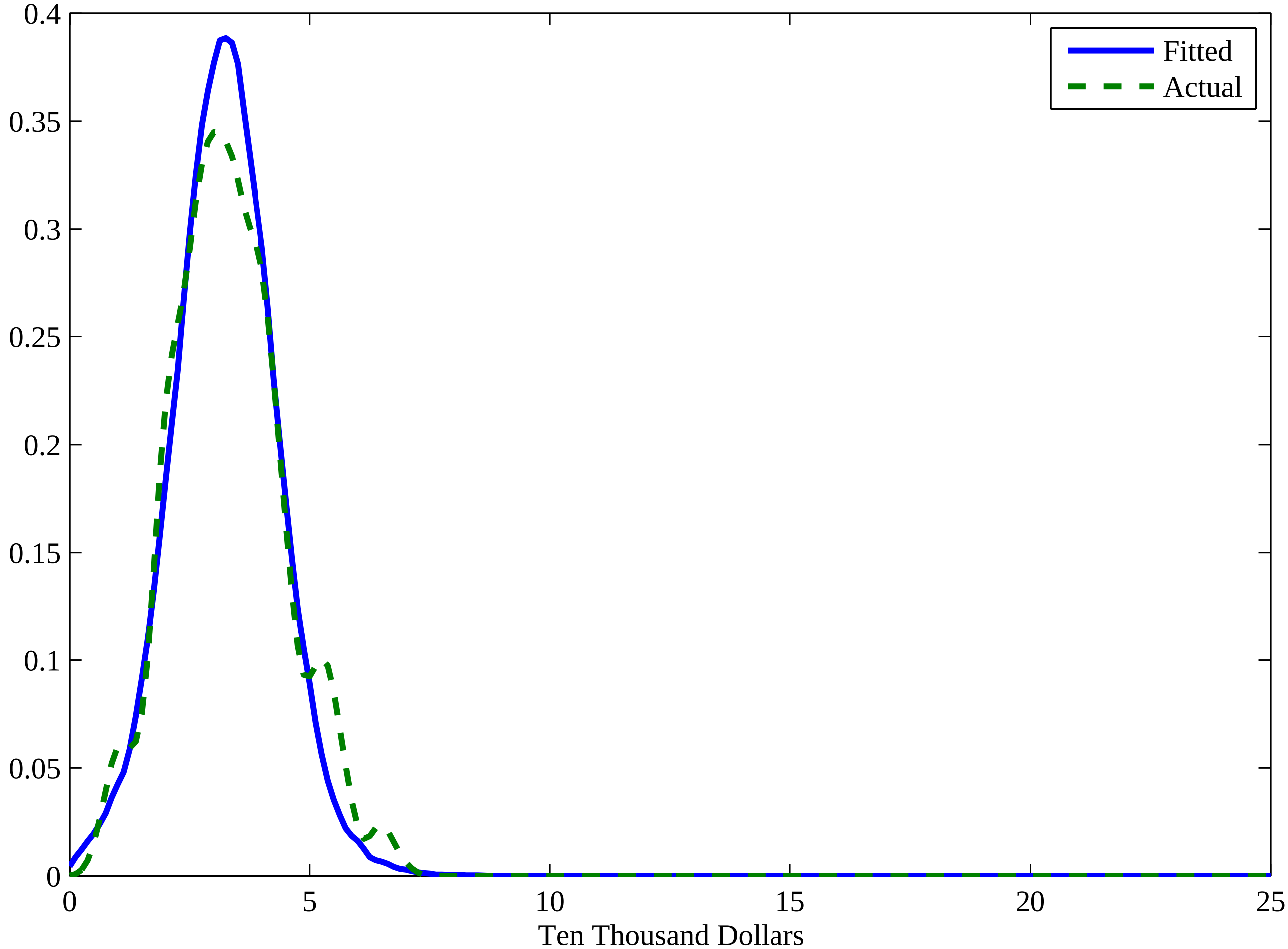


Let Y_0 denote earnings at age 25.

Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.20

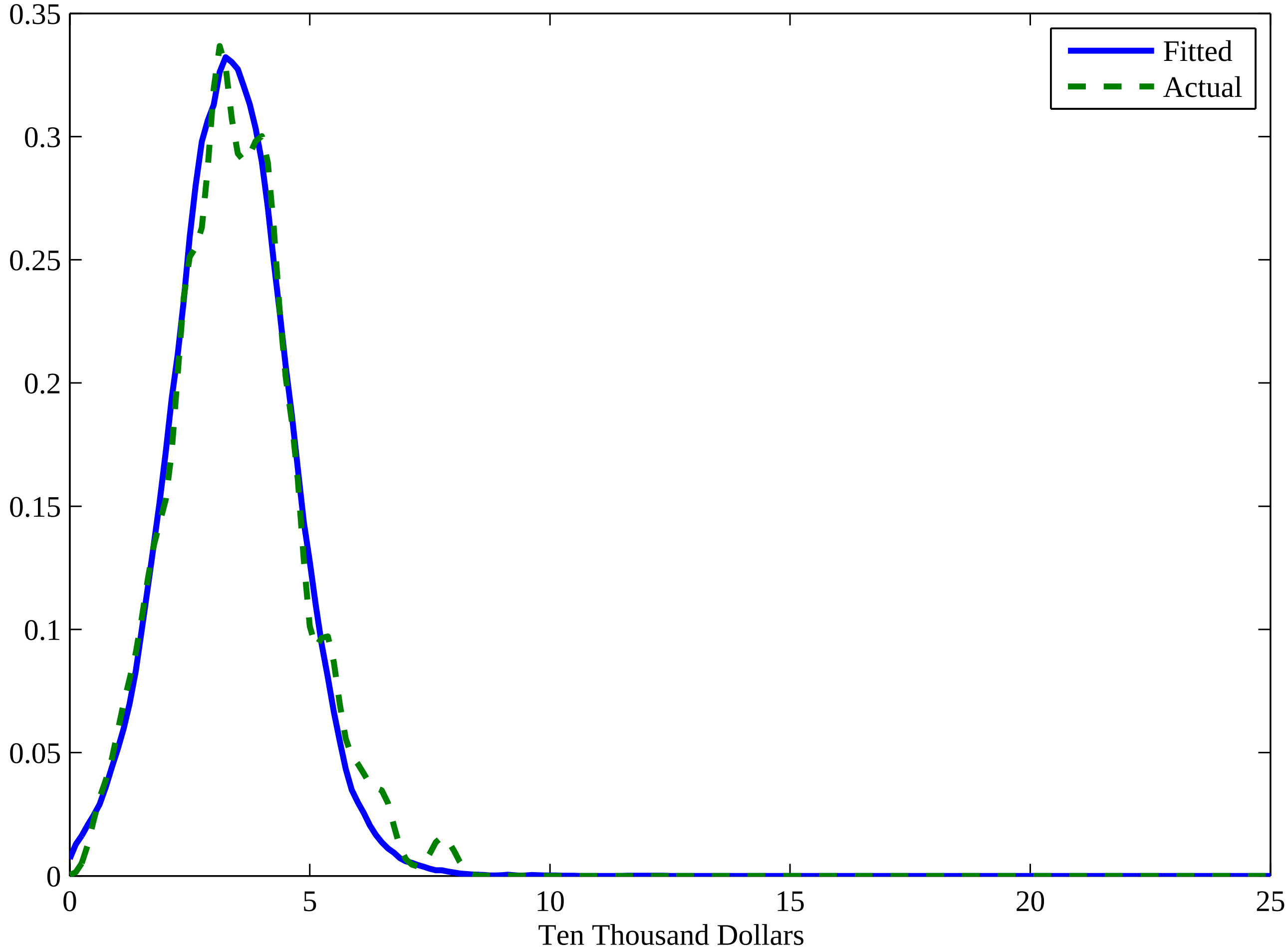
Densities of earnings at age 26 High School Sample NLS/66



Let Y_0 denote earnings at age 26.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

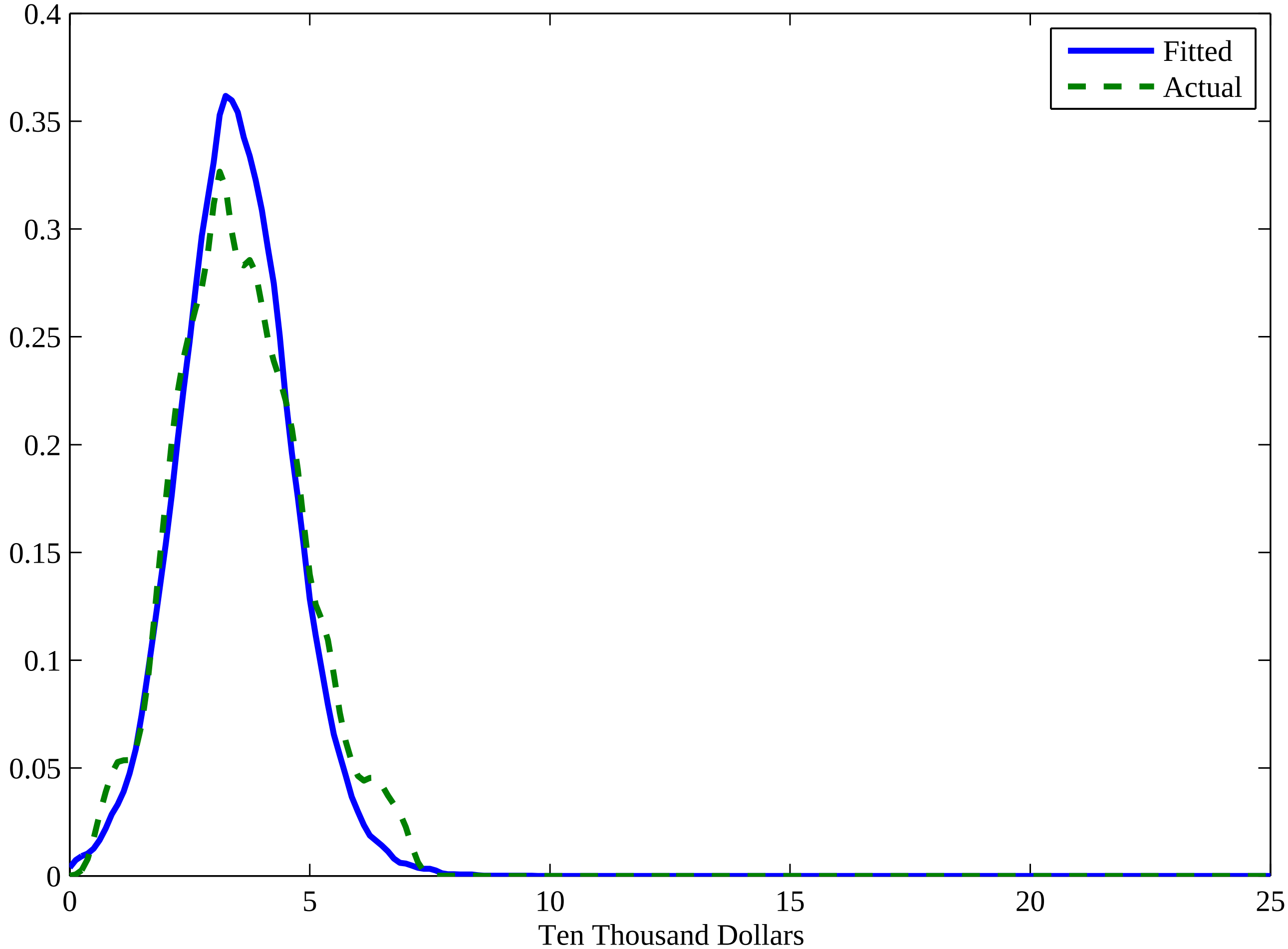
Web Data Appendix Figure II-2.21

Densities of earnings at age 27 High School Sample NLS/66



Let Y_0 denote earnings at age 27.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

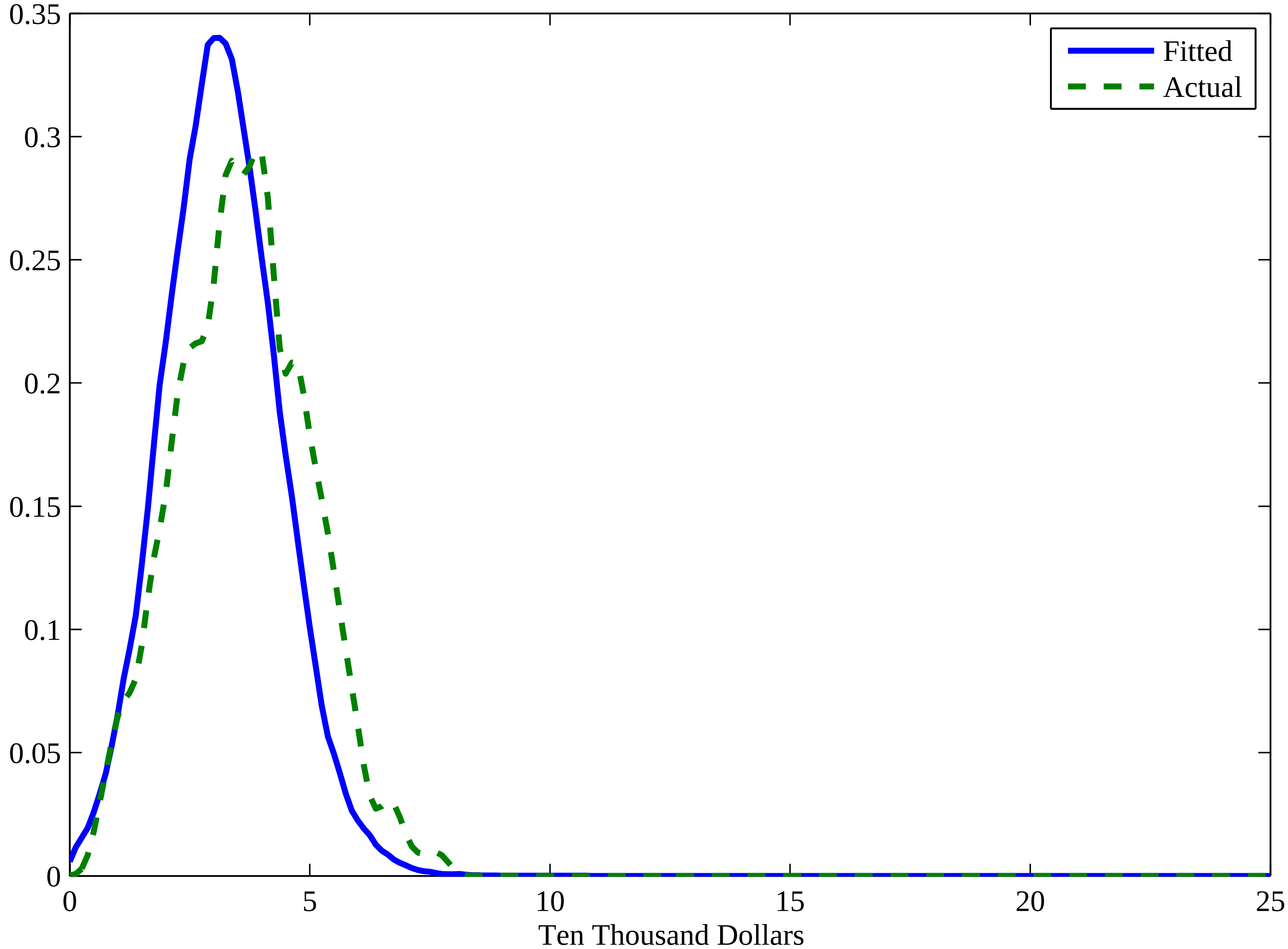
Web Data Appendix Figure II-2.22
Densities of earnings at age 28 High School Sample NLS/66



Let Y_0 denote earnings at age 28.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.23

Densities of earnings at age 29 High School Sample NLS/66

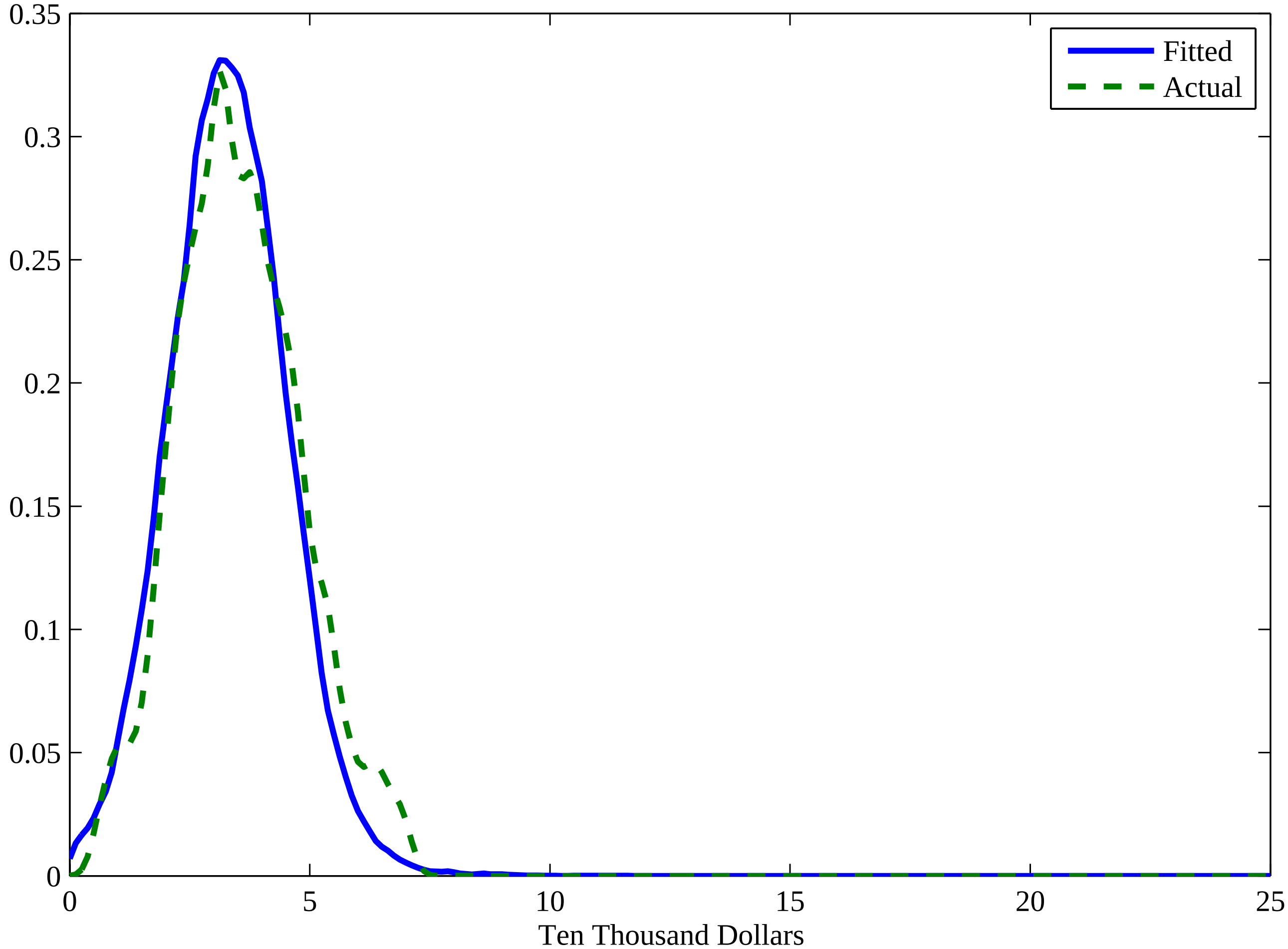


Let Y_0 denote earnings at age 29.

Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

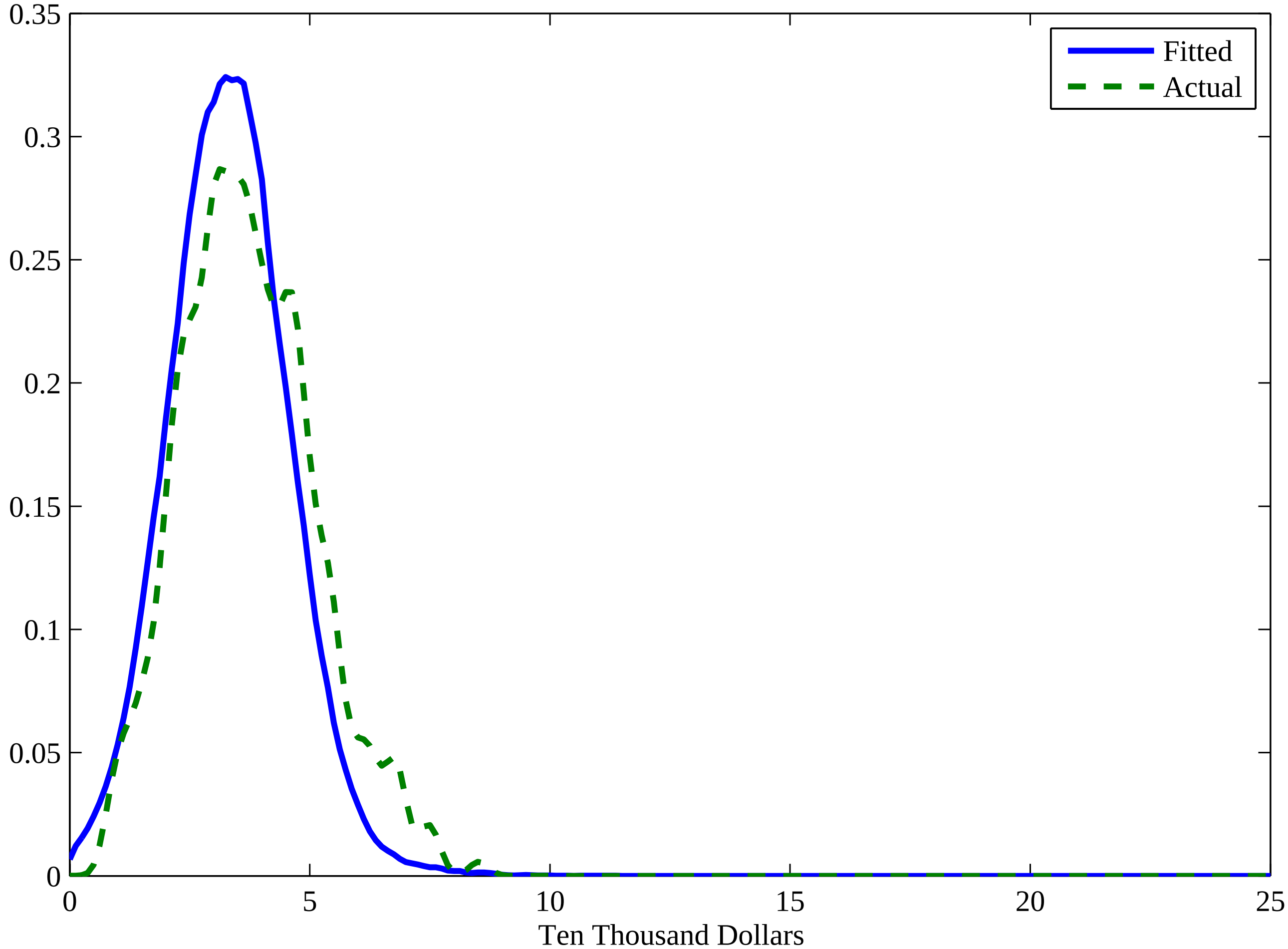
Web Data Appendix Figure II-2.24

Densities of earnings at age 30 High School Sample NLS/66



Let Y_0 denote earnings at age 30.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

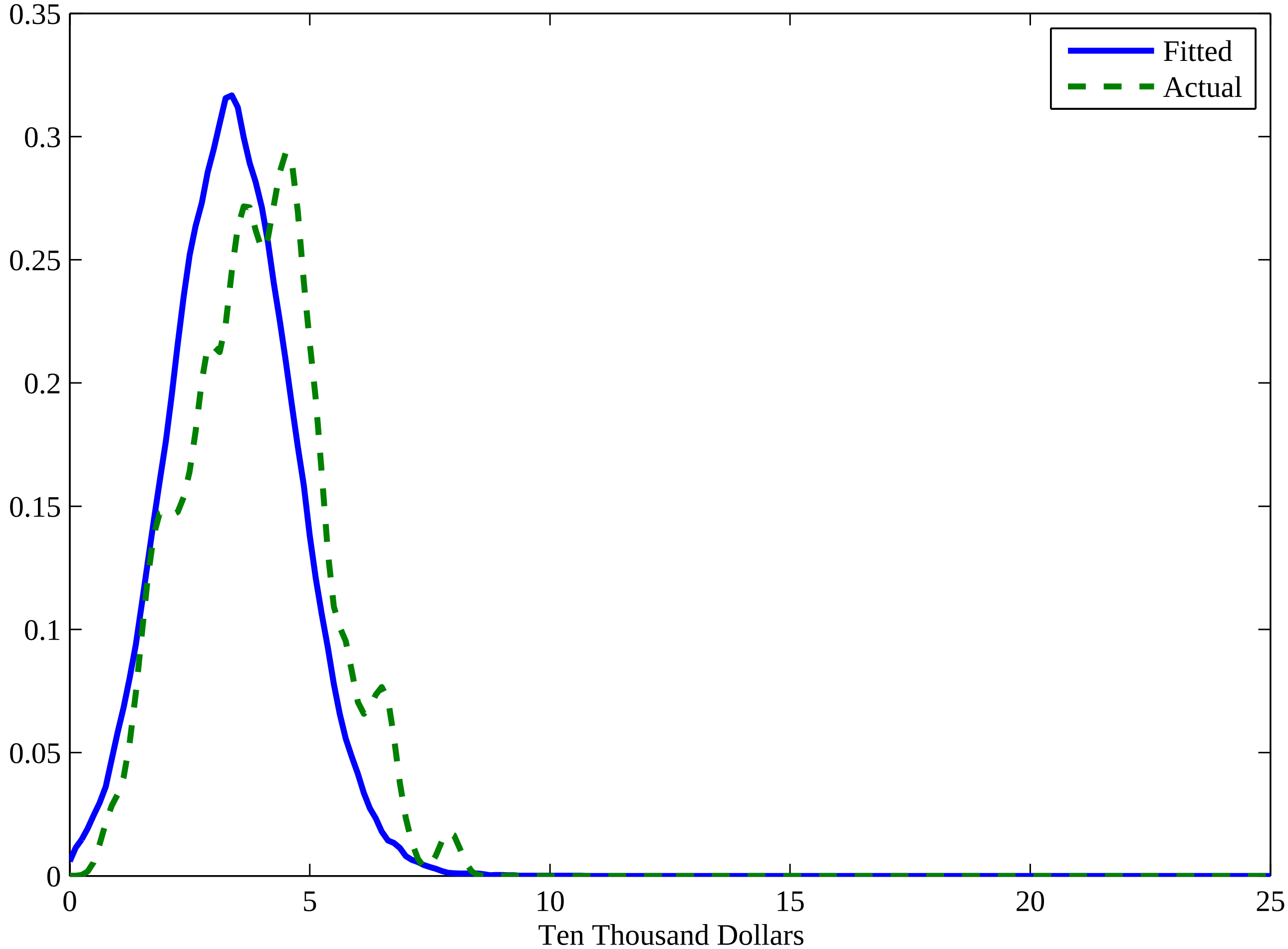
Web Data Appendix Figure II-2.25
Densities of earnings at age 31 High School Sample NLS/66



Let Y_0 denote earnings at age 31.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

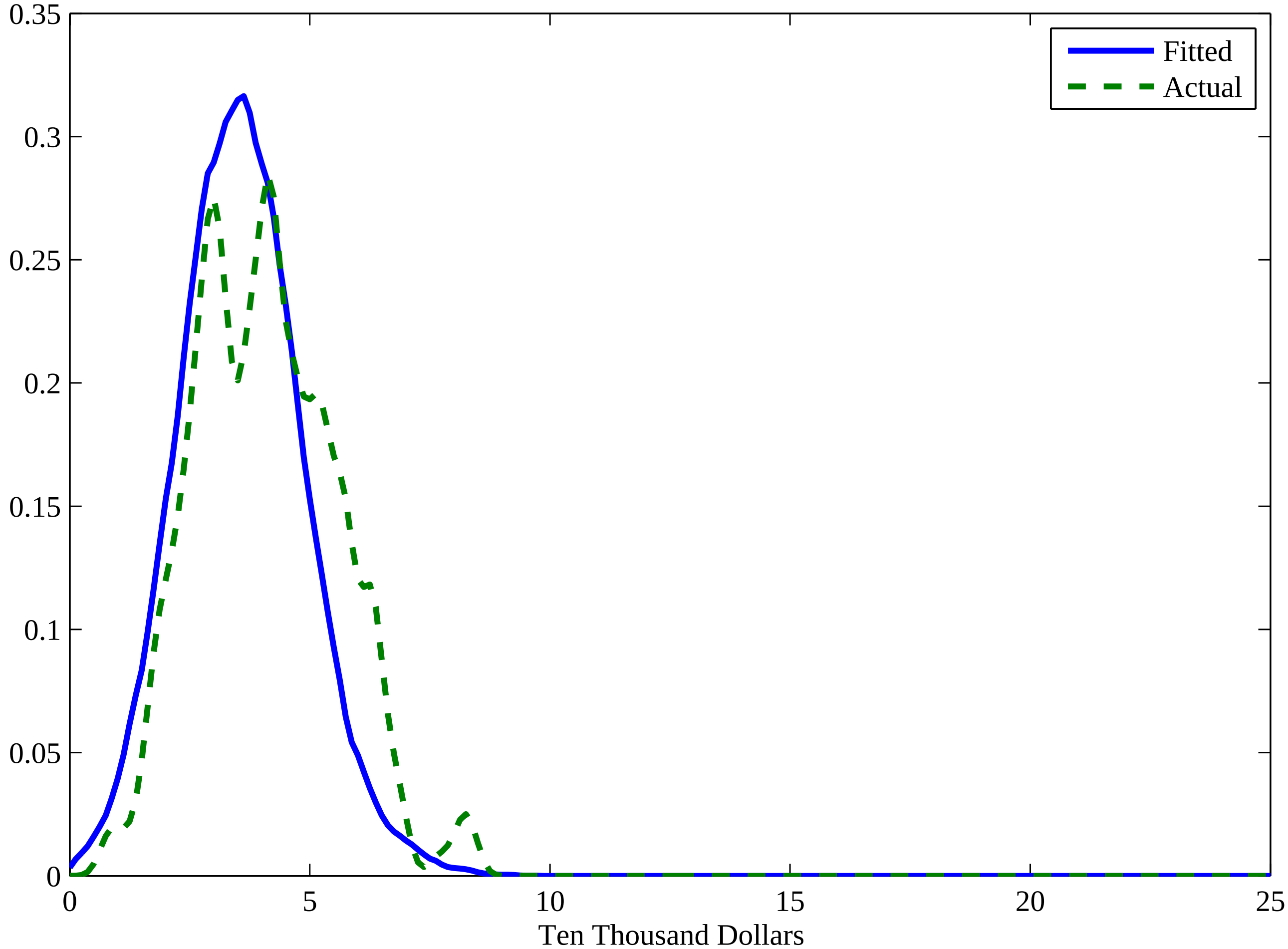
Web Data Appendix Figure II-2.26

Densities of earnings at age 32 High School Sample NLS/66



Let Y_0 denote earnings at age 32.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

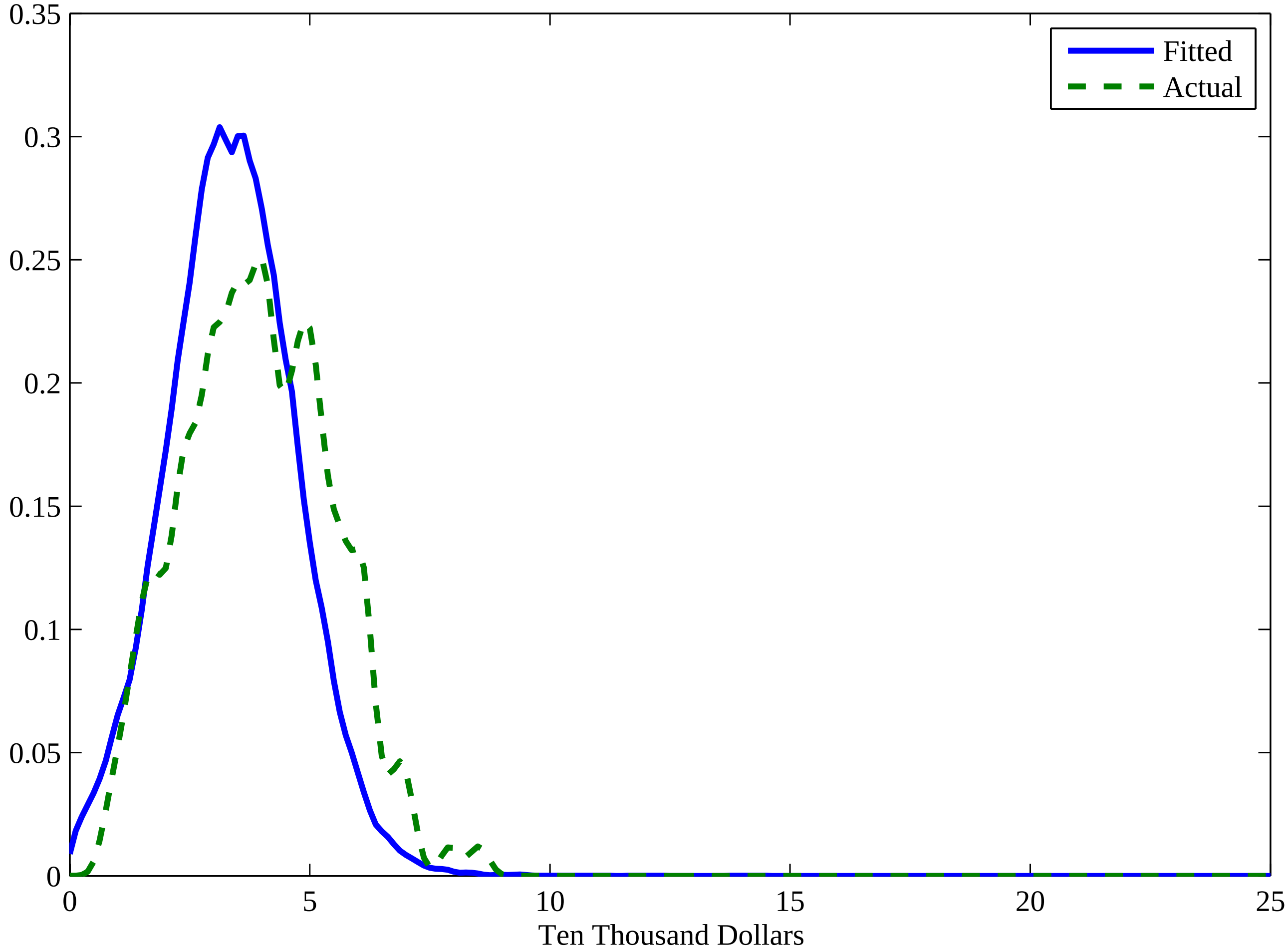
Web Data Appendix Figure II-2.27
Densities of earnings at age 33 High School Sample NLS/66



Let Y_0 denote earnings at age 33.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.28

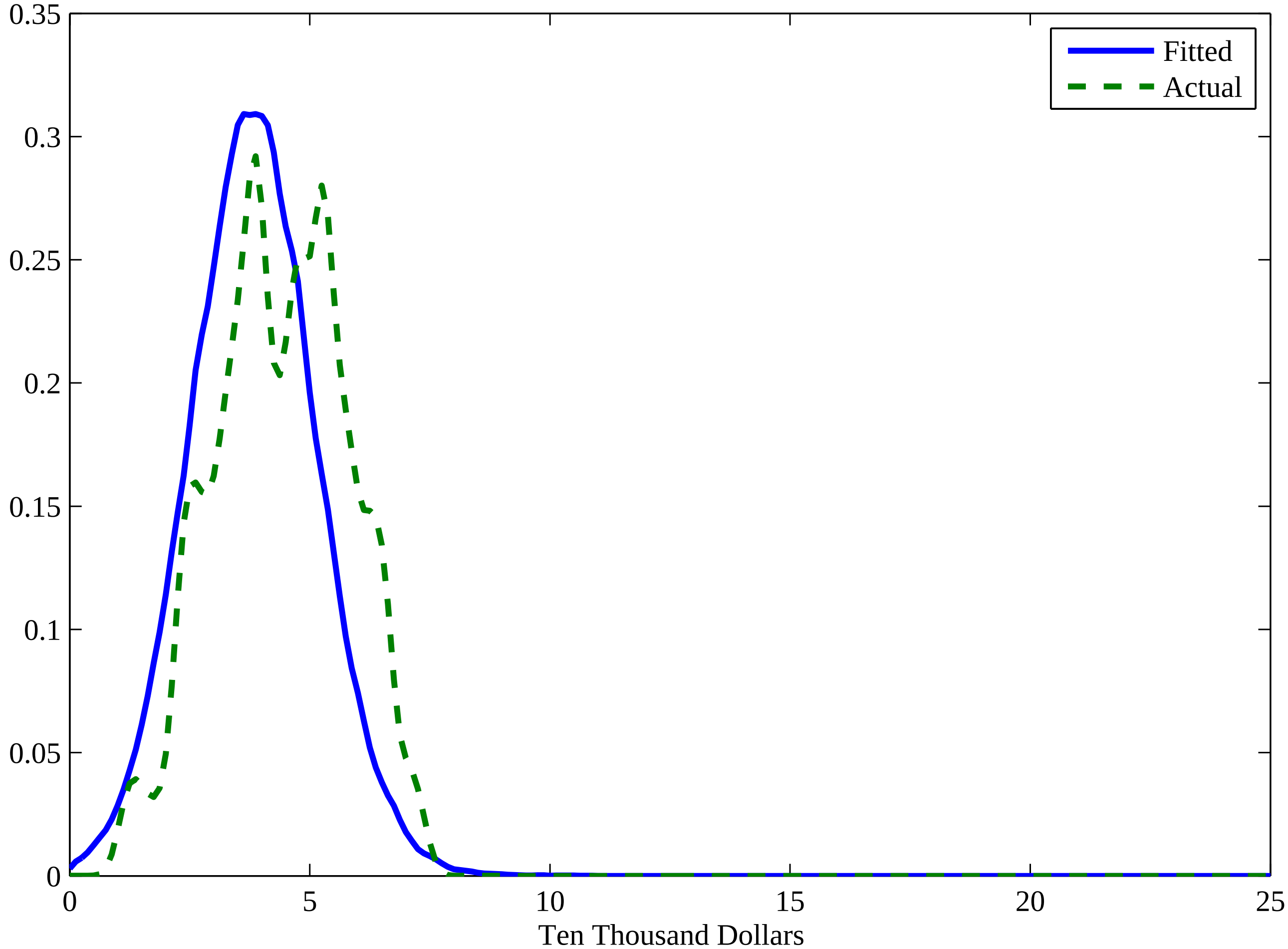
Densities of earnings at age 34 High School Sample NLS/66



Let Y_0 denote earnings at age 34.

Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

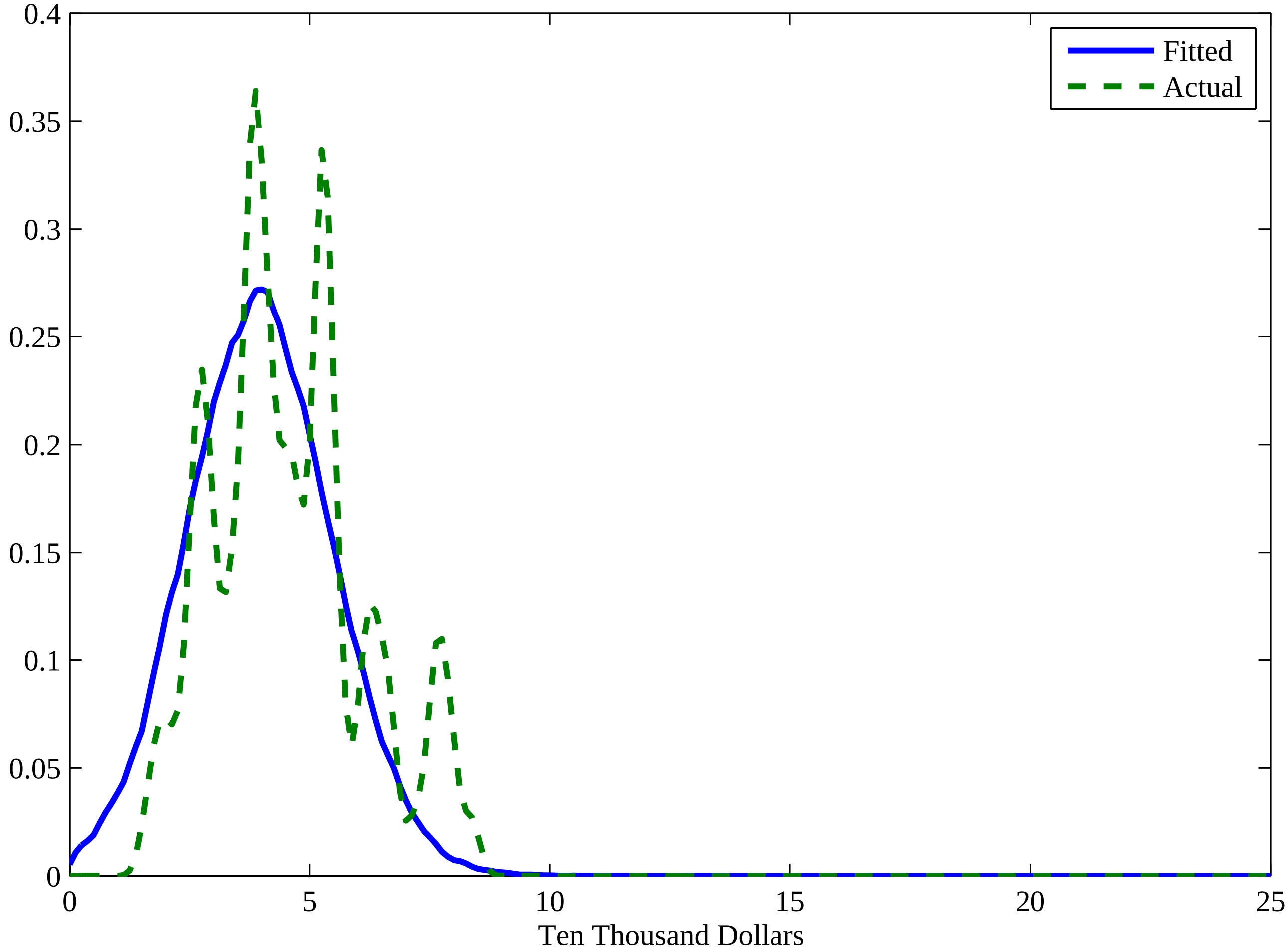
Web Data Appendix Figure II-2.29
Densities of earnings at age 35 High School Sample NLS/66



Let Y_0 denote earnings at age 35.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-2.30

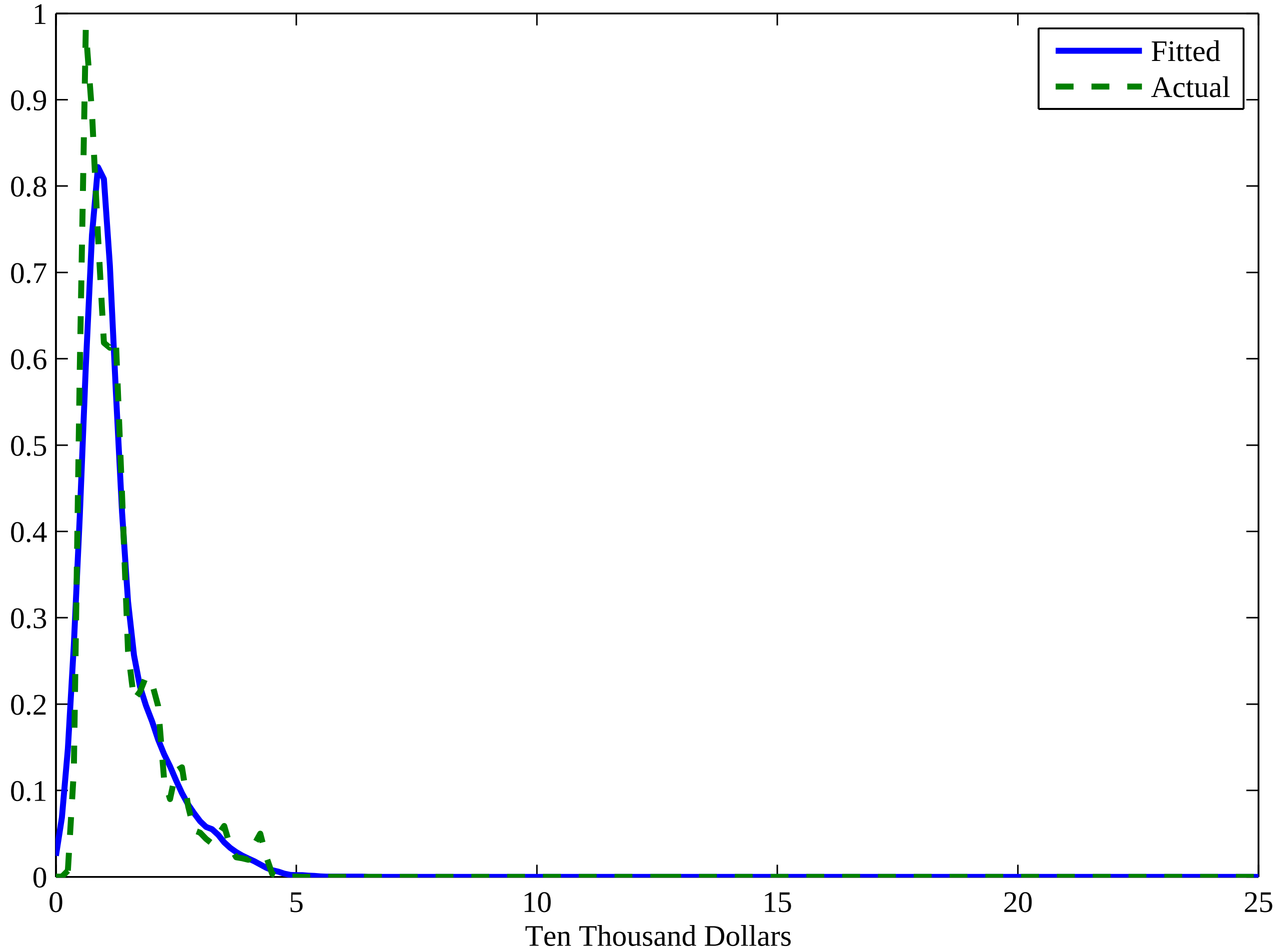
Densities of earnings at age 36 High School Sample NLS/66



Let Y_0 denote earnings at age 36.
Here we plot the density functions $f(y_0 \mid S = 0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.1

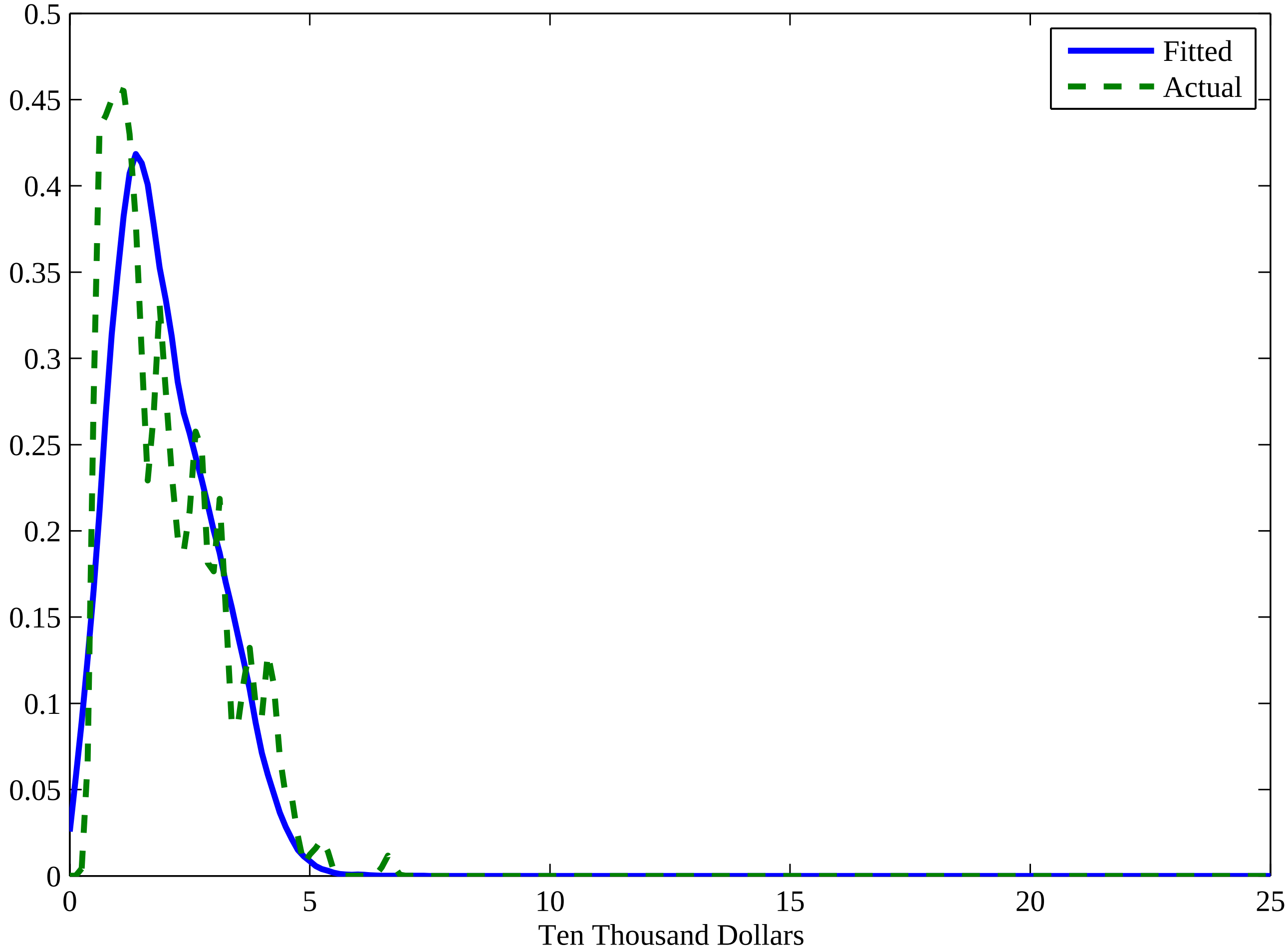
Densities of earnings at age 22 College Sample NLSY/79



Let Y_1 denote earnings at age 22.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.2

Densities of earnings at age 23 College Sample NLSY/79

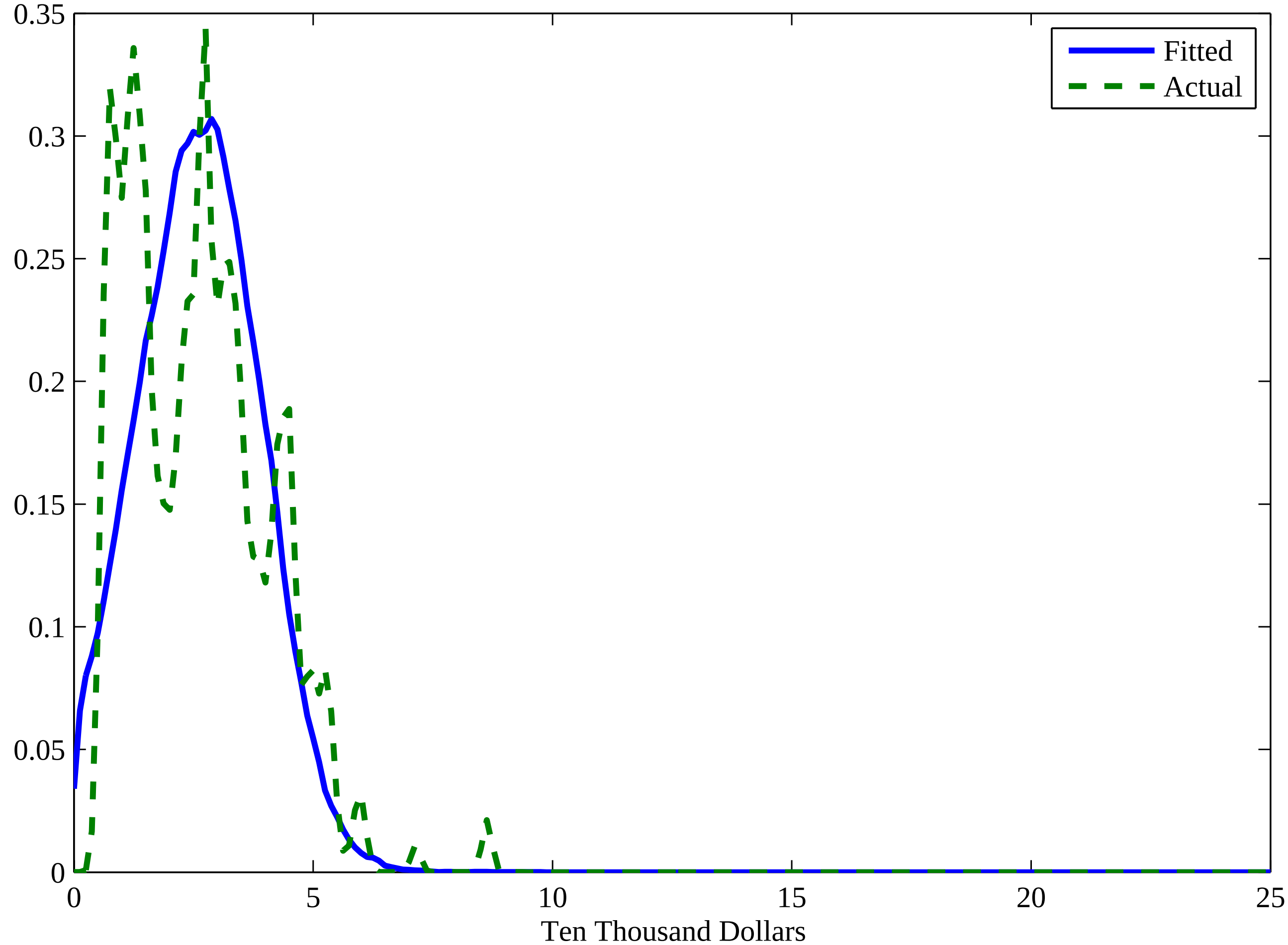


Let Y_1 denote earnings at age 23.

Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.3

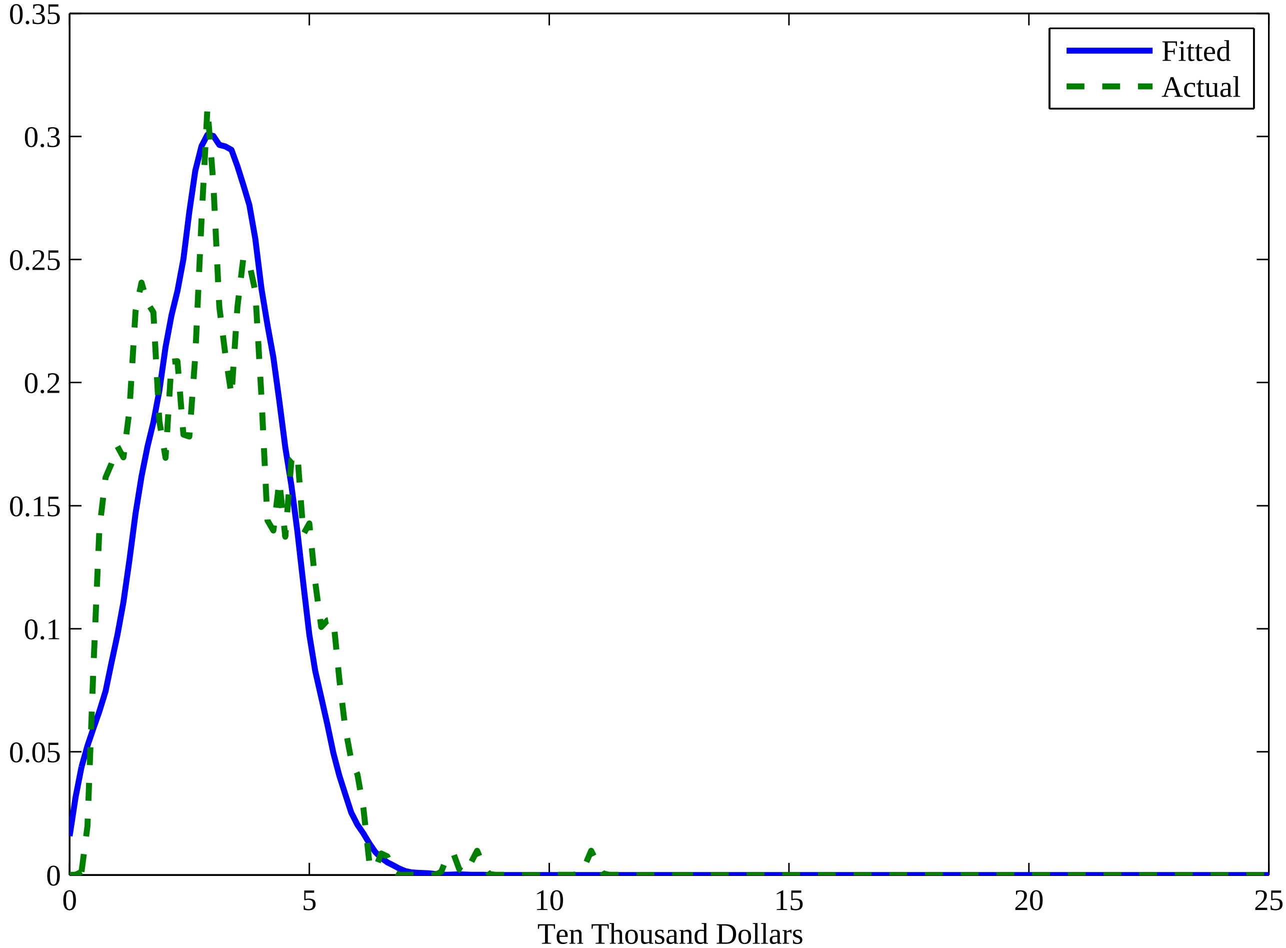
Densities of earnings at age 24 College Sample NLSY/79



Let Y_1 denote earnings at age 24.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.4

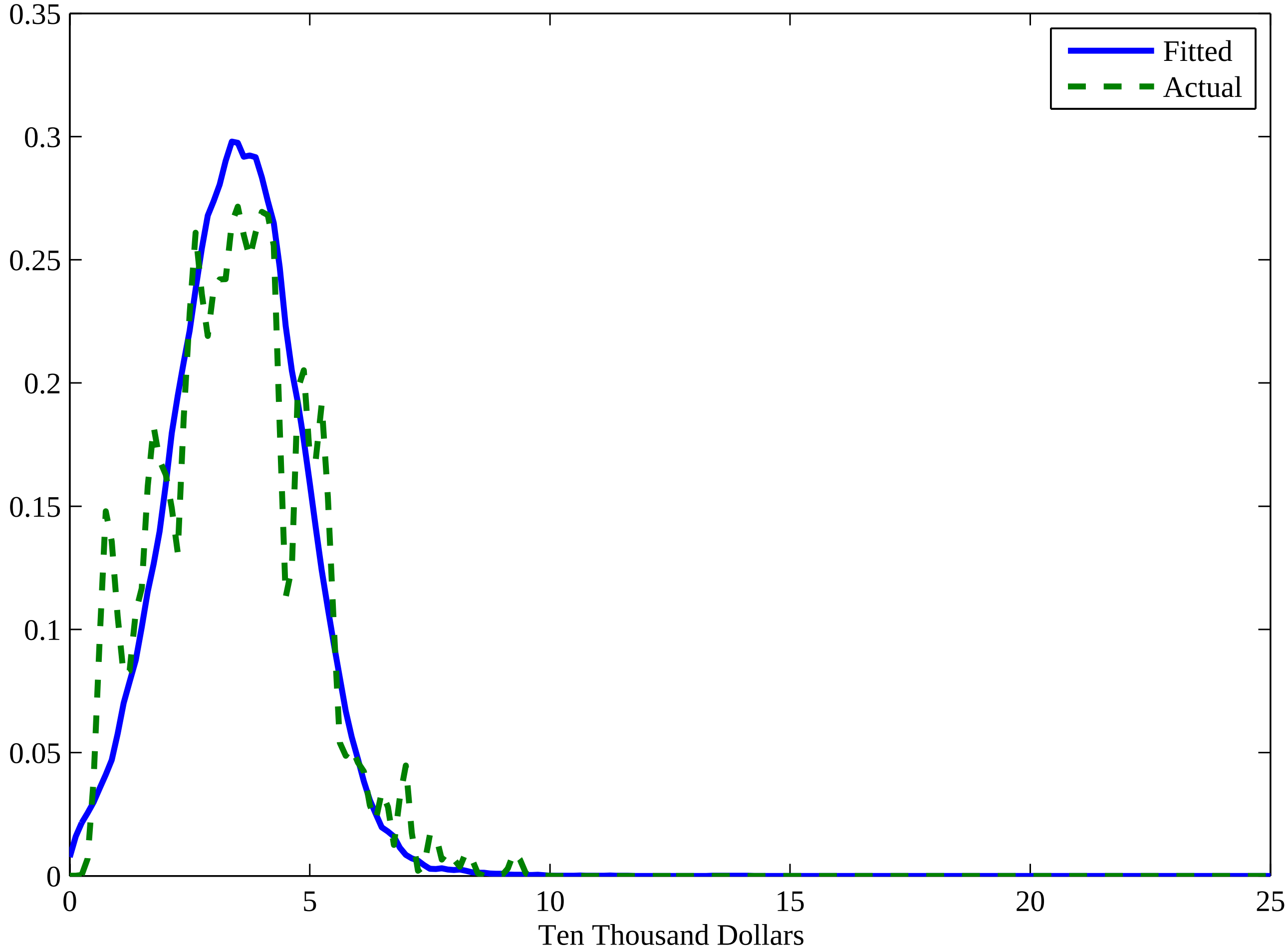
Densities of earnings at age 25 College Sample NLSY/79



Let Y_1 denote earnings at age 25.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.5

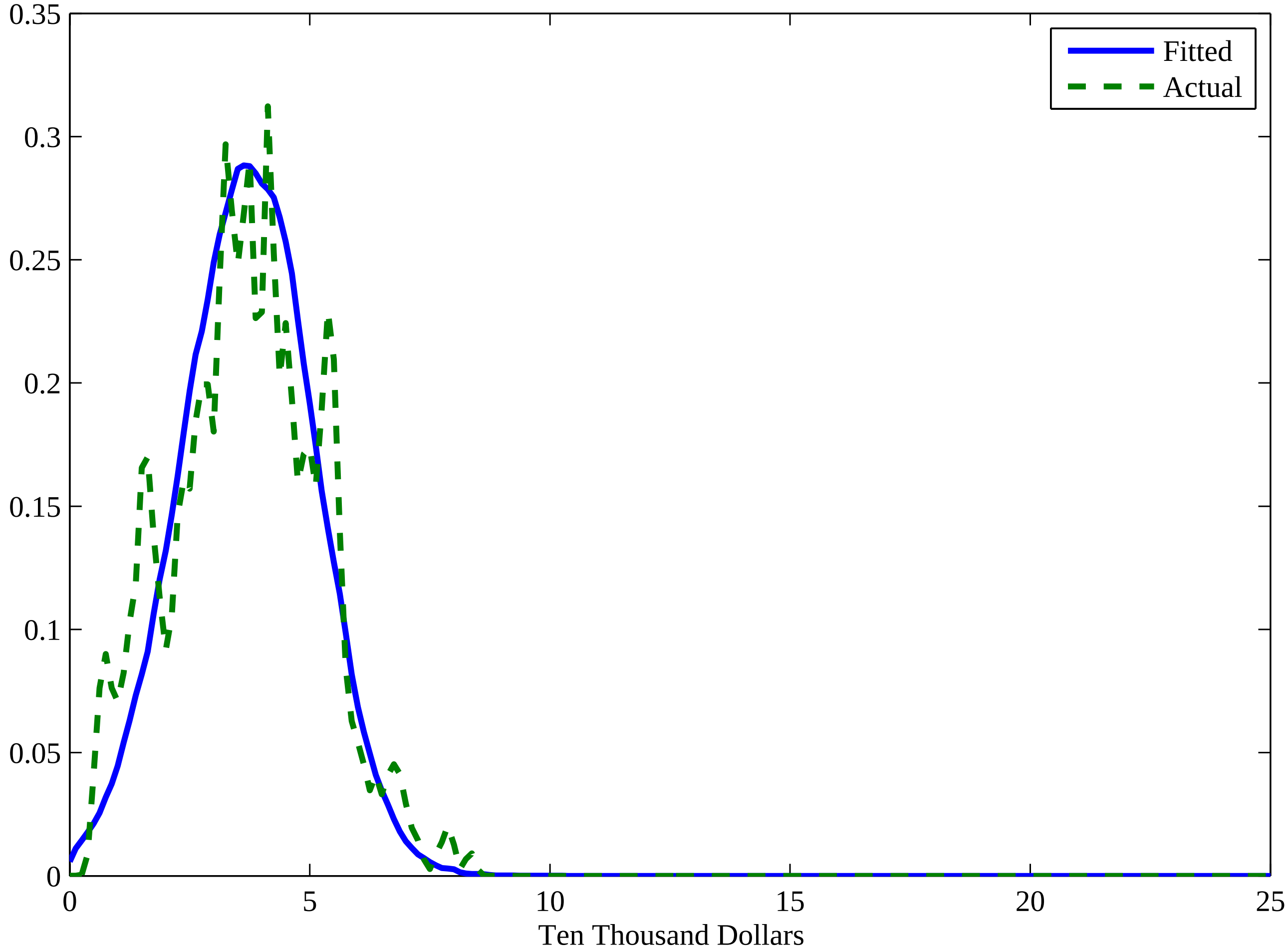
Densities of earnings at age 26 College Sample NLSY/79



Let Y_1 denote earnings at age 26.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

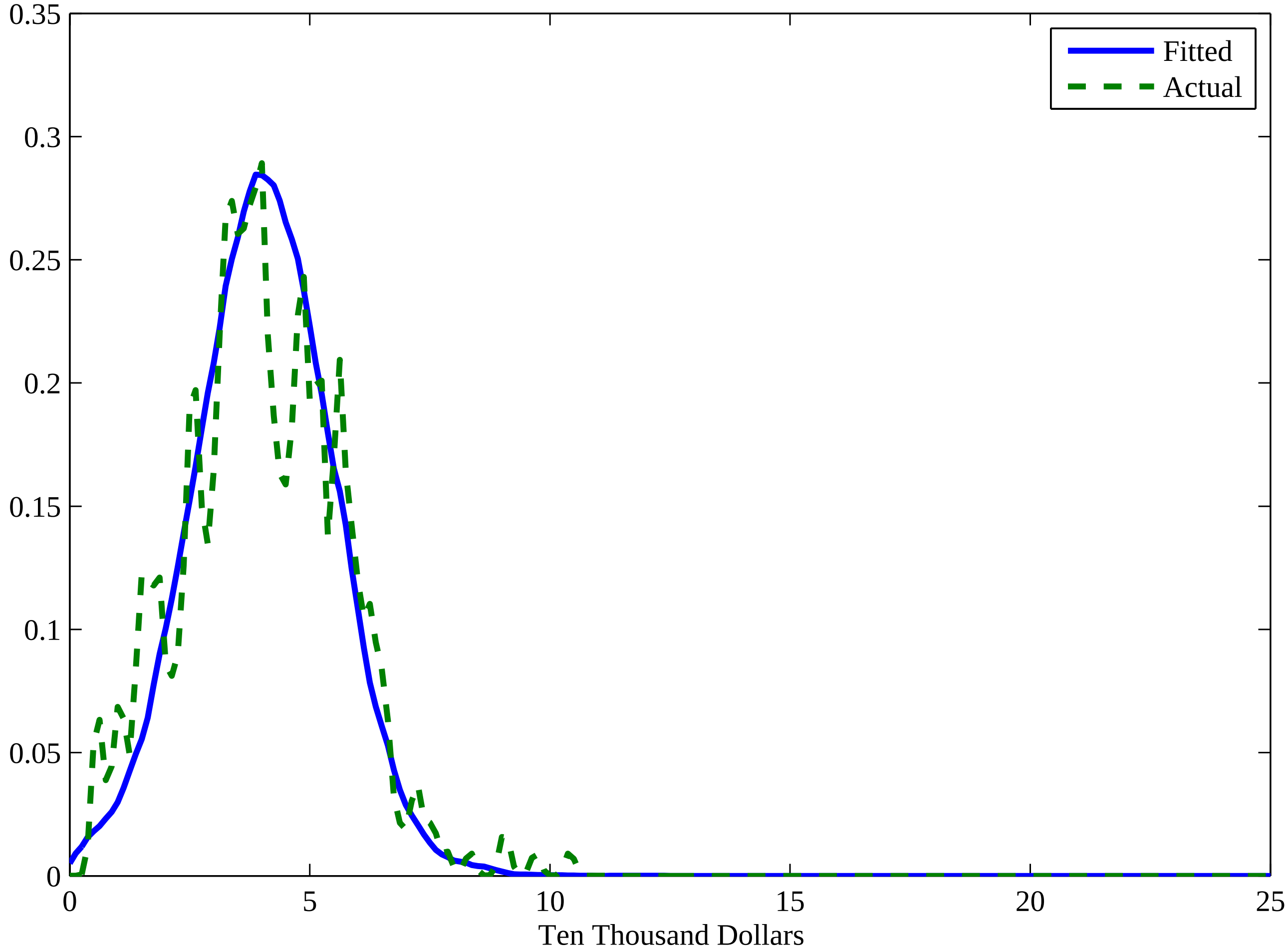
Web Data Appendix Figure II-3.6

Densities of earnings at age 27 College Sample NLSY/79



Let Y_1 denote earnings at age 27.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

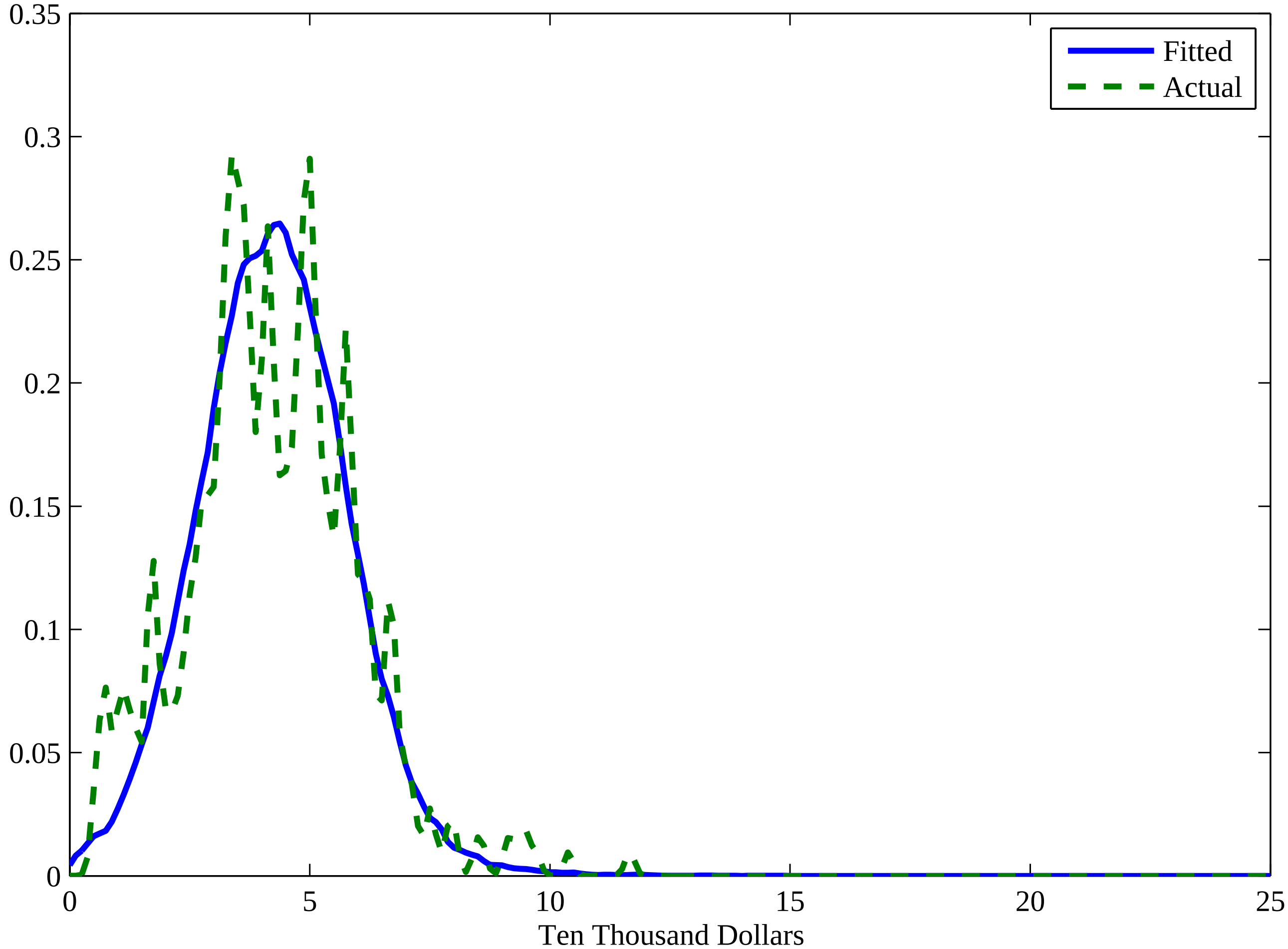
Web Data Appendix Figure II-3.7
Densities of earnings at age 28 College Sample NLSY/79



Let Y_1 denote earnings at age 28.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.8

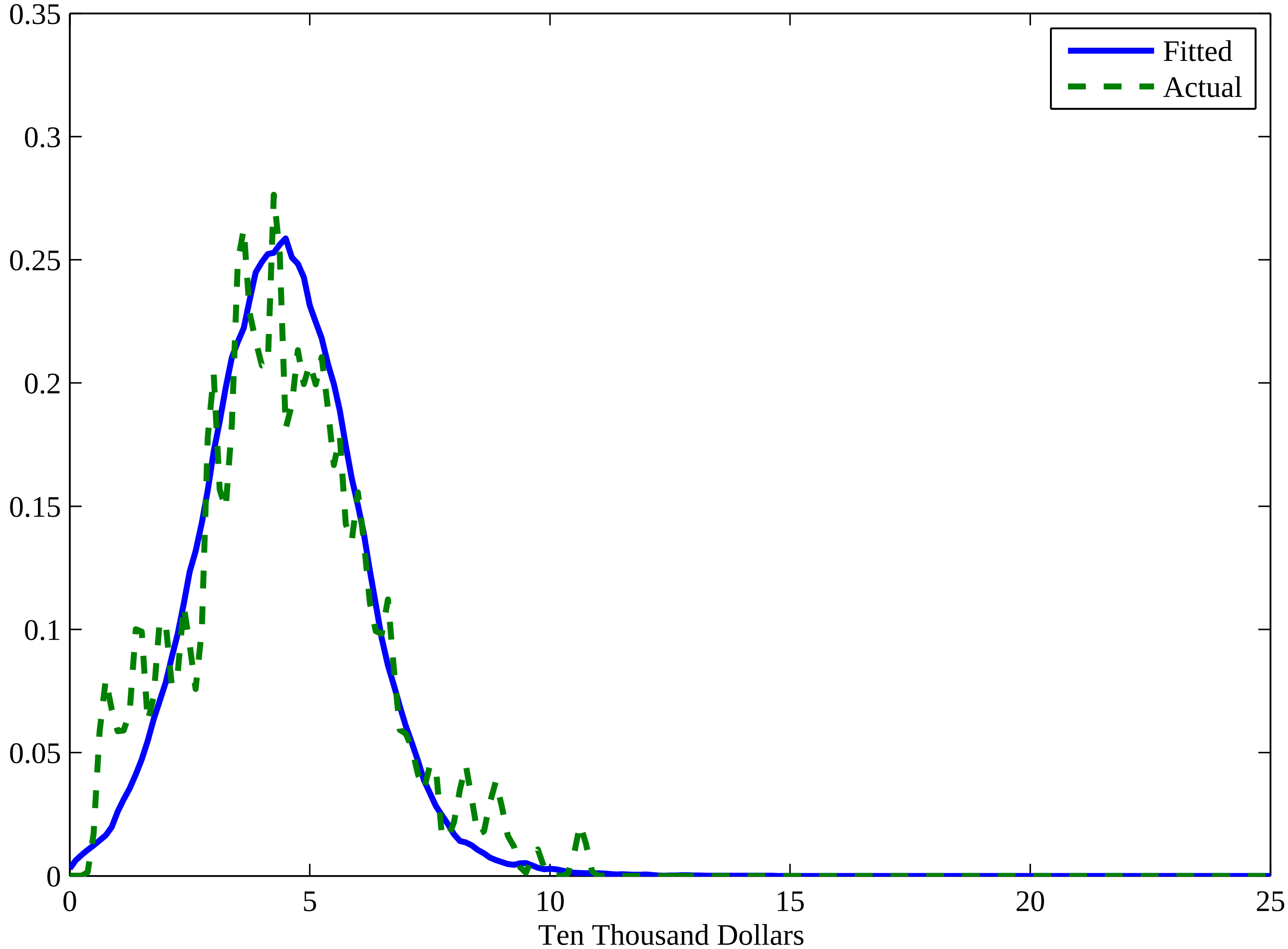
Densities of earnings at age 29 College Sample NLSY/79



Let Y_1 denote earnings at age 29.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.9

Densities of earnings at age 30 College Sample NLSY/79

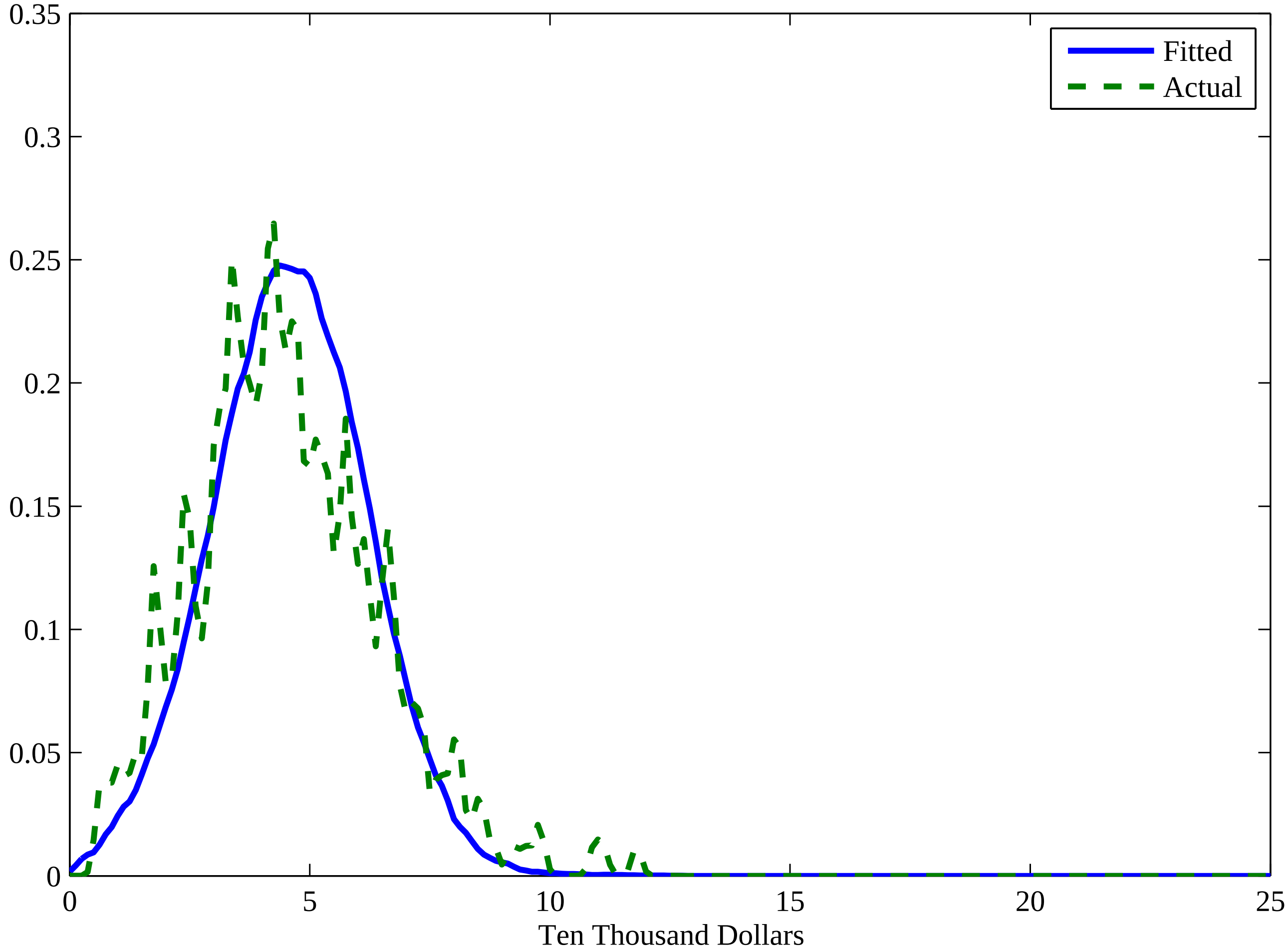


Let Y_1 denote earnings at age 30.

Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.10

Densities of earnings at age 31 College Sample NLSY/79

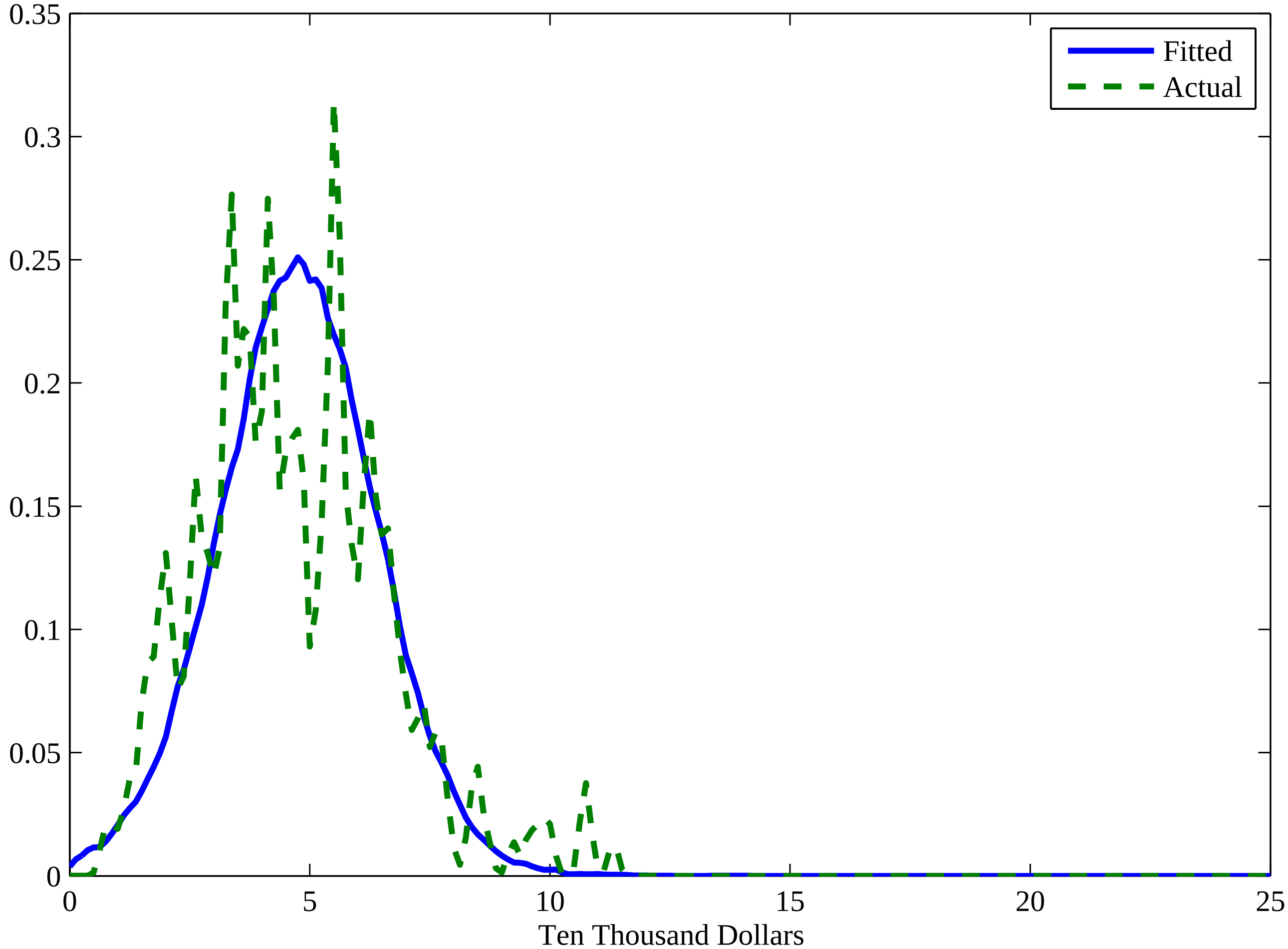


Let Y_1 denote earnings at age 31.

Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

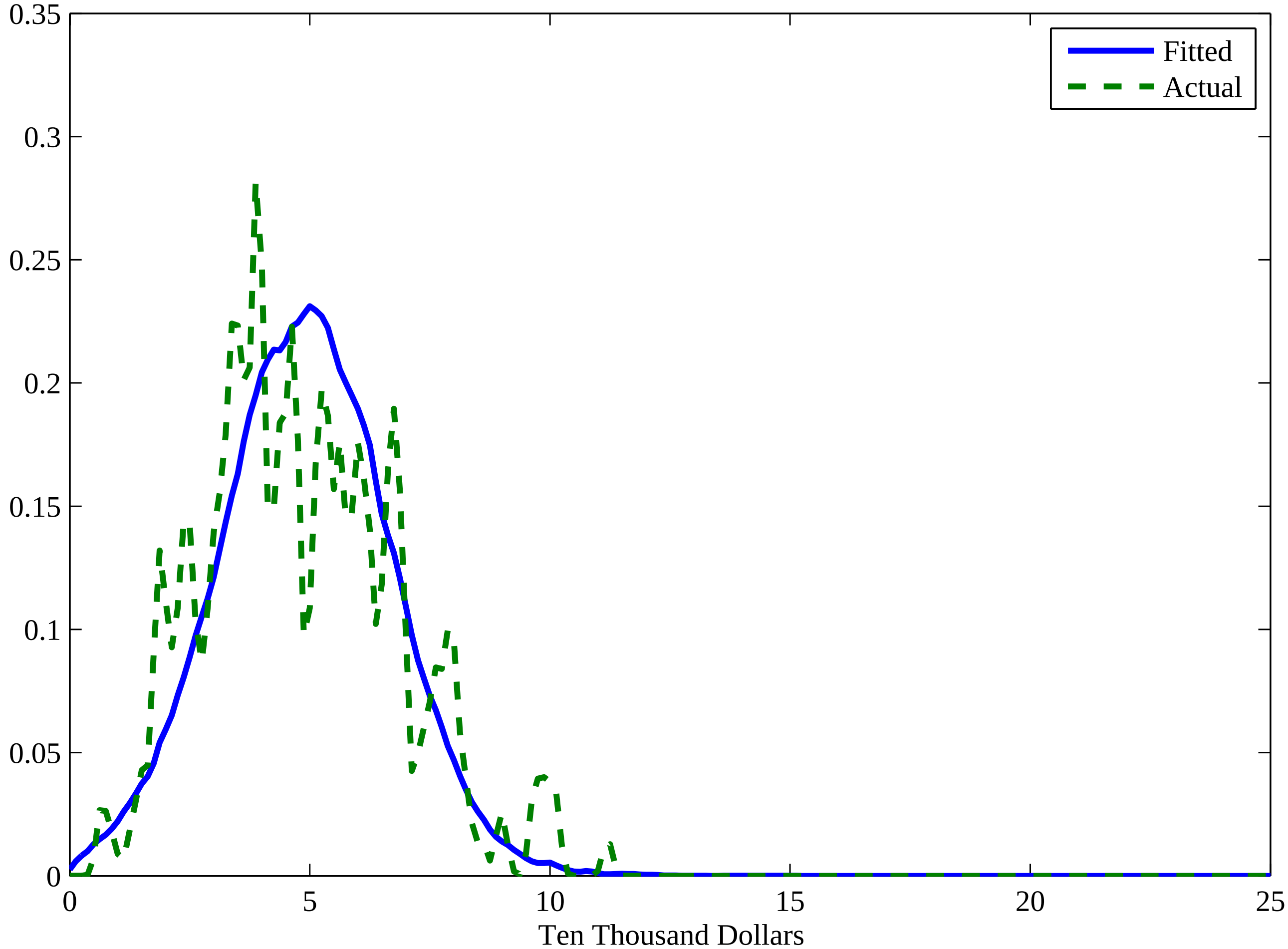
Web Data Appendix Figure II-3.11

Densities of earnings at age 32 College Sample NLSY/79



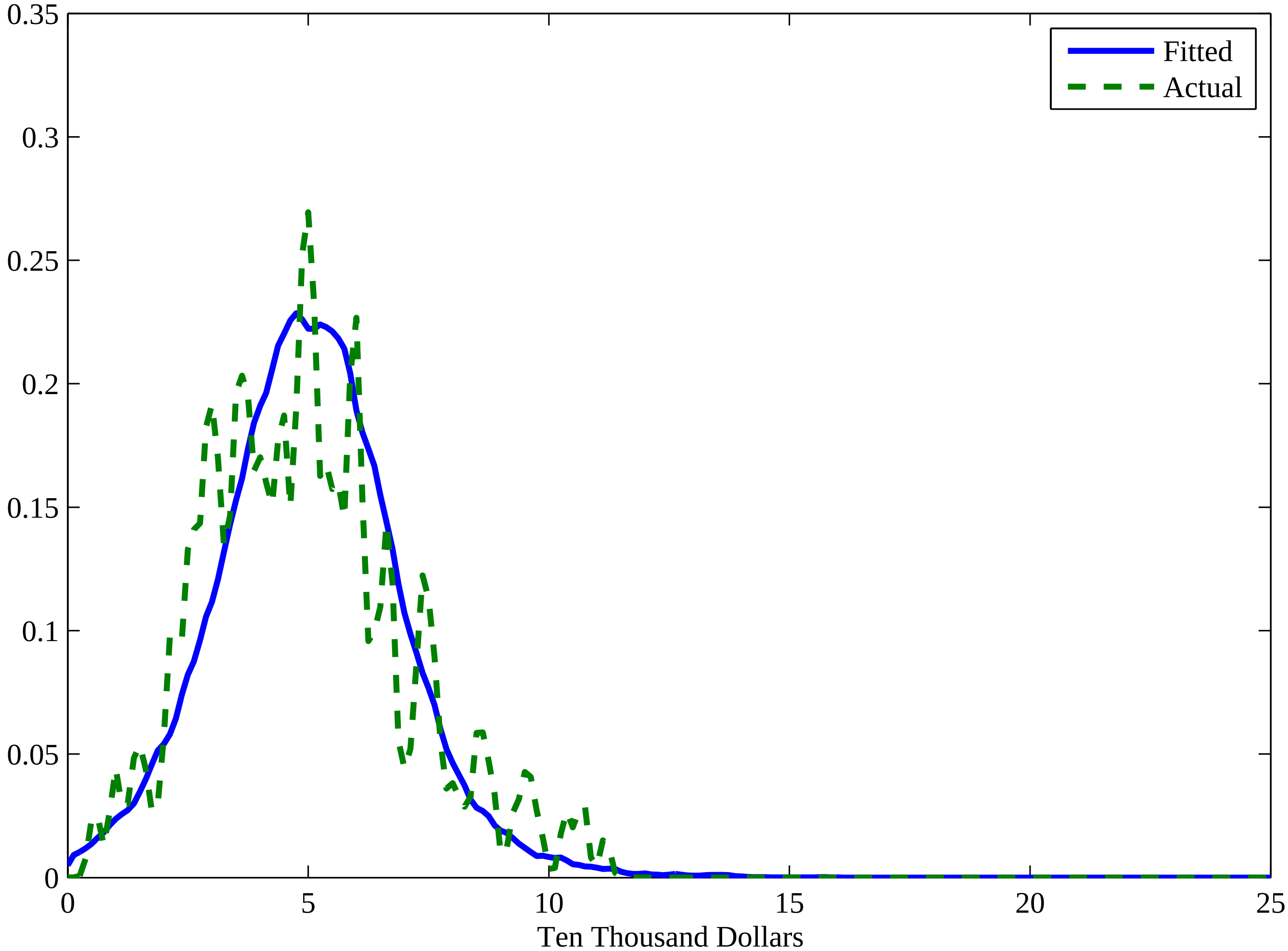
Let Y_1 denote earnings at age 32.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.12
Densities of earnings at age 33 College Sample NLSY/79



Let Y_1 denote earnings at age 33.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

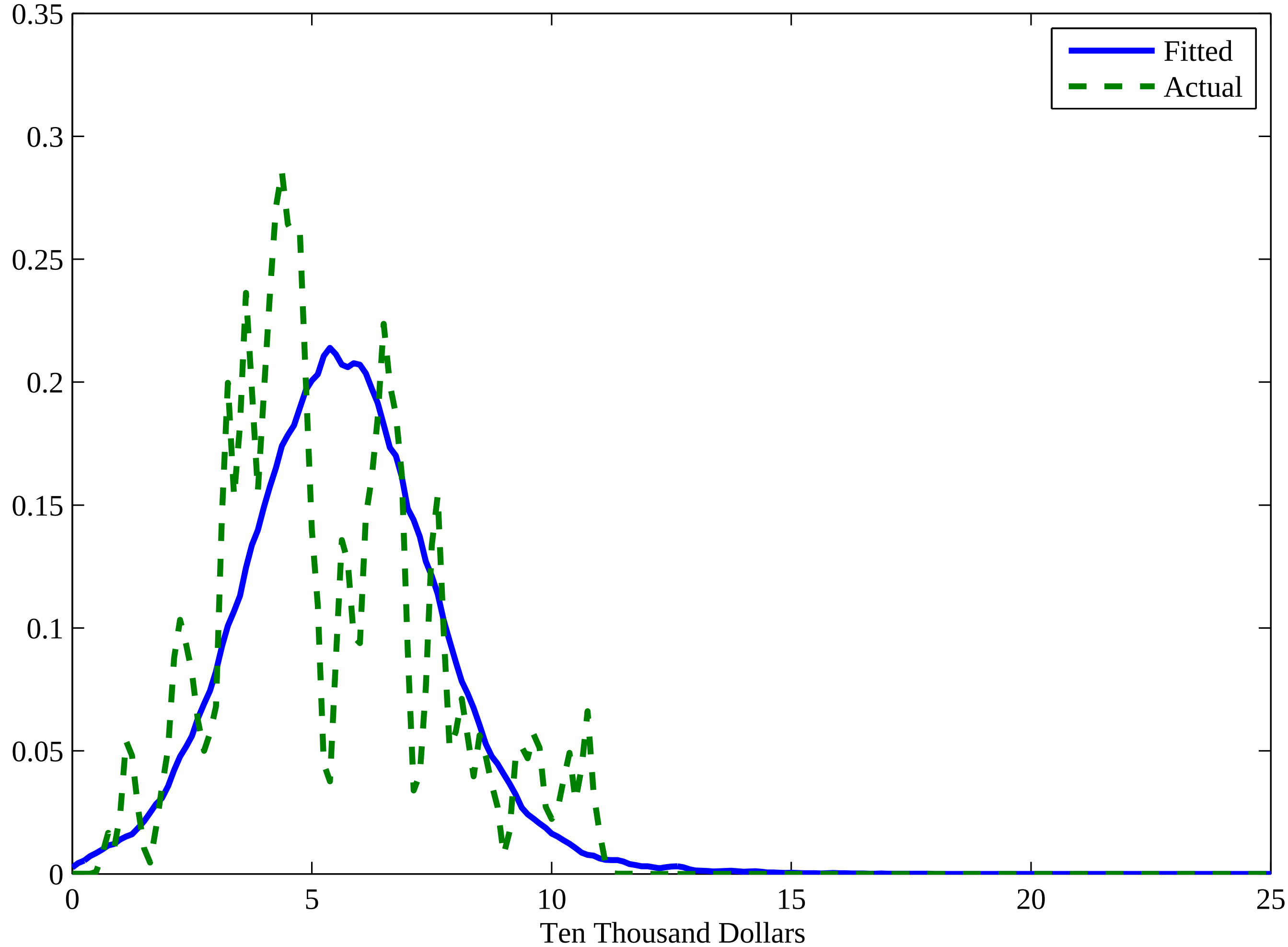
Web Data Appendix Figure II-3.13
Densities of earnings at age 34 College Sample NLSY/79



Let Y_1 denote earnings at age 34.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.14

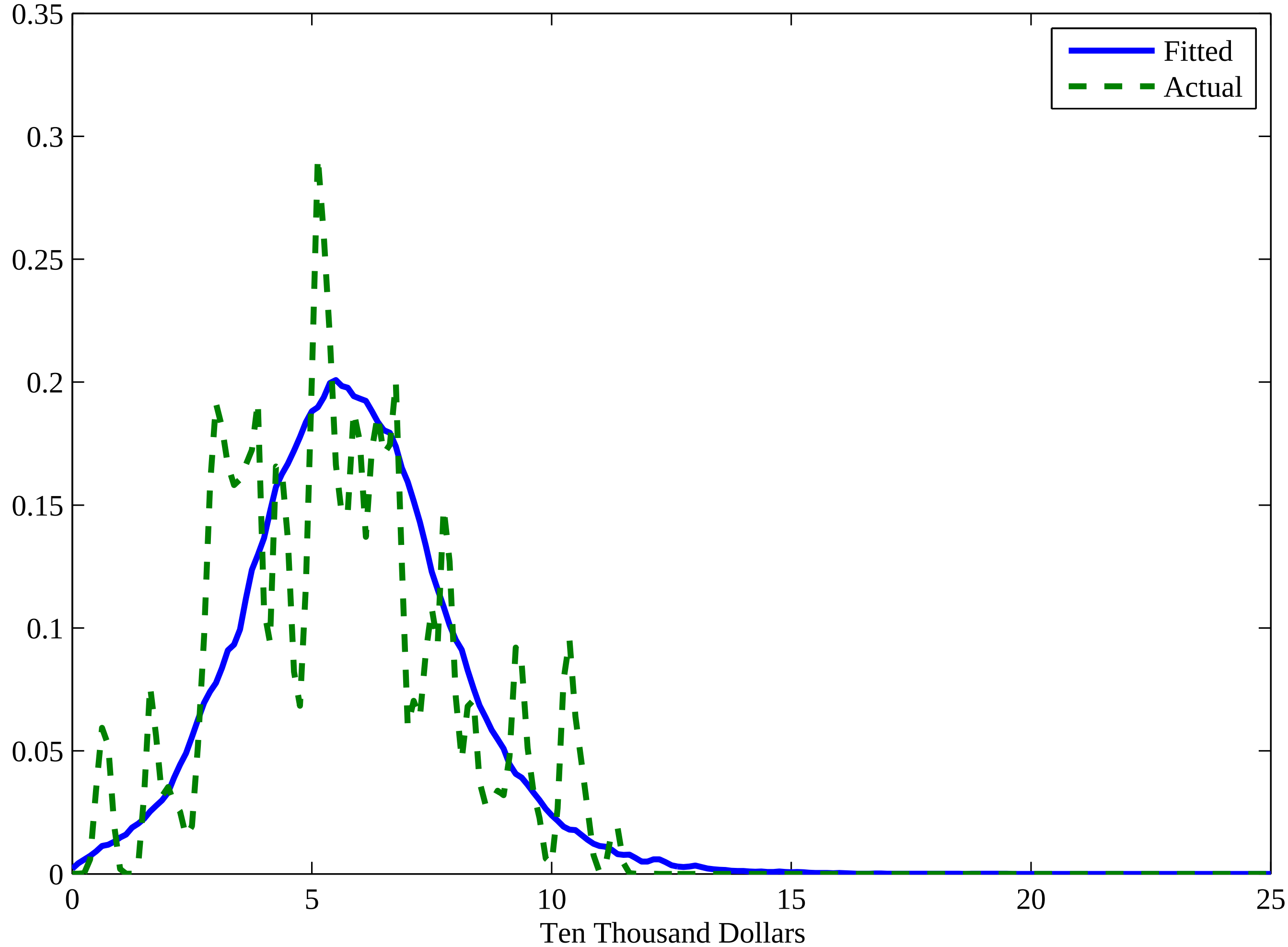
Densities of earnings at age 35 College Sample NLSY/79



Let Y_1 denote earnings at age 35.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

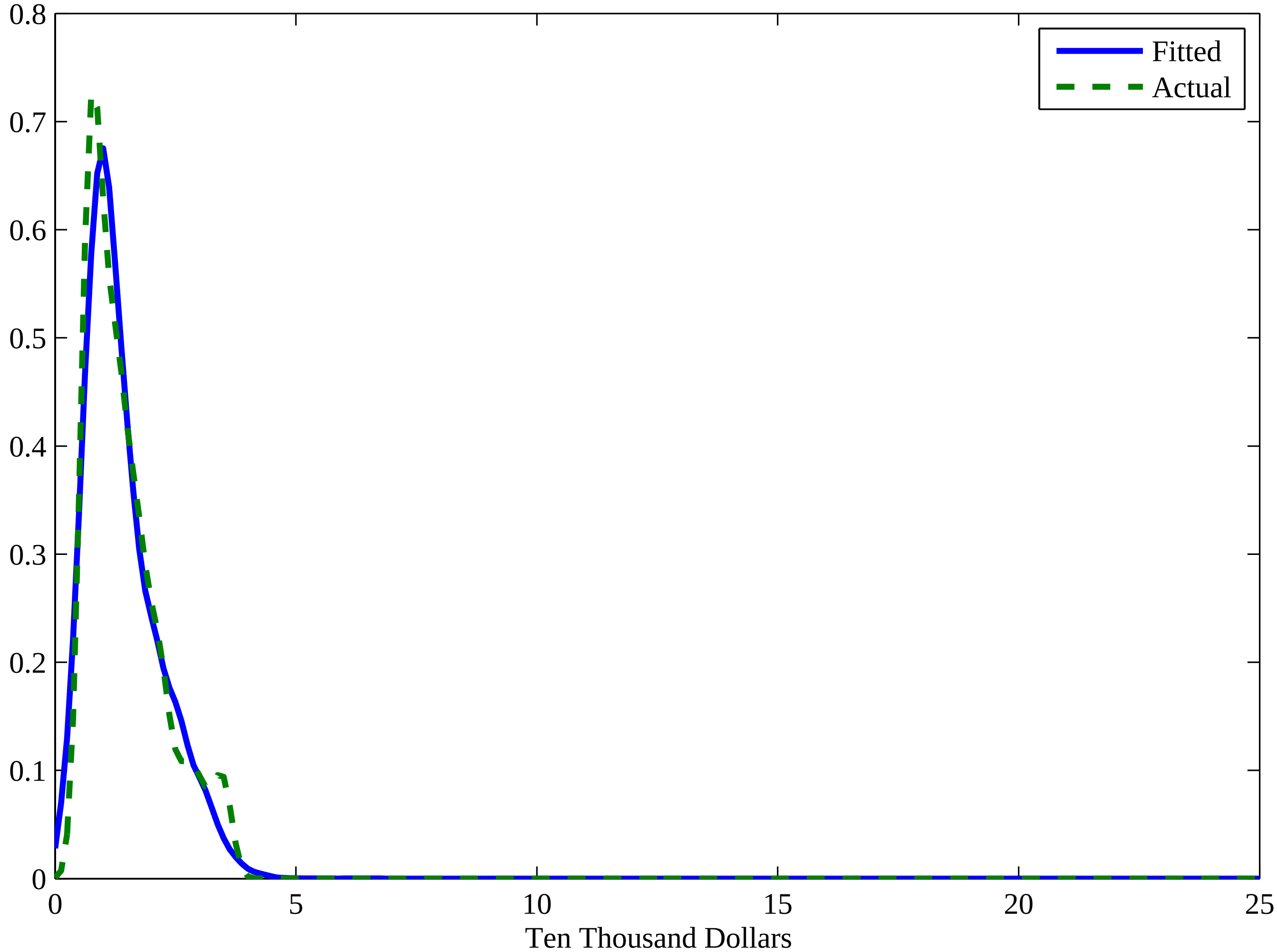
Web Data Appendix Figure II-3.15

Densities of earnings at age 36 College Sample NLSY/79



Let Y_1 denote earnings at age 36.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

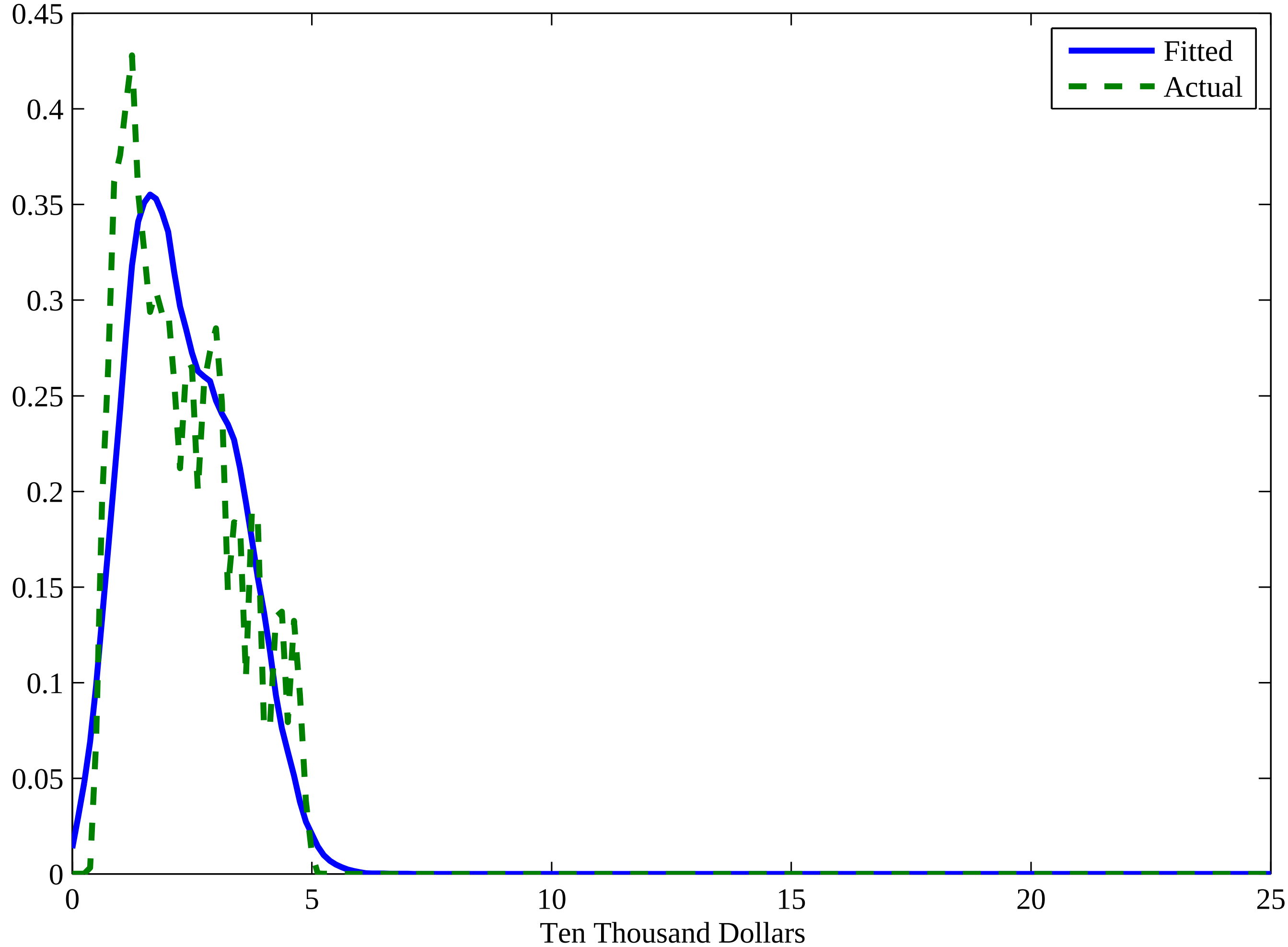
Web Data Appendix Figure II-3.16
Densities of earnings at age 22 College Sample NLS/66



Let Y_1 denote earnings at age 22.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.17

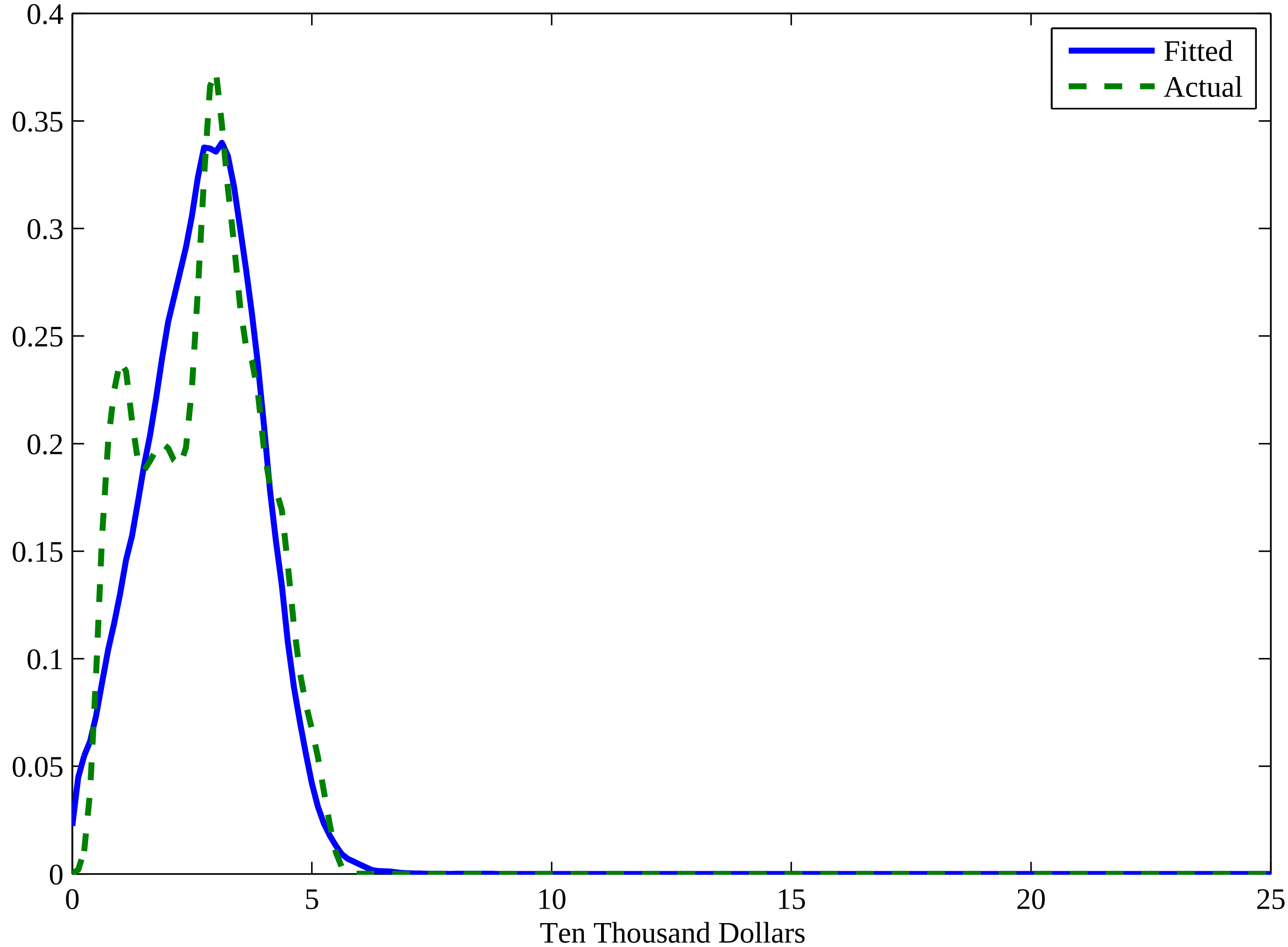
Densities of earnings at age 23 College Sample NLS/66



Let Y_1 denote earnings at age 23.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

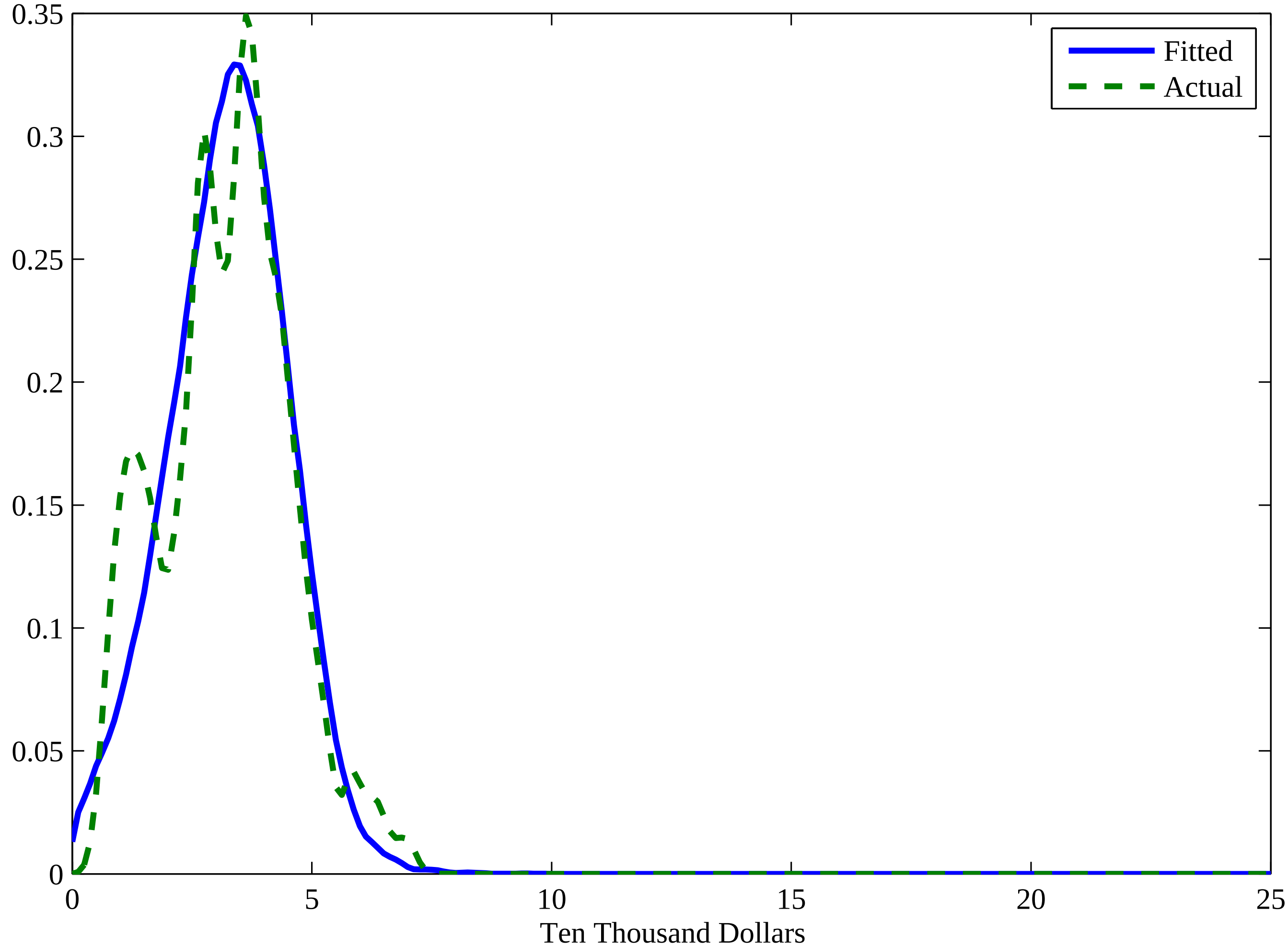
Web Data Appendix Figure II-3.18

Densities of earnings at age 24 College Sample NLS/66



Let Y_1 denote earnings at age 24.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

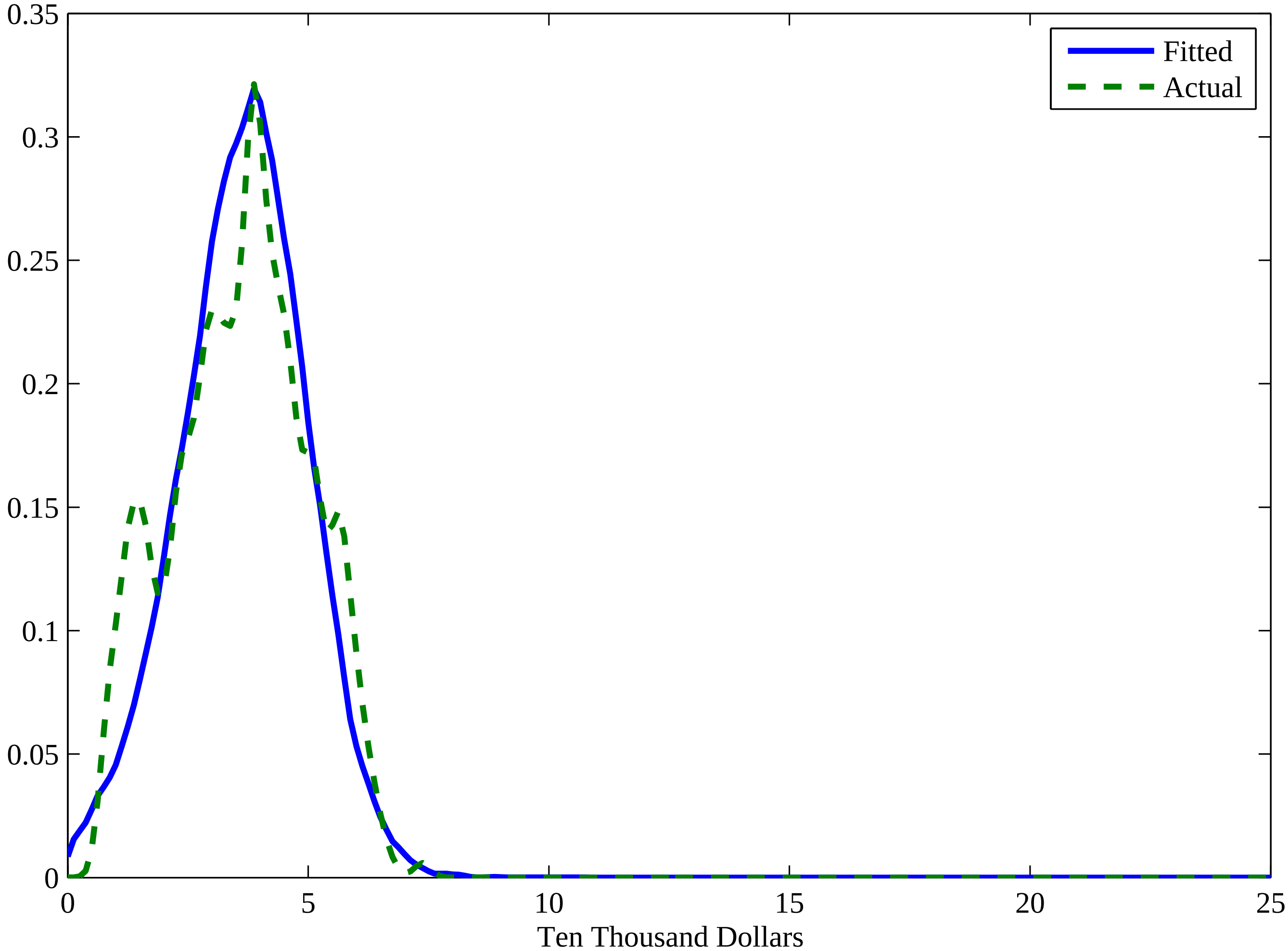
Web Data Appendix Figure II-3.19
Densities of earnings at age 25 College Sample NLS/66



