Adding Uncertainty to a Roy Economy with Two Sectors

James J. Heckman
The University of Chicago

Nuffield College, Oxford University
This draft, August 7, 2005

S denotes different sectors.

$S = 0$ denotes choice of the high school sector, and $S = 1$ denotes choice of the college sector.

C reflects the cost associated with choosing the college sector.

$$\begin{aligned} Y_1 &= \sum_{t=0}^{T} \frac{Y_{1it}}{(1+r)^t} \\ Y_0 &= \sum_{t=0}^{T} \frac{Y_{0it}}{(1+r)^t}, \end{aligned}$$

Y_1, Y_0 and C are *ex post* realizations of cost and returns.

$\mathcal{I}_0$ denotes the information set of the agent at time period $t = 0$.

$$S = \begin{cases} 1, & \text{if } E\left(Y_1 - Y_0 - C \mid \mathcal{I}_0\right) \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Essential Idea

Suppose, contrary to what is possible, analyst observes Y_0, Y_1 and C.

Ideal data set

Observe two different lifetimes

Construct $Y_1 - Y_0 - C$ from *ex post* lifetime data.

Information set $\mathcal{I}_0$ of the agent. We seek to construct $E(Y_1 - Y_0 - C \mid \mathcal{I}_0)$. Suppose we assume we know the right information set $\left(\widetilde{\mathcal{I}}_0 = \mathcal{I}_0\right)$.

We then obtain

$$V_{\widetilde{\mathcal{I}}_0} = (Y_1 - Y_0 - C) - E\left(Y_1 - Y_0 - C \mid \widetilde{\mathcal{I}}_0\right)$$

Our test is to determine if S depends on $V_{\widetilde{\mathcal{I}}_0}$.

Test for correct specification of $\mathcal{I}_0$: test if the coefficient on $V_{\widetilde{\mathcal{I}}_0}$ in a discrete choice equation for S is different from zero.

Search among candidate information sets $\widetilde{\mathcal{I}}_0$ to determine which ones satisfy the requirement that the generated $V_{\widetilde{\mathcal{I}}_0}$ does not predict S.

Procedure is in the form of a Sims (1972) version of a Wiener-Granger causality test.

It is also a test for misspecification of the information.

Components of $V_{\widetilde{\mathcal{I}}_0}$ that do not predict S are called intrinsic components of uncertainty.

Procedure as stated not practical.

We do not observe Y_1 and Y_0 together for anyone.

Specifics of the **BASIC STRATEGY** in complete market case.

Linear in parameters model:

$$
\begin{aligned}
Y_{0it} &= X_{it}\beta_{0t} + v_{0it} \qquad t = 0, ..., T \\
Y_{1it} &= X_{it}\beta_{1t} + v_{1it} \\
C_i &= Z_i\gamma + v_{iC}.
\end{aligned}
$$

$\theta = (\theta_1, \theta_2, \ldots, \theta_L)$, θ_i and θ_j mutually independent random variables for $i \neq j$.

$$
\begin{aligned}
v_{0it} &= \theta_i\alpha_{0t} + \varepsilon_{0it} \\
v_{1it} &= \theta_i\alpha_{1t} + \varepsilon_{1it},
\end{aligned}
$$

where α_{0t} and α_{1t} are vectors.

$$C_i = Z_i\gamma + \theta_i\alpha_C + \varepsilon_{iC}.$$

Choice Equation:

$$
\begin{aligned}
I &= E\left(\sum_{t=0}^{T}\frac{(X_{it}\beta_{1t}+\theta_i\alpha_{1t}+\varepsilon_{1it})-(X_{it}\beta_{0t}+\theta_i\alpha_{0t}+\varepsilon_{0it})}{(1+r)^t}-(Z_i\gamma+\theta_i\alpha_C+\varepsilon_{iC})\;\middle|\;\mathcal{I}_0\right)\\
S &= 1 \text{ if } I\geq 0;\; S=0 \text{ otherwise.}
\end{aligned}
$$

Let $\odot$ denote the Hadamard product $(a\odot b=(a_1b_1,\ldots,a_Lb_L))$

For candidate information set $\widetilde{\mathcal{I}}_0$,

$$
\begin{aligned}
I \quad = \quad & \sum_{t=0}^{T} \frac{E\left(X_{it} \mid \widetilde{\mathcal{I}}_0\right)}{(1+r)^t}(\beta_{1t}-\beta_{0t}) + \sum_{t=0}^{T} \frac{\left[X_{it} - E\left(X_{it} \mid \widetilde{\mathcal{I}}_0\right)\right]}{(1+r)^t}(\beta_{1t}-\beta_{0t}) \odot \Delta_X \\
& +E(\theta_i \mid \widetilde{\mathcal{I}}_0)\left[\sum_{t=0}^{T} \frac{(\alpha_{1t}-\alpha_{0t})}{(1+r)^t} - \alpha_C\right] + [\theta_i - E(\theta_i \mid \widetilde{\mathcal{I}}_0)]\left\{\left[\sum_{t=0}^{T} \frac{(\alpha_{1t}-\alpha_{0t})}{(1+r)^t} - \alpha_C\right] \odot \Delta_\theta\right\} \\
& + \sum_{t=0}^{T} \frac{E(\varepsilon_{1it}-\varepsilon_{0it} \mid \widetilde{\mathcal{I}}_0)}{(1+r)^t} + \sum_{t=0}^{T} \frac{[(\varepsilon_{1it}-\varepsilon_{0it}) - E(\varepsilon_{1it}-\varepsilon_{0it} \mid \widetilde{\mathcal{I}}_0)]}{(1+r)^t}\Delta_{\varepsilon_t} \\
& -E\left(Z_i \mid \widetilde{\mathcal{I}}_0\right)\gamma - \left[Z_i - E\left(Z_i \mid \widetilde{\mathcal{I}}_0\right)\right] \odot \Delta_Z - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_0\right) - \left[\varepsilon_{iC} - E\left(\varepsilon_{iC} \mid \widetilde{\mathcal{I}}_0\right)\right]\Delta_{\varepsilon_C}
\end{aligned}
$$

A test of the validity of information set $\widetilde{\mathcal{I}}_0$ is that $\Delta_X = 0$; $\Delta_\theta = 0; \Delta_Z = 0; \Delta_{\varepsilon_C} = 0$ and $\Delta_{\varepsilon_t} = 0, \forall t$. Components associated with zero Δ's are the unforecastable elements.

- We test what components of the future income process unobservable to the econometrician enter the agents information set and are *acted on* by the agent at the time of his schooling decision and so are not components of uncertainty but rather components of heterogeneity.

- This procedure can be generalized so θ becomes θ_t, a hidden state Markov process.

How is the Model Identified?

Problem: We need to construct counterfactuals. We only observe earnings in college for people who choose college and earnings in high school for people who choose high school. We can never form the covariance between college and high school in the raw data.

Solution: Extension of Factor Models to Nonlinear Settings (see Goldberger and Jöreskog, *MIMIC*, 1972; *LISREL*, Jöreskog, 1977 for linear versions)

Consider a simple example that motivates the main idea.
We observe

$$\begin{aligned} &(Y_{1,1}, \ldots, Y_{1,T}) \text{ for } s = 1 \\ &(Y_{0,1}, \ldots, Y_{0,T}) \text{ for } s = 0 \end{aligned}$$

Let net utility of $s = 1$ be represented by the index.

$$I = \mu_I(X, Z) + U_I, \quad Z \text{ instruments}$$

$$\begin{aligned} Y_{1,t} &= \mu_{1,t}(X) + U_{1,t} \quad t = 1, \ldots, T \\ Y_{0,t} &= \mu_{0,t}(X) + U_{0,t} \quad t = 1, \ldots, T \end{aligned}$$

We know $F\left(Y_{1,1},\ldots,Y_{1,T}\mid S=1,X,Z\right)$ and $F\left(Y_{0,1},\ldots,Y_{0,T}\mid S=0,X,Z\right)$.

Under conditions on regressors and support of Z (Heckman (1990), Heckman and Smith (1998), Carneiro, Heckman and Hansen (2003)), we can identify from these distributions

$$\mu_{0,t}\left(X\right),\mu_{1,t}\left(X\right),\mu_{I}\left(X,Z\right)\quad\text{(up to scale) } t=1,\ldots,T$$

and the joint distribution of

$$\begin{gathered}F\left(Y_{1,1},\ldots,Y_{1,T},I\mid X,Z\right)\\ F\left(Y_{0,1},\ldots,Y_{0,T},I\mid X,Z\right)\end{gathered}$$

with the scale of I not determined (must be normalized). I^{*} is the normalized index.

Motivation on the Nonparametric Identification of The Joint Distribution of Outcomes and The Binary Choice Equation

Motivate why $F(Y_0, I^* \mid X, Z)$ is identified. (Take simple case: two potential outcomes; $T = 0$ so one period model for simplicity only)

From Cosslett (1983), Manski (1988) and Matzkin (1992)

Can identify $\dfrac{\mu_I(X,Z)}{\sigma_I}$ from $\Pr(S = 1 \mid X, Z) = \Pr(\mu_I(X,Z) + U_I \geq 0 \mid X, Z)$ (Support conditions and continuous regressors).

Can identify distribution of $\dfrac{U_I}{\sigma_I}$.

From this information and

$$F(Y_0 \mid S = 0, X, Z) = \Pr(Y_0 \leq y_0 \mid \mu_I(X, Z) + U_I \leq 0, X, Z)$$

Form

$$F(Y_0 \mid S = 0, X, Z) \Pr(S = 0 \mid X, Z) = \Pr(Y_0 \leq y_0, I^* \leq 0 \mid X, Z)$$

Follow analysis of Heckman (1990), Heckman and Smith (1998) and Carneiro, Hansen and Heckman (2003).

Left hand side known.

Right hand side:

$$\Pr\left(Y_0 \le y_0, \frac{U_I}{\sigma_I} < -\frac{\mu_I(X,Z)}{\sigma_I} \mid X, Z\right)$$

Since we know $\dfrac{\mu_I(X,Z)}{\sigma_I}$, we can vary it for each fixed X.
If $\mu_I(X,Z)$ gets small ($\mu_I(X,Z) \to -\infty$), recover the marginal distribution Y and

$$Y_0 = \mu_0(X) + U_0 \quad \therefore \quad \text{can identify}$$

$$\Pr\left(U_0 \le y_0 - \mu_0(X), \frac{U_I}{\sigma_I} \le \frac{-\mu_I(X,Z)}{\sigma_I} \mid X, Z\right)$$

X and Z can be varied and y_0 is a number.

Trace out joint distribution of $\left(U_0, \frac{U_I}{\sigma_I}\right)$.

$\therefore$ Recover joint distribution of

$$(Y_0, I^*) = \left(\mu_0(X) + U_0, \frac{\mu_I(X,Z) + U_I}{\sigma_I}\right).$$

Three key ingredients.

1. The independence of (U_0, U_I) and (X, Z).
2. The assumption that we can set $\dfrac{\mu_I(X,Z)}{\sigma_I}$ to be very small (so we get the marginal distribution of Y_0 and hence $\mu_0(X)$).
3. The assumption that $\dfrac{\mu_I(X,Z)}{\sigma_I}$ can be varied independently of $\mu_0(X)$.

Trace out the joint distribution of $\left(U_0, \frac{U_I}{\sigma_I}\right)$. Result generalizes easily to the vector case. (Carneiro, Hansen, Heckman, *IER*, 2003)

Another way to see this is to write:

$$F(Y_0 \mid S=0, X, Z)\,\Pr(S=0 \mid X, Z)$$

This is a function of $\mu_0(X)$ and $\dfrac{\mu_I(X,Z)}{\sigma_I}$ (Index sufficiency)

Varying the $\mu_0(X)$ and $\dfrac{\mu_I(X,Z)}{\sigma_I}$ traces out the distribution of $U_0, \dfrac{U_I}{\sigma_I}$.

This means effectively that we observe $\left(\frac{I}{\sigma_I}, Y_1\right), \left(\frac{I}{\sigma_I}, Y_0\right)$

We do not observe $\left(\frac{I}{\sigma_I}, Y_0, Y_1\right)$

Using Factor Analysis, Can Construct Joint Distributions of Counterfactuals

Example: One factor model.

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

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

$$E(\theta) = 0, \quad \text{and} \quad E\left(\theta^2\right) = \sigma_\theta^2.$$
$$E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_{I^*}) = 0$$
$$Var(\varepsilon_0) = \sigma_{\varepsilon_0}^2, Var(\varepsilon_1) = \sigma_{\varepsilon_1}^2$$
$$Var(\varepsilon_I) = \sigma_{\varepsilon_I}^2$$

1 Factor Models: A Brief Digression

$$E(\theta) = 0; \qquad E(\varepsilon_i) = 0;\ i = 1, \ldots, 5$$

$$R_1 = \alpha_1\theta + \varepsilon_1, \quad R_2 = \alpha_2\theta + \varepsilon_2, \quad R_3 = \alpha_3\theta + \varepsilon_3,$$
$$R_4 = \alpha_4\theta + \varepsilon_4, \quad R_5 = \alpha_5\theta + \varepsilon_5, \quad \varepsilon_i \perp\!\!\!\perp \varepsilon_j$$

$$\begin{aligned} Cov(R_1, R_2) &= \alpha_1\alpha_2\sigma_\theta^2 \\ Cov(R_1, R_3) &= \alpha_1\alpha_3\sigma_\theta^2 \\ Cov(R_2, R_3) &= \alpha_2\alpha_3\sigma_\theta^2 \end{aligned}$$

Normalize $\alpha_1 = 1$

$$\frac{Cov(R_2, R_3)}{Cov(R_1, R_2)} = \alpha_3$$

$\therefore$ We know σ_θ^2 from $Cov\left(R_1, R_2\right)$. From $Cov\left(R_1, R_3\right)$ we know

$$\alpha_3, \alpha_4, \alpha_5.$$

Can get the variances of the ε_i from variances of the R_i

$$Var(R_i) = \alpha_i^2 \sigma_\theta^2 + \sigma_{\varepsilon_i}^2.$$

If $T = 2$, all we can identify is $\alpha_1 \alpha_2 \sigma_\theta^2$.

If $\alpha_1 = 1$, $\sigma_\theta^2 = 1$, we identify α_2.