Let Y_1 denote earnings at age 25.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

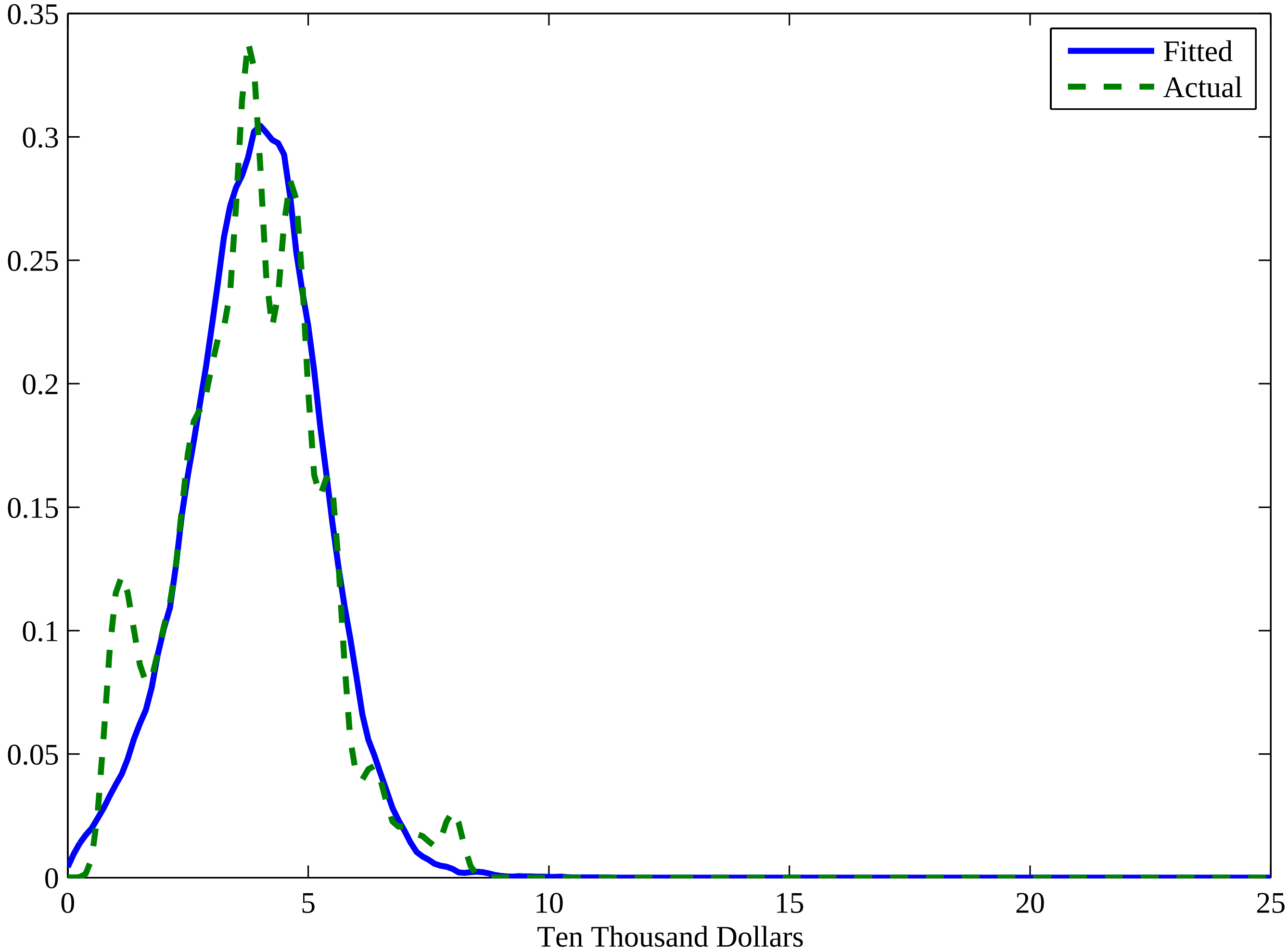
Web Data Appendix Figure II-3.20

Densities of earnings at age 26 College Sample NLS/66



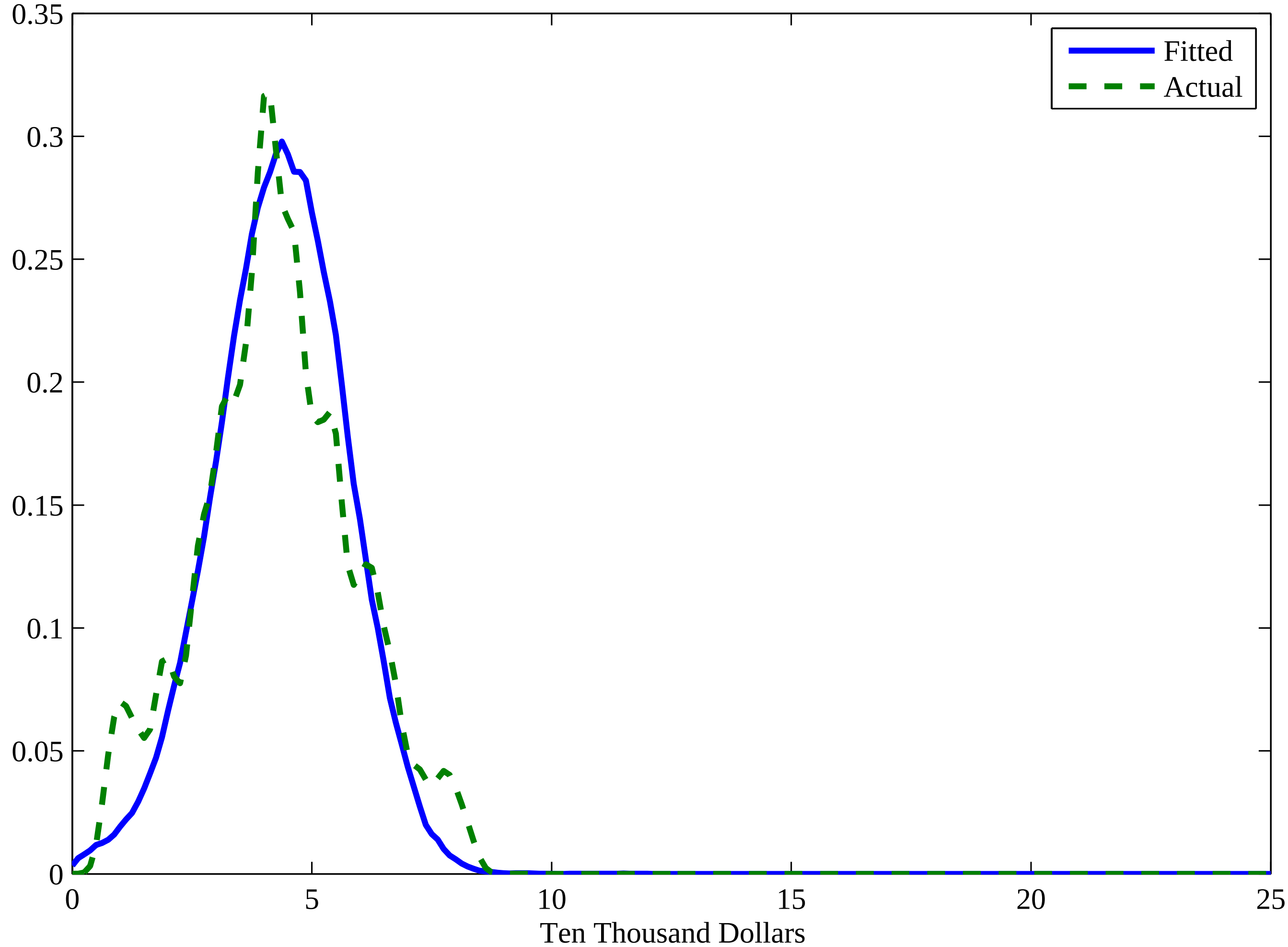
Let Y_1 denote earnings at age 26.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.21
Densities of earnings at age 27 College Sample NLS/66



Let Y_1 denote earnings at age 27.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

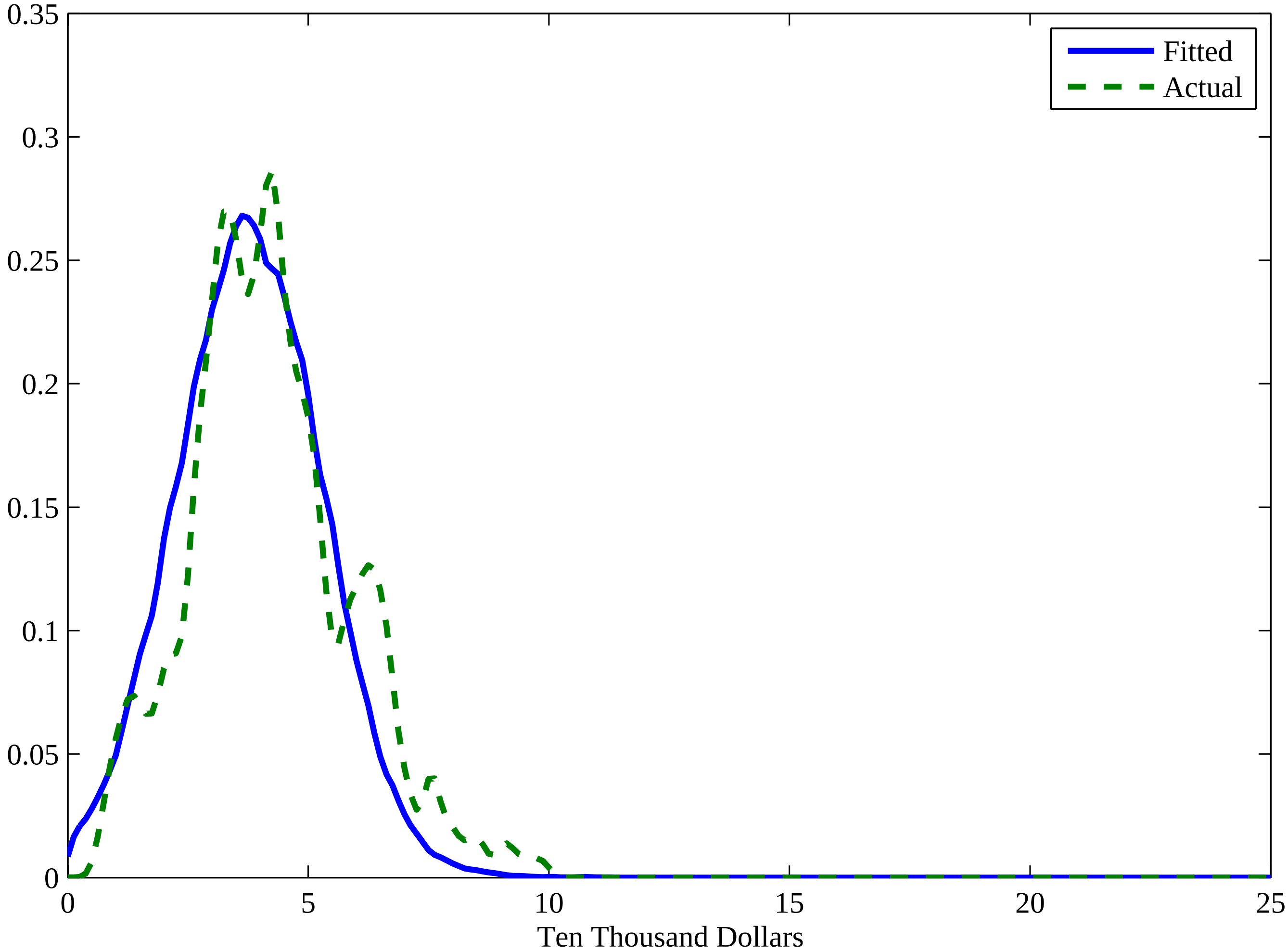
Web Data Appendix Figure II-3.22
Densities of earnings at age 28 College Sample NLS/66



Let Y_1 denote earnings at age 28.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.23

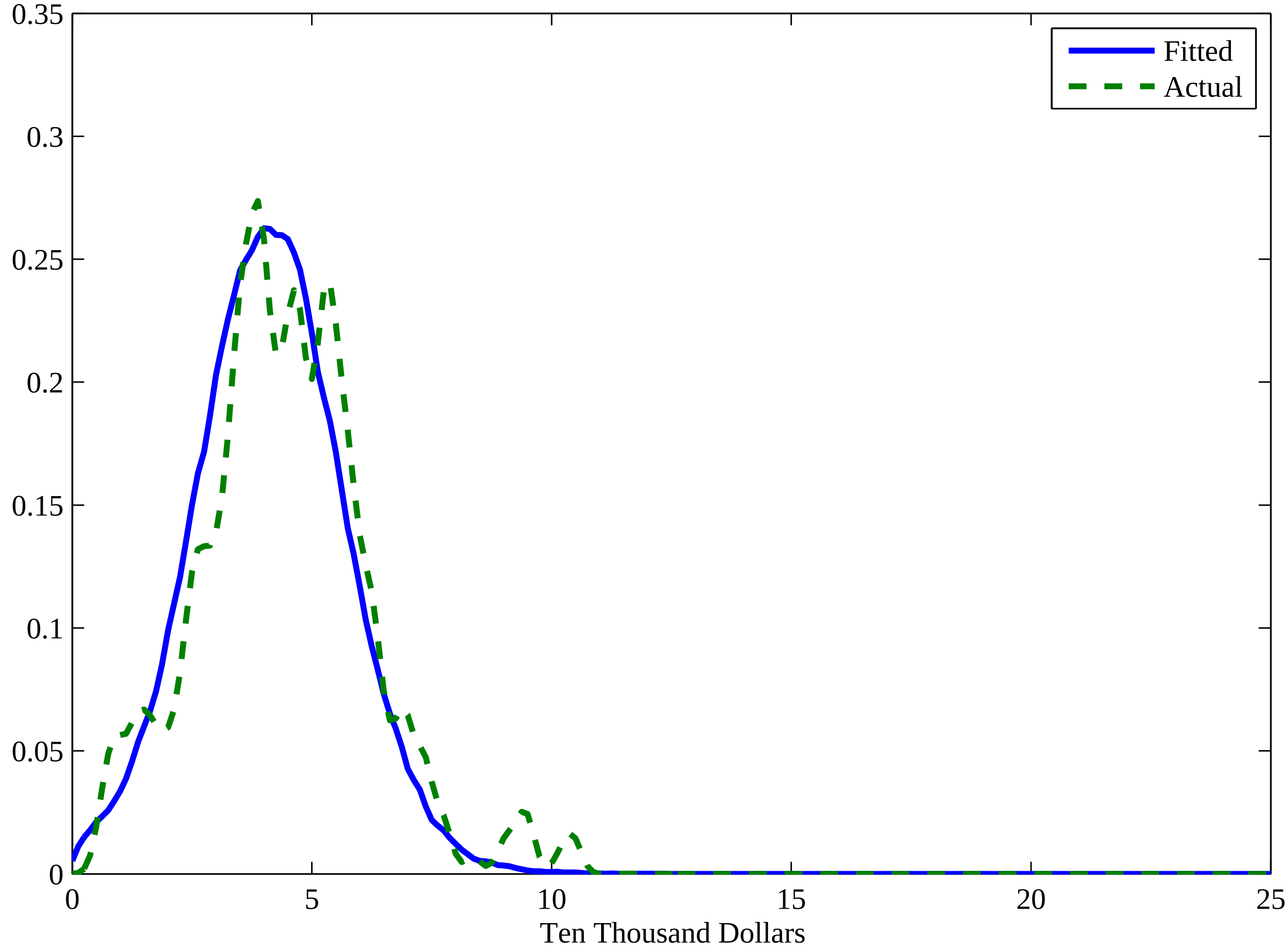
Densities of earnings at age 29 College Sample NLS/66



Let Y_1 denote earnings at age 29.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

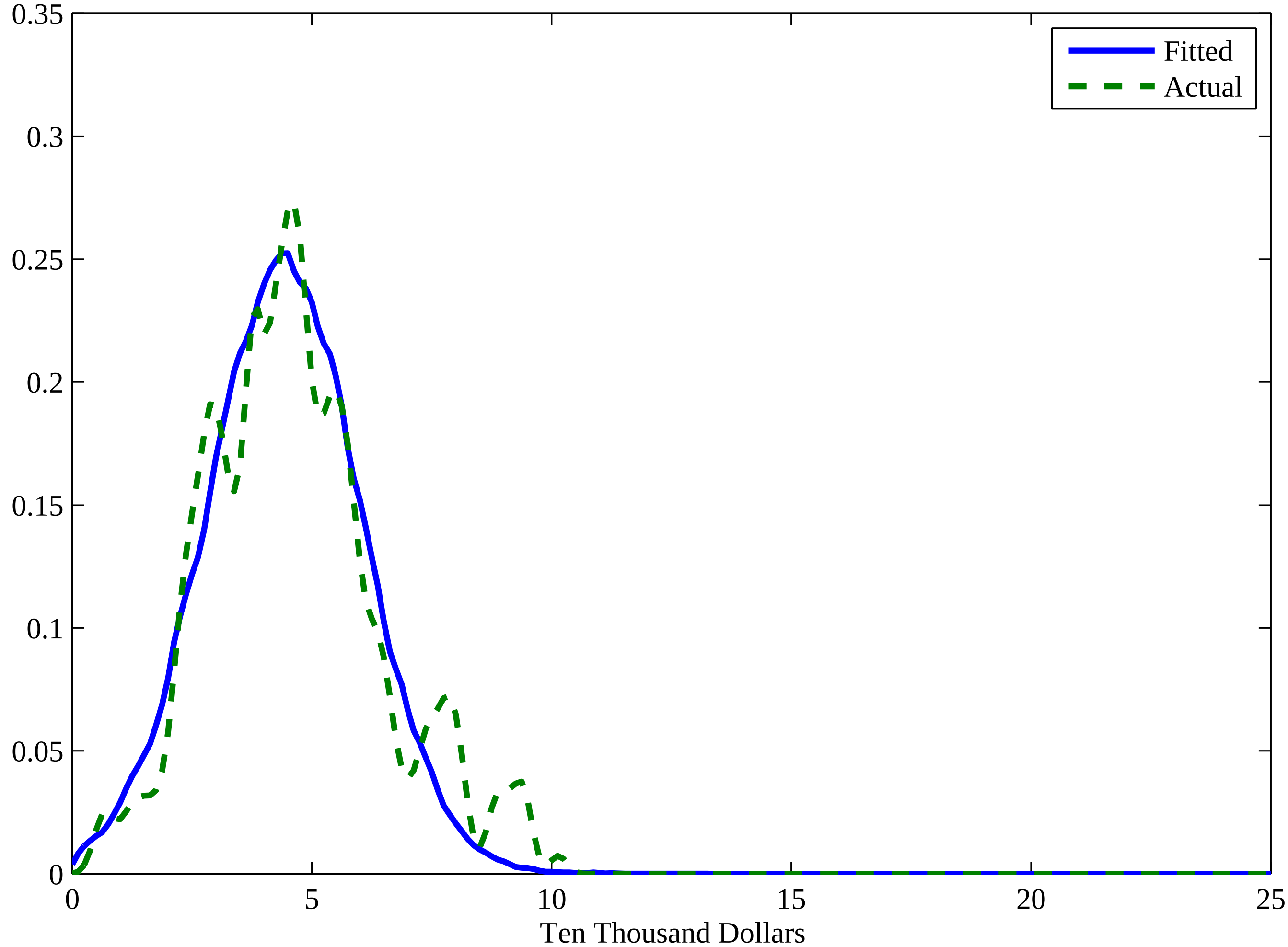
Web Data Appendix Figure II-3.24

Densities of earnings at age 30 College Sample NLS/66



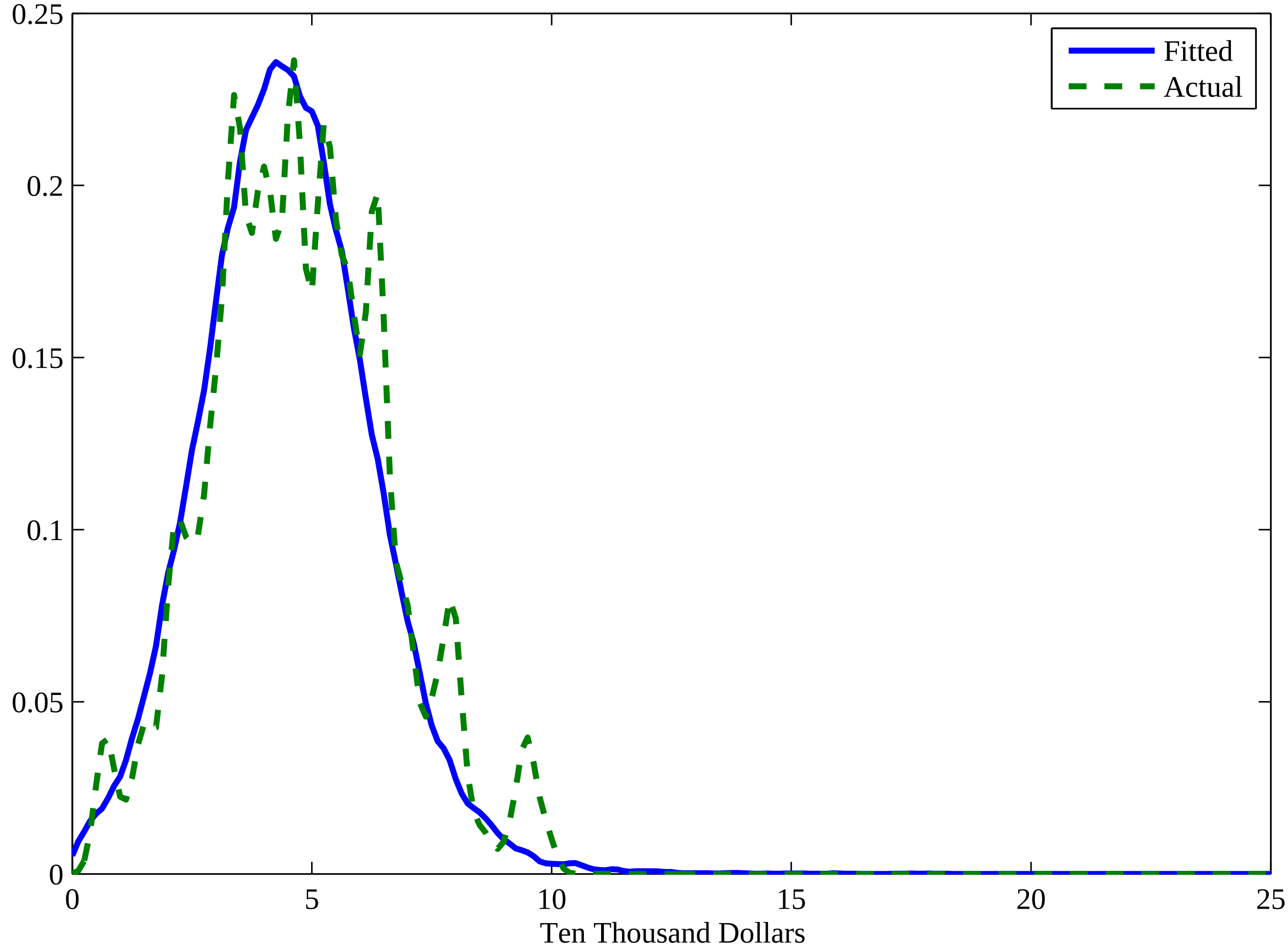
Let Y_1 denote earnings at age 30.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-3.25
Densities of earnings at age 31 College Sample NLS/66



Let Y_1 denote earnings at age 31.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

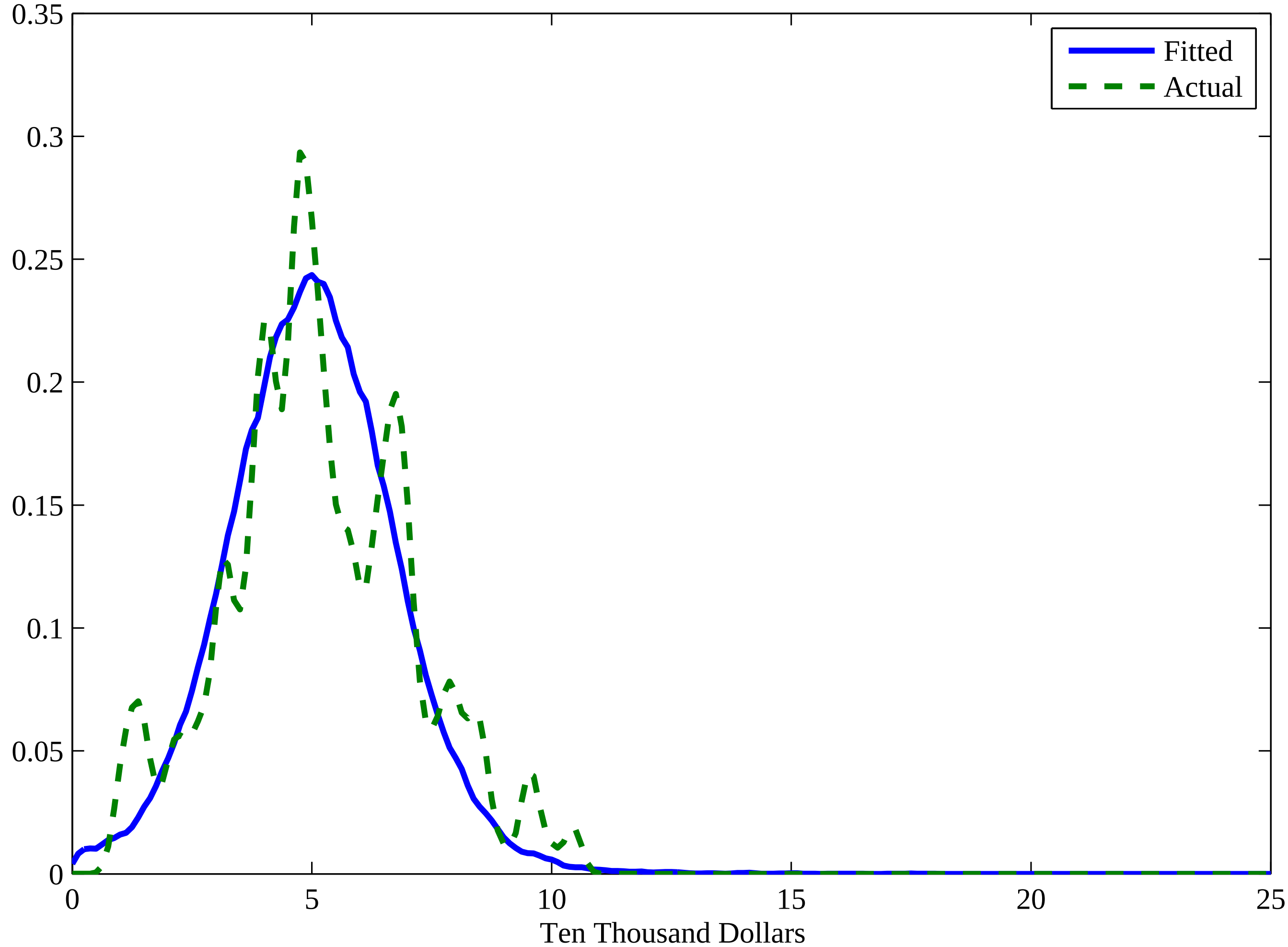
Web Data Appendix Figure II-3.26
Densities of earnings at age 32 College Sample NLS/66



Let Y_1 denote earnings at age 32.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

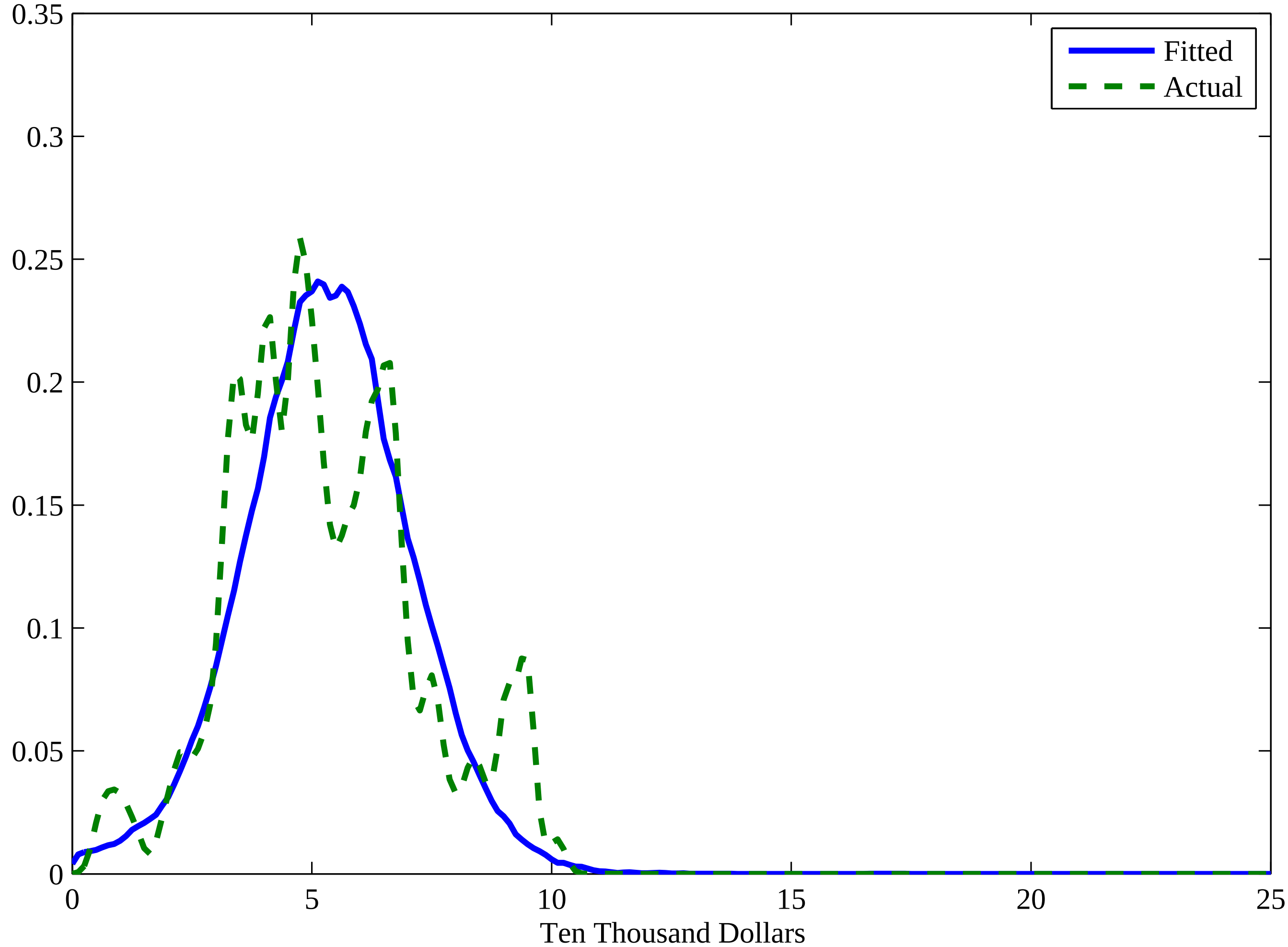
Web Data Appendix Figure II-3.27

Densities of earnings at age 33 College Sample NLS/66



Let Y_1 denote earnings at age 33.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

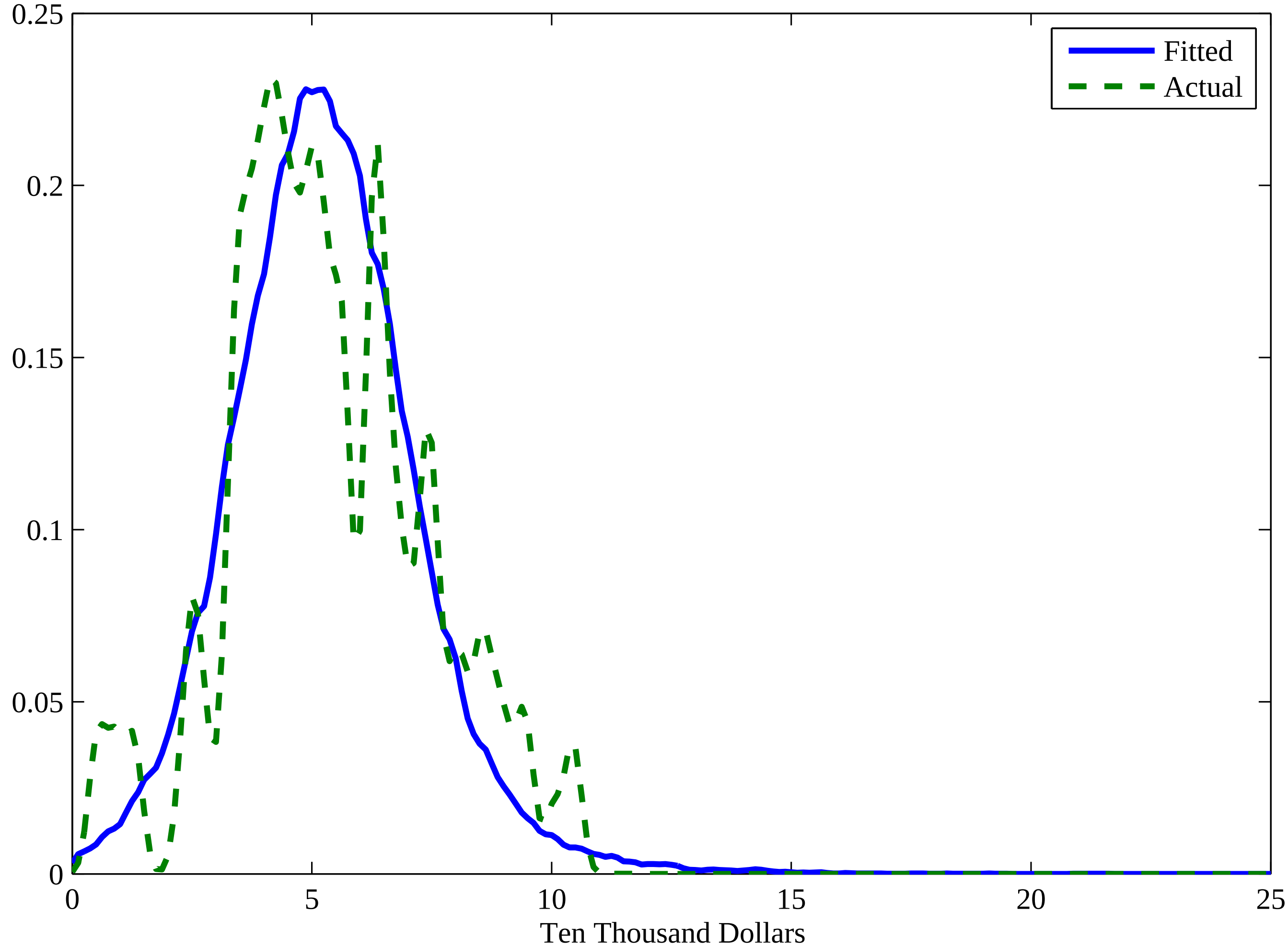
Web Data Appendix Figure II-3.28
Densities of earnings at age 34 College Sample NLS/66



Let Y_1 denote earnings at age 34.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

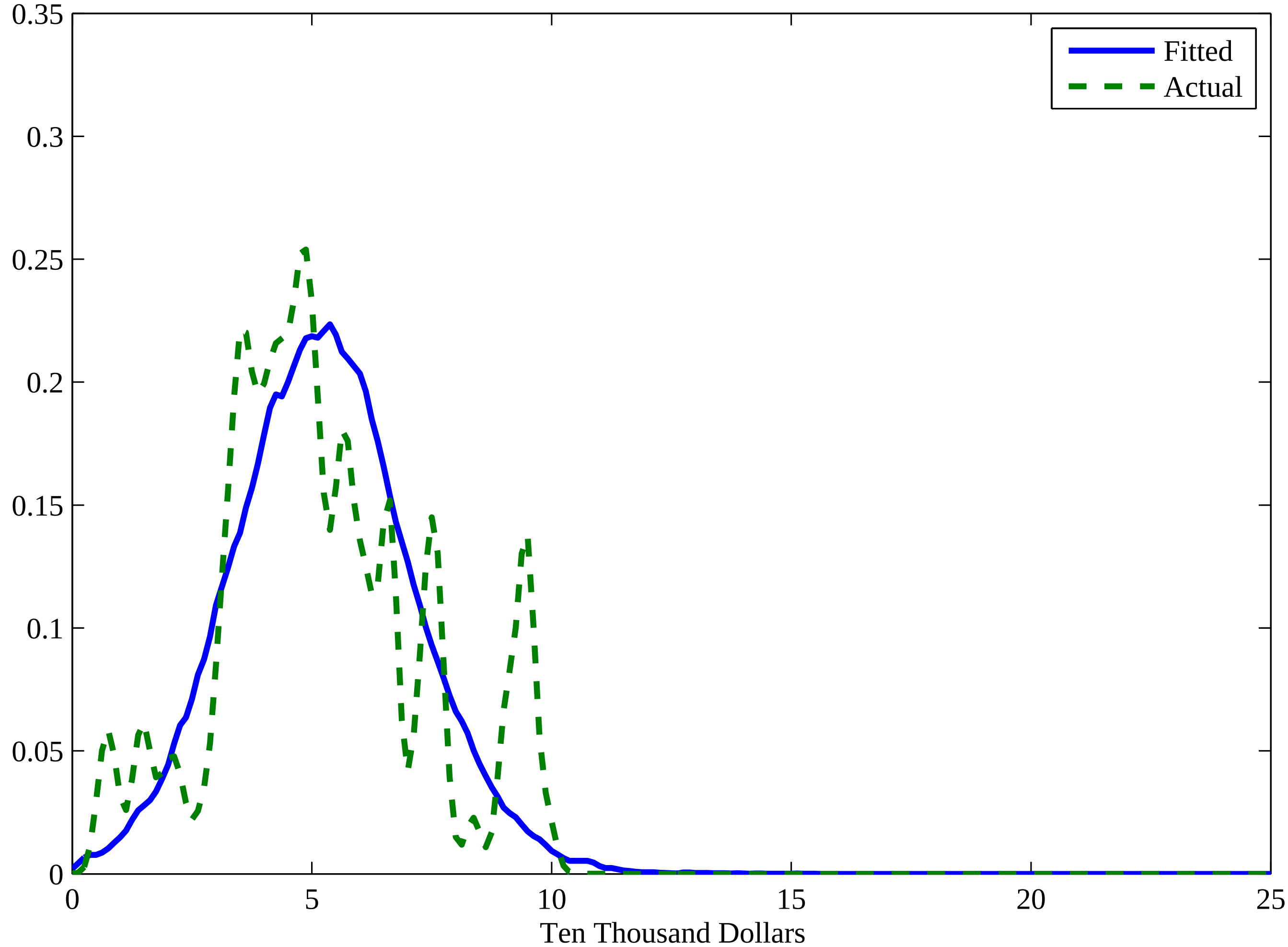
Web Data Appendix Figure II-3.29

Densities of earnings at age 35 College Sample NLS/66



Let Y_1 denote earnings at age 35.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

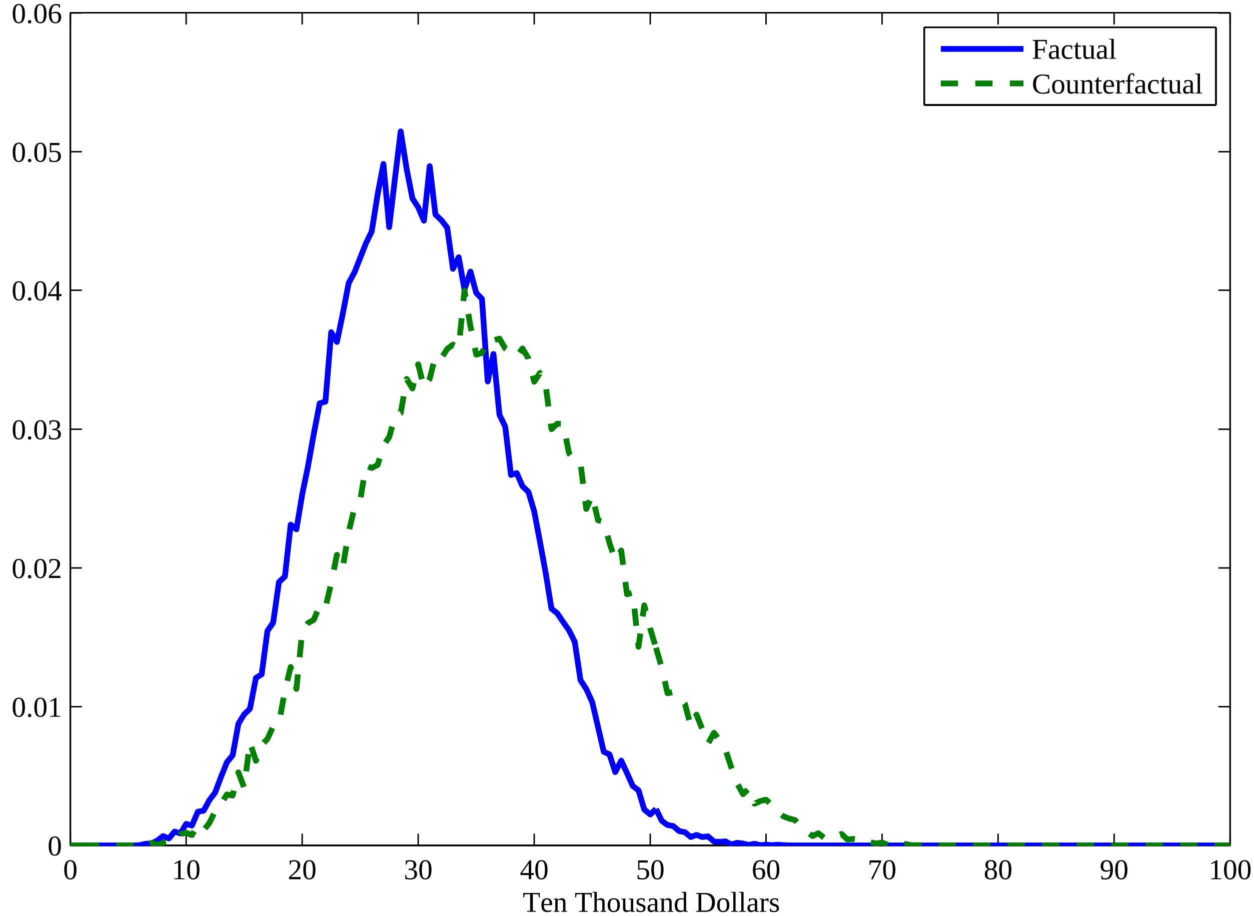
Web Data Appendix Figure II-3.30
Densities of earnings at age 36 College Sample NLS/66



Let Y_1 denote earnings at age 36.
Here we plot the density functions $f(y_1 \mid S = 1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line).

Web Data Appendix Figure II-4.1

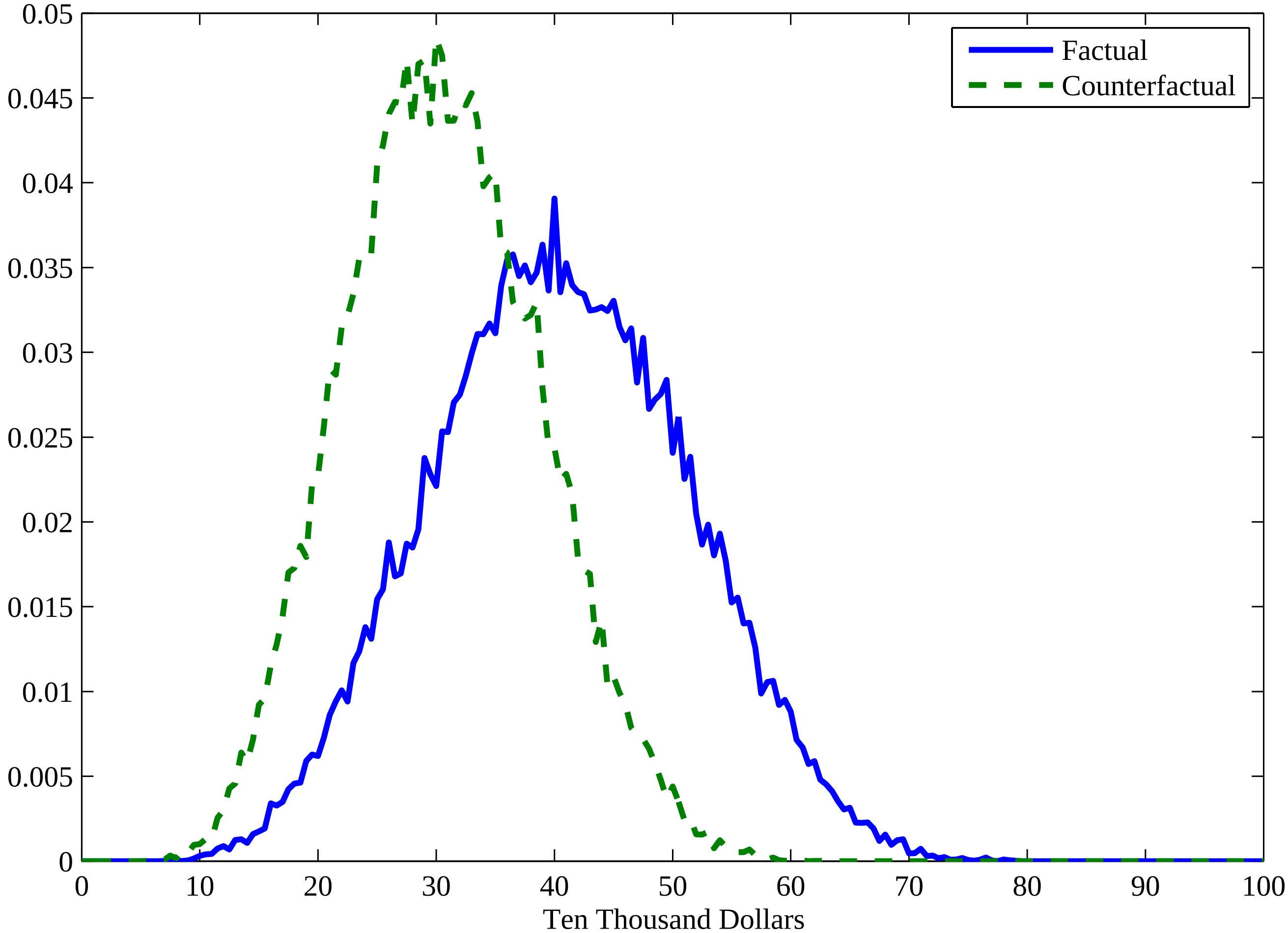
Densities of present value of earnings High School Sample NLSY/79



Present value of earnings from age 22 to 36 for High School Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_0|S=0)$ (the solid curve), against the counterfactual density function $f(y_1|S=0)$ (the dashed line). We use kernel density estimation to smooth these functions.

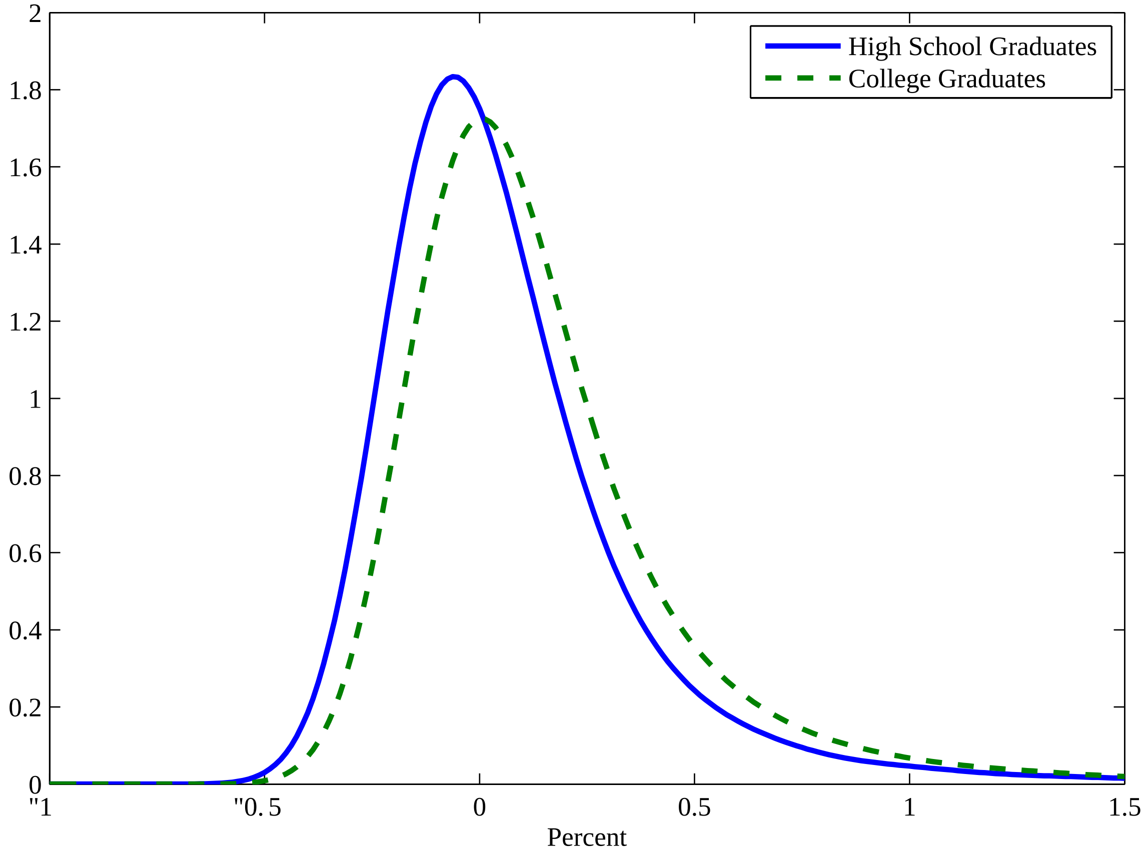
Web Data Appendix Figure II-4.2

Densities of present value of earnings College Sample NLSY/79



Present value of earnings from age 22 to 36 for College Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_1|S = 1)$ (the solid curve), against the counterfactual density function $f(y_0|S = 1)$ (the dashed line). We use kernel density estimation to smooth these functions.

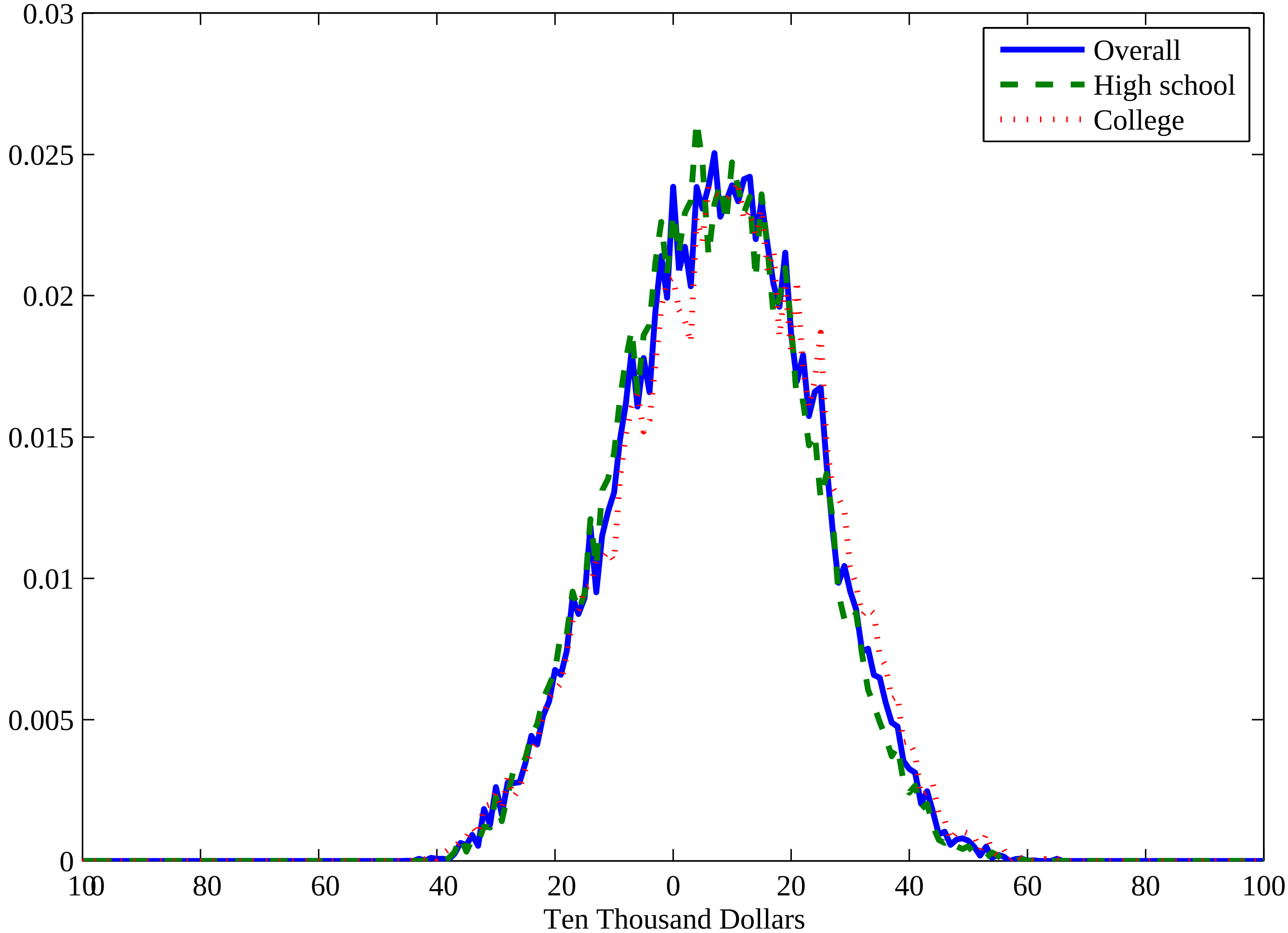
Web Data Appendix Figure II-4.3
Densities of Returns to College NLSY/79 Sample



Let Y_0, Y_1 denote the present value of earnings from age 22 to age 36 in the high school and college sectors, respectively. Define ex post returns to college as the ratio $R = (Y_1 - Y_0)/Y_0$. Let $f(r)$ denote the density function of the random variable R. The solid line is the density of ex post returns to college for high school graduates, that is $f(r|S = 0)$. The dashed line is the density of ex post returns to college for college graduates, that is, $f(r|S = 1)$. This assumes that the agent chooses schooling without knowing θ_3 and $\{\varepsilon_{0,t}, \varepsilon_{1,t}\}_{t=1}^{T}$.

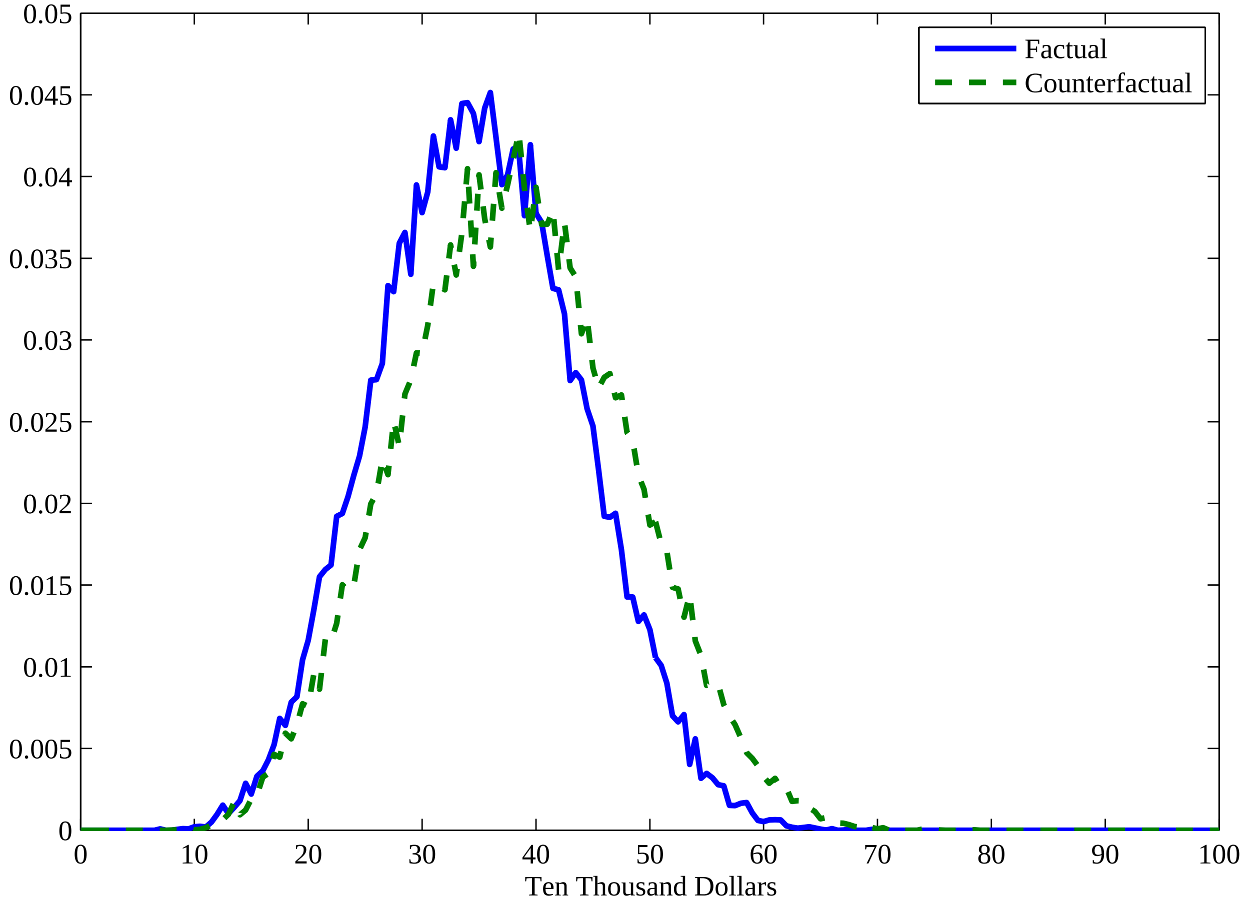
Web Data Appendix Figure II-4.4

Densities of monetary value of psychic cost, 1979



Web Data Appendix Figure II-4.5

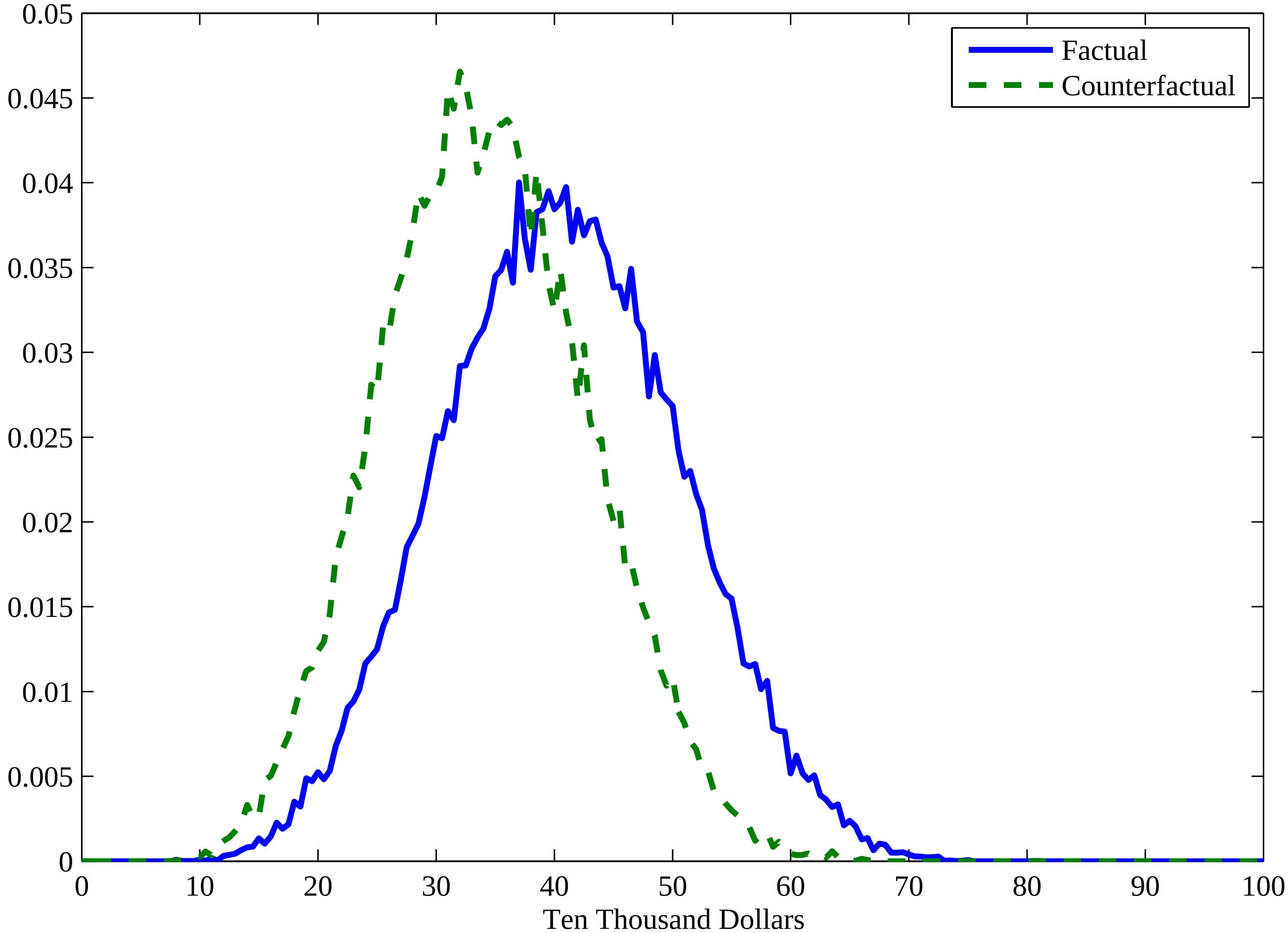
Densities of present value of earnings High School Sample NLS/66



Present value of earnings from age 22 to 36 for High School Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_0|S=0)$ (the solid curve), against the counterfactual density function $f(y_1|S=0)$ (the dashed line). We use kernel density estimation to smooth these functions.

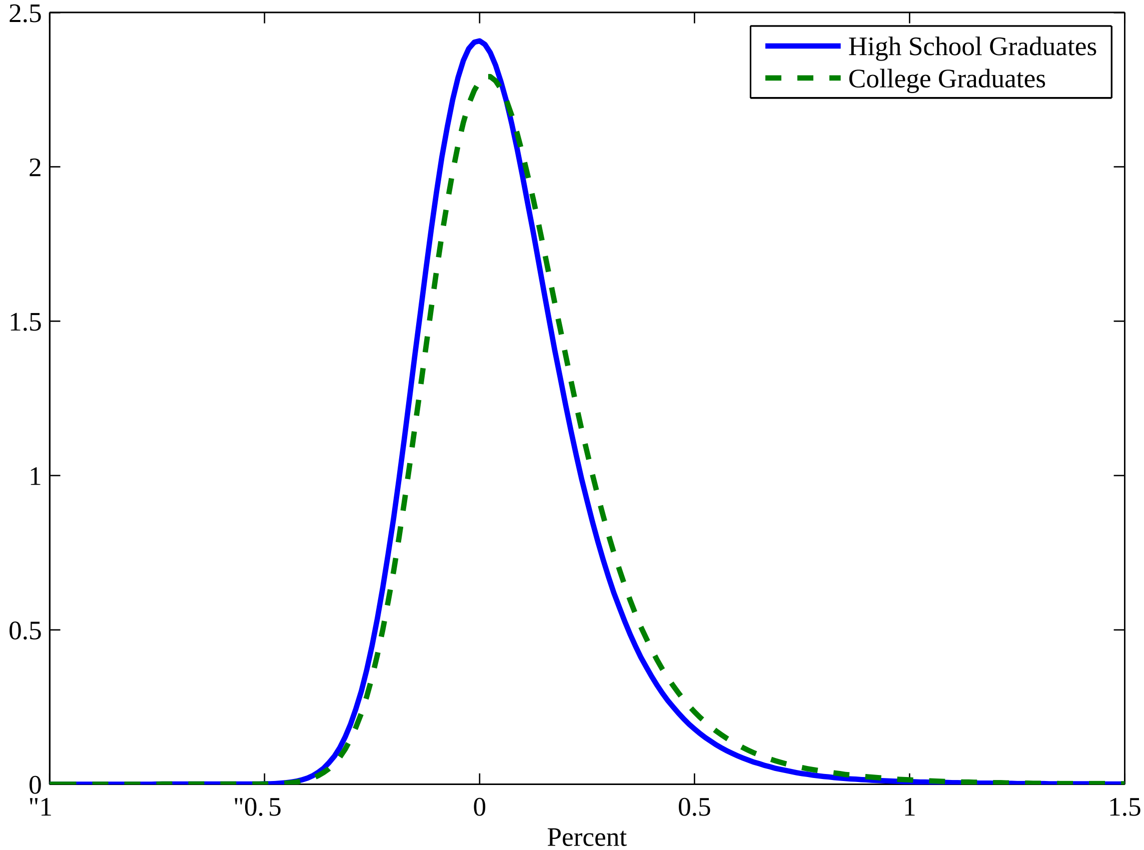
Web Data Appendix Figure II-4.6

Densities of present value of earnings College Sample NLS/66



Present value of earnings from age 22 to 36 for College Graduates discounted using an interest rate of 5%. Here we plot the factual density function $f(y_1|S=1)$ (the solid curve), against the counterfactual density function $f(y_0|S=1)$ (the dashed line). We use kernel density estimation to smooth these functions.

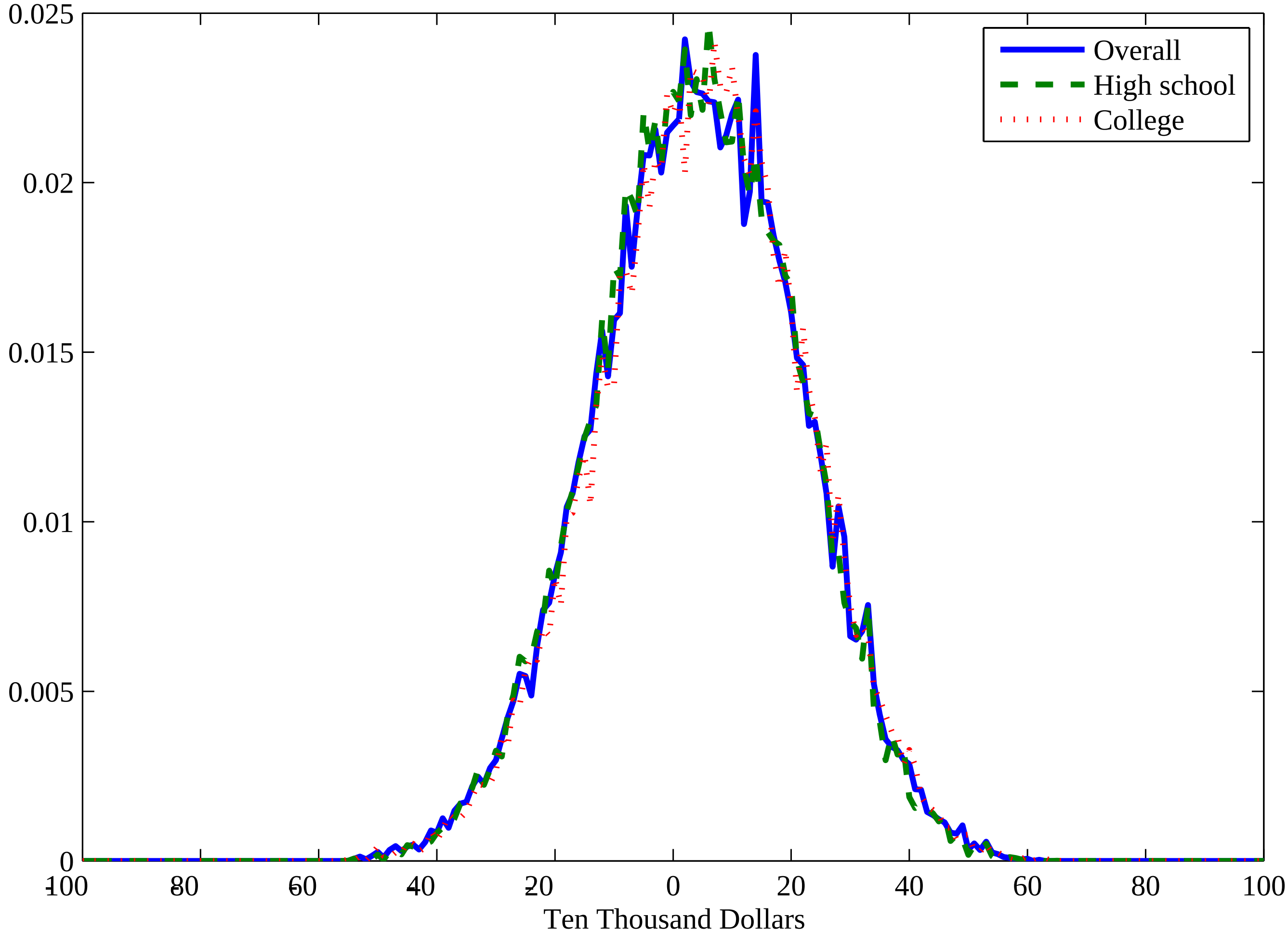
Web Data Appendix Figure II-4.7
Densities of Returns to College NLS/66 Sample



Let Y_0, Y_1 denote the present value of earnings from age 22 to age 36 in the high school and college sectors, respectively. Define ex post returns to college as the ratio $R = (Y1 - Y0)/Y0$. Let $f(r)$ denote the density function of the random variable R. The solid line is the density of ex post returns to college for high school graduates, that is $f(r|S = 0)$. The dashed line is the density of ex post returns to college for college graduates, that is, $f(r|S = 1)$. This assumes that the agent chooses schooling without knowing θ_3 and $\{\varepsilon_{0,t}, \varepsilon_{1,t}\}_{t=1}^{T}$.

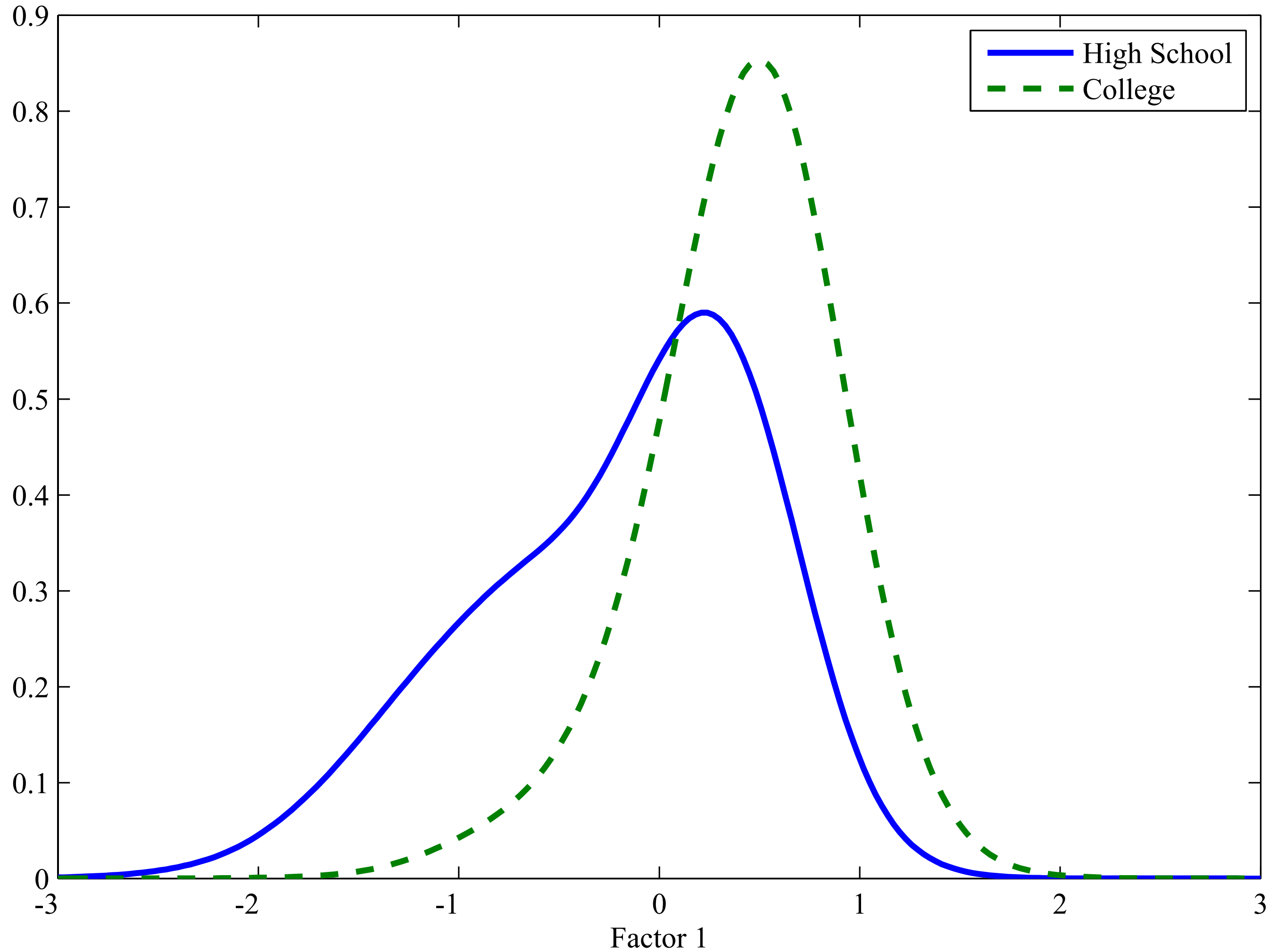
Web Data Appendix Figure II-4.8

Densities of monetary value of psychic cost, 1966



Web Data Appendix Figure II-5.1

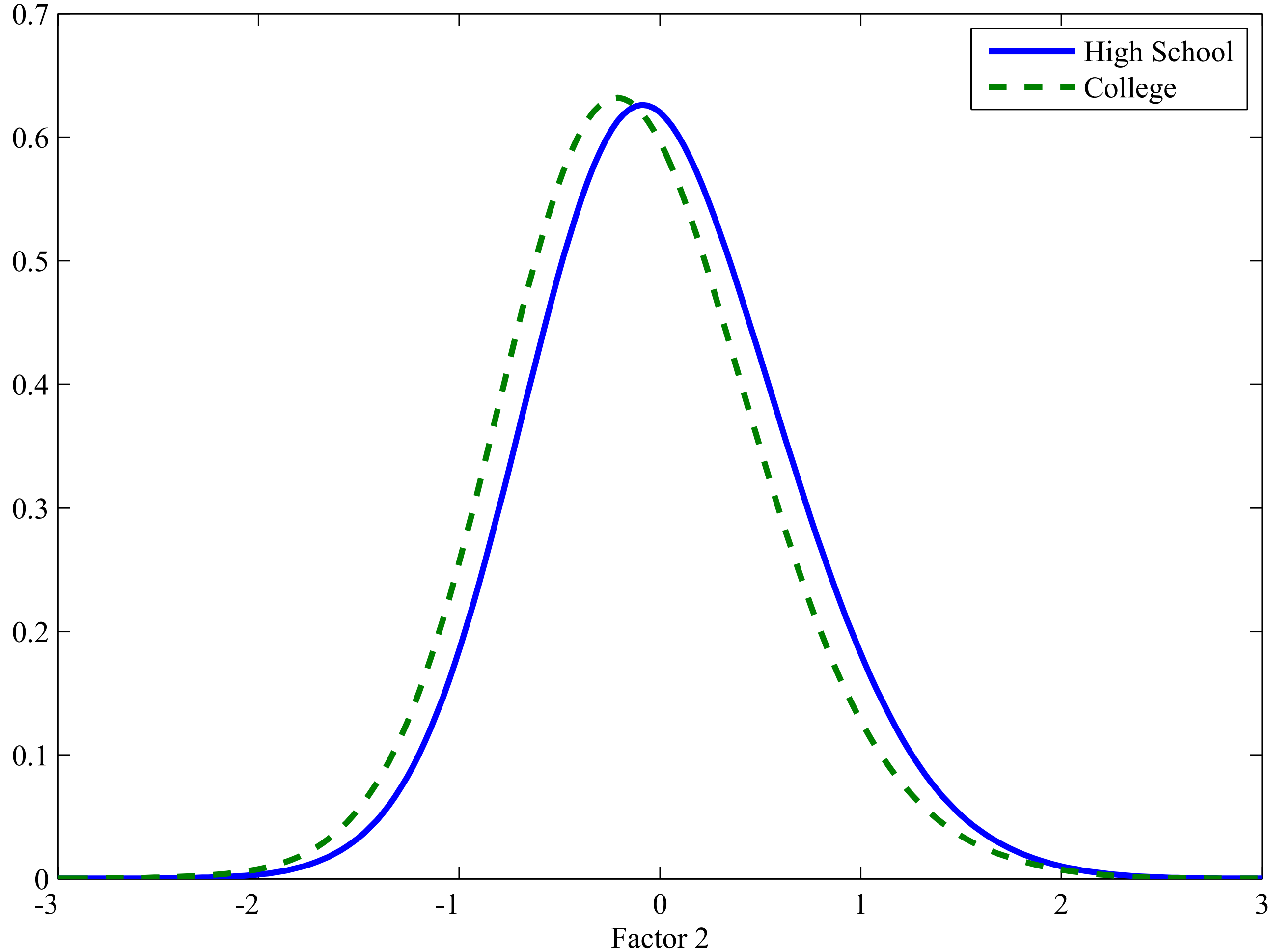
Densities of Factor 1 by schooling level NLSY/79



Let f(θ_1) denote the probability density function of factor θ_1. We allow f(θ_1) to be a mixture of normals. The solid line plots the density of factor 1 conditional on choosing the high school sector, that is, f(θ_1 | S = 0). The dashed line plots the density of factor 1 conditional on choosing the college sector, that is, f(θ_1 | S = 1).