2 Factors:

$$\theta_1 \perp\!\!\!\perp \theta_2$$

$$\varepsilon_i \perp\!\!\!\perp \varepsilon_j \quad \forall i, j$$

$$\begin{aligned}
R_1 &= \alpha_{11}\theta_1 + (0)\theta_2 + \varepsilon_1 \\
R_2 &= \alpha_{21}\theta_1 + (0)\theta_2 + \varepsilon_2 \\
R_3 &= \alpha_{31}\theta_1 + \alpha_{32}\theta_2 + \varepsilon_3 \\
R_4 &= \alpha_{41}\theta_1 + \alpha_{42}\theta_2 + \varepsilon_4 \\
R_5 &= \alpha_{51}\theta_1 + \alpha_{52}\theta_2 + \varepsilon_5
\end{aligned}$$

Let $\alpha_{11} = 1$, $\alpha_{32} = 1$.

$$\begin{aligned} Cov\left(R_1, R_2\right) &= \alpha_{21}\sigma^2_{\theta_1} \\ Cov\left(R_1, R_3\right) &= \alpha_{31}\sigma^2_{\theta_1} \\ Cov\left(R_2, R_3\right) &= \alpha_{21}\alpha_{31}\sigma^2_{\theta_1} \end{aligned}$$

Form ratio of $\dfrac{Cov\left(R_2, R_3\right)}{Cov\left(R_1, R_2\right)} = \alpha_{31}$, $\quad\therefore$ we identify $\alpha_{31}, \alpha_{21}, \sigma^2_{\theta_1}$, as before.

$$\begin{aligned} Cov\left(R_1, R_4\right) &= \alpha_{41}\sigma^2_{\theta_1}, \qquad \therefore \text{we get } \sigma^2_{\theta_1} \therefore \text{we get } \alpha = 1 \\ &\vdots \\ Cov\left(R_1, R_k\right) &= \alpha_{k1}\sigma^2_{\theta_1} \end{aligned}$$

$\therefore$ we identify α_{k1} for all k and $\sigma^2_{\theta_1}$.

$$
\begin{aligned}
Cov\left(R_3, R_4\right) - \alpha_{31}\alpha_{41}\sigma^2_{\theta_1} &= \alpha_{42}\sigma^2_{\theta_2} \\
Cov\left(R_3, R_5\right) - \alpha_{31}\alpha_{51}\sigma^2_{\theta_1} &= \alpha_{52}\sigma^2_{\theta_2} \\
Cov\left(R_4, R_5\right) - \alpha_{41}\alpha_{51}\sigma^2_{\theta_1} &= \alpha_{52}\alpha_{42}\sigma^2_{\theta_2},
\end{aligned}
$$

By same logic,

$$
\frac{Cov\left(R_4, R_5\right) - \alpha_{41}\alpha_{51}\sigma^2_{\theta_1}}{Cov\left(R_3, R_4\right) - \alpha_{31}\alpha_{41}\sigma^2_{\theta_1}} = \alpha_{52}
$$

$\therefore$ get $\sigma^2_{\theta_2}$ of "2" loadings.

If we have dedicated measurements do not need a normalization on R.

$$\begin{aligned} M_1 &= \theta_1 + \varepsilon_{1M} \\ M_2 &= \theta_2 + \varepsilon_{2M} \end{aligned}$$

$$\begin{aligned} Cov\,(R_1, M) &= \alpha_{11}\sigma^2_{\theta_1} \\ Cov\,(R_2, M) &= \alpha_{21}\sigma^2_{\theta_1} \\ Cov\,(R_3, M) &= \alpha_{31}\sigma^2_{\theta_1} \end{aligned}$$

$$\begin{aligned} Cov\,(R_1, R_2) &= \alpha_{11}\alpha_{12}\sigma^2_{\theta_1}, \\ Cov\,(R_1, R_3) &= \alpha_{11}\alpha_{13}\sigma^2_{\theta_1}, \qquad \therefore \alpha_{12}\sigma^2_{\theta_1}, \end{aligned}$$

$\therefore$ We can get $\alpha_{12}, \sigma^2_{\theta_1}$ and the other factors.

General Case

$$\underset{T\times 1}{R} = \underset{T\times 1}{M} + \underset{T\times K}{\Lambda}\underset{K\times 1}{\theta} + \underset{T\times 1}{\varepsilon}$$

θ are factors, ε uniquenesses

$$E(\varepsilon) = 0$$

$$Var(\varepsilon\varepsilon') = D = \begin{pmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma^2_{\varepsilon_T} \end{pmatrix}$$

$$E(\theta) = 0$$

$$Var(R) = \Lambda\Sigma_\theta\Lambda' + D \qquad \Sigma_\theta = E(\theta\theta')$$

The only source of information on Λ and Σ_θ is from the covariances.

Associated with each variance of R_i is a $\sigma^2_{\varepsilon_i}$.

Each variance contributes one new parameter.

How many unique covariance terms do we have?

$\frac{T(T-1)}{2}$ This is the data.

We have T uniquenesses; TK elements of Λ.

$\frac{K(K-1)}{2}$ elements of Σ_θ. $\frac{K(K-1)}{2} + TK$ parameters (Σ_θ, Λ).

Observe that if we multiply Λ by an orthogonal matrix C, $(CC' = I)$, we have

$$Var\,(R) = \Lambda C\,[C'\Sigma_\theta C]\,C'\Lambda' + D$$

C is a "rotation". Cannot separate ΛC from Λ.

Model not identified against orthogonal transformations in the general case.

Some common assumptions:

(i) $\theta_i \perp\!\!\!\perp \theta_j, \forall\ i \neq j$

$$\Sigma_\theta = \begin{pmatrix} \sigma^2_{\theta_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\theta_2} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma^2_{\theta_K} \end{pmatrix}$$

joined with

(ii)

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \alpha_{21} & 0 & 0 & 0 & \cdots & 0 \\ \alpha_{31} & 1 & 0 & 0 & \cdots & 0 \\ \alpha_{41} & \alpha_{42} & 0 & 0 & \cdots & 0 \\ \alpha_{51} & \alpha_{52} & 1 & 0 & \cdots & 0 \\ \alpha_{61} & \alpha_{62} & \alpha_{63} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & 1 & & \vdots \end{pmatrix}$$

We know that we can identify of the Λ, Σ_θ parameters.

$$\underbrace{\frac{K(K-1)}{2} + TK}_{\text{\# of free parameters}} \leq \underbrace{\frac{T(T-1)}{2}}_{\text{data}}$$

"Ledermann Bound"

Can get more information by looking at higher order moments.

Recovering the Factor Loadings in the Roy Model (Go back to simple case)

The Case when there is information only on Y_0 for $I < 0$ and Y_1 for $I > 0$

Can identify $F(U_0, U_{I^*})$ and $F(U_1, U_{I^*})$, $\therefore$ can identify

$$\begin{aligned} Cov(U_0, U_{I^*}) &= \alpha_0 \alpha_{I^*} \sigma_\theta^2 \\ Cov(U_1, U_{I^*}) &= \alpha_1 \alpha_{I^*} \sigma_\theta^2. \end{aligned}$$

Scale of the unobserved I is normalized

$$\sigma_\theta^2 = 1 \qquad \alpha\theta = k\alpha\frac{\theta}{k}$$

Normalize some α_j to one.

$$\alpha_{I^*} = 1$$

Identify α_1 and α_0 from the known covariances above.

Since

$$Cov\left(U_1, U_0\right) = \alpha_1 \alpha_0 \sigma_\theta^2$$

we can identify covariance between Y_1 and Y_0

Do not observe the pair (Y_1, Y_0)

Access to more observations (say from panel data $T > 0$)

$$\begin{aligned} \frac{Cov\left(Y_{1t'}, I^*\right)}{Cov\left(Y_{1t'}, Y_{1t}\right)} &= \alpha_{1t} \\ \frac{Cov\left(Y_{0t'}, I^*\right)}{Cov\left(Y_{0t'}, Y_{0t}\right)} &= \alpha_{0t} \end{aligned}$$

Crucial Idea of Identification

We never observe (Y_1, Y_0) as a pair, both Y_0 and Y_1 are linked to S through the choice equation.

From S we can generate I^*

We essentially 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^*.

Adding a Measurement Equation Helps to Identify the Model

Suppose we have a measurement for θ observed whether $S = 1$ or $S = 0$

Measured ability M is

$$M = \mu_M(X) + U_M.$$

Assume that

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

We assume $\alpha_M \neq 0$. Can form

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

$\alpha_M = 1.$

Can form the ratios

$$\text{Identify } \alpha_0: \qquad \frac{Cov\left(U_0, U_{I^*}\right)}{Cov\left(U_M, U_{I^*}\right)} = \alpha_0$$

$$\text{Recover } \alpha_1: \qquad \frac{Cov\left(U_1, U_{I^*}\right)}{Cov\left(U_M, U_{I^*}\right)} = \alpha_1$$

$$Cov\left(U_M, U_0\right) = \alpha_0 \sigma_\theta^2$$

Can identify α_{I^*}.

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

2 Intuition on Identification of the Normal Case Model

Generalized Roy versions of model:

$$M = \mu(X) + \theta_1 \alpha_{1,M} + \theta_2 \alpha_{2,M} + \varepsilon_M$$

(Measurement: A test score equation)

$$\left.\begin{array}{l} Y_1^1 = \mu_1^1(X) + \theta_1 \alpha_{1,1}^1 + \theta_2 \alpha_{2,1}^1 + \varepsilon_1^1 \\ Y_2^1 = \mu_2^1(X) + \theta_1 \alpha_{1,2}^1 + \theta_2 \alpha_{2,2}^1 + \varepsilon_2^1 \end{array}\right\} \text{College earnings}$$

$$\left.\begin{array}{l} Y_1^0 = \mu_1^0(X) + \theta_1 \alpha_{1,1}^0 + \theta_2 \alpha_{2,1}^0 + \varepsilon_1^0 \\ Y_2^0 = \mu_2^0(X) + \theta_1 \alpha_{1,2}^0 + \theta_2 \alpha_{2,2}^0 + \varepsilon_2^0 \end{array}\right\} \text{High School earnings}$$

Cost

$$C = Z\gamma + \theta_1 \alpha_{1C} + \theta_2 \alpha_{2C} + \varepsilon_C$$

Decision Rule Under Perfect Certainty:
(Assume $r = 0$)

$$\begin{aligned}
I &= \mu_1^1(X) + \mu_2^1(X) + \theta_1\left(\alpha_{1,1}^1 + \alpha_{1,2}^1\right) \\
&\quad + \theta_2\left(\alpha_{2,1}^1 + \alpha_{2,2}^1\right) + \varepsilon_1^1 + \varepsilon_2^1 \\
&\quad - \left[\begin{array}{c} \mu_1^0(X) + \mu_2^0(X) + \theta_1\left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) \\ + \theta_2\left(\alpha_{2,1}^0 + \alpha_{2,2}^0\right) + \varepsilon_1^0 + \varepsilon_2^0 \end{array}\right] \\
&\quad - Z\gamma - \theta_1\alpha_{1C} - \theta_2\alpha_{2C} - \varepsilon_C \\
&= \mu_1^1(X) + \mu_2^1(X) - \left[\mu_1^0(X) + \mu_2^0(X) + Z\gamma\right] \\
&\quad + \theta_1\left[\left(\alpha_{1,1}^1 + \alpha_{1,2}^1\right) - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1C}\right] \\
&\quad + \theta_2\left[\left(\alpha_{2,1}^1 + \alpha_{2,2}^1\right) - \left(\alpha_{2,1}^0 + \alpha_{2,2}^0\right) - \alpha_{2C}\right] \\
&\quad + \left(\varepsilon_1^1 + \varepsilon_2^1\right) - \left(\varepsilon_1^0 + \varepsilon_2^0\right) - \varepsilon_C
\end{aligned}$$

In Reduced Form

$$I = \varphi(X, Z) + \alpha_{I,1}\theta_1 + \alpha_{I,2}\theta_2 + \varepsilon_I.$$

Set $U_I = \alpha_{I,1}\theta_1 + \alpha_{I,2}\theta_2 + \varepsilon_I$.

$\therefore$ we can write

$$\begin{aligned}
Y_1^1 &= \mu_1^1(X) + U_1^1 \\
Y_2^1 &= \mu_2^1(X) + U_2^1 \\
Y_1^0 &= \mu_1^0(X) + U_1^0 \\
Y_2^0 &= \mu_2^0(X) + U_2^0
\end{aligned}$$

U_1^1, U_2^1 etc. match the error terms previously shown.

$$\begin{aligned}
U_1^1 &= \theta_1\alpha_{1,1}^1 + \theta_2\alpha_{2,1}^1 + \varepsilon_1^1 \text{ etc.} \\
U_M &= \theta_1\alpha_{1,M} + \theta_2\alpha_{2,M} + \varepsilon_M
\end{aligned}$$

$$E\left(Y_1^1 \mid X, Z, I > 0\right) = \mu_1^1(X) + \frac{Cov(U_1^1, I)}{Var(I)}\lambda()$$

Using standard sample selection bias arguments we can identify beside the means,

$$\mu_1^1(X), \mu_2^1(X), \mu_2^0(X), \mu_2^0(X), \text{ the following parameters:}$$

$$\begin{array}{l} Cov\left(U_1^1, U_2^1\right), Var\left(U_1^1\right), Var\left(U_2^1\right) \\ Cov\left(U_1^1, U_M\right), Cov\left(U_2^1, U_M\right), Var\left(U_M\right) \\ Cov\left(U_1^0, U_2^0\right), Var\left(U_1^0\right), Var\left(U_2^0\right) \\ Cov\left(U_1^0, U_M\right), Cov\left(U_2^0, U_M\right) \end{array}$$

Normal Case:

$$(\theta, \varepsilon) \perp\!\!\!\perp (X, Z) \quad : \quad (\theta, \varepsilon) \text{ normal.}$$

$$\begin{aligned} & \Pr\left(S = 1 \mid X, Z, \theta_1, \theta_2\right) \\ = \quad & \Phi\left[\frac{1}{\sigma_{\varepsilon_I}}\left[\begin{array}{c} \mu_1^1\left(X\right) + \mu_2^1\left(X\right) - \left[\mu_1^0\left(X\right) + \mu_2^0\left(X\right)\right] \\ -Z\gamma + \theta_1\alpha_{I,1} + \theta_2\alpha_{I,2} \end{array}\right]\right] \end{aligned}$$

Fact:

If $S = \mathbf{1}[X\beta + \theta > V]$, $X \perp\!\!\!\perp (\theta, V)$

θ, V are normal, $\theta \perp\!\!\!\perp V$, $E(\theta) = 0, E(V) = 0$

$$\begin{aligned}\Pr(S = 1 \mid X, \theta) &= \Phi\left(\frac{X\beta + \theta}{\sigma_V}\right)\\ \Pr(S = 1 \mid X) &= \Phi\left(\frac{X\beta}{(\sigma_V^2 + \sigma_\theta^2)^{\frac{1}{2}}}\right)\end{aligned}$$

Why? $S = \mathbf{1}[X\beta > V - \theta]$. Rest follows from independence (between $V - \theta$, and X, and normality).

Unconditional Probability:

$$\Pr(S=1 \mid X,Z) = \Phi\left[\frac{\mu_1^1(X)+\mu_2^1(X)-[\mu_1^0(X)+\mu_2^0(X)]-Z\gamma}{\left(\sigma_{\varepsilon_I}^2+\alpha_{I,1}^2\sigma_{\theta_1}^2+\alpha_{I,2}^2\sigma_{\theta_2}^2\right)^{1/2}}\right]$$

Observe that if we know $\mu_1^1(X), \mu_2^1(X), \mu_1^0(X), \mu_2^0(X)$ we know

$$\left[\mu_1^1(X)+\mu_2^1(X)\right]-\left[\mu_1^0(X)+\mu_2^0(X)\right].$$

If $Z\gamma$ not perfectly collinear with this term (e.g. one or more elements of X not in Z) we can identify

$$\left(\sigma_{\varepsilon_I}^2+\alpha_{I,1}^2\sigma_{\theta_1}^2+\alpha_{I,2}^2\sigma_{\theta_2}^2\right)^{\frac{1}{2}}$$

$\therefore$ we also identify γ (get absolute scale on costs).

Suppose agents do not know θ_2 or the future $\varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^0, \varepsilon_2^0$ but know ε_c and θ_1.

Then if what they know is set at mean zero, (they use rational expectations in a linear decision rule) and their mean forecast is the population mean,

$$\sigma_{\varepsilon_I}^2 = \sigma_{\varepsilon_c}^2$$

and $\alpha_{I,2} = 0$.