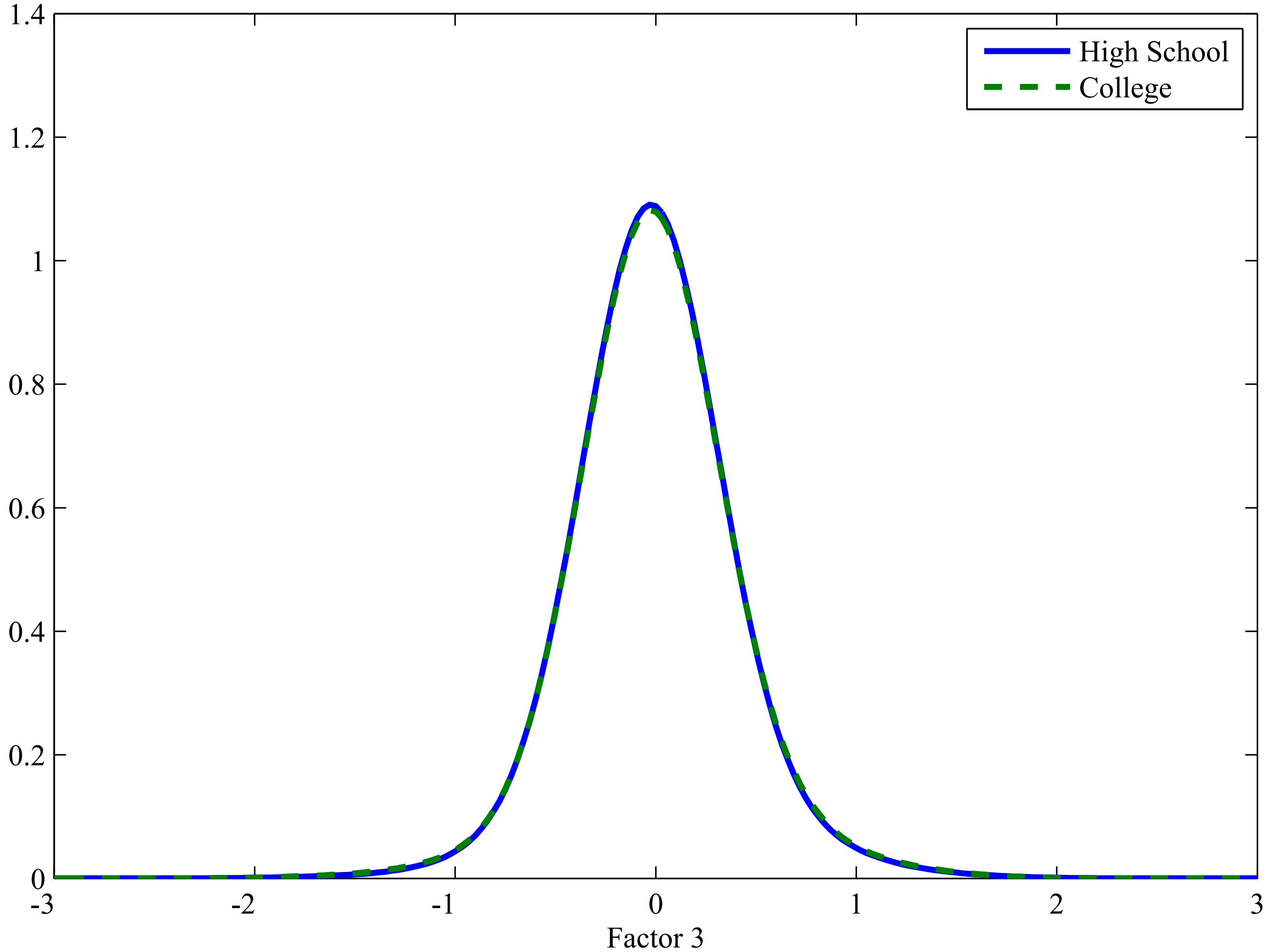
Web Data Appendix Figure II-5.2

Densities of Factor 2 by schooling level NLSY/79



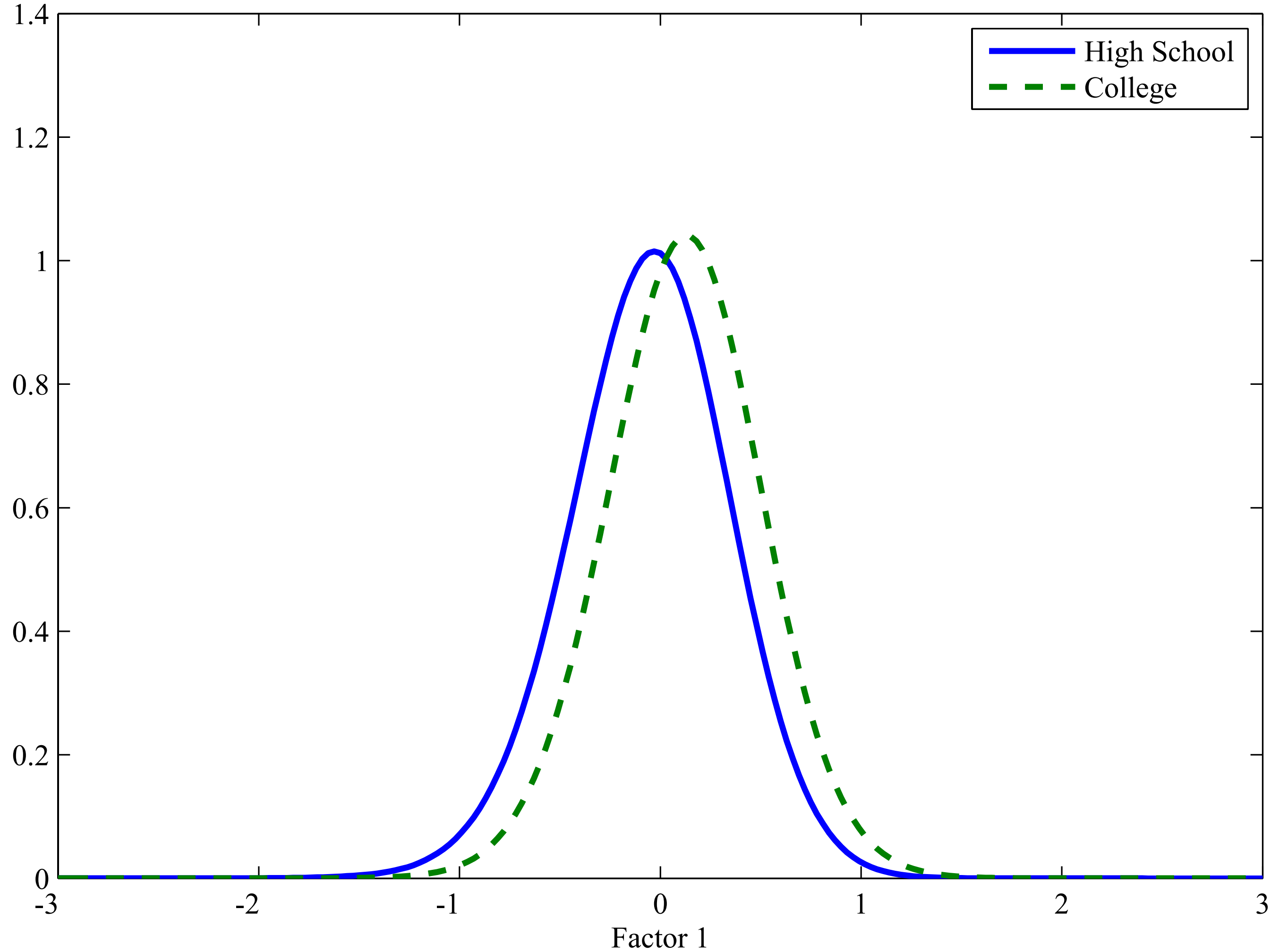
Let $f(\theta_2)$ denote the probability density function of factor θ_2. We allow $f(\theta_2)$ to be a mixture of normals. The solid line plots the density of factor 2 conditional on choosing the high school sector, that is, $f(\theta_2 \mid S = 0)$. The dashed line plots the density of factor 2 conditional on choosing the college sector, that is, $f(\theta_2 \mid S = 1)$.

Web Data Appendix Figure II-5.3
Densities of Factor 3 by schooling level NLSY/79



Let f(θ_3) denote the probability density function of factor θ_3. We allow f(θ_3) to be a mixture of normals. The solid line plots the density of factor 3 conditional on choosing the high school sector, that is, f(θ_3 | S = 0). The dashed line plots the density of factor 3 conditional on choosing the college sector, that is, f(θ_3 | S = 1).

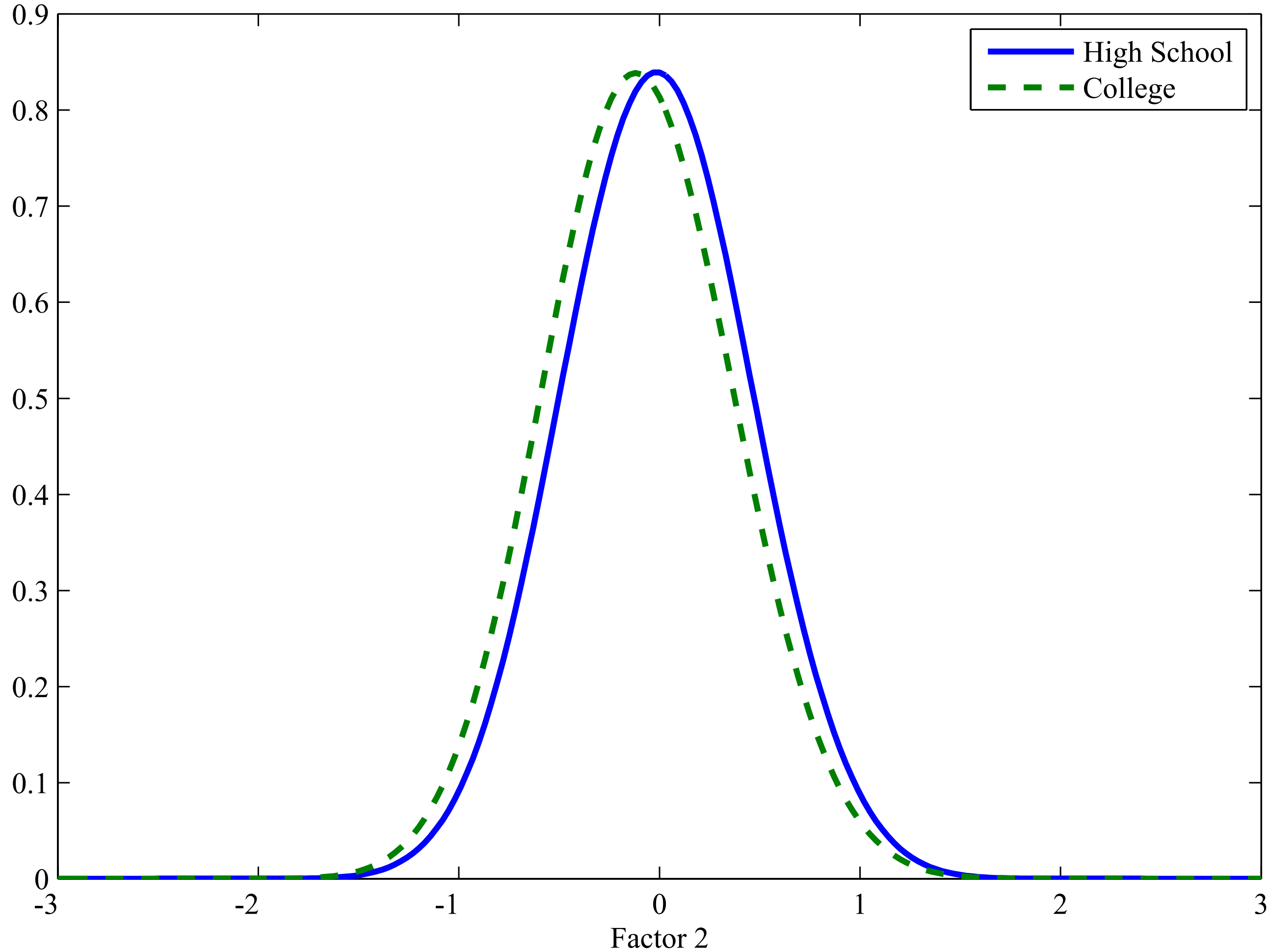
Web Data Appendix Figure II-5.4
Densities of Factor 1 by schooling level NLS/66



Let $f(\theta_1)$ denote the probability density function of factor θ_1. We allow $f(\theta_1)$ to be a mixture of normals. The solid line plots the density of factor 1 conditional on choosing the high school sector, that is, $f(\theta_1 \mid S = 0)$. The dashed line plots the density of factor 1 conditional on choosing the college sector, that is, $f(\theta_1 \mid S = 1)$.

Web Data Appendix Figure II-5.5

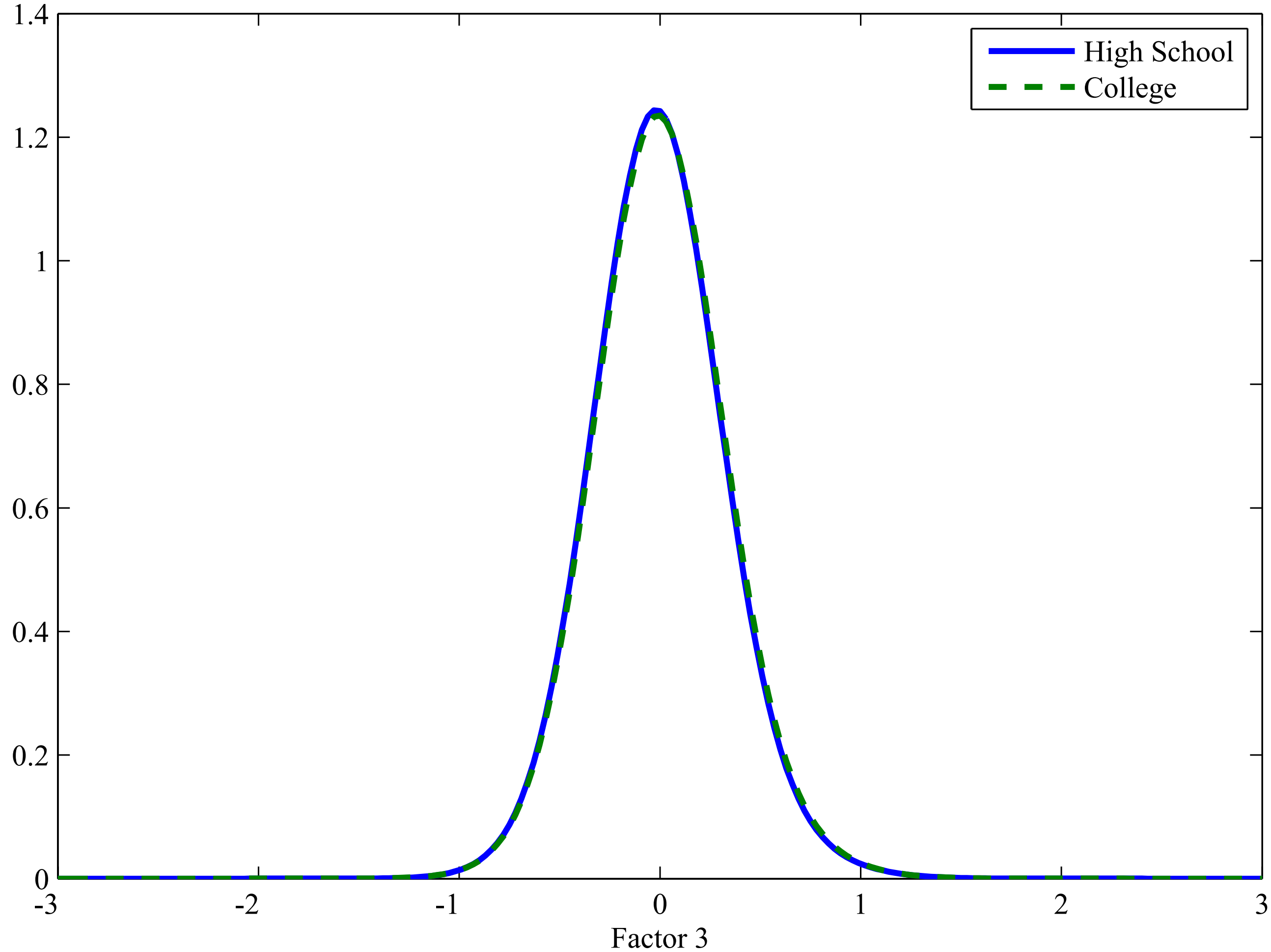
Densities of Factor 2 by schooling level NLS/66



Let f(θ_2) denote the probability density function of factor θ_2. We allow f(θ_2) to be a mixture of normals. The solid line plots the density of factor 2 conditional on choosing the high school sector, that is, f(θ_2 | S = 0). The dashed line plots the density of factor 2 conditional on choosing the college sector, that is, f(θ_2 | S = 1).

Web Data Appendix Figure II-5.6

Densities of Factor 3 by schooling level NLS/66



Let f(θ_3) denote the probability density function of factor θ_3. We allow f(θ_3) to be a mixture of normals. The solid line plots the density of factor 3 conditional on choosing the high school sector, that is, f(θ_3 | S = 0). The dashed line plots the density of factor 3 conditional on choosing the college sector, that is, f(θ_3 | S = 1).

Web Supplement III

Identification Analysis of Cunha, Heckman and Navarro

III-1 Distinguishing between heterogeneity and uncertainty

In the literature on earnings dynamics (*e.g.* Lillard and Willis, 1978), it is common to estimate an earnings equation of the sort

$$Y_{i,t} = \boldsymbol{X}_{i,t}\boldsymbol{\beta} + S_i\tau + v_{i,t}, \tag{III-1}$$

where $Y_{i,t}, \boldsymbol{X}_{i,t}, S_i, v_{i,t}$ denote (for person i at time t) the realized earnings, observable characteristics, educational attainment, and unobservable characteristics, respectively, from the point of view of the observing economist. We use bold characters to denote vectors and distinguish them from scalars. The variables generating outcomes realized at time t may or may not have been known to the agents at the time they made their schooling decisions.

Often the error term $v_{i,t}$ is decomposed into two or more components. For example, it is common to specify that

$$v_{i,t} = \phi_i + \varepsilon_{i,t}. \tag{III-2}$$

The term ϕ_i is a person-specific effect. The error term $\varepsilon_{i,t}$ is generally assumed to follow an ARMA (p, q) process (see, *e.g.* MaCurdy, 1982) such as $\varepsilon_{i,t} = \rho\varepsilon_{i,t-1} + m_{i,t}$, where $m_{i,t}$ is a mean zero innovation independent of $\boldsymbol{X}_{i,t}$ and the other error components. The components $\boldsymbol{X}_{i,t}, \phi_i$, and $\varepsilon_{i,t}$ all contribute to measured *ex post* variability across persons. However, the literature is silent about the difference between heterogeneity and uncertainty, the unforecastable part of earnings as measured from a given age—what Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns, and Michelson (1972) call 'luck.'

An alternative specification of the error process postulates a factor structure for earnings,

$$v_{i,t} = \boldsymbol{\theta}_i\boldsymbol{\alpha}_t + \delta_{i,t}, \tag{III-3}$$

where $\boldsymbol{\theta}_i$ is a vector of skills (*e.g.* ability, initial human capital, motivation, and the like), $\boldsymbol{\alpha}_t$ is a vector of skill prices, and the $\delta_{i,t}$ are mutually independent mean zero shocks independent of $\boldsymbol{\theta}_i$. See Hause (1980) and Heckman and Scheinkman (1987) for analysis of such a model. Any process in the form of equation (III-2) can be written in terms of (III-3). The latter specification is more directly interpretable as a pricing equation than (III-2) and is a natural starting point for human capital analyses. It is the one used in this paper.

Depending on the available market arrangements for coping with risk, the predictable components of $v_{i,t}$ will have a different effect on choices and economic welfare than the unpredictable components, if people are risk averse and cannot fully insure against uncertainty. Statistical decompositions based on (III-1), (III-2), and (III-3) or versions of them describe *ex post* variability but tell us nothing about which components of

(III-1) or (III-3) are forecastable by agents *ex ante*. Is ϕ_i unknown to the agent? $\varepsilon_{i,t}$? Or $\phi_i + \varepsilon_{i,t}$? Or $m_{i,t}$? In representation (III-3), the entire vector $\boldsymbol{\theta}_i$, components of the $\boldsymbol{\theta}_i$, the $\delta_{i,t}$, or all of these may or may not be known to the agent at the time schooling choices are made.

The methodology presented in this paper provides a framework with which it is possible to identify components of life cycle outcomes that are forecastable and acted on at the time decisions are taken from ones that are not. The essential idea of the method can be illustrated in the case of educational choice, the problem we study in our empirical work. In order to choose between high school and college, say at age 19, agents forecast future earnings (and other returns and costs) for each schooling level. Using information about educational choices at age 19, together with the *ex post* realization of earnings and costs that are observed at later ages, it is possible to estimate and test which components of future earnings and costs are forecast by the agent at age 19. This can be done provided we know, or can estimate, the earnings of agents under both schooling choices and provided we specify the market environment under which they operate as well as their preferences over outcomes. For certain market environments where separation theorems are valid, so that consumption decisions are made independently of the wealth maximizing decision, it is not necessary to know agent preferences to decompose realized earnings outcomes in this fashion. Our method uses choice information to extract *ex ante* or forecast components of earnings and to distinguish them from realized earnings. The difference between forecast and realized earnings allows us to identify the distributions of the components of uncertainty facing agents at the time they make their schooling decisions.

To be more precise, consider a version of the generalized Roy (1951) economy with two sectors.[3] Let S_i denote different schooling levels. $S_i = 0$ denotes choice of the high school sector for person i, and $S_i = 1$ denotes choice of the college sector. Each person chooses to be in one or the other sector but cannot be in both. Let the two potential outcomes be represented by the pair $(Y_{0,i}, Y_{1,i})$, only one of which is observed by the analyst for any agent. Denote by C_i the direct cost of choosing sector 1, which is associated with choosing the college sector (*e.g.* tuition and non-pecuniary costs of attending college expressed in monetary values).

$Y_{1,i}$ is the *ex post* present value of earnings in the college sector, discounted over horizon T for a person choosing at a fixed age, assumed for convenience to be zero,

$$Y_{1,i} = \sum_{t=0}^{T} \frac{Y_{1,i,t}}{(1+r)^t},$$

[3]See Heckman (1990) and Heckman and Smith (1998) for discussions of the generalized Roy model. In this paper we assume only two schooling levels for expositional simplicity, although our methods apply more generally.

and $Y_{0,i}$ is the *ex post* present value of earnings in the high school sector at age zero,

$$Y_{0,i} = \sum_{t=0}^{T} \frac{Y_{0,i,t}}{(1+r)^t},$$

where r is the one-period risk-free interest rate. $Y_{1,i}$ and $Y_{0,i}$ can be constructed from time series of *ex post* potential earnings streams in the two states: $(Y_{0,i,0}, \ldots, Y_{0,i,T})$ for high school and $(Y_{1,i,0}, \ldots, Y_{1,i,T})$ for college. A practical problem is that we only observe one or the other of these streams. This partial observability creates a fundamental identification problem which we address in this paper.

The variables $Y_{1,i}$, $Y_{0,i}$, and C_i are *ex post* realizations of returns and costs, respectively. At the time agents make their schooling choices, these may be only partially known to the agent, if at all. Let $\mathcal{I}_{i,0}$ denote the information set of agent i at the time the schooling choice is made, which is time period $t = 0$ in our notation. Under a complete markets assumption with all risks diversifiable (so that there is risk-neutral pricing) or under a perfect foresight model with unrestricted borrowing or lending but full repayment, the decision rule governing sectoral choices at decision time '0' is

$$S_i = \begin{cases} 1, \text{ if } E\left(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}\right) \geq 0 \\ 0, \text{ otherwise.}^4 \end{cases} \tag{III-4}$$

Under perfect foresight, the postulated information set would include $Y_{1,i}$, $Y_{0,i}$, and C_i. In either model of information, the decision rule is simple: one attends school if the expected gains from schooling are greater than or equal to the expected costs. Under either set of assumptions, a separation theorem governs choices. Agents maximize expected wealth independently of how they consume it.

The decision rule is more complicated in the absence of full risk diversifiability and depends on the curvature of utility functions, the availability of markets to spread risk, and possibilities for storage. (See Heckman and Navarro (2004), and Navarro (2004) for a more extensive discussion.) In more realistic economic settings, the components of earnings and costs required to forecast the gain to schooling depend on higher moments than the mean. In this paper we use a model with a simple market setting to motivate the identification analysis of a more general environment we analyze elsewhere (Carneiro, Hansen, and Heckman, 2003)

Suppose that we seek to determine $\mathcal{I}_{i,0}$. This is a difficult task. Typically we can only partially identify $\mathcal{I}_{i,0}$ and generate a list of candidate variables that belong in the information set. We can usually only estimate the distributions of the unobservables in $\mathcal{I}_{i,0}$ (from the standpoint of the econometrician) and not individual person-specific information sets. To fix ideas, we start the analysis discussing identification of $\mathcal{I}_{i,0}$ for each

[4]If there are aggregate sources of risk, full insurance would require a linear utility function.

person, but in our empirical work we only partially identify person-specific $\mathcal{I}_{i,0}$ and instead identify the distributions of the remaining unobserved components.

To motivate the objectives of our analysis we offer the following heuristic discussion. We seek to decompose the 'returns coefficient' in an earnings-schooling model into components that are known at the time schooling choices are made and components that are not known. For simplicity we assume that, for person *i*, returns are the same at all levels of schooling. Write discounted lifetime earnings of person *i* as

$$Y_i = \rho_0 + \rho_{1,i} S_i + J_i, \tag{III-5}$$

where $\rho_{1,i}$ is the person-specific *ex post* return, S_i is years of schooling, and J_i is a mean zero unobservable. We seek to decompose $\rho_{1,i}$ into two components $\rho_{1,i} = \eta_i + \nu_i$, where η_i is a component known to the agent when he/she makes schooling decisions and ν_i is revealed after the choice is made. Schooling choices are assumed to depend on what is known to the agent at the time decisions are made, $S_i = \lambda\left(\eta_i, \boldsymbol{Z}_i, \tau_i\right)$, where the $\boldsymbol{Z}_i$ are other observed determinants of schooling and τ_i represents additional factors unobserved by the analyst but known to the agent. We seek to determine what components of *ex post* lifetime earnings Y_i enter the schooling choice equation.

If η_i is known, it enters λ. Otherwise it does not. Component ν_i and any measurement errors in $Y_{1,i}$ or $Y_{0,i}$ should not be determinants of schooling choices. Neither should future skill prices that are unknown at the time agents make their decisions. If agents do not use η_i in making their schooling choices, even if they know it, η_i would not enter the schooling choice equation. Determining the correlation between realized Y_i and schooling choices based on *ex ante* forecasts enables us to identify components known to agents making their schooling decisions. Even if we cannot identify $\rho_{1,i}$, η_i, or ν_i for each person, under conditions specified in this paper we can identify their distributions.

Suppose that the model for schooling can be written in linear in parameters form:

$$S_i = \lambda_0 + \lambda_1 \eta_i + \lambda_2 \nu_i + \boldsymbol{\lambda}_3 \boldsymbol{Z}_i + \tau_i, \tag{III-6}$$

where τ_i has mean zero and is independent of $\boldsymbol{Z}_i$. $\boldsymbol{Z}_i$ is assumed to be independent of η_i and ν_i. The $\boldsymbol{Z}_i$ and the τ_i proxy costs and may also be correlated with J_i in (III-5).[5] In this framework, the goal of the analysis is to determine if $\lambda_2 = 0$, *i.e.*, to determine if agents pick schooling based on *ex post* shocks to returns and, if they do, the relative magnitude of the variance of η_i to that of ν_i.

Application of $\boldsymbol{Z}_i$ as an instrument for S_i in outcome equation (III-5) does not enable us to decompose

[5] Card (2001) presents a perfect certainty model that can be written in this form.

$\rho_{1,i}$ into forecastable and unforecastable components. Only if agents do not use η_i in making their schooling decisions does the instrumental variable (IV) method recover the population mean of $\rho_{1,i}$. In that case, standard random coefficient models can identify the variance of $(\eta_i + \nu_i)$ which is assumed to be independent of S_i.[6]

Notice that even under the most favorable conditions for application of the IV method, we are only able to recover the *ex post* mean and total *ex post* variability of $\rho_{1,i} = \eta_i + \nu_i$. We cannot, however, decompose $\mathrm{Var}(\eta_i + \nu_i)$ into its components. That is, we are not able to assign the proportion of the variance in the return that is due to η_i and that due to ν_i. Since we cannot identify how much of the *ex post* return to schooling is unknown to the agent at the time he makes his decision, we cannot solve the stated problem using just the instrumental variable method.[7]

Our procedure is not based on the method of instrumental variables. Rather, it exploits certain covariances that arise under different information structures. To see how the method works, simplify the model down to two schooling levels. Suppose, contrary to what is possible, that the analyst observes $Y_{0,i}$, $Y_{1,i}$, and C_i. Such information would come from an ideal data set in which we could observe two different lifetime earnings streams for the same person in high school and in college as well as the costs they pay for attending college. From such information we could construct $Y_{1,i} - Y_{0,i} - C_i$. If we knew the information set $\mathcal{I}_{i,0}$ of the agent, we could also construct $E(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0})$. Under the correct model of expectations, we could form the residual

$$V_{\mathcal{I}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E(Y_{1,i} - Y_{0,i} - C_i \mid \mathcal{I}_{i,0}),$$

and from the *ex ante* college choice decision, we could determine whether S_i depends on $V_{\mathcal{I}_{i,0}}$. It should not if we have specified $\mathcal{I}_{i,0}$ correctly. In terms of the model of equations (III-5) and (III-6), if there are no direct costs of schooling, $E(Y_{1,i} - Y_{0,i} \mid \mathcal{I}_{i,0}) = \eta_i$, and $V_{\mathcal{I}_{i,0}} = \nu_i$. A test for correct specification of candidate information set $\widetilde{\mathcal{I}}_{i,0}$ is a test of whether S_i depends on $V_{\widetilde{\mathcal{I}}_{i,0}}$, where $V_{\widetilde{\mathcal{I}}_{i,0}} = (Y_{1,i} - Y_{0,i} - C_i) - E\left(Y_{1,i} - Y_{0,i} - C_i \mid \widetilde{\mathcal{I}}_{i,0}\right)$.

More precisely, the information set is valid if $S_i \perp\!\!\!\perp V_{\widetilde{\mathcal{I}}_{i,0}} \mid \widetilde{\mathcal{I}}_{i,0}$, where $X \perp\!\!\!\perp Y \mid Z$ means X is independent of Y given Z. In terms of the simple model of (III-5) and (III-6), ν_i should not enter the schooling choice equation ($\lambda_2 = 0$). A test of misspecification of $\widetilde{\mathcal{I}}_{i,0}$ is a test of whether the coefficient of $V_{\widetilde{\mathcal{I}}_{i,0}}$ is statistically significantly different from zero in the schooling choice equation.

More generally, $\widetilde{\mathcal{I}}_{i,0}$ is the correct information set if $V_{\widetilde{\mathcal{I}}_{i,0}}$ does not help to predict schooling. We can search among candidate information sets $\widetilde{\mathcal{I}}_{i,0}$ to determine which ones satisfy the requirement that the generated $V_{\widetilde{\mathcal{I}}_{i,0}}$ does not predict S_i and what components of $Y_{1,i} - Y_{0,i} - C_i$ (and $Y_{1,i} - Y_{0,i}$) are predictable at the age

[6] One can use the residuals from $Y_i - \widehat{\rho}_0 - \widehat{\rho}_1 S_i = \widehat{U}_i$ to decompose the variance components, where instrumental variables are used to generate the coefficient estimates. For the instrumental variable method in this case, see Heckman and Vytlacil (1998).

[7] For further discussion of the IV method applied to separate heterogeneity from uncertainty see Cunha and Heckman (2007).

for the specified information set.[8] For a properly specified $\widetilde{\mathcal{I}}_{i,0}$, $V_{\widetilde{\mathcal{I}}_{i,0}}$ should not cause (predict) schooling choices. The components of $V_{\widetilde{\mathcal{I}}_{i,0}}$ that are unpredictable are called intrinsic components of uncertainty, as defined in this paper.

Usually, we cannot determine the exact content of $\mathcal{I}_{i,0}$ known to each agent. If we could, we would perfectly predict S_i given our decision rule. More realistically, we might find variables that proxy $\mathcal{I}_{i,0}$ or their distribution. Thus, in the example of equations (III-5) and (III-6) we would seek to determine the distribution of ν_i and the allocation of the variance of $\rho_{1,i}$ to η_i and ν_i rather than trying to estimate $\rho_{1,i}$, η_i, or ν_i for each person. This is the strategy pursued in this paper for a two-choice model of schooling.

Inference

The procedure just described is not practical for general models of educational outcomes. We do not know all of the information possessed by the agent. We do not observe $Y_{1,i,t}$ and $Y_{0,i,t}$ together for anyone. We must solve the problem of constructing counterfactuals. This entails solving the selection problem.

One conventional way to solve the selection problem is to invoke a 'common coefficient' assumption,

$$Y_{1,i,t} = \varphi_t(\boldsymbol{X}_{i,t}) + Y_{0,i,t}, \qquad t = 0, \ldots, T,$$

where $\varphi_t(\boldsymbol{X}_{i,t})$ is the same for everyone with the same $\boldsymbol{X}_{i,t}$. A special case is where $\varphi_t(\boldsymbol{X}_{i,t}) = \varphi$, a constant. This specification assumes that for each person i, the earnings in college at age t differ from the earnings in high school by a constant, or a constant conditional on $\boldsymbol{X}_{i,t}$. Under standard assumptions, conventional econometric methods such as matching, instrumental variables, or control functions recover $\varphi_t(\boldsymbol{X}_{i,t})$ for everyone (see Heckman and Robb, 1986, reprinted 2000, for discussions of alternative assumptions).

A common coefficient returns to schooling assumption for all groups with the same values of $\boldsymbol{X}_{i,t}$ rules out comparative advantage in the labor market that has been shown to be empirically important (see Heckman, 2001, and Carneiro, Heckman, and Vytlacil, 2005). The common coefficient assumption can be tested nonparametrically and is decisively rejected (Heckman, Smith, and Clements, 1997). An alternative and weaker assumption is that ranks in the distribution of $Y_{1,i,t}$ can be mapped into ranks in the distribution of $Y_{0,i,t}$ (*e.g.* the best in the $Y_{1,i,t}$ distribution is the best in the $Y_{0,i,t}$ distribution or the best in one is the worst in the other). We present evidence against that assumption below.

An alternative approach is to use matching. Given matching variables $\boldsymbol{Q}_i$, we can form counterfactual

[8] This procedure is a Sims (1972) version of a Wiener-Granger causality test.

marginal distributions from observed distributions using the matching assumption that

$$\begin{aligned} F(Y_{1,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 1, \boldsymbol{Q}_i) &= F(Y_{1,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 0, \boldsymbol{Q}_i) \\ &= F(Y_{1,i,t} \mid \boldsymbol{X}_{i,t}, \boldsymbol{Q}_i), \quad t = 0, \ldots, T. \end{aligned}$$

If the matching assumptions are valid, we can construct counterfactuals for everyone since the first distribution is observed and the second is the distribution of the counterfactual (what persons who do not attend college would have earned if they had attended college). By a parallel analysis of $F(Y_{0,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 0, \boldsymbol{Q}_i)$, we can construct $F(Y_{0,i,t} \mid \boldsymbol{X}_{i,t}, S_i = 1, \boldsymbol{Q}_i) = F(Y_{0,i,t} \mid \boldsymbol{X}_{i,t}, \boldsymbol{Q}_i)$ for everyone, $t = 0, \ldots, T$. This is the distribution of high school outcomes for those who attend college. The marginal distributions acquired from matching are not enough to construct the distribution of returns $Y_{1,i} - Y_{0,i}$ because they do not identify the covariance or dependence between $Y_{1,i,t}$ and $Y_{0,i,t}$, unless it is assumed that the only dependence across the $Y_{1,i,t}$ and $Y_{0,i,t}$ is due to $\boldsymbol{Q}_i$ and/or $\boldsymbol{X}_{i,t}$, and the parameters of this dependence can be determined from the marginal distributions, or else special assumptions about dependence across outcomes are invoked.

Matching makes strong assumptions about the richness of the data available to analysts and does not, in general, identify joint distributions of counterfactual returns and hence the distribution of the rate of return. It assumes that the return to the marginal person is the same as the return to the average person conditional on the matching variables (Heckman and Navarro, 2004).

Either matching or *IV* solves the selection problem under their assumed identifying conditions. Neither method provides a way for identifying the information agents act on *ex ante* when there are important unobserved (by the econometrician) components. In this paper, we build on Carneiro, Hansen, and Heckman (2003) and use the factor structure representation (III-3) to construct the missing counterfactual earnings data.

To understand the essential idea underlying our method, consider the following linear in parameters model:

$$\begin{aligned} Y_{0,i,t} &= \boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + v_{0,i,t}, \qquad t = 0, \ldots, T, \\ Y_{1,i,t} &= \boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + v_{1,i,t}, \\ C_i &= \boldsymbol{Z}_i\boldsymbol{\gamma} + v_{i,C}. \end{aligned}$$

We assume that the life cycle of the agent ends after period T. Linearity of outcomes in terms of parameters is convenient but not essential to our method.

Suppose that there exists a vector of factors $\boldsymbol{\theta}_i = (\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,L})$ such that $\theta_{i,k}$ and $\theta_{i,j}$ are mutually independent random variables for $k, j = 1, \ldots, L, k \neq j$. Assume we can represent the error term in earnings at age t for agent i in the following manner:

$$\begin{aligned} \upsilon_{0,i,t} &= \boldsymbol{\theta}_i \boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}, \\ \upsilon_{1,i,t} &= \boldsymbol{\theta}_i \boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}, \end{aligned}$$

where $\boldsymbol{\alpha}_{0,t}$ and $\boldsymbol{\alpha}_{1,t}$ are vectors and $\boldsymbol{\theta}_i$ is a vector distributed independently across persons. The $\varepsilon_{0,i,t}$ and $\varepsilon_{1,i,t}$ are mutually independent of each other and independent of the $\boldsymbol{\theta}_i$. We can also decompose the cost function C_i in a similar fashion:

$$C_i = \boldsymbol{Z}_i \boldsymbol{\gamma} + \boldsymbol{\theta}_i \boldsymbol{\alpha}_C + \varepsilon_{i,C}.$$

All of the statistical dependence across potential outcomes and costs is generated by $\boldsymbol{\theta}$, $\boldsymbol{X}$, and $\boldsymbol{Z}$. Thus, if we could match on $\boldsymbol{\theta}_i$ (as well as $\boldsymbol{X}$ and $\boldsymbol{Z}$), we could use matching to infer the distribution of counterfactuals and capture all of the dependence across the counterfactual states through the $\boldsymbol{\theta}_i$. However, in general, not all of the required elements of $\boldsymbol{\theta}_i$ are observed.

The parameters $\boldsymbol{\alpha}_C$ and $\boldsymbol{\alpha}_{s,t}$ for $s = 0, 1$, and $t = 0, \ldots, T$ are the factor loadings. $\varepsilon_{i,C}$ is independent of the $\boldsymbol{\theta}_i$ and the other ε components. In this notation, the choice equation can be written as:

$$\begin{aligned} I_i &= E\left(\begin{array}{c} \sum_{t=0}^{T} \frac{\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i \boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - \left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i \boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+r)^t} - \\ \left(\boldsymbol{Z}_i \boldsymbol{\gamma} + \boldsymbol{\theta}_i \boldsymbol{\alpha}_C + \varepsilon_{iC}\right) \end{array} \middle| \, \mathcal{I}_{i,0} \right) \\ S_i &= 1 \text{ if } I_i \geq 0;\; S_i = 0 \text{ otherwise.} \end{aligned} \tag{III-7}$$

The sum inside the parentheses is the discounted earnings of agent i in college minus the discounted earnings of the agent in high school. The second term is cost. Constructing (III-7) entails making a counterfactual comparison. Even if the earnings of one schooling level are observed over the lifetime using panel data, the earnings in the counterfactual state are not. After the schooling choice is made, some components of the $\boldsymbol{X}_{i,t}$, the $\boldsymbol{\theta}_i$, and the $\varepsilon_{i,t}$ may be revealed (*e.g.* unemployment rates, macro shocks) to both the observing economist and the agent, although different components may be revealed to each and at different times. Examining alternative information sets, one can determine which ones produce models for outcomes that fit the data best in terms of producing a model that predicts date $t = 0$ schooling choices and at the same time passes our test for misspecification of predicted earnings and costs. Some components of the error terms may be known or not known at the date schooling choices are made. The unforecastable components are intrinsic uncertainty as we have defined it.[9]

[9] As pointed out to us by Lars Hansen, the term 'heterogeneity' is somewhat unfortunate. Under this term, we include trends

To formally characterize our empirical procedure, it is useful to introduce some additional notation. Let $\odot$ denote the Hadamard product ($\boldsymbol{a} \odot \boldsymbol{b} = (\mathrm{a}_1 b_1, \ldots, \mathrm{a}_L b_L)$) for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ of length L. Let $\boldsymbol{\Delta}_{X_t}$, $t = 0, ..., T$, $\boldsymbol{\Delta}_Z$, $\boldsymbol{\Delta}_\theta$, Δ_{ε_C}, Δ_{ε_t}, denote coefficient vectors associated with the $\boldsymbol{X}_t$, $t = 0, ..., T$, the $\boldsymbol{Z}$, the $\boldsymbol{\theta}$, the $\varepsilon_{1,t} - \varepsilon_{0,t}$, and the ε_C, respectively. These coefficients will be estimated to be nonzero in a schooling choice equation if there is a deviation between the proposed information set and the actual information set used by agents. For a proposed information set $\widetilde{\mathcal{I}}_{i,0}$ which may or may not be the true information set on which agents act we can define the proposed choice index $\widetilde{I}_i$ in the following way:

$$
\begin{aligned}
\widetilde{I}_i \;=\; & \sum_{t=0}^{T} \frac{E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} (\boldsymbol{\beta}_{1,t} - \boldsymbol{\beta}_{0,t}) + \sum_{t=0}^{T} \frac{\left[\boldsymbol{X}_{i,t} - E\left(\boldsymbol{X}_{i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t} (\boldsymbol{\beta}_{1,t} - \boldsymbol{\beta}_{0,t}) \odot \boldsymbol{\Delta}_{X_t} \\
& + E(\boldsymbol{\theta}_i \mid \widetilde{\mathcal{I}}_{i,0}) \left[\sum_{t=0}^{T} \frac{(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t})}{(1+r)^t} - \boldsymbol{\alpha}_C\right] \\
& + \left[\boldsymbol{\theta}_i - E\left(\boldsymbol{\theta}_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \left\{\left[\sum_{t=0}^{T} \frac{(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t})}{(1+r)^t} - \boldsymbol{\alpha}_C\right] \odot \boldsymbol{\Delta}_\theta\right\} + \sum_{t=0}^{T} \frac{E\left(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)}{(1+r)^t} \\
& + \sum_{t=0}^{T} \frac{\left[(\varepsilon_{1,i,t} - \varepsilon_{0,i,t}) - E\left(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right]}{(1+r)^t} \Delta_{\varepsilon_t} - E\left(\boldsymbol{Z}_i \mid \widetilde{\mathcal{I}}_{i,0}\right) \boldsymbol{\gamma} \\
& - \left[\boldsymbol{Z}_i - E\left(\boldsymbol{Z}_i \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \boldsymbol{\gamma} \odot \boldsymbol{\Delta}_Z - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right) - \left[\varepsilon_{iC} - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_{i,0}\right)\right] \Delta_{\varepsilon_C}.
\end{aligned}
\tag{III-8}
$$

To conduct our test, we fit a schooling choice model based on the proposed model (III-8). We estimate the parameters of the model including the $\boldsymbol{\Delta}$ parameters. This decomposition for $\widetilde{I}_i$ assumes that agents know the $\boldsymbol{\beta}$, the $\boldsymbol{\gamma}$, and the $\boldsymbol{\alpha}$. If it is not correct, the presence of additional unforecastable components due to unknown coefficients affects the interpretation of the estimates.