What can we identify?
Is the model testable?

What information do we have about covariances?

Suppose we have two dedicated measurement systems for θ_1 and θ_2.

$$\left.\begin{aligned} M_1^1 &= \theta_1 &+\ \ & \varepsilon_{1,M}^1 \\ M_2^1 &= \alpha_{2,M}^1 \theta_1 &+\ \ & \varepsilon_{2,M}^1 \\ M_3^1 &= \alpha_{3,M}^1 \theta_1 &+\ \ & \varepsilon_{3,M}^1 \end{aligned}\right\} \text{Cognitive Ability}$$

$$\left.\begin{aligned} M_1^2 &= \theta_2 &+\ \ & \varepsilon_{1,M}^2 \\ M_2^2 &= \alpha_{2,M}^2 \theta_2 &+\ \ & \varepsilon_{2,M}^2 \\ M_3^2 &= \alpha_{3,M}^2 \theta_2 &+\ \ & \varepsilon_{3,M}^2 \end{aligned}\right\} \text{Noncognitive Ability}$$

(See *e.g.* Heckman, Urzua and Stixrud, 2004)

Observe from M^1 system we get

$$Var(\theta_1), \alpha^1_{2,M}, \alpha^1_{3,M}$$

From M^2 system we get

$$Var(\theta_2), \alpha^2_{2,M}, \alpha^2_{3,M}$$

Then

$$\begin{aligned} Cov\left(U_1^1, M_1^1\right) &= \alpha_{1,1}^1 \sigma_{\theta_1}^2 \\ Cov\left(U_2^1, M_1^1\right) &= \alpha_{1,2}^1 \sigma_{\theta_1}^2 \end{aligned}$$

$\therefore$ we get all of the factor loadings in Y^1 on θ_1.

Using M_1^2 we get $\alpha_{2,1}^1, \alpha_{2,2}^1$ and we get variances of uniquenesses $Var\left(\varepsilon_1^1\right), Var\left(\varepsilon_2^1\right)$.

By similar reasoning, we get

$$\begin{gathered} \alpha_{1,1}^0, \alpha_{2,1}^0, \alpha_{1,2}^0, \alpha_{2,2}^0 \\ Var\left(\varepsilon_1^0\right), Var\left(\varepsilon_2^1\right) \end{gathered}$$

Observe from

$$Cov\left(I, M_1^1\right) = \sigma_{\theta_1}^2 \left[\alpha_{1,1}^1 + \alpha_{1,2}^1 - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1,C}\right]$$

$\therefore$ We can get α_{1C} up to scale σ_I, since we know everything else by the previous reasoning.

From

$$Cov\left(I, M_1^2\right) = \sigma_{\theta_2}^2 \left[\alpha_{2,1}^1 + \alpha_{2,2}^1 - \left(\alpha_{2,1}^0 + \alpha_{2,2}^0\right) - \alpha_{2,C}\right]$$

$\therefore$ we get α_{2C} up to scale σ_I.

From $\Pr\left(S = 1 \mid X, Z\right)$, we can identify $\sigma_{\varepsilon_I}^2$ using previous reasoning

Therefore we have that we can identify everything in the model if there is one X not in Z since we can identify the terms in the numerator.

But, can we test the model?

In the previous notation, we have that for a test of whether θ_2 belongs in the model

$$\Pr\left(S=1 \mid X,Z\right)=\Phi\left[\frac{\mu_1^1\left(X\right)+\mu_2^1\left(X\right)-\left[\mu_1^0\left(X\right)+\mu_2^0\left(X\right)\right]-Z\gamma}{\left(\sigma_{\varepsilon_I}^2+\alpha_{I,1}^2\sigma_{\theta_1}^2+\alpha_{I,2}^2\sigma_{\theta_2}^2\Delta_{\theta_2}\right)^{\frac{1}{2}}}\right]$$

Apparently, we can test the null

$$H_0:\Delta_{\theta_2}=0$$

$\therefore$ apparently we can test if θ_2 components enter or not.

The problem with this test is that if $\sigma^2_{\varepsilon_c} \neq 0$, we can always adjust its value to fit the model perfectly well. If we have a pure Roy model, the test is clean. A pure Roy model assumes $\sigma^2_{\varepsilon_c} = 0$.

Notice, however, that we can also tolerate $\gamma \neq 0$ so long as $\sigma^2_{\varepsilon_c} = 0$. Thus we can depart from the Roy model somewhat.

Basic point: we don't observe costs directly. $\therefore$we do not get a clean measurement on $\sigma^2_{\varepsilon_c}$. We can identify σ^2_I but the problem is that $\sigma^2_{\varepsilon_c}$ can be adjusted.

Correct Test:

Form

$$\begin{aligned} Cov\left(\frac{I}{\sigma_I}, U_1^1\right) &= \frac{\sigma_{\theta_1}^2}{\sigma_I}\alpha_{1,1}^1\left[\alpha_{1,1}^1 + \alpha_{1,2}^1 - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1,C}\right] \\ &+\Delta_{\theta_2}\sigma_{\theta_2}^2\alpha_{1,2}^1\left[\alpha_{1,1}^1 + \alpha_{1,2}^1 - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1,C}\right] \end{aligned}$$

we can compute the left hand side under the null. (with exclusion and normality). We identify all components of right hand side by a separate argument (from measurement systems).

Thus under the null that $\Delta_{\theta_2} = 0$, we can identify $\sigma_{\varepsilon_c}^2$.

$\therefore$ we construct a test under null:

$$Cov\left(\frac{I}{\sigma_I}, U_1^1\right) - \frac{\sigma_{\theta_1}^2\alpha_{1,1}^1\left[\alpha_{1,1}^1 + \alpha_{1,2}^1 - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1,c}\right]}{\sigma_I} = 0$$

We know both terms under the null. (We do not use the information on $Cov\left(\frac{I}{\sigma_I}, U_1^1\right)$. Departures are evidence that agents know θ_2.

It is assumed that if the agent knows θ_1 but not θ_2, he sets

$$E(\theta_2) = 0.$$

This is justified by linearity of the criterion and rational expectations, assuming $E(\theta_2 \mid \mathcal{I}_0) = 0$.

Then we can test among models by deciding

- Which model fits the data better?

Average effect (we estimate the average probability):

$$\int \Pr\left(S=1 \mid X, Z, \theta_1, \Delta_{\theta_2}, \theta_2\right) f\left(\theta_1\right) f\left(\theta_2\right) d\theta.$$

(we test $\Delta_{\theta_2}=0$)

This is what is done in the Hicks lecture.
Don't need normality.

Recovering the Distributions Nonparametrically

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 the random variables possess nonvanishing (a.e.) *characteristic functions, then the densities $f(\theta), f(v_1),$ and $f(v_2)$ are identified.*

Proof. See Kotlarski (1967). ∎

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

System can be rewritten as

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

Applying Kotlarski's theorem, identify the densities of $\theta, \dfrac{\varepsilon_{I^*}}{\alpha_{I^*}}, \dfrac{\varepsilon_0}{\alpha_0}, \dfrac{\varepsilon_1}{\alpha_1}, \varepsilon_M$.

We know α_{I^*}, α_0 and α_1. Can identify the densities of $\theta, \varepsilon_{I^*}, \varepsilon_0, \varepsilon_1, \varepsilon_M$. Recover the joint distribution of (Y_1, Y_0)

$$F(Y_1, Y_0 \mid X) = \int F(Y_1, Y_0 \mid \theta, X)\, dF(\theta).$$

$F(\theta)$ is known

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

$F(Y_1 \mid \theta, X)$ and $F(Y_0 \mid \theta, X)$ identified

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

Can identify the number of factors generating dependence among the Y_1, Y_0, C, S and M.

- Crucial idea: even though we never observe (y_1, y_0) as a pair, both y_0 and y_1 are linked to S through the choice equation (I) or a measurement equation (M).

- Can extend to nonseparable models (Heckman, Matzkin, Navarro and Urzua, 2004)

- For the other market structures the decision rule is no longer linear (solution to a dynamic programing problem). That is

$$I = E_{\mathcal{I}_0}\left(V_1\left(X, \theta, \varepsilon_{1,1}, a_0; \phi\right) - V_0\left(X, \theta, \varepsilon_{0,1}, a_0; \phi\right) - Z'\gamma - \theta'\lambda - \varepsilon_{\text{cost}}\right)$$

- The argument still goes through using external measurements like the test equations instead of the choice equation as common identifying relationships.

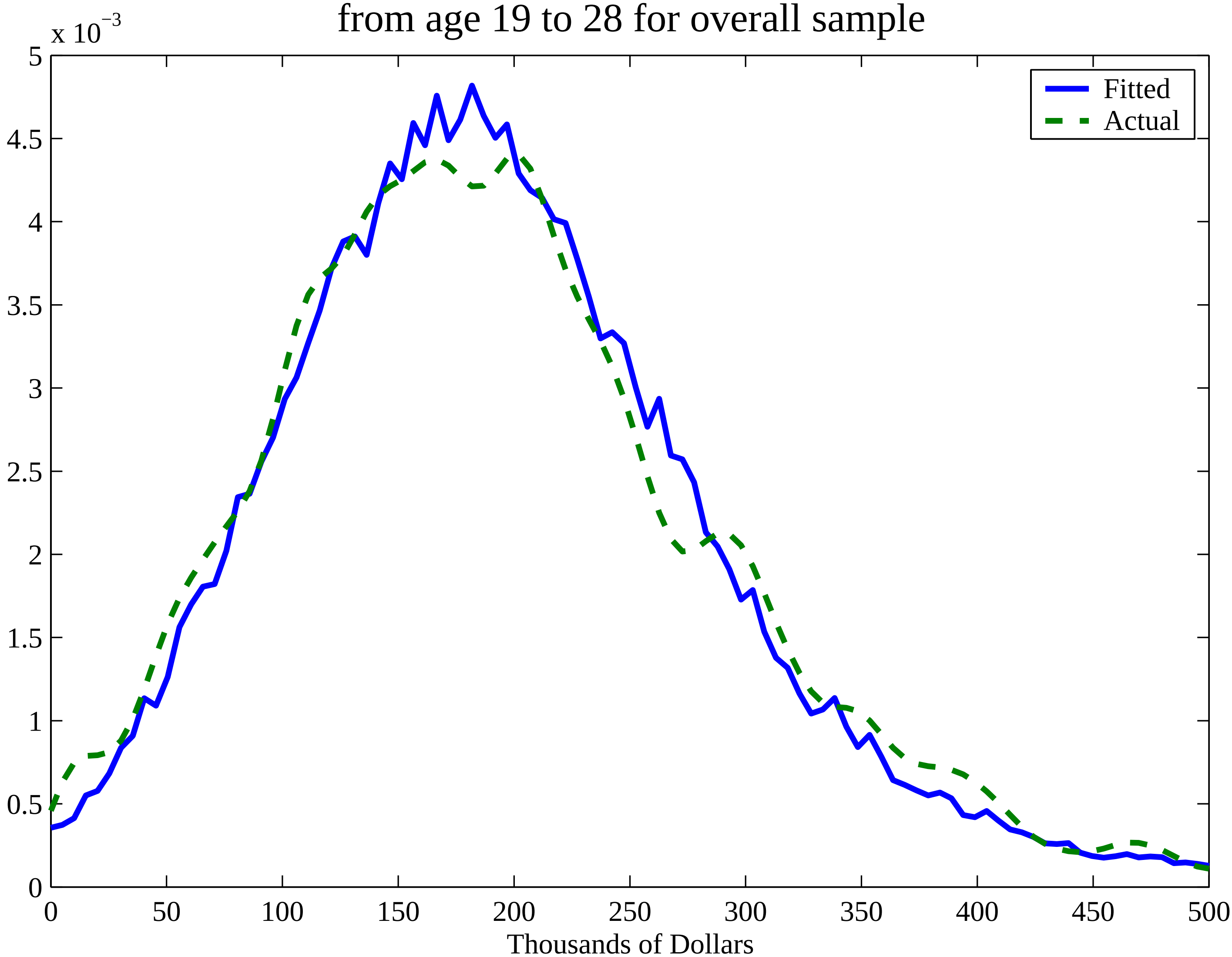
- Alternatively, we can also identify the factor loadings using nonsymmetric θ (nonlinear factor analysis).

- Can fit model, determine the number of factors and generate counterfactuals.

- ϕ is identified. Obvious if we use consumption data. Also true without consumption. Under large support conditions, factors and uniquenesses nonparametrically identified are means. We maintain separabililty.

- Can extend to multiple periods and multiple schooling levels (Heckman and Navarro, 2004)

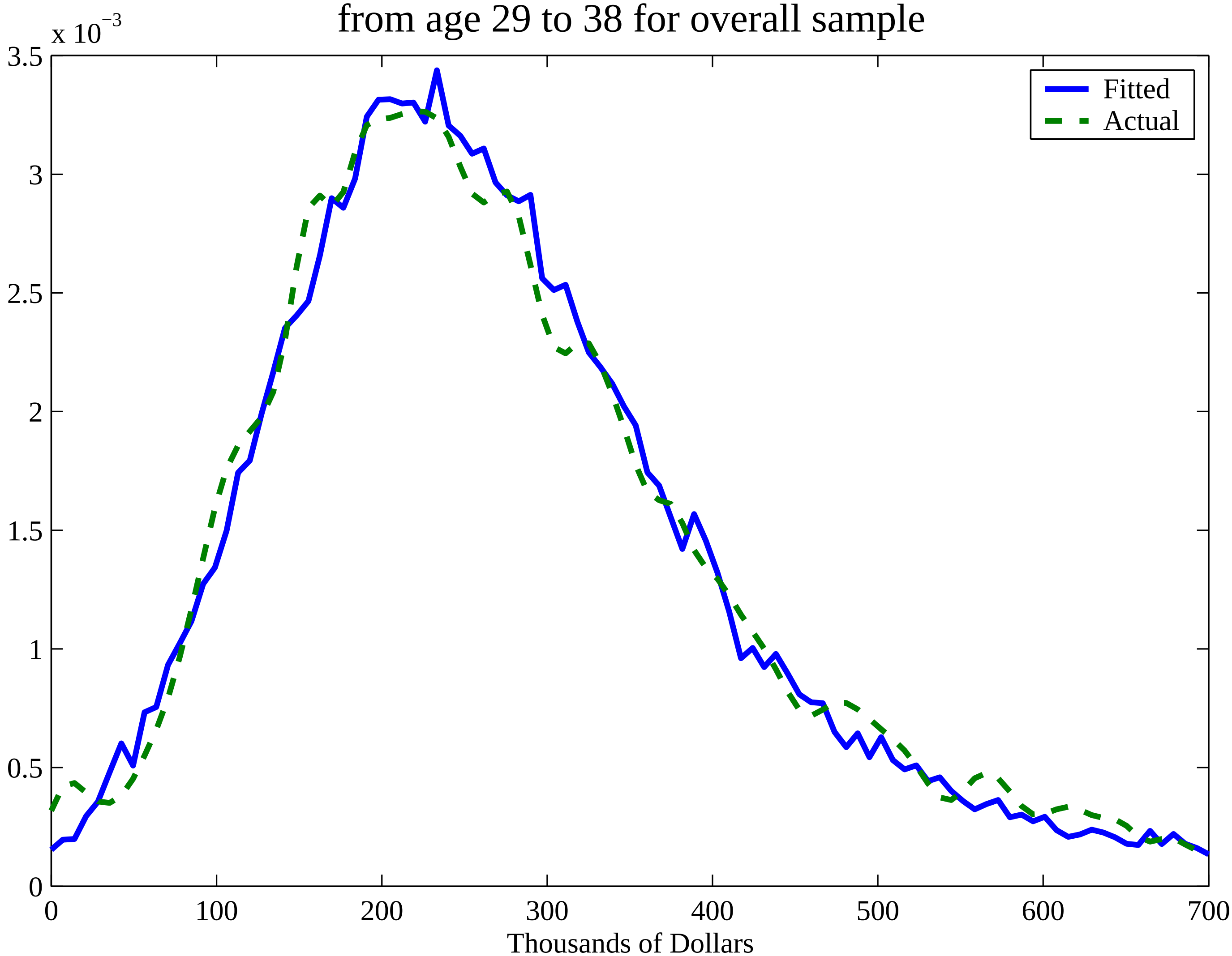
Figure 1.1
Densities of fitted and actual present value of earnings from age 19 to 28 for overall sample



Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(y) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

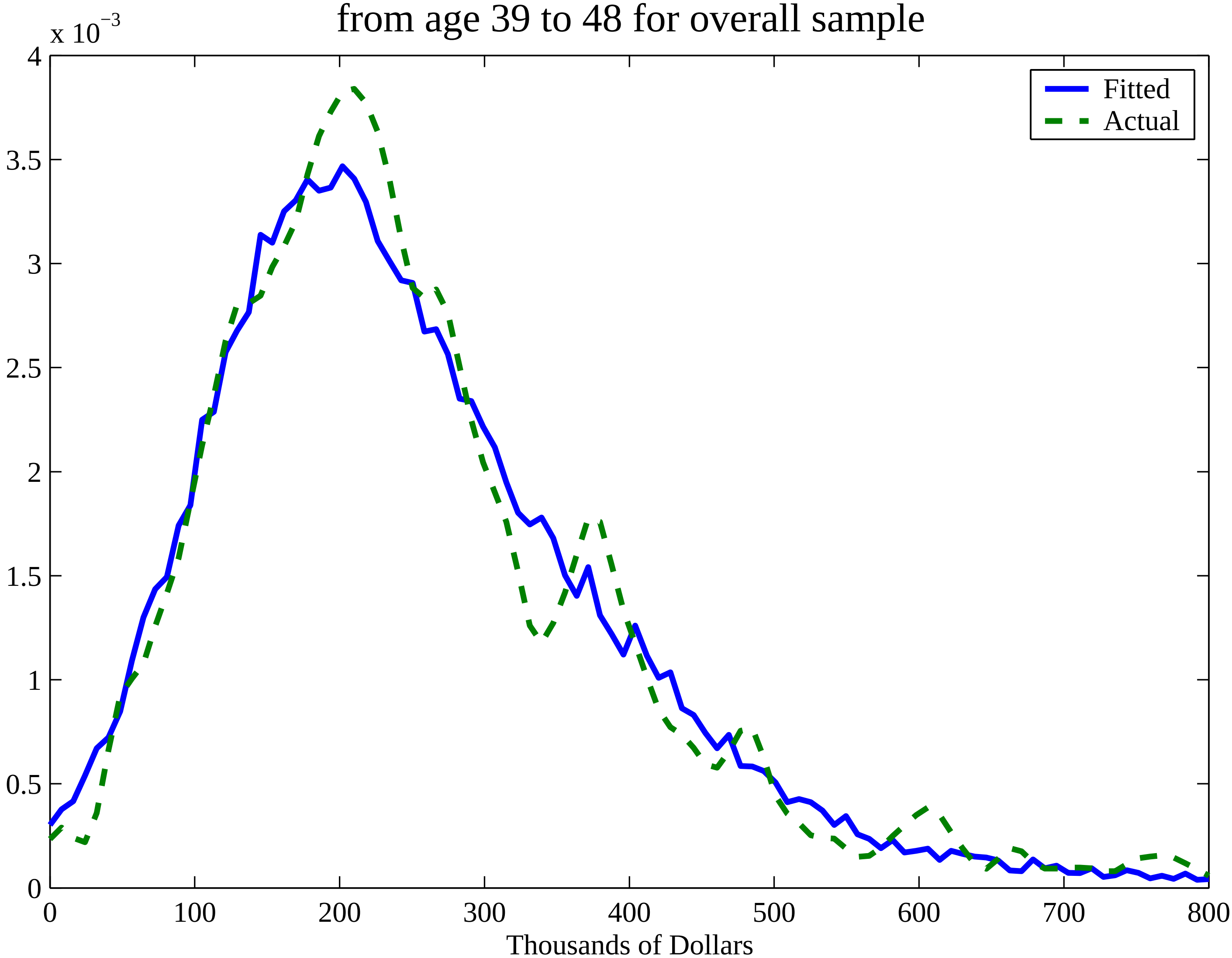
Figure 1.2

Densities of fitted and actual present value of earnings from age 29 to 38 for overall sample



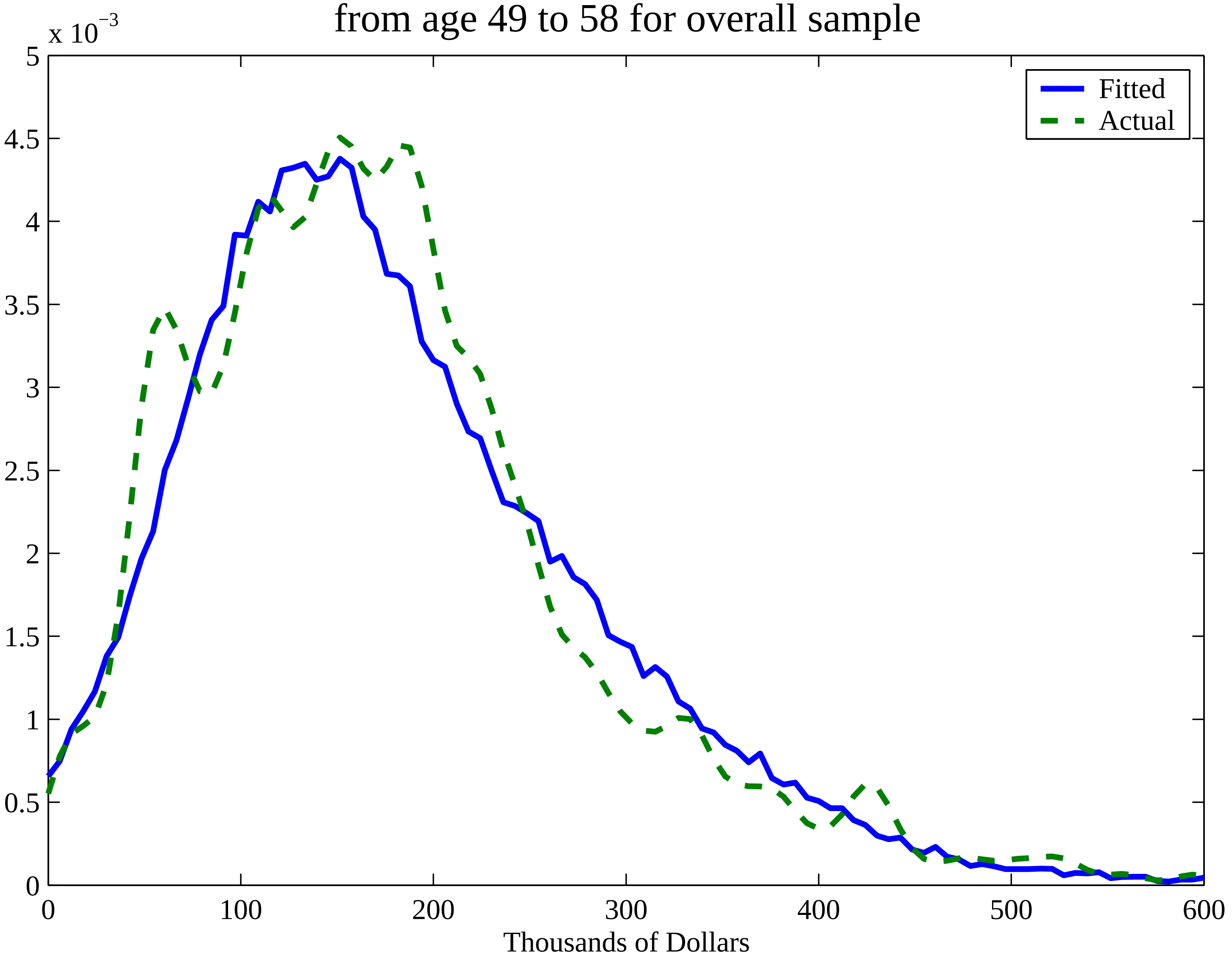
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(y) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 1.3
Densities of fitted and actual present value of earnings from age 39 to 48 for overall sample



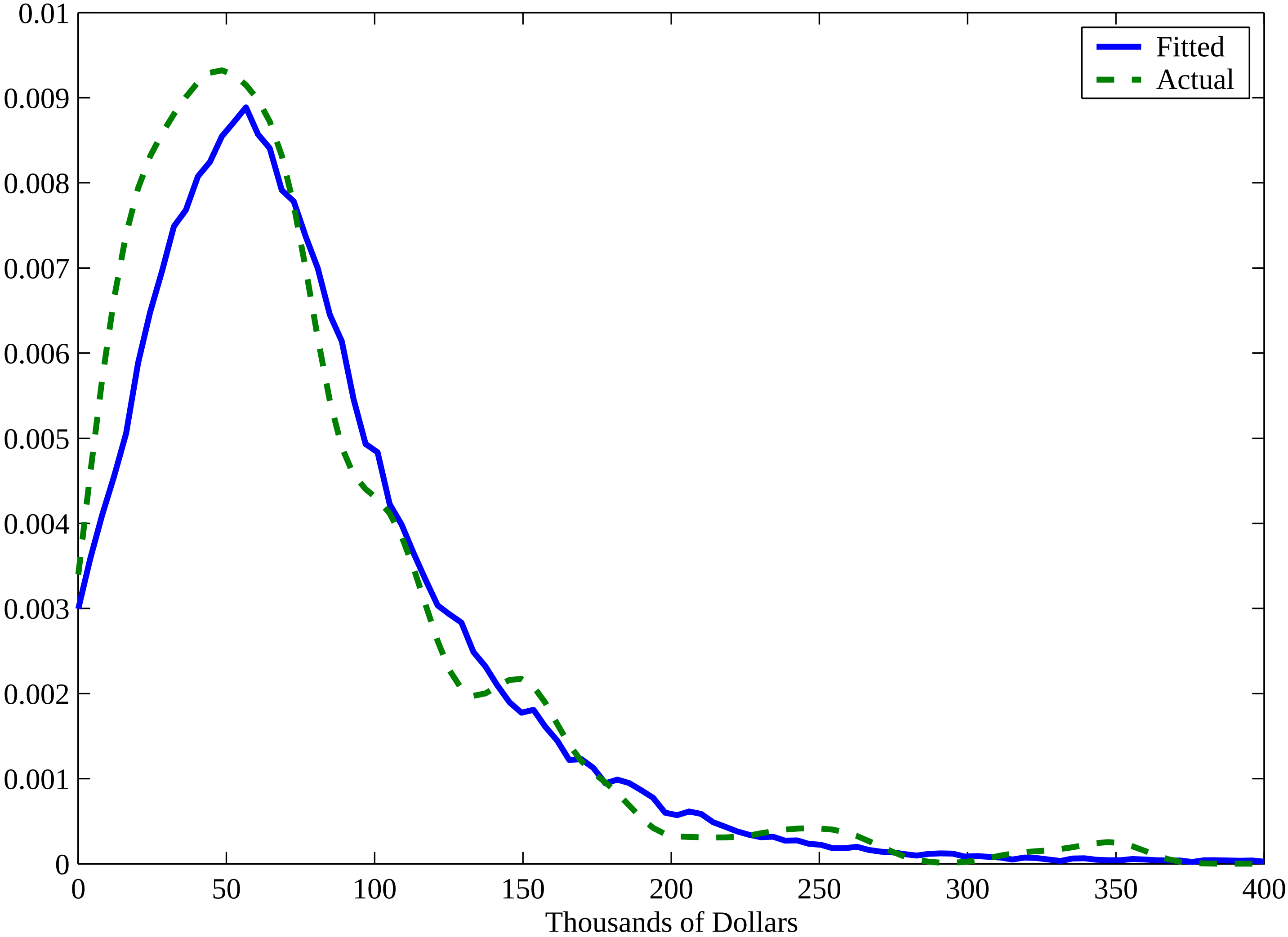
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(y) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 1.4
Densities of fitted and actual present value of earnings from age 49 to 58 for overall sample



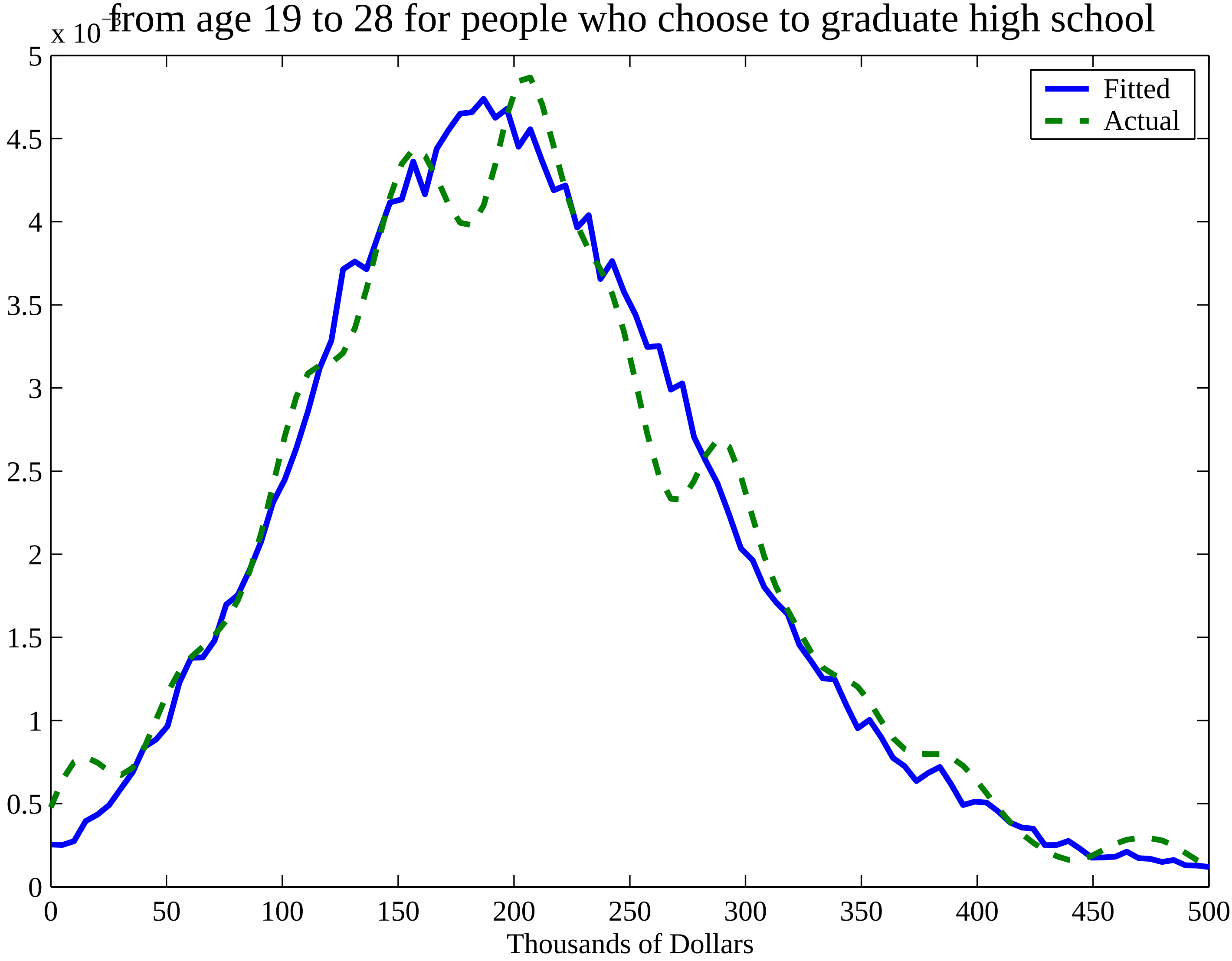
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. Let (Y_0, Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(y) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 1.5
Densities of fitted and actual present value of earnings
from age 59 to 65 for overall sample



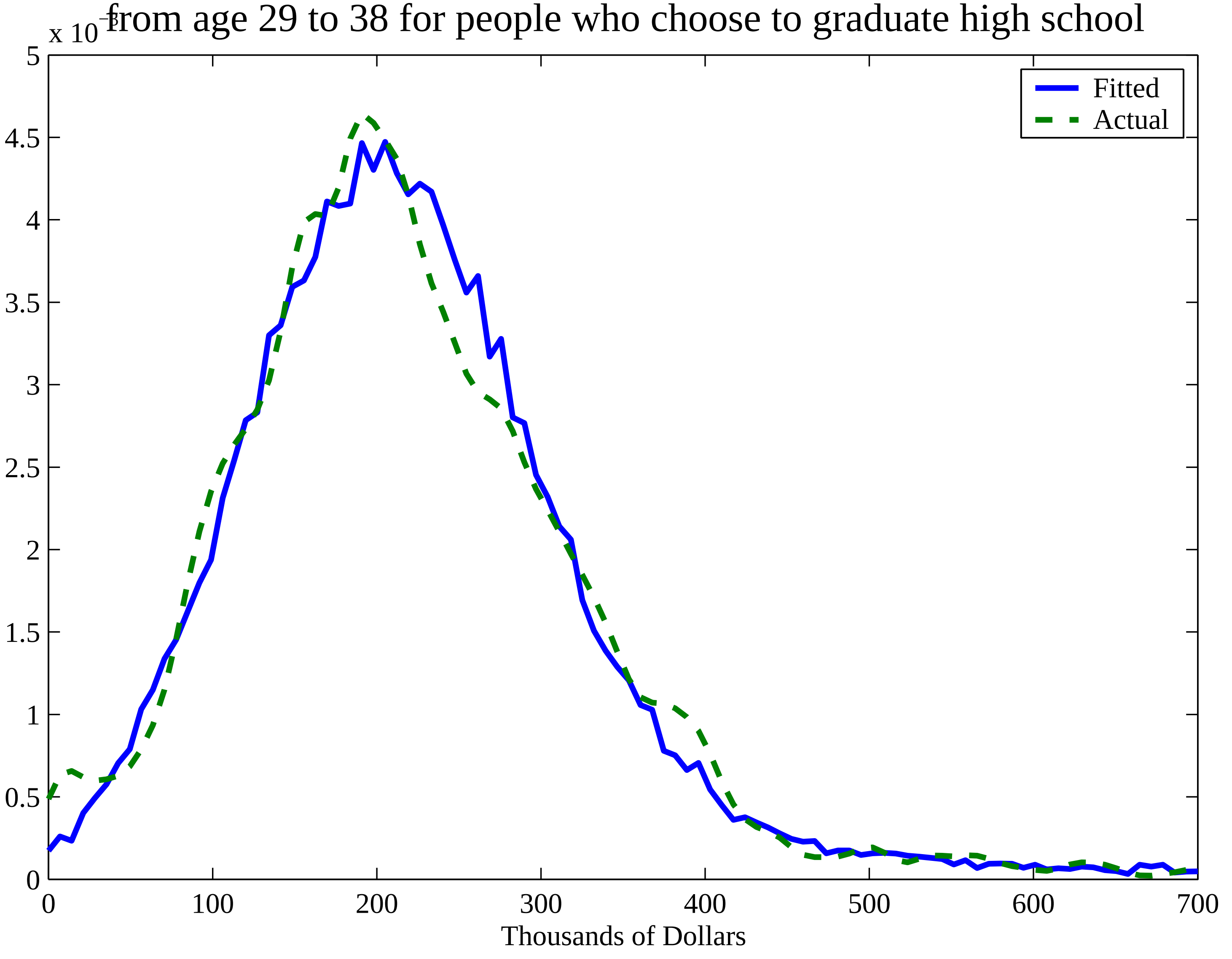
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. Let (Y_0,Y_1) denote potential outcomes in high school and college sectors, respectively. Let S=0 denote high school sector, and S=1 denote college sector. Define observed earnings as $Y=SY_1+(1-S)Y_0$. Finally, let f(y) denote the density function of observed earnings. Here we plot the density functions f generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.1
Densities of fitted and actual present value of earnings
from age 19 to 28 for people who choose to graduate high school



Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

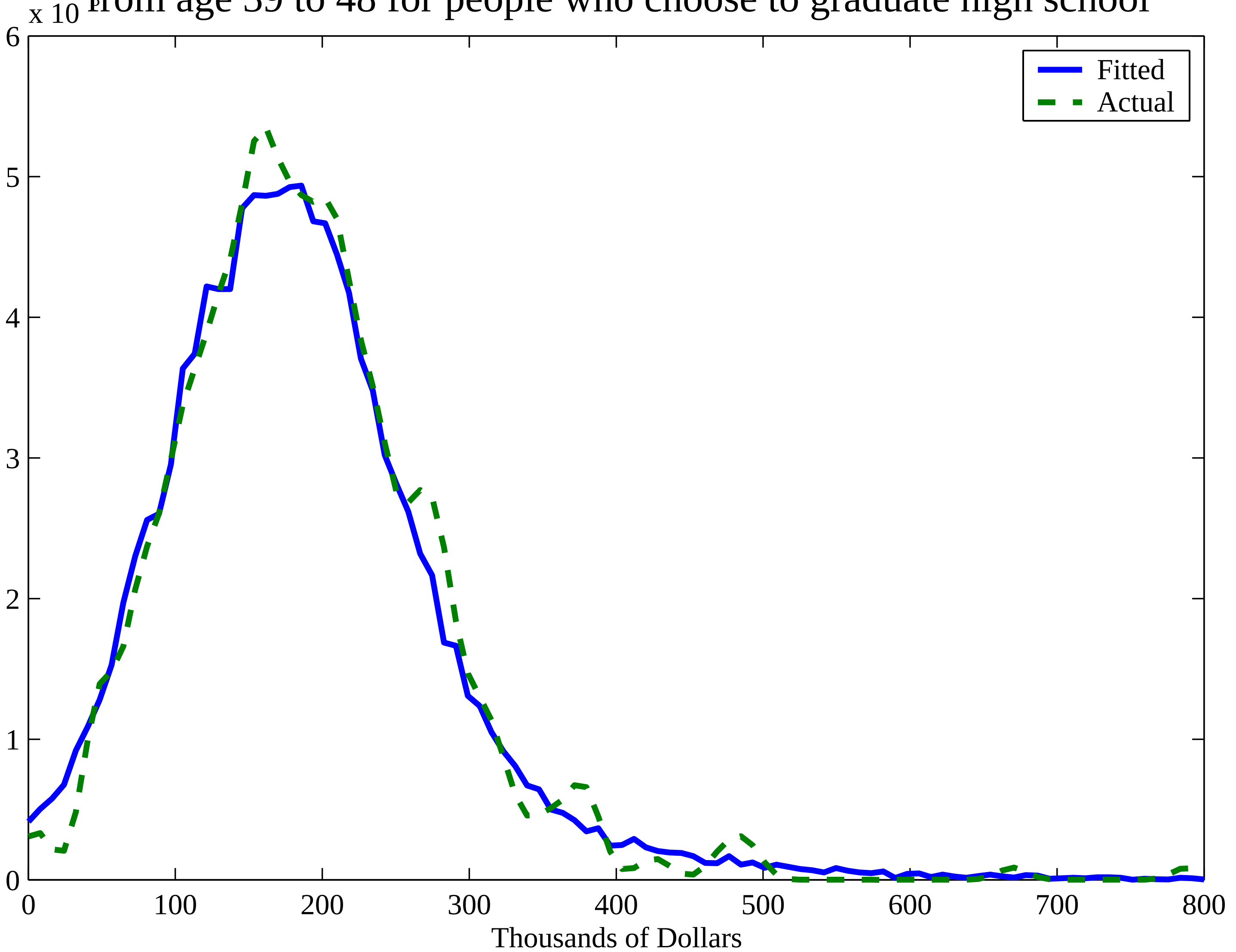
Figure 2.2
Densities of fitted and actual present value of earnings
from age 29 to 38 for people who choose to graduate high school



Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

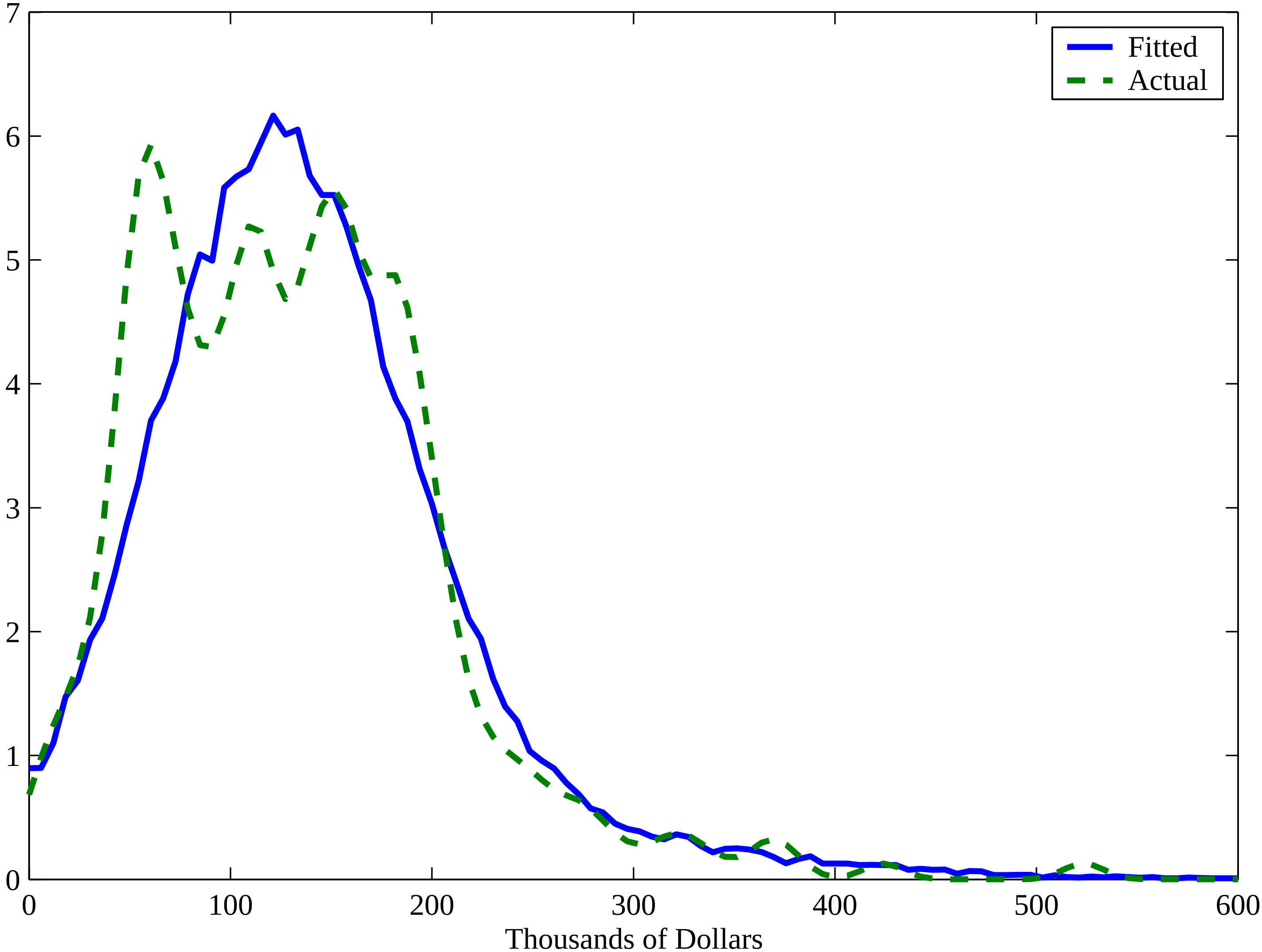
Figure 2.3

Densities of fitted and actual present value of earnings from age 39 to 48 for people who choose to graduate high school



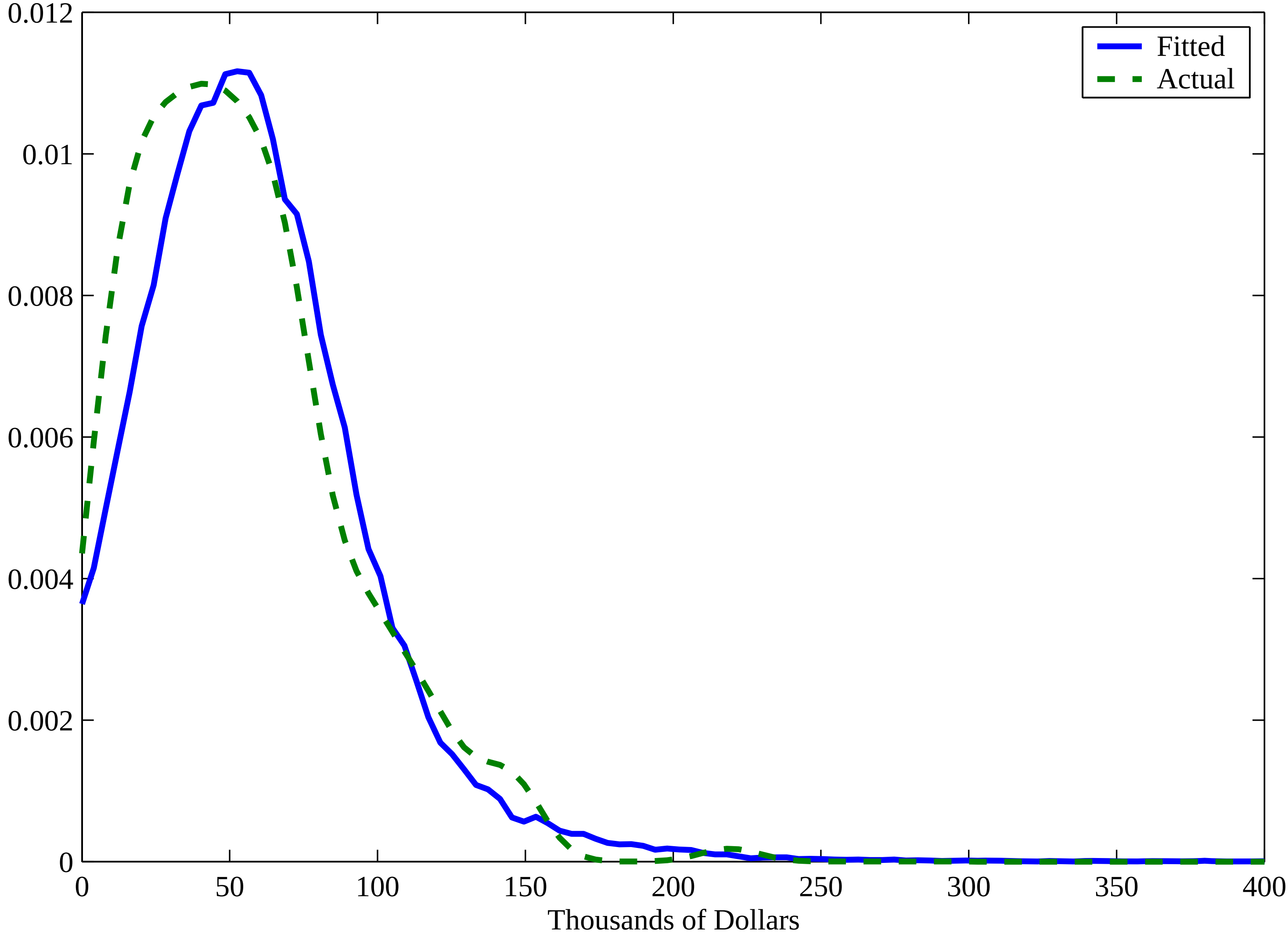
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.4
Densities of fitted and actual present value of earnings
from age 49 to 58 for people who choose to graduate high school