A test of no misspecification of information set $\widetilde{\mathcal{I}}_{i,0}$ is a joint test of the hypothesis that $\boldsymbol{\Delta}_{X_t} = 0$, $\boldsymbol{\Delta}_\theta = 0$, $\boldsymbol{\Delta}_Z = 0$, $\Delta_{\varepsilon_C} = 0$, and $\Delta_{\varepsilon_t} = 0$, $t = 0, \ldots, T$. That is, when $\widetilde{\mathcal{I}}_{i,0} = \mathcal{I}_{i,0}$, then $\boldsymbol{\Delta}_{X_t} = 0$, $\boldsymbol{\Delta}_\theta = 0$, $\boldsymbol{\Delta}_Z = 0$, $\Delta_{\varepsilon_C} = 0$, $\Delta_{\varepsilon_t} = 0$, $t = 0, \ldots, T$, and the proposed choice index $\widetilde{I}_i = I_i$.

In a correctly specified model, the components associated with zero $\boldsymbol{\Delta}_j$ are the unforecastable elements or the elements which, even if known to the agent, are not acted on in making schooling choices. To illustrate the application of our method, assume for simplicity that the $\boldsymbol{X}_{i,t}$, the $\boldsymbol{Z}_i$, the $\varepsilon_{i,C}$, the $\boldsymbol{\beta}_{1,t}, \boldsymbol{\beta}_{0,t}$, the $\boldsymbol{\alpha}_{1,t}, \boldsymbol{\alpha}_{0,t}$, and $\boldsymbol{\alpha}_C$ are known to the agent, and the $\varepsilon_{j,i,t}$ are unknown and are set at their mean zero values. We can infer which components of the $\boldsymbol{\theta}_i$ are known and acted on in making schooling decisions if we postulate that some components of $\boldsymbol{\theta}_i$ are known perfectly at date $t = 0$ while others are not known at all, and their forecast values have mean zero given $\mathcal{I}_{i,0}$.

common across all people (e.g., macrotrends). The real distinction we are making is between components of realized earnings forecastable by agents at the time they make their schooling choices *vs.* components that are not forecastable.

If there is an element of the vector $\boldsymbol{\theta}_i$, say $\theta_{i,2}$ (factor 2), that has nonzero loadings (coefficients) in the schooling choice equation and a nonzero loading on one or more potential future earnings, then one can say that at the time the schooling choice is made, the agent knows the unobservable captured by factor 2 that affects future earnings. If $\theta_{i,2}$ does not enter the choice equation but explains future earnings, then $\theta_{i,2}$ is unknown (not predictable by the agent) at the age schooling decisions are made. An alternative interpretation is that the second component of $\left[\sum_{t=0}^{T} \frac{(\alpha_{1,t}-\alpha_{0,t})}{(1+r)^t} - \boldsymbol{\alpha}_C\right]$ is zero, *i.e.*, that even if the component is known, it is not acted on. We can only test for what the agent knows and acts on.

One plausible scenario is that $\varepsilon_{i,C}$ is known but the future $\varepsilon_{1,i,t}$ and $\varepsilon_{0,i,t}$ are not, have mean zero, and are insurable. If there are components of the $\varepsilon_{j,i,t}$ that are predictable at age $t = 0$, they will induce additional dependence between S_i and future earnings beyond the dependence induced by the $\boldsymbol{\theta}_i$. Under a perfect foresight assumption we can identify this extra dependence. We develop this point further in section III-2 after we introduce additional helpful notation. Our procedure can be generalized to consider all components of (III-8). We can test the predictive power of each subset of the overall possible information set at the date the schooling decision is being made.

The intuition underlying our testing procedure is thus very simple. The components that are forecastable and acted on in making schooling choices are captured by the components of *ex post* realizations that are known by the agents when they make their educational choices. In terms of the simple model of equations (III-5) and (III-6), by decomposing $\rho_{1,i}$ into η_i and ν_i so $\rho_{1,i} = \eta_i + \nu_i$, we determine how much of the *ex post* variability in $\rho_{1,i}$ is due to forecastable η_i and unforecastable ν_i. The predictable components will be estimated to have nonzero coefficients in the schooling choice equation. The uncertainty at the date the decision about college is being made is captured by the factors that the agent does not act on when making the decision of whether or not to attend college.[10]

A similar but distinct idea motivates the Flavin (1981) test of the permanent income hypothesis and her measurement of unforecastable income innovations. She picks a particular information set $\widetilde{\mathcal{I}}_{i,0}$ (permanent income constructed from an assumed ARMA (p, q) time series process for income, where she estimates the coefficients given a specified order of the AR and MA components) and tests if $V_{\widetilde{\mathcal{I}}_{i,0}}$ (our notation) predicts consumption. Her test of 'excess sensitivity' can be interpreted as a test of the correct specification of the ARMA process that she assumes generates $\widetilde{\mathcal{I}}_{i,0}$ which is unobserved (by the economist), although she does not state it that way. Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2004) extend her analysis but, like her, maintain an *a priori* specification of the stochastic process generating $\mathcal{I}_{i,0}$. Blundell, Pistaferri, and Preston (2004) claim to test for 'partial insurance.' In fact their procedure can be

[10]This test has been extended to a nonlinear setting, allowing for credit constraints, preferences for risk, and the like. See Cunha, Heckman, and Navarro (2004) and Navarro (2004).

viewed as a test of their specification of the stochastic process generating the agent's information set. More closely related to our work is the analysis of Pistaferri (2001), who uses the distinction between expected starting wages (to measure expected returns) and realized wages (to measure innovations) in a consumption analysis.

In the context of our factor structure representation, the contrast between our approach to identifying components of intrinsic uncertainty and the approach followed in the literature is as follows. The traditional approach would assume that the $\boldsymbol{\theta}_i$ are known to the agent while the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are not.[11] Our approach allows us to determine which components of $\boldsymbol{\theta}_i$ and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known and acted on at the time schooling decisions are made.

Assuming that the problems raised by selection on S_i are solved by the methods exposited in the next section and their vector generalizations, we can estimate the distributions of the components of (III-3) and the coefficients on the factors $\boldsymbol{\theta}_i$ from panel data on earnings. This statistical decomposition does not tell us which components of (III-3) are known at the time agents make their schooling decisions. If some of the components of $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ are known to the agent at the date schooling decisions are made and enter (III-8), then additional dependence between S_i and future $Y_{1,i} - Y_{0,i}$ due to the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$, beyond that due to $\boldsymbol{\theta}_i$, would be estimated.

It is helpful to contrast the dependence between S_i and future $Y_{0,i,t}, Y_{1,i,t}$ arising from $\boldsymbol{\theta}_i$ and the dependence between S_i and the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$. Some of the $\boldsymbol{\theta}_i$ in the *ex post* earnings equation may not appear in the choice equation. Under other information sets, some additional dependence between S_i and $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$ may arise. The contrast between the sources generating realized earnings outcomes and the sources generating dependence between S_i and realized earnings is the essential idea in this paper. The method can be generalized to deal with nonlinear preferences and imperfect market environments.[12]

[11] The analysis of Hartog and Vijverberg (2002) exemplifies this approach and uses variances of *ex post* income to proxy *ex ante* variability.

[12] In a model with complete autarky with preferences G, ignoring costs,

$$I_i = \sum_{t=0}^{T} E\left[\frac{G\left(\mathbf{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - G\left(\mathbf{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \,\middle|\, \widetilde{\mathcal{I}}_{i,0} \right],$$

where ρ is the time rate of discount, we can make a similar decomposition but it is more complicated given the nonlinearity in G. For this model we could do a Sims noncausality test where

$$\begin{aligned} V_{\widetilde{\mathcal{I}}_{i,0}} &= \sum_{t=0}^{T} \frac{G\left(\mathbf{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - G\left(\mathbf{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \\ &\quad - \sum_{t=0}^{T} E\left[\frac{G\left(\mathbf{X}_{i,t}\boldsymbol{\beta}_{1,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{1,t} + \varepsilon_{1,i,t}\right) - G\left(\mathbf{X}_{i,t}\boldsymbol{\beta}_{0,t} + \boldsymbol{\theta}_i\boldsymbol{\alpha}_{0,t} + \varepsilon_{0,i,t}\right)}{(1+\rho)^t} \,\middle|\, \widetilde{\mathcal{I}}_{i,0} \right]. \end{aligned}$$

This requires some specification of G. See Carneiro, Hansen, and Heckman (2003), who assume $G(Y) = \ln Y$ and that the equation for $\ln Y$ is linear in parameters. Cunha, Heckman, and Navarro (2004) and Navarro (2004) generalize that framework to a model with imperfect capital markets where some lending and borrowing is possible.

III-2 Identifying counterfactual distributions and extracting components of unpredictable uncertainty using factor models: a semiparametric analysis

To motivate our econometric procedures, it is useful to work with a slightly more abstract notation and a simpler set up. Omit the individual i subscript to simplify the notation and suppose that there is one period only ($T = 0$) so $Y_1 = Y_{1,0}$, $Y_0 = Y_{0,0}$. We relax this assumption later in this section but initially use this framework to focus on the main econometric ideas motivating our solution of the selection problem. Assume that (Y_0, Y_1) have finite means and can be expressed in terms of conditioning variables $\boldsymbol{X}$. Write

$$Y_0 = \mu_0(\boldsymbol{X}) + U_0, \tag{III-9a}$$

$$Y_1 = \mu_1(\boldsymbol{X}) + U_1, \tag{III-9b}$$

where $E(U_0 \mid \boldsymbol{X}) = E(U_1 \mid \boldsymbol{X}) = 0$, $E(Y_0 \mid \boldsymbol{X}) = \mu_0(\boldsymbol{X})$, and $E(Y_1 \mid \boldsymbol{X}) = \mu_1(\boldsymbol{X})$. The *ex post* gain for an individual who moves from $S = 0$ to $S = 1$ is $Y_1 - Y_0$.

Write index I as a net utility,

$$I = Y_1 - Y_0 - C, \tag{III-10}$$

where C is the cost of participation in sector 1. We write $C = \mu_C(\boldsymbol{Z}) + U_C$, where the $\boldsymbol{Z}$ are determinants of cost. We may write

$$I = \mu_I(\boldsymbol{X}, \boldsymbol{Z}) + U_I. \tag{III-11}$$

Under perfect certainty,

$$\mu_I(\boldsymbol{X}, \boldsymbol{Z}) = \mu_1(\boldsymbol{X}) - \mu_0(\boldsymbol{X}) - \mu_C(\boldsymbol{Z}) \quad \text{and} \quad U_I = U_1 - U_0 - U_C.$$

More generally, we define U_I as the error in the choice equation and it may or may not include all future U_1, U_0, or U_C. Similarly, $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ may only be based on expectations of future $\boldsymbol{X}$ and $\boldsymbol{Z}$ at the time schooling decisions are made. We write

$$S = 1 \text{ if } I \geq 0; \; S = 0 \text{ otherwise.} \tag{III-12}$$

A major advantage of our approach over previous work on estimating components of uncertainty facing agents is that we control for the econometric consequences of endogeneity in the choice of S and thereby avoid self-selection biases. The choice equation is also a source of identifying information for extracting forecastable components. This paper builds on recent research by Carneiro, Hansen, and Heckman (2003) that solves the problem of constructing counterfactuals by identifying the joint distribution of (Y_0, Y_1) conditional on S (or I) using a factor structure model. These models generalize the LISREL models of Jöreskog (1977) and the MIMIC models of Jöreskog and Goldberger (1975) to produce counterfactual distributions. We now exposit the main idea underlying our method, working with a one-factor model to simplify the exposition. Carneiro, Hansen, and Heckman (2003) develop the general multifactor model we use in our empirical analysis.

III-2.1 Identifying counterfactual distributions

Identifying the joint distribution of potential outcomes is a difficult problem because we do not observe both components of (Y_0, Y_1) for anyone. Thus, one cannot directly form the joint distribution of potential outcomes (Y_0, Y_1). Heckman and Honoré (1990) show that if (i) $C = 0$ for every person, (ii) decision rule (III-12) applies in an environment of perfect certainty, (iii) there are distinct variables in $\mu_1(\boldsymbol{X})$ and $\mu_0(\boldsymbol{X})$, (iv) $\boldsymbol{X}$ is independent of (U_1, U_0), and other mild regularity restrictions are satisfied, then one can identify the joint distribution of (Y_0, Y_1) given $\boldsymbol{X}$, even without additional $\boldsymbol{Z}$ variables. In this case the agents choose S solely in terms of the differences in potential outcomes. However, in an environment of uncertainty or if C varies across people and contains some variables unobserved by the analyst, this method breaks down. We present a more general analysis without maintaining the perfect certainty assumption.

As shown by Heckman (1990), Heckman and Smith (1998), and Carneiro, Hansen, and Heckman (2003), under the assumptions that (i) $(\boldsymbol{Z}, \boldsymbol{X})$ are statistically independent from (U_0, U_1, U_I), (ii) $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ is a nontrivial function of $\boldsymbol{Z}$ given $\boldsymbol{X}$, (iii) $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ has full support, and (iv) the elements of the pairs $(\mu_0(\boldsymbol{X}), \mu_I(\boldsymbol{X}, \boldsymbol{Z}))$ and $(\mu_1(\boldsymbol{X}), \mu_I(\boldsymbol{X}, \boldsymbol{Z}))$ can be varied independently of each other, then one can identify the joint distributions of (U_0, U_I), (U_1, U_I) up to a scale σ_I^* for U_I and also $\mu_0(\boldsymbol{X})$, $\mu_1(\boldsymbol{X})$, and $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$, the last expression up to scale σ_I.[13] Thus, one can identify the joint distributions of (Y_0, I^*) and (Y_1, I^*) given $\boldsymbol{X}$ and $\boldsymbol{Z}$ where $I^* = I/\sigma_I$. As a by-product we identify the mean functions. One cannot recover the joint distribution of (Y_0, Y_1) or (Y_0, Y_1, I^*) given $\boldsymbol{X}$ and $\boldsymbol{Z}$ without further assumptions. We provide an intuitive motivation for why $F(Y_0, I^*)$ and $F(Y_1, I^*)$ are identified in in Appendix 1 of Cunha, Heckman, and Navarro (2005). Once we estimate these distributions, we perform factor analysis on (Y_0, I^*) and (Y_1, I^*).

The factor structure approach provides a solution to the problem of constructing counterfactual distributions. We show the essential idea. Suppose that the unobservables follow a one-factor structure (*i.e.*, θ is

[13] Full support means the support of $\mu_I(\boldsymbol{X}, \boldsymbol{Z})$ matches (or contains) the support of U_I. (See Heckman and Honoré, 1990, and Carneiro, Hansen, and Heckman, 2003, for more precise formulations of these conditions.) The support of a random variable is the set of values where it has a positive density.

a scalar). Carneiro, Hansen, and Heckman (2003) generalize these methods to the multifactor case. We can extend these methods to nonseparable models using the analysis reported in Heckman, Matzkin, Navarro, and Urzua (2004), but we do not do so in this paper.

We assume that all of the dependence across (U_0, U_1, U_{I^*}) is generated by a scalar factor θ,

$$\begin{aligned} U_0 &= \theta\alpha_0 + \varepsilon_0, \\ U_1 &= \theta\alpha_1 + \varepsilon_1, \\ U_{I^*} &= \theta\alpha_{I^*} + \varepsilon_{I^*}. \end{aligned}$$

We assume that θ is statistically independent of $(\varepsilon_0, \varepsilon_1, \varepsilon_I)$ and satisfies $E(\theta) = 0$ and $E\left(\theta^2\right) = \sigma^2_\theta$. All the ε's are mutually independent with $E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_{I^*}) = 0$, $\mathrm{Var}(\varepsilon_0) = \sigma^2_{\varepsilon_0}$, $\mathrm{Var}(\varepsilon_1) = \sigma^2_{\varepsilon_1}$, and $\mathrm{Var}(\varepsilon_I) = \sigma^2_{\varepsilon_I}$ (the ε terms are called uniquenesses in factor analysis). Because the factor loadings may be different, the factor may affect outcomes and choices differently and may even have different signs in different equations.

To show how one can recover the joint distribution of (Y_0, Y_1) using factor models, we break the argument into two parts. First we show how to recover the factor loadings, factor variance, and the variances of the uniquenesses. This part is like traditional factor analysis except that some latent variables (*e.g.* I^*) are only observed up to scale so their scale must be normalized. Then, we show how to construct joint distributions of counterfactuals.

III-2.2 Recovering the factor loadings

We consider identification of the model when the analyst has different types of information about the choices and characteristics of the agent.

The case when there is information on Y_0 for $I < 0$ and Y_1 for $I > 0$ and the decision rule is (III-12)

Under the conditions stated in section III-2.1 and the papers referenced there, after conditioning on $\boldsymbol{X}$ and controlling for selection, one can identify $F(U_0, U_{I^*})$ and $F(U_1, U_{I^*})$. From these distributions one can identify the left hand side of

$$\mathrm{Cov}(U_0, U_{I^*}) = \alpha_0\alpha_{I^*}\sigma^2_\theta$$

and

$$\mathrm{Cov}(U_1, U_{I^*}) = \alpha_1\alpha_{I^*}\sigma^2_\theta.$$

The scale of the unobserved I is normalized, a standard condition for discrete choice models. A second normalization that we need to impose is $\sigma_\theta^2 = 1$. This is required since the factor is not observed and we must set its scale. That is, since $\alpha\theta = k\alpha\frac{\theta}{k}$ for any constant k, we need to set the scale by normalizing the variance of θ. We could alternatively normalize some α_j to one. Finally, we set $\alpha_{I^*} = 1$, an assumption we can relax, as noted below.

Under these conditions, we can identify α_1 and α_0 from the known covariances above. From the first covariance, we identify α_0. From the second, we identify α_1. From the normalization, we know σ_θ^2. Since

$$\operatorname{Cov}(U_1, U_0) = \alpha_1\alpha_0\sigma_\theta^2,$$

we can identify the covariance between Y_1 and Y_0 even though we do not observe the pair (Y_1, Y_0) for anyone. We then use the variances $\operatorname{Var}(U_1), \operatorname{Var}(U_0)$ and the normalization $\operatorname{Var}(U_{I^*}) = 1$ to recover the variance of the uniquenesses $\sigma_{\varepsilon_0}^2, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_{I^*}}^2$.

The fact that we needed to normalize both $\sigma_\theta^2 = 1$ and $\alpha_{I^*} = 1$ is a consequence of our assumption that we have only one observation for Y_1 and Y_0. If we have access to more observations on life cycle earnings from panel data, as we do in our empirical work, we can use $(Y_{0,0}, \ldots Y_{0,T}, Y_{1,0}, \ldots, Y_{1,T})$ to relax one normalization, say $\sigma_\theta^2 = 1$, since then we can form, conditional on $\boldsymbol{X}$ and $\boldsymbol{Z}$, the left hand side of

$$\frac{\operatorname{Cov}(U_{1,t'}, U_{1,t})}{\operatorname{Cov}(U_{1,t'}, U_{I^*})} = \alpha_{1,t}$$

and

$$\frac{\operatorname{Cov}(U_{0,t'}, U_{0,t})}{\operatorname{Cov}(U_{0,t'}, U_{I^*})} = \alpha_{0,t},$$

and recover σ_θ^2 from, say, $\operatorname{Cov}(U_{1,t}, U_{I^*}) = \alpha_{1,t}\sigma_\theta^2$. Identification of the variances of the uniquenesses follows as before.

The central idea motivating our identification strategy is that even though we never observe (Y_0, Y_1) as a pair, both Y_0 and Y_1 are linked to S through the choice equation. From S we can generate I^*, using standard methods in discrete choice analysis. From this analysis we effectively observe (Y_0, I^*) and (Y_1, I^*). The common dependence of Y_0 and Y_1 on I^* secures identification of the joint distribution of Y_0, Y_1, I^*. We next develop a complementary strategy based on the same idea where, in addition to a choice equation, we have a measurement equation observed for all observations whether or not Y_1 or Y_0 is observed. The measurement may be a test score which is a proxy for 'ability' θ. This measurement plays the role of I^* and, in certain respects, identification with a measurement of this type is more transparent and more traditional.

Adding a measurement equation

Suppose that we have access to a measurement for θ that is observed whether $S = 1$ or $S = 0$ in addition to data on outcomes S and Y_0 or Y_1. In educational statistics, a test score is often used to proxy ability. Suppose that the analyst has access to one ability test M for each person. Measured ability M is

$$M = \mu_M(\boldsymbol{X}) + U_M.$$

Assume that

$$U_M = \theta\alpha_M + \varepsilon_M,$$

where ε_M is mutually independent from $(\varepsilon_0, \varepsilon_1, \varepsilon_I)$, and θ.[14] We assume $\alpha_M \neq 0$. With this additional information we can form

$$\begin{aligned}
\operatorname{Cov}(M, Y_0|\boldsymbol{X}, \boldsymbol{Z}) &= \operatorname{Cov}(U_M, U_0) = \alpha_M\alpha_0\sigma_\theta^2, \\
\operatorname{Cov}(M, Y_1|\boldsymbol{X}, \boldsymbol{Z}) &= \operatorname{Cov}(U_M, U_1) = \alpha_M\alpha_1\sigma_\theta^2, \\
\operatorname{Cov}(M, I^*|\boldsymbol{X}, \boldsymbol{Z}) &= \operatorname{Cov}(U_M, U_{I^*}) = \alpha_M\alpha_{I^*}\sigma_\theta^2.
\end{aligned}$$

Conditioning on $(\boldsymbol{X}, \boldsymbol{Z})$, we can recover the error terms for the unobservables U_0, U_{I^*} and U_M using the preceding arguments. If we impose the normalization $\alpha_M = 1$, which can be interpreted as requiring that higher levels of measured ability are associated with higher levels of factor θ, we can form the ratio

$$\frac{\operatorname{Cov}(U_0, U_{I^*})}{\operatorname{Cov}(U_M, U_{I^*})} = \alpha_0$$

and identify α_0. In a similar fashion, we can form

$$\frac{\operatorname{Cov}(U_1, U_{I^*})}{\operatorname{Cov}(U_M, U_{I^*})} = \alpha_1$$

and we can recover α_1. From

$$\operatorname{Cov}(U_M, U_0) = \alpha_0\sigma_\theta^2,$$

we can obtain σ_θ^2. Finally, we can identify α_{I^*} based on information from

$$\operatorname{Cov}(U_M, U_{I^*}) = \alpha_{I^*}\sigma_\theta^2,$$

[14]For simplicity, we assume that this is a continuous measurement. Discrete measurements can also be used. See Carneiro, Hansen, and Heckman (2003).

so we can obtain α_{I^*} up to scale. Thus, with one measurement, one choice equation and two outcomes we can identify σ_θ^2 and α_{I^*} up to scale. We can use the identified variances $\text{Var}(U_0)$, $\text{Var}(U_1)$, $\text{Var}(U_{I^*}) = 1$, and $\text{Var}(U_M)$ to recover the variance of the uniquenesses $\sigma^2_{\varepsilon_0}, \sigma^2_{\varepsilon_1}, \sigma^2_{\varepsilon_{I^*}}$, and $\sigma^2_{\varepsilon_M}$. Thus, having access to a measurement (M) and choice data with decision rule (III-10)–(III-12) allows us to estimate the covariances among the counterfactual states.[15]

But how to identify the distributions? Traditional factor analysis assumes normality. We present a more general nonparametric analysis. Allowing for nonnormality is essential for getting acceptable empirical results as we note below.

III-2.3 Recovering the distributions nonparametrically

Given the identification of factor loadings, factor variances, and uniquenesses, we show how to identify the marginal distributions of θ and $\varepsilon_0, \varepsilon_1, \varepsilon_{I^*}$ nonparametrically (the last one up to scale). The method is based on a theorem by Kotlarski (1967). For completeness, we state his theorem.

Theorem 1 Suppose that we have two random variables T_1 and T_2 that satisfy:

$$\begin{aligned} T_1 &= \theta + v_1 \\ T_2 &= \theta + v_2 \end{aligned}$$

with θ, v_1, v_2 mutually statistically independent, $E(\theta) < \infty$, $E(v_1) = E(v_2) = 0$, that the conditions for Fubini's Theorem are satisfied for each random variable, and that the random variables possess nonvanishing (almost everywhere) characteristic functions. Then, the densities $f_\theta, f_{v_1}, f_{v_2}$ are identified.

Proof See Kotlarski (1967). □

Applied to the current context, we have a choice equation, two outcome equations, and a measurement equation.[16] Assume that we normalize $\alpha_M = 1$ so that all factor loadings, factor variances, and variances of

[15] We cannot dispense with the choice equation unless we have data on $F(Y_0, M \mid \boldsymbol{X}, \boldsymbol{Z})$ and $F(Y_1, M \mid \boldsymbol{X}, \boldsymbol{Z})$. Recall that, in most cases, we observe data that allows us to construct $F(Y_0, M \mid \boldsymbol{X}, \boldsymbol{Z}, S = 0)$ and $F(Y_1, M \mid \boldsymbol{X}, \boldsymbol{Z}, S = 1)$. The required information for dispensing with the choice equation might be obtained when we have limit sets $\bar{\mathcal{Z}}_u$ and $\bar{\mathcal{Z}}_l$ such that $\Pr(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}) = 1$ for $\boldsymbol{z} \in \bar{\mathcal{Z}}_u$ and $\Pr(S = 0 \mid \boldsymbol{X}, \boldsymbol{Z}) = 0$ for $\boldsymbol{z} \in \bar{\mathcal{Z}}_l$. Then we can replace I with M and do factor analysis(see Carneiro, Hansen, and Heckman, 2001).

[16] Again, for the sake of simplicity, we assume that M is continuous but our methods work for discrete measurements. (See Carneiro, Hansen, and Heckman, 2003).

uniquenesses are known. The system is

$$
\begin{aligned}
I^* &= \mu_{I^*}(\boldsymbol{X}, \boldsymbol{Z}) + \theta\alpha_{I^*} + \varepsilon_{I^*}, \\
Y_0 &= \mu_0(\boldsymbol{X}) + \theta\alpha_0 + \varepsilon_0, \\
Y_1 &= \mu_1(\boldsymbol{X}) + \theta\alpha_1 + \varepsilon_1, \\
M &= \mu_M(\boldsymbol{X}) + \theta + \varepsilon_M.
\end{aligned}
$$

Note that this system can be rewritten as

$$
\begin{aligned}
\frac{I^* - \mu_{I^*}(\boldsymbol{X}, \boldsymbol{Z})}{\alpha_{I^*}} &= \theta + \frac{\varepsilon_{I^*}}{\alpha_{I^*}}, \\
\frac{Y_0 - \mu_0(\boldsymbol{X})}{\alpha_0} &= \theta + \frac{\varepsilon_0}{\alpha_0}, \\
\frac{Y_1 - \mu_1(\boldsymbol{X})}{\alpha_1} &= \theta + \frac{\varepsilon_1}{\alpha_1}, \\
M - \mu_M(\boldsymbol{X}) &= \theta + \varepsilon_M.
\end{aligned}
$$

Applying Kotlarski's theorem to any pair of equations, we conclude that we can identify the densities of $\theta, \frac{\varepsilon_{I^*}}{\alpha_{I^*}}, \frac{\varepsilon_0}{\alpha_0}, \frac{\varepsilon_1}{\alpha_1}, \varepsilon_M$. Since we know α_{I^*}, α_0, and α_1, we can identify the densities of $\theta, \varepsilon_{I^*}, \varepsilon_0, \varepsilon_1, \varepsilon_M$.[17] Thus, we can identify the distributions of all of the error terms. Finally, to recover the joint distribution of (Y_1, Y_0), note that

$$
F(Y_1, Y_0 \mid \boldsymbol{X}) = \int F(Y_1, Y_0 \mid \theta, \boldsymbol{X})\, dF_\theta(\theta).
$$

From Kotlarski's Theorem, $F_\theta(\theta)$ is known. Because of the factor structure, Y_1, Y_0, and S are independent once we condition on θ, so it follows that

$$
F(Y_1, Y_0 \mid \theta, \boldsymbol{X}) = F(Y_1 \mid \theta, \boldsymbol{X})\, F(Y_0 \mid \theta, \boldsymbol{X}).
$$

But $F(Y_1 \mid \theta, \boldsymbol{X})$ and $F(Y_0 \mid \theta, \boldsymbol{X})$ are identified once we condition on the factors since

$$
\begin{aligned}
F(Y_1 \mid \theta, \boldsymbol{X}, S = 1) &= F(Y_1 \mid \theta, \boldsymbol{X}) \\
F(Y_0 \mid \theta, \boldsymbol{X}, S = 0) &= F(Y_0 \mid \theta, \boldsymbol{X}).
\end{aligned}
$$

Note further that if θ were known to the analyst, our procedure would be equivalent to matching on θ which is equivalent, for identification, to matching on the propensity score $\Pr(S = 1 \mid \boldsymbol{X}, \boldsymbol{Z}, \theta)$.[18] Our method

[17] Recall that U_I is only known up to scale σ_I.

[18] Carneiro, Hansen, and Heckman (2003) discuss the matching relationship between factor and matching models. For a discussion

generalizes matching by allowing the variables that would produce the conditional independence assumed in matching to be unobserved by the analyst.

The discussion in this section is for a one-factor model. In our empirical work, we use a multifactor model where the factors are used to characterize earnings dynamics and possible dependence between future ε and S. Carneiro, Hansen, and Heckman (2003) provide the analysis we need for the general multifactor case. The key idea is that, with enough measurements, outcomes and choice equations, we can identify the number of factors generating dependence among the Y_1, Y_0, C, S, and M and the distributions of the factors.[19]

III-2.4 Models with multiple factors and tests for full insurance versus perfect certainty

The empirical work in CHN is based on a 5 period ($t = 0, \ldots, 4$) version of equations (III-1) and (III-8). In fitting the model, we introduce the possibility of additional sources of dependence in the choice equation (III-8), distinct from the dependence arising from some or all of the components of $\boldsymbol{\theta}$. This additional dependence may be generated from future $(\varepsilon_{1,i,t}, \varepsilon_{0,i,t})$, $t = 0, \ldots, T$, that affect schooling choices.

From the covariances between S_i (or I_i^*) and $Y_{0,i,t}$ and $Y_{1,i,t}$, $t = 0, \ldots, T$, under certain conditions, we can identify additional sources of dependence between $(Y_{0,i,t}, Y_{1,i,t})$ and I_i^* apart from $\boldsymbol{\theta}_i$ arising from the dependence of $\varepsilon_{0,i,t}$ and $\varepsilon_{1,i,t}$ with $\sum_{t=0}^{T} \frac{E(\varepsilon_{1,i,t} - \varepsilon_{0,i,t} | \widetilde{\mathcal{I}}_{i,0})}{(1+r)^t}$. In our empirical specification discussed below, there are multiple earnings outcomes in each schooling state, a choice equation and a vector of measurement equations to tie down the distribution of $\boldsymbol{\theta}_i$ and the distributions of the $\{\varepsilon_{0,i,t}, \varepsilon_{1,i,t}\}_{t=0}^{T}$.

To see how additional sources of dependence might arise in fitting the data, consider a model with perfect foresight. Following the analysis in section III-2.2 and in the papers cited there, we can estimate

$$\begin{aligned} \operatorname{Cov}\left(Y_{j,i,t}, I_i^* \mid \boldsymbol{X}, \boldsymbol{Z}\right) &= \frac{\boldsymbol{\alpha}'_{j,t}}{\sigma_I^*} \Sigma_\Theta \left[\frac{\sum_{t=0}^{T} \left(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t}\right)}{(1+r)^t} - \boldsymbol{\alpha}_C \right] + \left(\frac{1}{\sigma_I^*}\right) \frac{\operatorname{Var}\left(\varepsilon_{j,i,t}\right)}{(1+r)^t}, \\ & t = 0, \ldots, T; \; j = 0, 1, \end{aligned}$$

where Σ_Θ is the variance-covariance matrix of the $\boldsymbol{\theta}_i$. Conditional on $\boldsymbol{X}$ and $\boldsymbol{Z}$, dependence between $Y_{j,i,t}$ and I_i^* can arise from two sources: from the $\boldsymbol{\theta}_i$ and from the $\varepsilon_{j,i,t}$. Under complete markets, if the $\varepsilon_{j,i,t}$ are unknown at date $t = 0$ and have mean zero given $\mathcal{I}_{i,0}$, the second term on the right hand side vanishes and

of factor models and control functions, see Heckman and Navarro (2004).

[19] A precise statement of what is 'enough' information is given in Carneiro, Hansen, and Heckman (2003). See their discussion of the Ledermann bound. The key idea is that the number of factors has to be small relative to the number of measurements, outcomes and choice equations. This bound can be relaxed if there are *a priori* restrictions on the factor loadings beyond innocuous normalizations. Using nonnormality one can also relax the Ledermann bound.

the factors $\boldsymbol{\theta}_i$ capture any dependence between $Y_{j,i,t}$ and S_i.

Using limit set arguments, as in Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2004), we can identify the $\boldsymbol{\alpha}_{j,t}$, $j = 0, 1$, $t = 0, \ldots, T$, the distribution of $\boldsymbol{\theta}_i$ and the distributions of the $\varepsilon_{j,i,t}$ from earnings data alone in the limit sets.[20] Under either complete markets or under perfect foresight, we can identify $\boldsymbol{\alpha}_C$ up to scale σ_I^* from the covariances between $Y_{j,i,t}$, and I_i^*, provided a rank condition is satisfied. In the case of scalar θ_i, we can identify α_C for a fixed scale of I_i^* from the preceding equation for perfect foresight as

$$\frac{1}{\alpha_{j,t}\sigma_\theta^2}\left[-\operatorname{Cov}\left(Y_{j,i,t}, I_i^* \mid \boldsymbol{X}, \boldsymbol{Z}\right) + \frac{\operatorname{Var}\left(\varepsilon_{j,i,t}\right)}{\left(\sigma_I^*\right)(1+r)^t} + \frac{\alpha_{j,t}}{\sigma_I^*}\sigma_\theta^2 \frac{\sum_{t=0}^{T}\left(\alpha_{1,t} - \alpha_{0,t}\right)}{(1+r)^t}\right] = \frac{\alpha_C}{\sigma_I^*}.$$

Since we know all of the ingredients on the left hand side, we can identify α_C up to scale σ_I^*. If there is an element of $\boldsymbol{X}$ not in $\boldsymbol{Z}$, we can identify the scale σ_I^* (See equation (III-7)). Since α_C is overidentified if $T > 0$, we can test between a perfect foresight model and a complete contingent claims model by checking if the same α_C is estimated for different $\operatorname{Cov}\left(Y_{j,i,t}, I^*\right)$ terms.[21] In the complete contingent claims model with uncertainty, the middle term in the brackets would be zero for all $\varepsilon_{j,i,t}$.[22]

III-3 Perfect Foresight

CHN interpret the α_t as prices and the θ as quantities. Their interpretation assumes that the agent has perfect foresight about future prices. In this paper, we view the factor structure as an approximation of an income process which involves uncertainty in terms of prices and quantities. Thus, we do *not* assume perfect foresight in the model, and we use the factor structure as a method for decomposing realized earnings into anticipated and unanticipated components of quantities and prices.

[20]Footnote 15 defines the limit sets. See Carneiro, Hansen, and Heckman (2003) for a more complete discussion of identification in limit sets.

[21]This procedure would break down only if $\dfrac{\frac{\operatorname{Var}\left(\varepsilon_{j,i,t}\right)}{(1+r)^t}}{\alpha_{j,t}\Sigma_\Theta \sum_{t=0}^{T}\frac{\left(\alpha_{1,t}-\alpha_{0,t}\right)}{(1+r)^t}}$ is constant across all t.

[22]This testing procedure generalizes to the case of vector $\boldsymbol{\theta}$ provided that a rank condition

$$\boldsymbol{\alpha}'_{j,t}\Sigma_\Theta \frac{\sum_{t=0}^{T}\left(\boldsymbol{\alpha}_{1,t} - \boldsymbol{\alpha}_{0,t}\right)}{(1+r)^t} \neq 0$$

holds for a collection of L terms of the covariances of $Y_{j,i,t}$ with I_i^* where L is the number of factors.

III-4 A Heuristic Discussion of Identification of the Model

It is useful to provide an intuitive discussion of identification based on normal errors. Normality joined with the assumption of expected value income maximization produces closed form solutions. See Carneiro, Hansen, and Heckman (2003) for proofs of semi-parametric identification of the distributions of the factors θ and uniquenesses ε without the normality assumption. The identification strategy used in our paper is not based on normality. In this section of the appendix, we use the notation of the main paper and not the distinct, but related, notation of CHN.

III-4.1 Test Scores

First consider identification of the test score equations. Test scores are available for all agents and are determined before they make their college decisions. There is no selection bias in the test score equations. Three assumptions are crucial in securing identification through factor models. First, the explanatory variables X^M are independent of θ_1 and ε_k^M, for $k = 1, \ldots, K$. Second, the factor θ_1 is independent of ε_k^M, for $k = 1, \ldots, K$. Third, the uniqueness ε_k^M is independent from ε_l^M for any $k \neq l$, for $k, l = 1, \ldots, K$. The first assumption, along with standard rank conditions, allows β_k^M to be consistently estimated from a simple OLS regression of M_k against X^M. Given the β_k^M, we can construct differences $M_k - X^M\beta_k^M$ and compute the covariances:

$$\operatorname{Cov}\left(M_1 - X^M\beta_1^M, M_2 - X^M\beta_2^M\right) = \alpha_1^M\alpha_2^M\sigma_{\theta_1}^2, \tag{III-13}$$

$$\operatorname{Cov}\left(M_1 - X^M\beta_1^M, M_3 - X^M\beta_3^M\right) = \alpha_1^M\alpha_3^M\sigma_{\theta_1}^2, \tag{III-14}$$

$$\operatorname{Cov}\left(M_2 - X^M\beta_2^M, M_3 - X^M\beta_3^M\right) = \alpha_2^M\alpha_3^M\sigma_{\theta_1}^2. \tag{III-15}$$

The left-hand sides of (III-13), (III-14), and (III-15) can be computed from the data. The right-hand sides of (III-13), (III-14), and (III-15) are implied by the factor model. As is common in the factor literature, we need to normalize one of the factor loadings to set the scale of the factor. Let $\alpha_1^M = 1$. If we take the ratio of (III-15) to (III-13) we identify α_3^M. Analogously, the ratio of (III-15) to (III-14) allows us to recover α_2^M. Given the normalization of $\alpha_1^M = 1$ and identification of α_2^M, we identify $\sigma_{\theta_1}^2$ from (III-13). Finally, we can identify the variance of ε_k^M from the variance of $M_k - X^M\beta_k^M$. Because the factor θ_1 and uniquenesses ε_k are independently normally distributed random variables, we have identified their distribution.