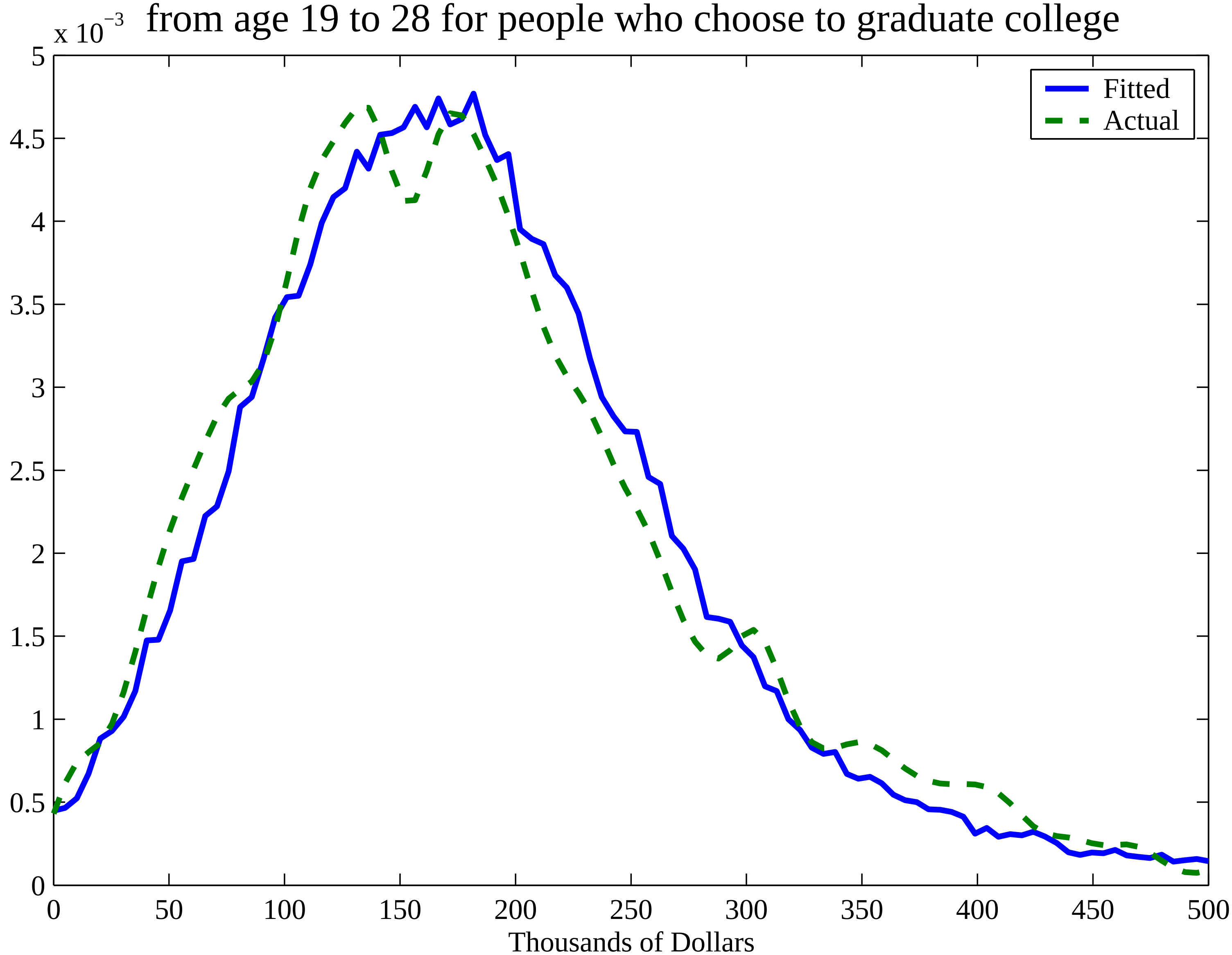
Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 2.5
Densities of fitted and actual present value of earnings
from age 59 to 65 for people who choose to graduate high school



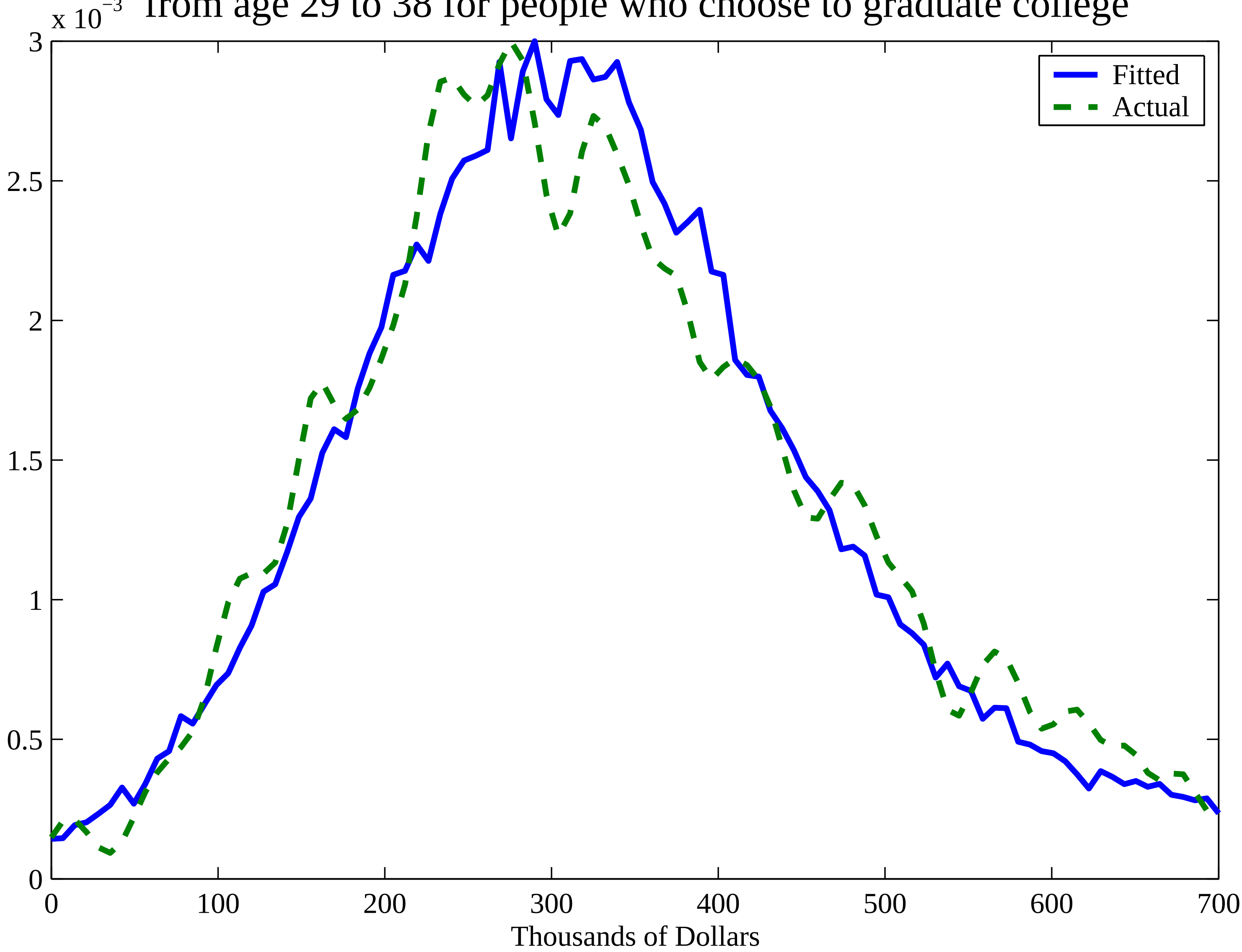
Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. Earnings here are Y_0. Here we plot the density functions $f(y_0|S=0)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.1
Densities of fitted and actual present value of earnings from age 19 to 28 for people who choose to graduate college



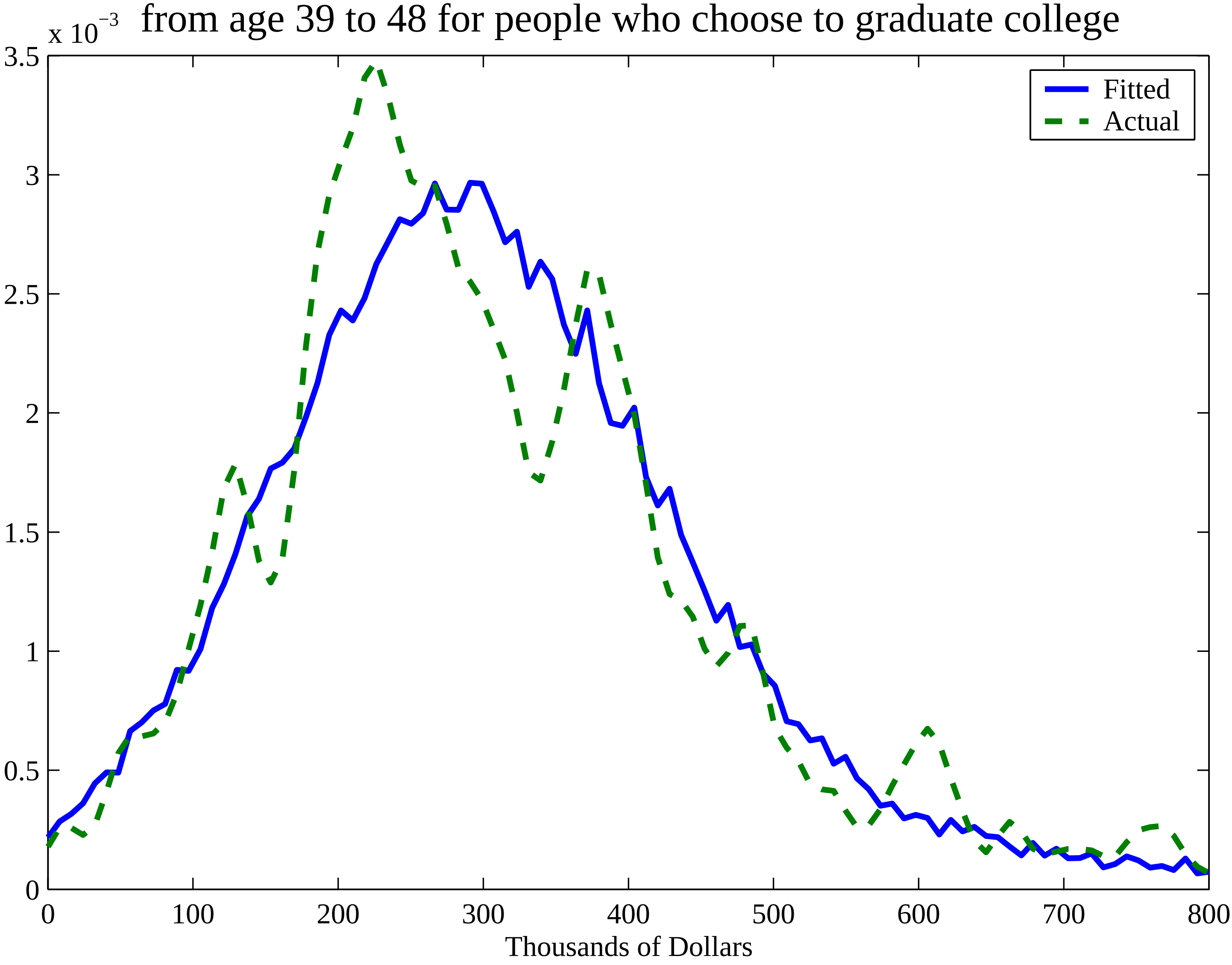
Present value of earnings from age 19 to 28 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.2
Densities of fitted and actual present value of earnings
from age 29 to 38 for people who choose to graduate college



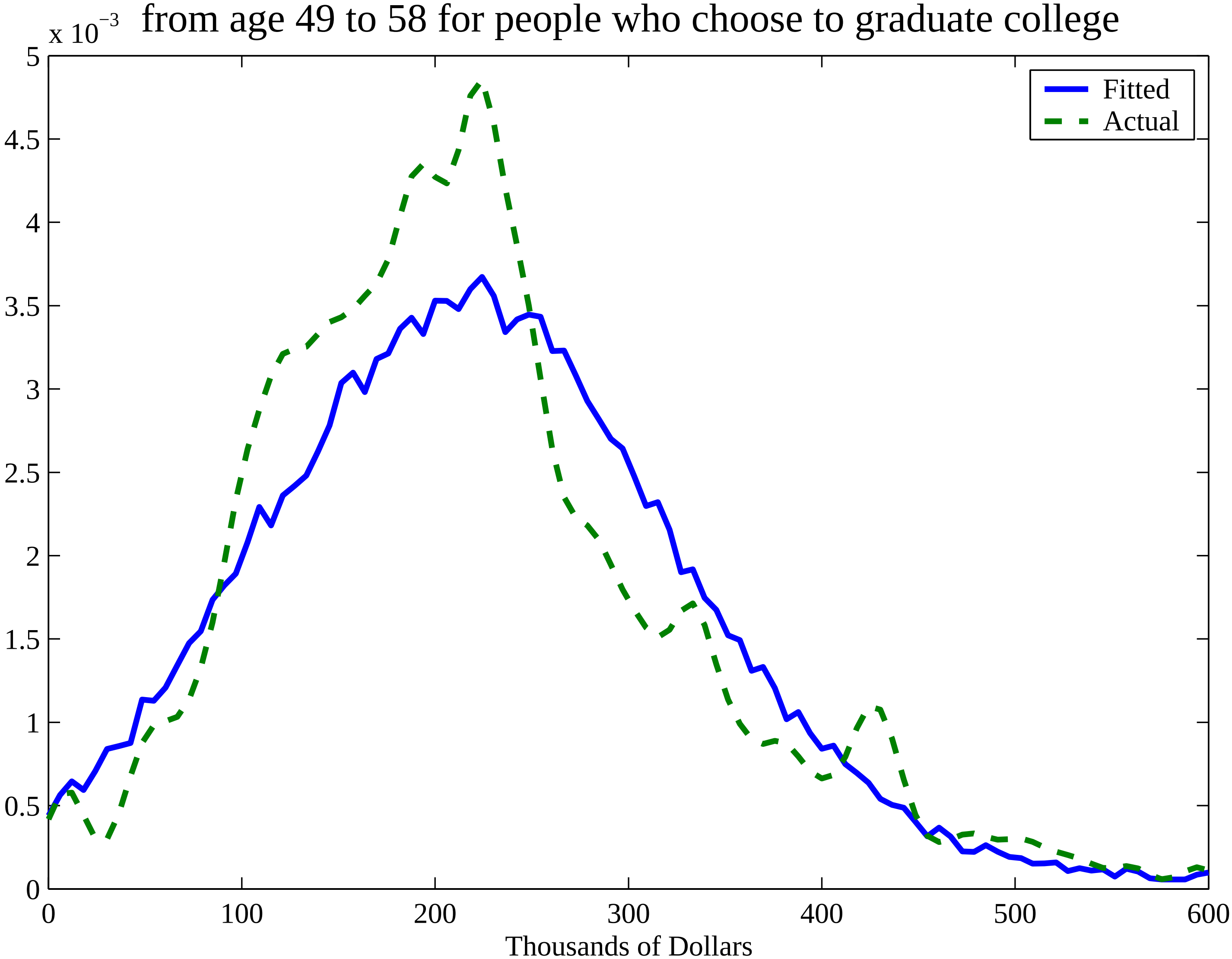
Present value of earnings from age 29 to 38 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.3
Densities of fitted and actual present value of earnings from age 39 to 48 for people who choose to graduate college