III-4.2 Earnings and Choice Equations

To establish identification of the objects of interest in earnings equations requires a little more work because of the selection problem. Our assumption of normally distributed factors and uniquenesses simplifies the analysis because we can use closed-form solutions to reduce the identification problem to the identification of a few parameters.

We rely on four key assumptions to secure identification. First, all of the observable explanatory variables X and Z are independent of the unobservable factors, θ_1 and θ_2, as well as uniquenesses $\varepsilon_{s,t}$ for all s, t. Second, θ_1 is independent of θ_2. Third, both θ_1 and θ_2 are independent of ε_C and $\varepsilon_{s,t}$ for all s, t. Fourth, $\varepsilon_{s,t}$ is independent from ε_C and $\varepsilon_{s',t'}$ for any pairs s, s' and t, t' such that $s \neq s'$ or $t \neq t'$. All of the dependence among $U_{0,t}, U_{1,t}$, and U_C is captured through the factors θ_1 and θ_2. As a consequence of these assumptions,

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2_{\theta_1} & 0 \\ 0 & \sigma^2_{\theta_2} \end{bmatrix} \right).$$

Because the loadings $\alpha_{1,s,t}$, $\alpha_{2,s,t}$, $\alpha_{1,C}$, and $\alpha_{2,C}$ can be freely specified, the factors θ can affect $U_{0,t}, U_{1,t}$, and U_C differently. The joint distribution of the labor earnings $Y_{0_{,t}}, Y_{1,t}$ conditional on X is

$$\begin{bmatrix} Y_{0,t} \\ Y_{1,t} \end{bmatrix} \mid X \sim N\left(\begin{bmatrix} X\beta_{0,t} \\ X\beta_{1,t} \end{bmatrix}, \begin{bmatrix} \alpha^2_{1,0,t}\sigma^2_{\theta_1} + \alpha^2_{2,0,t}\sigma^2_{\theta_2} + \sigma^2_{\varepsilon_{0,t}} & \alpha_{1,0,t}\alpha_{1,1,t}\sigma^2_{\theta_1} + \alpha_{2,0,t}\alpha_{2,1,t}\sigma^2_{\theta_2} \\ \alpha_{1,0,t}\alpha_{1,1,t}\sigma^2_{\theta_1} + \alpha_{2,0,t}\alpha_{2,1,t}\sigma^2_{\theta_2} & \alpha^2_{1,1,t}\sigma^2_{\theta_1} + \alpha^2_{2,1,t}\sigma^2_{\theta_2} + \sigma^2_{\varepsilon_{1,t}} \end{bmatrix} \right).$$

As a result, identification of the joint distribution $F(Y_{0,t}, Y_{1,t} \mid X)$ reduces to the identification of the parameters $\beta_{s,t}, \alpha_{k,s,t}, \sigma_{\varepsilon_{s,t}}$, and $\sigma^2_{\theta_j}$ for $s = 0, 1$; $t = 1, \ldots, T$ and $j = 1, 2$, and $k = 1, 2$. From the observed data and the factor structure assumption it follows that

$$E\left(Y_{1,t} \middle| X, S = 1 \right) = X\beta_{1,t} + \alpha_{1,1,t}E\left[\theta_1 \middle| X, S = 1 \right] + \alpha_{2,1,t}E\left[\theta_2 \middle| X, S = 1 \right] + E\left[\varepsilon_{1,t} \middle| X, S = 1 \right]. \qquad \text{(III-16)}$$

The event $S = 1$ corresponds to the event $I = E\left(\sum_{t=1}^{T} \left(\frac{1}{1+\rho}\right)^{t-1} (Y_{1,t} - Y_{0,t}) - C \middle| \mathcal{I} \right) \geq 0$. Assuming that $\varepsilon_{s,t}$ does not enter agent information sets, for the case $\{\theta_1, \theta_2\} \subset \mathcal{I}$ we obtain

$$E\left(\sum_{t=1}^{T} \left(\frac{1}{1+\rho} \right)^{t-1} (Y_{1,t} - Y_{0,t}) - C \middle| \mathcal{I} \right) = \mu_I(X, Z) + \alpha_{1,I}\theta_1 + \alpha_{2,I}\theta_2 - \varepsilon_C.$$

Let η be the linear combination of three independent normal random variables: $\eta = \alpha_{1,I}\theta_1 + \alpha_{2,I}\theta_2 - \varepsilon_C$.

Then, $\eta \sim N\left(0, \sigma_\eta^2\right)$, with $\sigma_\eta^2 = \alpha_{1,I}^2\sigma_{\theta_1}^2 + \alpha_{2,I}^2\sigma_{\theta_2}^2 + \sigma_{\varepsilon_c}^2$ and

$$S = 1 \Leftrightarrow \eta > -\mu_I(X, Z). \tag{III-17}$$

If we replace (III-17) in (III-16) and use the fact that $\varepsilon_{s,t}$ is independent of X, Z, and θ,

$$E\left(Y_{1,t}\middle| X, S = 1\right) = X\beta_1 + \alpha_{1,1,t}E\left[\theta_1 | X, \eta > -\mu_I(X, Z)\right] + \alpha_{2,1,t}E\left[\theta_2 | X, \eta > -\mu_I(X, Z)\right]. \tag{III-18}$$

Because θ_1, θ_2 and η are normal random variables,

$$\theta_j = \frac{\operatorname{Cov}\left(\theta_j, \eta\right)}{\operatorname{Var}(\eta)}\eta + \rho_j \text{ for } j = 1, 2, \tag{III-19}$$

where ρ_j is a mean zero, normal random variable independent from η. Because $\operatorname{Cov}(\theta_1, \eta) = \sigma_{\theta_1}^2\alpha_{1,I}$ and $\operatorname{Cov}(\theta_2, \eta) = \sigma_{\theta_2}^2\alpha_{2,I}$ it follows that

$$E\left[\theta_1 | X, \eta > -\mu_I(X, Z)\right] = \frac{\sigma_{\theta_1}^2\alpha_{1,I}}{\sigma_\eta^2}E\left[\eta\middle|\eta > -\mu_I(X, Z)\right]$$

and

$$E\left[\theta_2 | X, \eta > -\mu_I(X, Z)\right] = \frac{\sigma_{\theta_2}^2\alpha_{2,I}}{\sigma_\eta^2}E\left[\eta\middle|\eta > -\mu_I(X, Z)\right].$$

For any standard normal random variable μ, $E\left(\mu\middle|\mu \geq -c\right) = \frac{\phi(c)}{\Phi(c)}$ where $\phi(.)$ and $\Phi(.)$ are the density and distribution function of a standard normal random variable. Define, for $j = 0, 1$, $\pi_{j,t} = \left(\frac{\alpha_{1,j,t}\alpha_{1,I}\sigma_{\theta_1}^2 + \alpha_{2,j,t}\alpha_{2,I}\sigma_{\theta_2}^2}{\sigma_\eta}\right)$. These facts together allow us to rewrite $(III-16)$ as

$$E\left(Y_{1,t}\middle|\eta \leq -\mu_I(X, Z)\right) = X\beta_{1,t} + \pi_{1,t}\frac{\phi\left(\frac{\mu_I(X,Z)}{\sigma_\eta}\right)}{\Phi\left(\frac{\mu_I(X,Z)}{\sigma_\eta}\right)}. \tag{III-20}$$

We can derive a similar expression for mean observed earnings in sector "0":

$$E\left(Y_{0,t}\middle|\eta > -\mu_I(X, Z)\right) = X\beta_{0,t} - \pi_{0,t}\frac{\phi\left(\frac{\mu_I(X,Z)}{\sigma_\eta}\right)}{\Phi\left(\frac{\mu_I(X,Z)}{\sigma_\eta}\right)}. \tag{III-21}$$

We can apply the two-step procedure developed in Heckman (1976) to identify $\beta_{0,t}, \beta_{1,t}, \pi_{0,t}$ and $\pi_{1,t}$. Given identification of $\beta_{s,t}$ for all s and t, we can construct the differences $Y_{s,t} - X\beta_{s,t}$ and compute the covariances

$$\operatorname{Cov}\left(M_1 - X^M\beta_1^M, Y_{0,t} - X\beta_{0,t}\right) = \alpha_{1,0,t}\sigma_{\theta_1}^2 \tag{III-22}$$

and

$$\operatorname{Cov}\left(M_1 - X^M\beta_1^M, Y_{1,t} - X\beta_{1,t}\right) = \alpha_{1,1,t}\sigma_{\theta_1}^2. \tag{III-23}$$

The left-hand sides of (III-22) and (III-23) are identified from the data. The right-hand sides are an implication of the factor model. We determined $\sigma_{\theta_1}^2$ from the analysis of the test scores. From equations (III-22) and (III-23), we can recover $\alpha_{1,0,t}$ and $\alpha_{1,1,t}$ for all t. Note that we can also identify $\alpha_{1,C}/\sigma_\eta$ by computing the covariance

$$\operatorname{Cov}\left(M_1 - X\beta_1^M, \frac{I - \mu_I(X,Z)}{\sigma_\eta}\right) = \frac{\sum_{t=1}^T \left(\frac{1}{1+\rho}\right)^{t-1}(\alpha_{1,1,t} - \alpha_{1,0,t}) - \alpha_{1,C}}{\sigma_\eta}\sigma_{\theta_1}^2. \tag{III-24}$$

Using (III-22) and (III-23), we can identify $\alpha_{1,1,t}$ and $\alpha_{1,0,t}$ for all t. The only remaining term to be identified is the ratio $\alpha_{1,C}/\sigma_\eta$, which can be identified from covariance equation (III-24).

Note that if $T \geq 2$, we can also identify the parameters related to factor θ_2, such as $\alpha_{2,s,t}$ and $\sigma_{\theta_2}^2$. To see this, first normalize $\alpha_{2,0,1} = 1$ and compute the covariances:

$$\operatorname{Cov}\left(Y_{0,1} - X\beta_{0,1}, Y_{0,2} - X\beta_{0,2}\right) - \alpha_{1,0,1}\alpha_{1,0,2}\sigma_{\theta_1}^2 = \alpha_{2,0,2}\sigma_{\theta_2}^2, \tag{III-25}$$

$$\operatorname{Cov}\left(Y_{0,1} - X\beta_{0,1}, \frac{I - \mu_I(X,Z)}{\sigma_\eta}\right) - \frac{\alpha_{1,0,1}\sigma_{\theta_1}^2 \sum\limits_{t=1}^T (\alpha_{1,1,t} - \alpha_{1,0,t} - \alpha_{1,C})}{\sigma_\eta} = \frac{\sigma_{\theta_2}^2 \sum\limits_{t=1}^T (\alpha_{2,1,t} - \alpha_{2,0,t} - \alpha_{2,C})}{\sigma_\eta},$$

$$\operatorname{Cov}\left(Y_{0,2} - X\beta_{0,2}, \frac{I - \mu_I(X,Z)}{\sigma_\eta}\right) - \frac{\alpha_{1,0,2}\sigma_{\theta_1}^2 \sum\limits_{t=1}^T (\alpha_{1,1,t} - \alpha_{1,0,t} - \alpha_{1,C})}{\sigma_\eta} = \frac{\alpha_{2,0,2}\sigma_{\theta_2}^2 \sum\limits_{t=1}^T (\alpha_{2,1,t} - \alpha_{2,0,t} - \alpha_{2,C})}{\sigma_\eta}.$$

The left-hand sides of (III-25), (III-26), and (III-26) are identified from the data. Computing the ratio of (III-26) to (III-26), we can recover $\alpha_{2,0,2}$. From (III-25) we can recover $\sigma_{\theta_2}^2$. We now add in the information on the covariances from the college earnings equation:

$$\mathrm{Cov}\left(Y_{1,1} - X\beta_{1,1}, Y_{1,2} - X\beta_{1,2}\right) - \alpha_{1,1,1}\alpha_{1,1,2}\sigma^2_{\theta_1} = \alpha_{2,1,1}\alpha_{2,1,2}\sigma^2_{\theta_2}, \tag{III-26}$$

$$\mathrm{Cov}\left(Y_{1,1} - X\beta_{1,1}, \frac{I-\mu_I(X,Z)}{\sigma_\eta}\right) - \frac{\alpha_{1,1}\sigma^2_{\theta_1}\sum_{t=1}^{T}\left(\alpha_{1,1,t} - \alpha_{1,0,t} - \alpha_{1,C}\right)}{\sigma_\eta} = \frac{\alpha_{2,1,1}\sigma^2_{\theta_2}\sum_{t=1}^{T}\left(\alpha_{2,1,t} - \alpha_{2,0,t} - \alpha_{2,C}\right)}{\sigma_\eta},$$

$$\mathrm{Cov}\left(Y_{1,2} - X\beta_{1,2}, \frac{I-\mu_I(X,Z)}{\sigma_\eta}\right) - \frac{\alpha_{1,1,2}\sigma^2_{\theta_1}\sum_{t=1}^{T}\left(\alpha_{1,1,t} - \alpha_{1,0,t} - \alpha_{1,C}\right)}{\sigma_\eta} = \frac{\alpha_{2,1,2}\sigma^2_{\theta_2}\sum_{t=1}^{T}\left(\alpha_{2,1,t} - \alpha_{2,0,t} - \alpha_{2,C}\right)}{\sigma_\eta}.$$

Computing the ratios of (III-27) to (III-26) and (III-27) to (III-26), we obtain $\alpha_{2,1,2}$ and $\alpha_{2,1,1}$ respectively. Finally, we use the information available from the data on earnings by schooling choice, $\mathrm{Var}\left(Y_{0,t}\middle| X, S=0\right)$ and $\mathrm{Var}\left(Y_{1,t}\middle| X, S=1\right)$, to compute $\sigma^2_{\varepsilon_{0,t}}$ and $\sigma^2_{\varepsilon_{1,t}}$, respectively. Note that we have identified all of the elements that characterize the joint distribution as specified in (III-16).

The identification analysis in this section uses the data on test scores in an essential way. However, with sufficiently long panel earnings data, it is possible to identify the model without test score data. See the analysis in Abbring and Heckman (2007).

Web Supplement IV

Determining *Ex Ante* and *Ex Post* Joint Distribution

Let $E(Y_s | \mathcal{I})$ denote the *ex ante* present value of truncated lifetime earnings at schooling level s. We seek to determine the means and the covariances between *ex ante* college and high-school earnings conditional on information set $\mathcal{I}$ that we estimate. Since θ_1 and θ_2 are known by the agent at the time schooling choices are made, the *ex ante* mean truncated present value of earnings is

$$E(Y_s | \mathcal{I}) = \sum_{t=1}^{T^*} \frac{X\beta_{s,t} + \theta_1 \alpha_{1,s,t} + \theta_2 \alpha_{2,s,t}}{(1+\rho)^{t-1}}, \quad s = 0, 1,$$

where $T^* = 36$ is the latest age at which we observe earnings and the first age we analyze (age 22) is denoted $t = 1$. Conditional on covariates X, the covariance between $E(Y_1 | \mathcal{I})$ and $E(Y_0 | \mathcal{I})$ is

$$\mathrm{Cov}\left(E\left(Y_1 \mid \mathcal{I}\right), E\left(Y_0 \mid \mathcal{I}\right)\right) = \sum_{j=1}^{2} \mathrm{Var}\left(\theta_j\right) \left(\sum_{t=1}^{T^*} \frac{\alpha_{j,1,t}}{(1+\rho)^{t-1}} \right) \left(\sum_{t=1}^{T^*} \frac{\alpha_{j,0,t}}{(1+\rho)^{t-1}} \right).$$

Tables IV-1-A and IV-1-B present the conditional distributions of the truncated (at T^*) present values of *ex ante* college earnings given *ex ante* high school earnings by decile for the NLSY/1979 and NLS/1966 samples, respectively. If the dependence across outcomes were perfect and positive, the diagonal elements would be '1' and the off-diagonal elements would be '0.' The estimated distributions exhibit strong negative dependence between the relative positions of individuals in the two distributions. For example, for the NLSY/1979 sample, 59.76% of the individuals who are in the first decile of the truncated high school present value of earnings distributions would be in the 10th decile of the college present value of earnings distributions. For the NLS/1966 sample, this figure is 74.22%. Comparing Tables IV-1-A and IV-1-B, the correlation between *ex ante* high school and *ex ante* college truncated present value of lifetime earnings strengthens slightly for the more recent cohort. For both cohorts, the evidence supports the conjecture of Willis and Rosen (1979) that comparative advantage is prevalent in the U.S. labor market. Persons who are good college educated workers make poor high school educated workers. CHN report similar findings.

We can also compute the covariance between the truncated present value of *ex post* college and high-school earnings conditional on X. Recall that agents learn θ_3 after schooling is completed. For both samples, the *ex post* covariance is

$$\mathrm{Cov}\left(Y_1, Y_0 | X\right) = \sum_{j=1}^{3} \mathrm{Var}\left(\theta_j\right) \left(\sum_{t=1}^{T^*} \frac{\alpha_{j,1,t}}{(1+\rho)^{t-1}} \right) \left(\sum_{t=1}^{T^*} \frac{\alpha_{j,0,t}}{(1+\rho)^{t-1}} \right).$$

Tables IV-2-A and IV-2-B display the conditional distributions of the present values of truncated *ex post* college earnings given truncated *ex post* high school earnings for the NLSY/1979 and NLS/1966 samples, respectively. The *ex post* correlations are roughly the same as those found for the *ex ante* distributions. The extreme negative dependence found in the *ex ante* distributions weakens slightly in the *ex post* distributions.

From knowledge of the joint distribution, we can compute the percentage of individuals who regret their schooling choice after all the information up to age 36 is in about both schooling choices.[23] These are reported by schooling level in Table IV-3. A higher fraction of the individuals who stop at high-school regret not graduating from college (16.6% in NLSY/1979 and 15.22% in NLS/1966). Around 14.5% of individuals who attend college regret not stopping their schooling upon high-school graduation, for the NLSY/1979 and NLS/1966.

[23]This calculation assumes that agents observe *both* counterfactual states *ex post* and may overstate the amount of information individuals would have *ex post*.

Web Data Appendix Table IV-1-A

Ex-Ante Conditional Distributions for the NLSY/79 (College Earnings Conditional on High School Earnings) $\Pr(d_i < Y_c < d_i + 1 \mid d_j < Y_h < d_j + 1)$ where d_i is the ith decile of the College Lifetime Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution Individuals fix unknown θ at their means, so $\theta_3 = 0$ Correlation $(Y_C, Y_H) = -0.8730$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0002	0.0000	0.0014	0.0053	0.0175	0.0323	0.0645	0.0987	0.1825	0.5976
2	0.0000	0.0020	0.0106	0.0248	0.0524	0.0808	0.0968	0.1598	0.2978	0.2750
3	0.0008	0.0072	0.0276	0.0488	0.0782	0.0976	0.1368	0.2304	0.2774	0.0952
4	0.0006	0.0234	0.0608	0.0814	0.0932	0.1350	0.1932	0.2230	0.1678	0.0216
5	0.0052	0.0430	0.0872	0.0900	0.1362	0.1880	0.2146	0.1744	0.0574	0.0040
6	0.0144	0.0662	0.1038	0.1400	0.1798	0.2166	0.1778	0.0856	0.0158	0.0000
7	0.0318	0.1108	0.1443	0.1895	0.2273	0.1693	0.0960	0.0276	0.0034	0.0000
8	0.0792	0.1609	0.2219	0.2711	0.1757	0.0698	0.0188	0.0020	0.0006	0.0000
9	0.1765	0.3325	0.2931	0.1437	0.0400	0.0112	0.0026	0.0002	0.0002	0.0000
10	0.6064	0.3216	0.0637	0.0078	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000

Web Data Appendix Table IV-1-B

Ex-Ante Conditional Distributions for the NLSY/66 (College Earnings Conditional on High School Earnings) $\Pr(d_i < Y_c < d_i + 1 \mid d_j < Y_h < d_j + 1)$ where d_i is the ith decile of the College Lifetime Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution Individuals fix unknown θ at their means, so $\theta_3 = 0$ Correlation $(Y_C, Y_H) = -0.9332$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0012	0.0020	0.0004	0.0004	0.0010	0.0010	0.0056	0.0396	0.2065	0.7422
2	0.0014	0.0000	0.0000	0.0010	0.0034	0.0210	0.0966	0.2664	0.4118	0.1984
3	0.0012	0.0002	0.0004	0.0048	0.0342	0.1148	0.2464	0.3142	0.2322	0.0516
4	0.0004	0.0004	0.0048	0.0326	0.1302	0.2442	0.2726	0.1942	0.1146	0.0060
5	0.0002	0.0014	0.0318	0.1218	0.2354	0.2658	0.1876	0.1256	0.0302	0.0002
6	0.0002	0.0130	0.1034	0.2494	0.2618	0.1862	0.1288	0.0514	0.0058	0.0000
7	0.0020	0.0774	0.2590	0.2864	0.1898	0.1216	0.0550	0.0088	0.0000	0.0000
8	0.0236	0.2616	0.3410	0.2042	0.1186	0.0436	0.0072	0.0000	0.0002	0.0000
9	0.1992	0.4510	0.2260	0.0966	0.0252	0.0018	0.0002	0.0000	0.0000	0.0000
10	0.7669	0.1961	0.0337	0.0028	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000

Web Data Appendix Table IV-2-A

Ex-Post Conditional Distributions for the NLSY/79 (College Earnings Conditional on High School Earnings) $\Pr(d_i < Y_c < d_i + 1 \mid d_j < Y_h < d_j + 1)$ where d_i is the ith decile of the College Lifetime Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution

Full Information Set Correlation$(Y_C, Y_H) = -0.8752$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0000	0.0000	0.0004	0.0029	0.0123	0.0263	0.0466	0.0936	0.2046	0.6132
2	0.0000	0.0010	0.0072	0.0220	0.0470	0.0762	0.1110	0.1818	0.2968	0.2570
3	0.0010	0.0074	0.0270	0.0454	0.0788	0.1172	0.1722	0.2250	0.2464	0.0796
4	0.0012	0.0160	0.0440	0.0784	0.1084	0.1528	0.2076	0.2184	0.1476	0.0256
5	0.0032	0.0368	0.0754	0.1108	0.1536	0.1816	0.1950	0.1596	0.0766	0.0074
6	0.0108	0.0668	0.1108	0.1644	0.1892	0.1950	0.1506	0.0850	0.0264	0.0010
7	0.0264	0.1129	0.1649	0.2084	0.2028	0.1547	0.0895	0.0326	0.0078	0.0000
8	0.0630	0.1994	0.2487	0.2195	0.1559	0.0794	0.0255	0.0068	0.0018	0.0000
9	0.1780	0.3341	0.2716	0.1409	0.0517	0.0177	0.0043	0.0012	0.0004	0.0000
10	0.5622	0.3368	0.0806	0.0157	0.0035	0.0012	0.0000	0.0000	0.0000	0.0000

Web Data Appendix Table IV-2-B

Ex-Post Conditional Distributions for the NLS/66 (College Earnings Conditional on High School Earnings)
$\Pr(d_i < Y_c < d_i + 1 \mid d_j < Y_h < d_j + 1)$ where d_i is the ith decile of the College Lifetime Ex-Ante Earnings Distribution and d_j is the jth decile of the High School Ex-Ante Lifetime Earnings Distribution
Full Information Set Correlation$(Y_C, Y_H) = -0.8869$

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0002	0.0016	0.0012	0.0006	0.0016	0.0067	0.0224	0.0668	0.2331	0.6657
2	0.0006	0.0002	0.0016	0.0064	0.0216	0.0526	0.1280	0.2388	0.3354	0.2148
3	0.0002	0.0012	0.0060	0.0250	0.0692	0.1360	0.2112	0.2526	0.2218	0.0768
4	0.0004	0.0046	0.0262	0.0694	0.1302	0.1938	0.2234	0.2024	0.1240	0.0256
5	0.0016	0.0156	0.0636	0.1326	0.1906	0.2166	0.1864	0.1262	0.0580	0.0086
6	0.0032	0.0452	0.1294	0.2028	0.2090	0.1832	0.1270	0.0734	0.0242	0.0026
7	0.0188	0.1112	0.2180	0.2306	0.1894	0.1260	0.0684	0.0310	0.0060	0.0006
8	0.0620	0.2348	0.2666	0.1966	0.1368	0.0650	0.0284	0.0082	0.0014	0.0002
9	0.2220	0.3639	0.2204	0.1190	0.0488	0.0192	0.0048	0.0016	0.0002	0.0000
10	0.6762	0.2320	0.0699	0.0178	0.0027	0.0010	0.0004	0.0000	0.0000	0.0000

Web Data Appendix Table IV-3
Percentage that Regret Schooling Choices

Schooling Group	NLS/1966	NLSY/1979
Percentage of High School Graduates who Regret Not Graduating from College	0.1522	0.1659
Percentage of College Graduates who Regret Graduating from College	0.1402	0.1495

Web Supplement V

Accounting for Schooling Choice

This appendix shows that the inequality decompositions reported in Section 3.3.4 in the text are barely affected by allowing for re-optimization of the schooling decision when uncertainty is shut down.

Web Data Appendix Table V-1
Gini Decomposition

	NLS/66	NLSY/79	%Growth
Factual Economy: Heterogeneity and Uncertainty[1]	0.1803	0.2088	15.85%
Counterfactual: Fixing Schooling Choices as in Factual Economy			
Heterogeneity Only[2]	0.1591	0.1825	14.73%
Counterfactual: Allowing Agents to Change Schooling Choices			
Heterogeneity Only[3]	0.1590	0.1825	14.80%

[1]Let $Y_{k,s,t,i}$ denote the earnings of an agent i, $i = 1, ..., n$, at age t, $t = 22, ..., 36$, in schooling level s, $s =$ high school, college, and cohort k, $k = NLS/1966, NLSY/1979$. We model earnings $Y_{k,s,t,i}$ as:

$$Y_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i} + \theta_{3,k,i}\alpha_{3,k,s,t,i} + \varepsilon_{k,s,t,i}. \tag{V-1}$$

The present value of earnings at schooling level s, $Y_{k,s,i}$, is $Y_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y_{k,s,t,i}}{(1+\rho)^{t-1}}$. The observed present value of earnings satisfies $Y_{k,i} = S_{k,i}Y_{k,1,i} + (1 - S_{k,i})Y_{k,0,i}$ where $S_{k,i} = 1$ if agent i in cohort k graduates college, and $S_{k,i} = 0$ otherwise. Let $C_{k,i}$ denote the direct costs for individual i in cohort k. The schooling choice is:

$$S_{k,i} = 1 \Leftrightarrow E\left(Y_{k,1,i} - Y_{k,0,i} - C_{k,i}\middle| \mathcal{I}_k\right) \geq 0 \tag{V-2}$$

This is the factual economy. In this row, we show the Gini coefficient for the observed present value of earnings $Y_{k,i}$.

[2]We simulate the economy by replacing (V-1) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are as determined in (V-2). In this row, we show the Gini coefficient for the observed present value of earnings $Y^h_{k,s,i}$ when we constrain schooling choices are $S_{k,i}$.

[3]We simulate the economy by replacing (V-1) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are then deterministic:

$$S^h_{k,i} = 1 \Leftrightarrow Y^h_{k,1,i} - Y^h_{k,0,i} - C_{k,i} \geq 0.$$

In this row, we show the Gini coefficient for the observed present value of earnings $Y^h_{k,s,i}$ when schooling choices are $S^h_{k,i}$.

Web Data Appendix Table V-2
The Theil Entropy Index T
Overall

	NLS/66	NLSY/79	% Change
Factual Economy: Heterogeneity and Uncertainty	0.0502	0.0693	37.98%
Counterfactual: Fixing Schooling Choices as in Factual Economy[1]			
Heterogeneity Only	0.0390	0.0522	33.76%
Counterfactual: Allowing Agents to Change Schooling Choices[2]			
Heterogeneity Only	0.0390	0.0521	33.61%

Within Schooling Groups

	NLS/66	NLSY/79	% Change
Factual Economy: Heterogeneity and Uncertainty	0.0491	0.0631	28.53%
Counterfactual: Fixing Schooling Choices as in Factual Economy[1]			
Heterogeneity Only	0.0378	0.0465	22.85%
Counterfactual: Allowing Agents to Change Schooling Choices[2]			
Heterogeneity Only	0.0379	0.0464	22.69%

Between Schooling Groups

	NLS/66	NLSY/79	% Change
Factual Economy: Heterogeneity and Uncertainty	0.0011	0.0062	447.37%
Counterfactual: Fixing Schooling Choices as in Factual Economy[1]			
Heterogeneity Only	0.0011	0.0057	394.22%
Counterfactual: Allowing Agents to Change Schooling Choices[2]			
Heterogeneity Only	0.0012	0.0057	392.18%

[1]Let $Y_{k,s,t,i}$ denote the earnings of an agent i, $i = 1, ..., n$, at age t, $t = 1, ..., T^*$, in schooling level s, s = high school, college, and cohort k, $k = NLS/1966, NLSY/1979$. We model earnings $Y_{k,s,t,i}$ as:

$$Y_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i} + \theta_{3,k,i}\alpha_{3,k,s,t,i} + \varepsilon_{k,s,t,i}. \quad \text{(V-3)}$$

The present value of earnings in schooling level s, $Y_{k,s,i}$, is $Y_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y_{k,s,t,i}}{(1+\rho)^{t-1}}$. The observed present value of earnings satisfies $Y_{k,i} = S_{k,i}Y_{k,1,i} + (1 - S_{k,i})Y_{k,0,i}$. Let $C_{k,i}$ denote the direct costs for individual i in cohort k. The schooling choice is:

$$S_{k,i} = 1 \Leftrightarrow E\left(Y_{k,1,i} - Y_{k,0,i} - C_{k,i}\middle| \mathcal{I}_k\right) \geq 0 \quad \text{(V-4)}$$

This is the factual economy. In this row, we show the Theil Entropy Index T for the observed present value of earnings $Y_{k,i}$.

[2]We simulate the economy by replacing (V-3) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are as determined in (V-4). In this row, we show the Theil Entropy Index T for the observed present value of earnings $Y^h_{k,s,i}$ when we constrain schooling choices are $S_{k,i}$.

[3]We simulate the economy by replacing (V-3) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are then deterministic:

$$S^h_{k,i} = 1 \Leftrightarrow Y^h_{k,1,i} - Y^h_{k,0,i} - C_{k,i} \geq 0.$$

In this row, we show the Theil Entropy Index T for the observed present value of earnings $Y^h_{k,s,i}$ when schooling choices are $S^h_{k,i}$.

Web Data Appendix Table V-3
Atkinson Index

	ε = 0.50			ε = 1.0		
	NLS/66	NLSY/79	%Change	NLS/66	NLSY/79	%Change
Factual Economy: Heterogeneity and Uncertainty[1]	0.0276	0.0389	0.4111	0.0586	0.0847	0.4446
Counterfactual: Fixing Schooling Choices as in Factual Economy						
Heterogeneity Only[2]	0.0213	0.0286	0.3437	0.0447	0.0604	0.3503
Counterfactual: Allowing Agents to Change Schooling Choices						
Heterogeneity Only[3]	0.0213	0.0286	0.3418	0.0447	0.0603	0.3485
	ε = 1.5			ε = 2.0		
	NLS/66	NLSY/79	%Change	NLS/66	NLSY/79	%Change
Factual Economy: Heterogeneity and Uncertainty[1]	0.0968	0.1467	0.5147	0.1627	0.2627	0.6149
Counterfactual: Fixing Schooling Choices as in Factual Economy						
Heterogeneity Only[2]	0.0716	0.0980	0.3687	0.1060	0.1506	0.4205
Counterfactual: Allowing Agents to Change Schooling Choices						
Heterogeneity Only[3]	0.0716	0.0979	0.3669	0.1059	0.1503	0.4185

[1]Let $Y_{k,s,t,i}$ denote the earnings of an agent i, $i = 1, ..., n$, at age t, $t = 1, ..., T$, in schooling level s, s = high school, college, and cohort k, $k = NLS/1966, NLSY/1979$. We model earnings $Y_{k,s,t,i}$ as:

$$Y_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i} + \theta_{3,k,i}\alpha_{3,k,s,t,i} + \varepsilon_{k,s,t,i}. \tag{V-5}$$

The present value of earnings in schooling level s, $Y_{k,s,i}$, is $Y_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y_{k,s,t,i}}{(1+\rho)^{t-1}}$. The observed present value of earnings satisfies $Y_{k,i} = S_{k,i}Y_{k,1,i} + (1 - S_{k,i})Y_{k,0,i}$. Let $C_{k,i}$ denote the psychic costs for individual i in cohort k. The schooling choice is:

$$S_{k,i} = 1 \Leftrightarrow E\left(Y_{k,1,i} - Y_{k,0,i} - C_{k,i}\middle| \mathcal{I}_k\right) \geq 0 \tag{V-6}$$

This is the factual economy. We then compute the average present value of earnings across all individuals in cohort k, $\mu_k = \frac{1}{n}\sum_{i=1}^{n} Y_{k,i}$. For a given inequality aversion parameter ϵ, we compute the level of permanent income $\bar{Y}_k(\epsilon)$ that generates the same welfare as the social welfare of the actual distribution in cohort k:

$$\frac{[\bar{Y}_k(\epsilon)]^{1-\epsilon} - 1}{1-\epsilon} = \frac{1}{n_k}\sum_{i=1}^{n_k} \frac{(Y_{k,i})^{1-\epsilon} - 1}{1-\epsilon}.$$

For each value of ϵ, the Atkinson Index is $A(\epsilon) = 1 - \frac{\bar{Y}_k(\epsilon)}{\mu_k}$. In this row, we show the Atkinson Index for the observed present value of earnings $Y_{k,i}$ for different values of ϵ.

[2]We simulate the economy by replacing (V-5) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are as determined in (V-6). In this row, we show the Atkinson Index for the observed present value of earnings $Y^h_{k,i}$ for different values of ϵ when we constrain schooling choices are $S_{k,i}$.

[3]We simulate the economy by replacing (V-5) with:

$$Y^h_{k,s,t,i} = \mu_{s,k}(X_k) + \theta_{1,k,i}\alpha_{1,k,s,t,i} + \theta_{2,k,i}\alpha_{2,k,s,t,i},$$

where $Y^h_{k,s,t,i}$ are the individual earnings when idiosyncratic uncertainty is completely shut down. The present value of earnings when only heterogeneity is accounted for is constructed in a similar manner: $Y^h_{k,s,i} = \sum_{t=1}^{T^*} \frac{Y^h_{k,s,t,i}}{(1+\rho)^{t-1}}$. The schooling choices are then:

$$S^h_{k,i} = 1 \Leftrightarrow E\left(Y^h_{k,1,i} - Y^h_{k,0,i} - C_{k,i} \middle| \mathcal{I}_k\right) \geq 0.$$

In this row, we show the Atkinson Index for the observed present value of earnings $Y^h_{k,i}$ for different values of ϵ when schooling choices are $S^h_{k,i}$.

Bibliography

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume 6*, Volume 6B. Amsterdam: Elsevier. Forthcoming.

Blundell, R., L. Pistaferri, and I. Preston (2004, October). Consumption inequality and partial insurance. Technical Report WP04/28, Institute for Fiscal Studies.

Blundell, R. and I. Preston (1998, May). Consumption inequality and income uncertainty. *Quarterly Journal of Economics 113*(2), 603–640.

Cameron, S. V. and J. J. Heckman (2001, June). The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy 109*(3), 455–99.

Card, D. (2001, September). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica 69*(5), 1127–1160.

Carneiro, P., K. Hansen, and J. J. Heckman (2001, Fall). Removing the veil of ignorance in assessing the distributional impacts of social policies. *Swedish Economic Policy Review 8*(2), 273–301.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2005). Understanding what instrumental variables estimate: Estimating marginal and average returns to education. Presented at Harvard University, 2001. Under review.

Cawley, J., K. Conneely, J. J. Heckman, and E. J. Vytlacil (1997). Cognitive ability, wages and meritocracy. In

B. Devlin, S. Fienberg, D. Resnick, and K. Roeder (Eds.), *Intelligence, Genes, and Success : Scientists Respond to The Bell Curve*, pp. 179–192. New York: Springer.

Cunha, F. and J. J. Heckman (2007). Identifying and estimating the distributions of *Ex Post* and *Ex Ante* returns to schooling: A survey of recent developments. *Labour Economics*. Forthcoming.

Cunha, F., J. J. Heckman, and S. Navarro (2004, March). Separating heterogeneity from uncertainty in an aiyagari-laitner economy. Presented at the Goldwater Conference on Labor Markets, Arizona.

Cunha, F., J. J. Heckman, and S. Navarro (2005, April). Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks Lecture. *Oxford Economic Papers 57*(2), 191–261.

Flavin, M. A. (1981, October). The adjustment of consumption to changing expectations about future income. *Journal of Political Economy 89*(5), 974–1009.

Hartog, J. and W. Vijverberg (2002, February). Do wages really compensate for risk aversion and skewness affection? Technical Report IZA DP No. 426, IZA, Bonn, Germany.

Hause, J. C. (1980, May). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica 48*(4), 1013–1029.

Heckman, J. J. (1976, December). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. (1990, May). Varieties of selection bias. *American Economic Review 80*(2), 313–318.

Heckman, J. J. (2001, August). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy 109*(4), 673–748.

Heckman, J. J. and B. E. Honoré (1990, September). The empirical content of the Roy model. *Econometrica 58*(5), 1121–1149.

Heckman, J. J., R. L. Matzkin, S. Navarro, and S. Urzua (2004). Nonseparable factor analysis. Unpublished manuscript, Department of Economics, University of Chicago.

Heckman, J. J. and S. Navarro (2004, February). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics 86*(1), 30–57.

Heckman, J. J. and R. Robb (1986). Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples*, pp. 63–107. New York: Springer-Verlag. Reprinted in 2000, Mahwah, NJ: Lawrence Erlbaum Associates.

Heckman, J. J. and R. Robb (2000). Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In H. Wainer (Ed.), *Drawing Inferences from Self-Selected Samples*, pp. 63–107. Mahwah, N.J.: Lawrence Erlbaum Associates.

Heckman, J. J. and J. Scheinkman (1987, April). The importance of bundling in a Gorman-Lancaster model of earnings. *Review of Economic Studies 54*(2), 243–355.

Heckman, J. J. and J. A. Smith (1998). Evaluating the welfare state. In S. Strom (Ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pp. 241–318. New York: Cambridge University Press.

Heckman, J. J., J. A. Smith, and N. Clements (1997, October). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies 64*(221), 487–536.

Heckman, J. J. and E. J. Vytlacil (1998, Fall). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources 33*(4), 974–987.

Jencks, C., M. Smith, H. Acland, M. J. Bane, D. K. Cohen, H. Gintis, B. Heyns, and S. Michelson (1972). *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.

Jöreskog, K. G. (1977). Structural equations models in the social sciences: Specification, estimation and testing. In P. Krishnaiah (Ed.), *Applications of Statistics*, pp. 265–287. New York: North-Holland.

Jöreskog, K. G. and A. S. Goldberger (1975, September). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association 70*(351), 631–639.

Kotlarski, I. I. (1967). On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics 20*, 69–76.

Lillard, L. A. and R. J. Willis (1978, September). Dynamic aspects of earning mobility. *Econometrica 46*(5), 985–1012.

MaCurdy, T. E. (1982, January). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*(1), 83–114.

Miller, S. (2004). *The National Longitudinal Surveys NLSY79 User's Guide 1979-2002*. Washington, DC: Bureau of Labor Statistics, U.S. Department of Labor.

Navarro, S. (2004). Semiparametric identification of factor models for counterfactual analysis. Unpublished manuscript, University of Chicago, Department of Economics.

Pistaferri, L. (2001, August). Superior information, income shocks, and the permanent income hypothesis. *Review of Economics and Statistics 83*(3), 465–476.

Roy, A. (1951, June). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Sims, C. A. (1972, September). Money, income, and causality. *American Economic Review 62*(4), 540–552.

Willis, R. J. and S. Rosen (1979, October). Education and self-selection. *Journal of Political Economy 87*(5, Part 2), S7–S36.