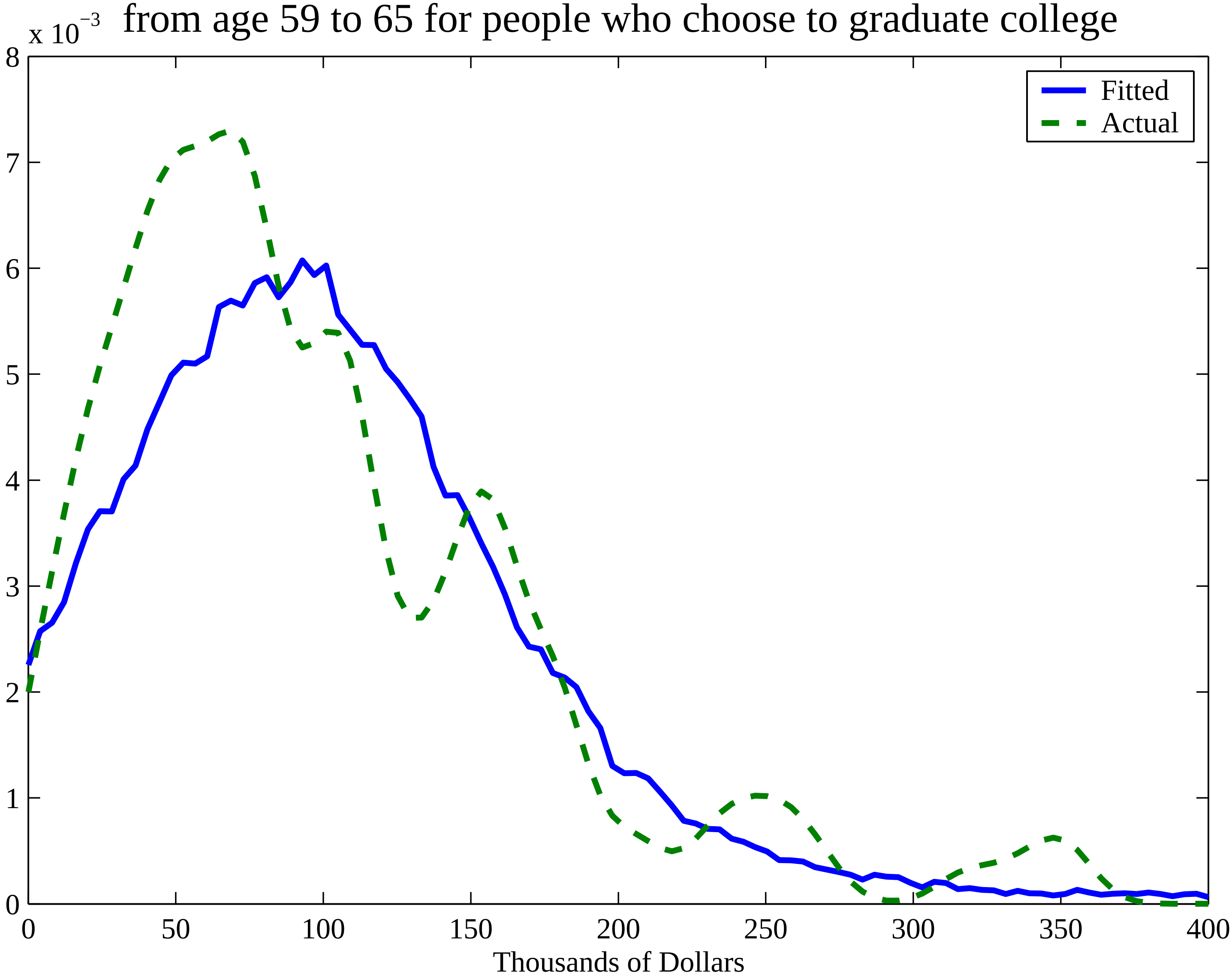
Present value of earnings from age 39 to 48 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 3.4
Densities of fitted and actual present value of earnings
from age 49 to 58 for people who choose to graduate college



Present value of earnings from age 49 to 58 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

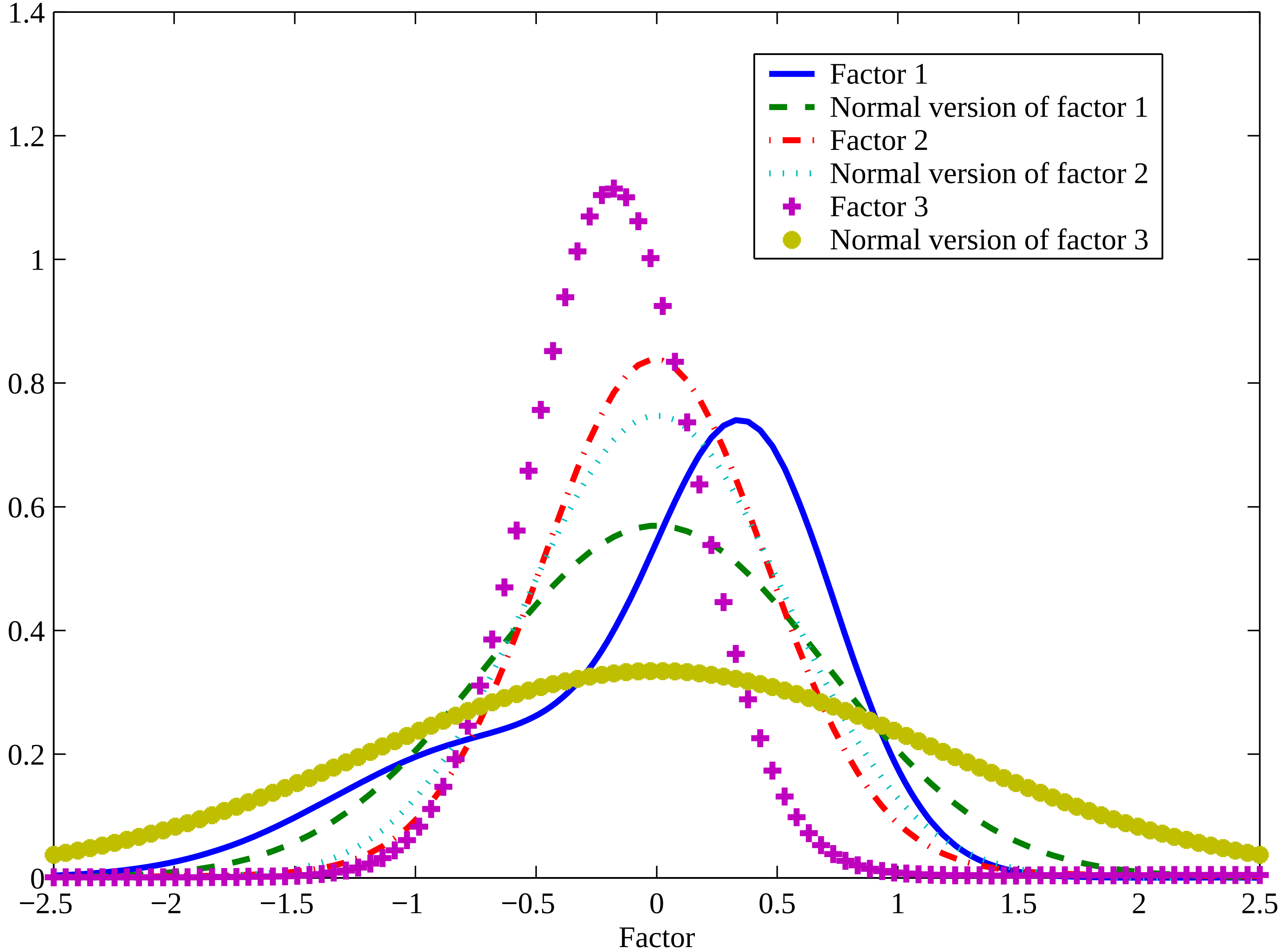
Figure 3.5
Densities of fitted and actual present value of earnings from age 59 to 65 for people who choose to graduate college



Present value of earnings from age 59 to 65 discounted using an interest rate of 3%. This plot is for Y_1. Here we plot the density functions $f(y_1|S=1)$ generated from the data (the solid curve), against that predicted by the model (the dashed line). We use kernel density estimation to smooth these functions.

Figure 4

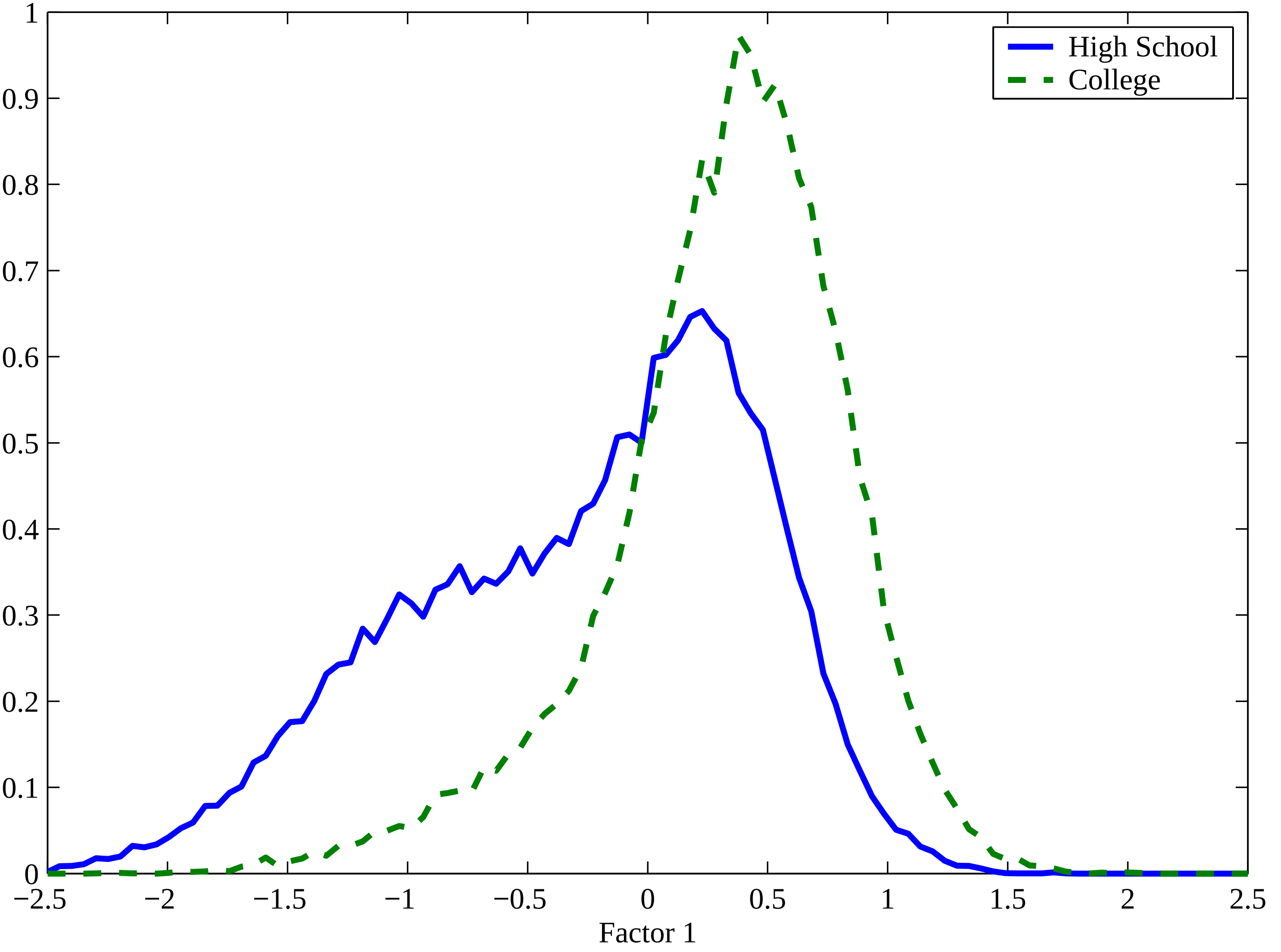
Densities of estimated factors and their normal equivalents



Let $f(\theta_1)$ denote the probability density function of factor θ_1. We allow $f(\theta_1)$ to be a mixture of normals. Assume $\mu_1=E(\theta_1)$ and $\sigma_1=Var(\theta_1)$. Let $\phi(\mu_1,\sigma_1)$ denote the density of a normal random variable with mean μ_1 and variance σ_1. The solid curve is the actual density of factor θ_1, $f(\theta_1)$, while the dashed curve is the density of a normal random variable with mean μ_1 and variance σ_1. We proceed similarly for factors 2 and 3.

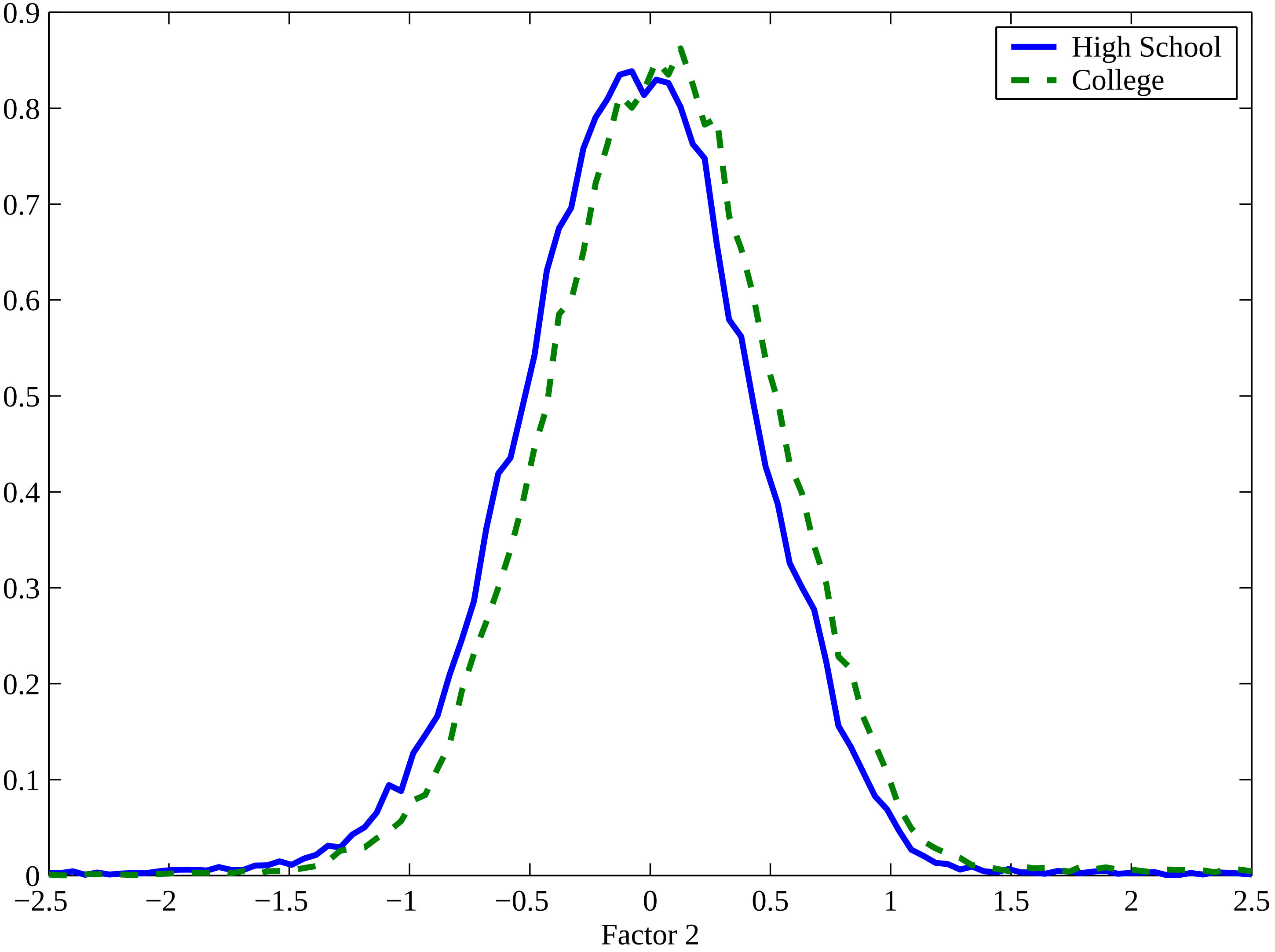
Figure 5.1

Densities of "ability" (factor 1) by schooling level



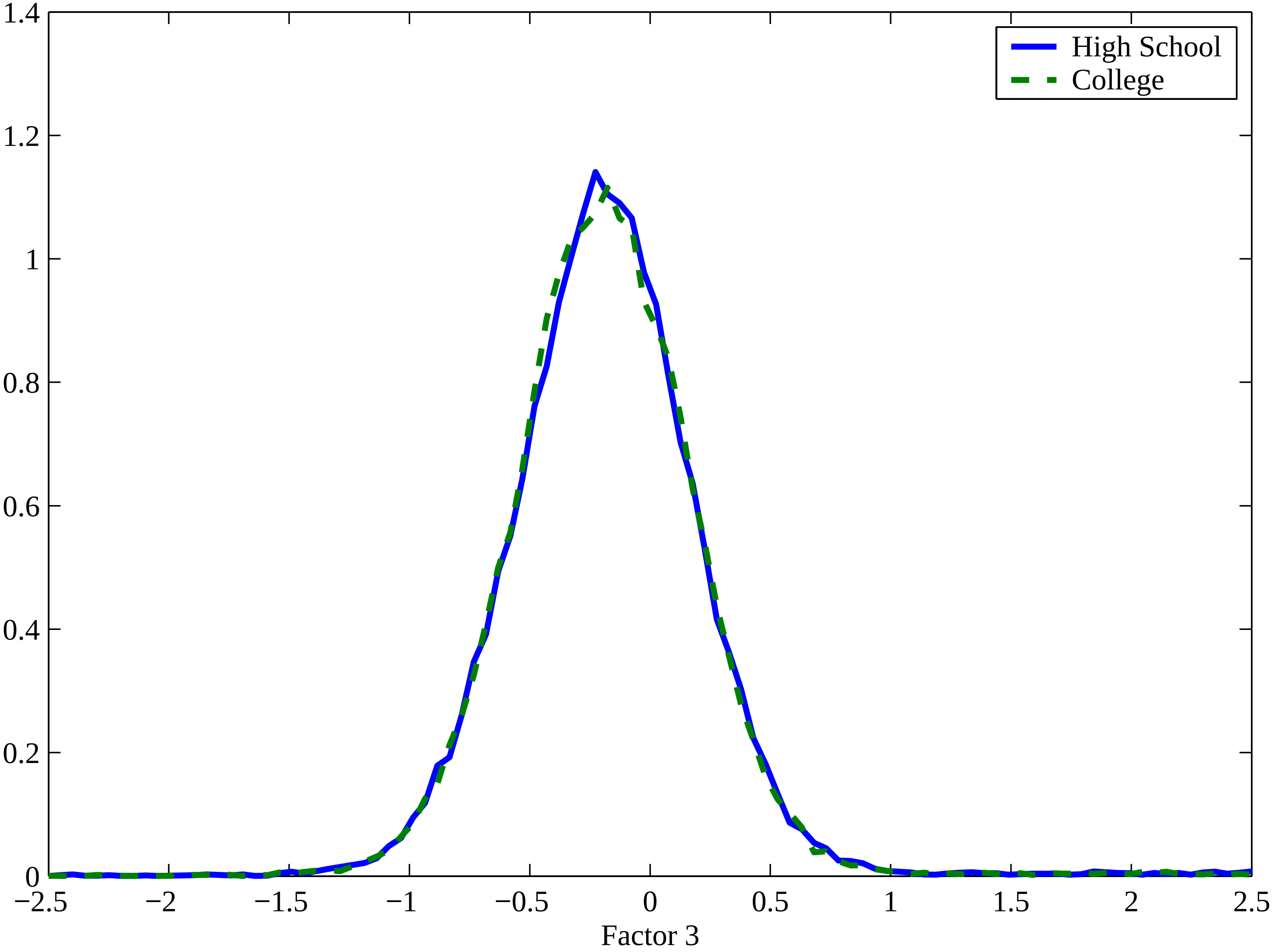
Let $f(\theta_1)$ denote the probability density function of factor θ_1. We allow $f(\theta_1)$ to be a mixture of normals. The solid line plots the density of factor 1 conditional on choosing the high school sector, that is, $f(\theta_1|\text{choice=high school})$. The dashed line plots the density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

Figure 5.2
Densities of factor 2 by schooling level



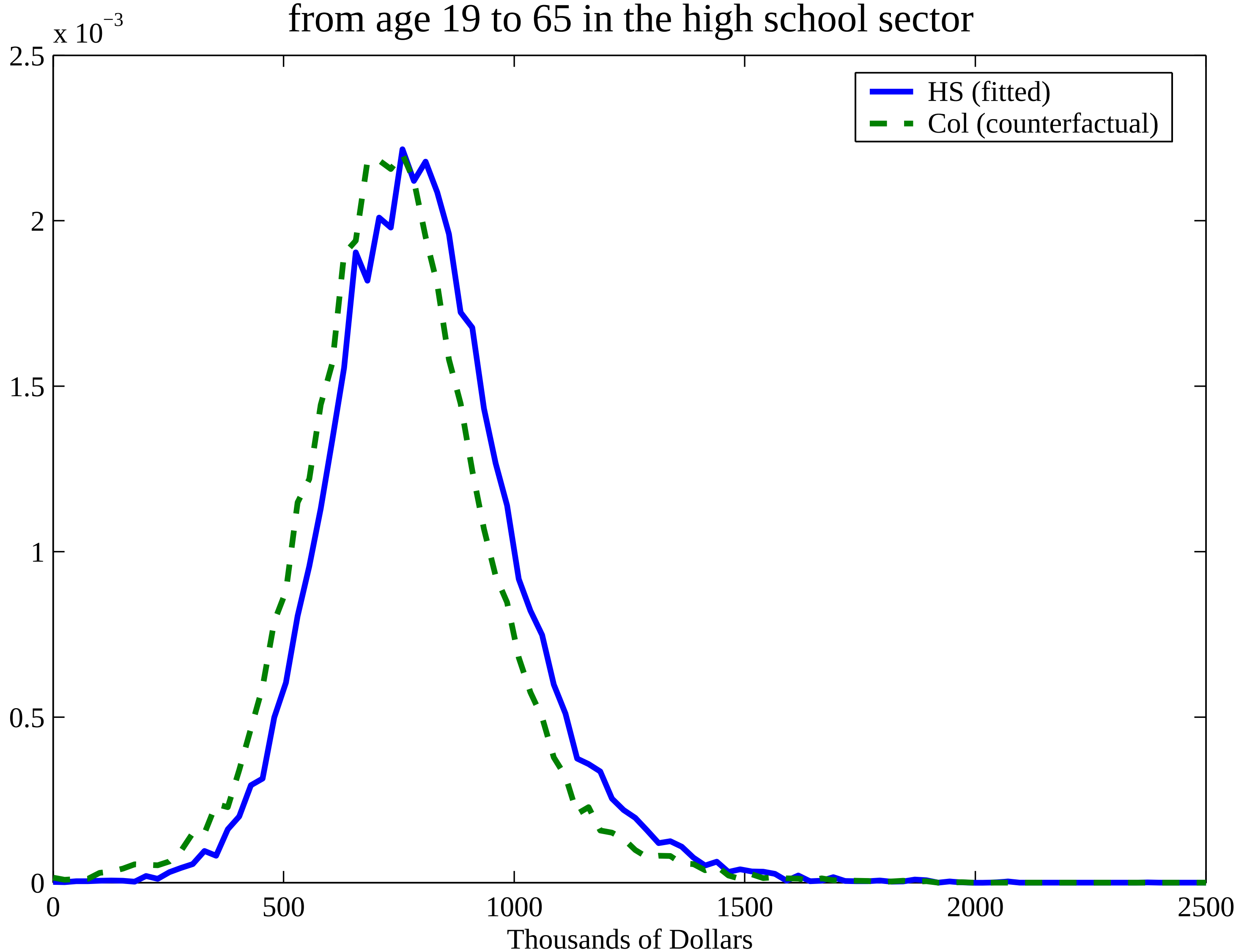
Let f(θ_2) denote the probability density function of factor θ_2. We allow f(θ_2) to be a mixture of normals. The solid line plots the density of factor 2 conditional on choosing the high school sector, that is, f(θ_2|choice=high school). The dashed line plots the density of factor 2 conditional on choosing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

Figure 5.3
Densities of factor 3 by schooling level



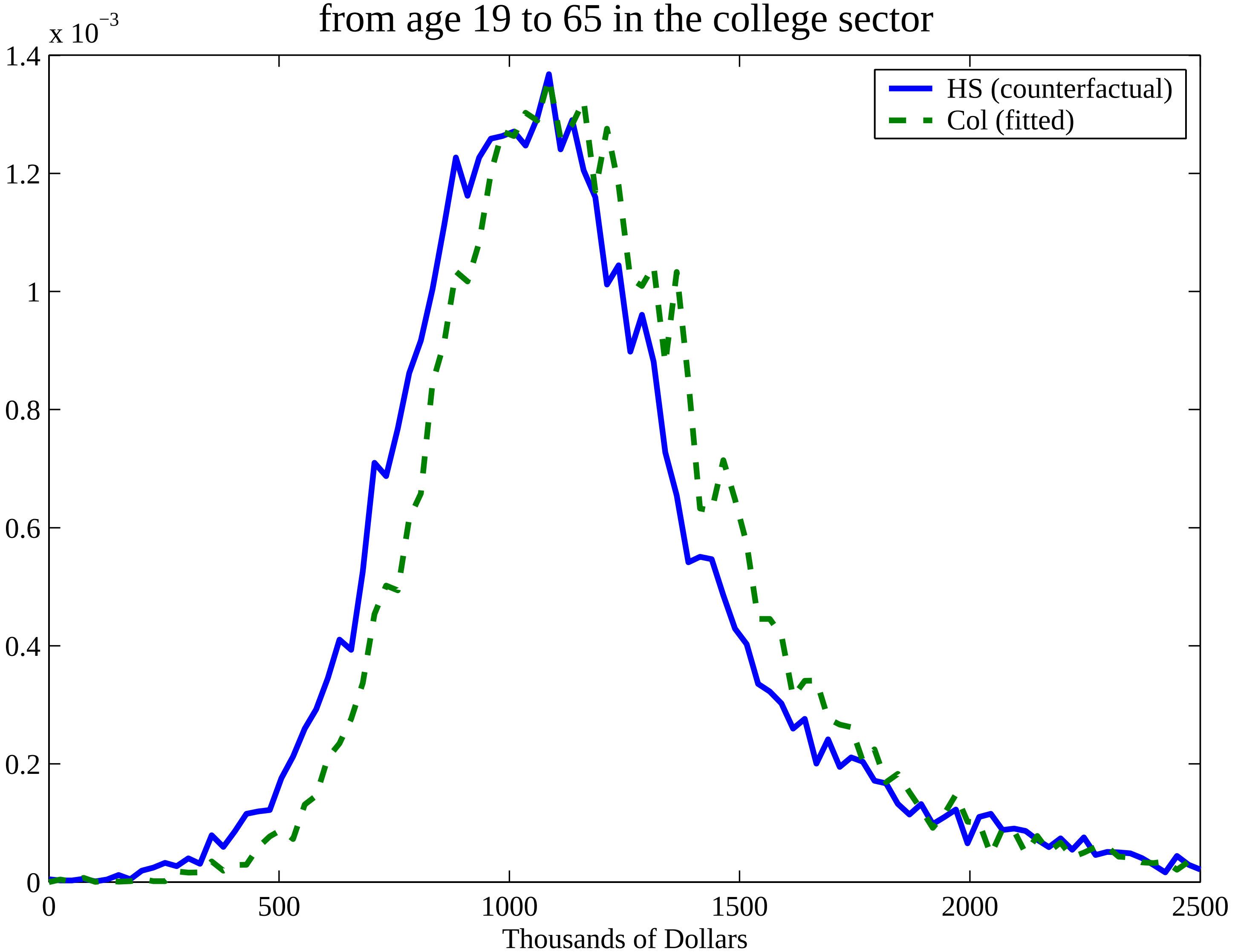
Let $f(\theta_3)$ denote the probability density function of factor θ_3. We allow $f(\theta_3)$ to be a mixture of normals. The solid line plots the density of factor 3 conditional on choosing the high school sector, that is, $f(\theta_3|\text{choice=high school})$. The dashed line plots the density of factor 3 conditional on choosing the college sector, that is, $f(\theta_3|\text{choice=college})$.

Figure 6.1
Densities of ex post present value of counterfactual and fitted earnings from age 19 to 65 in the high school sector



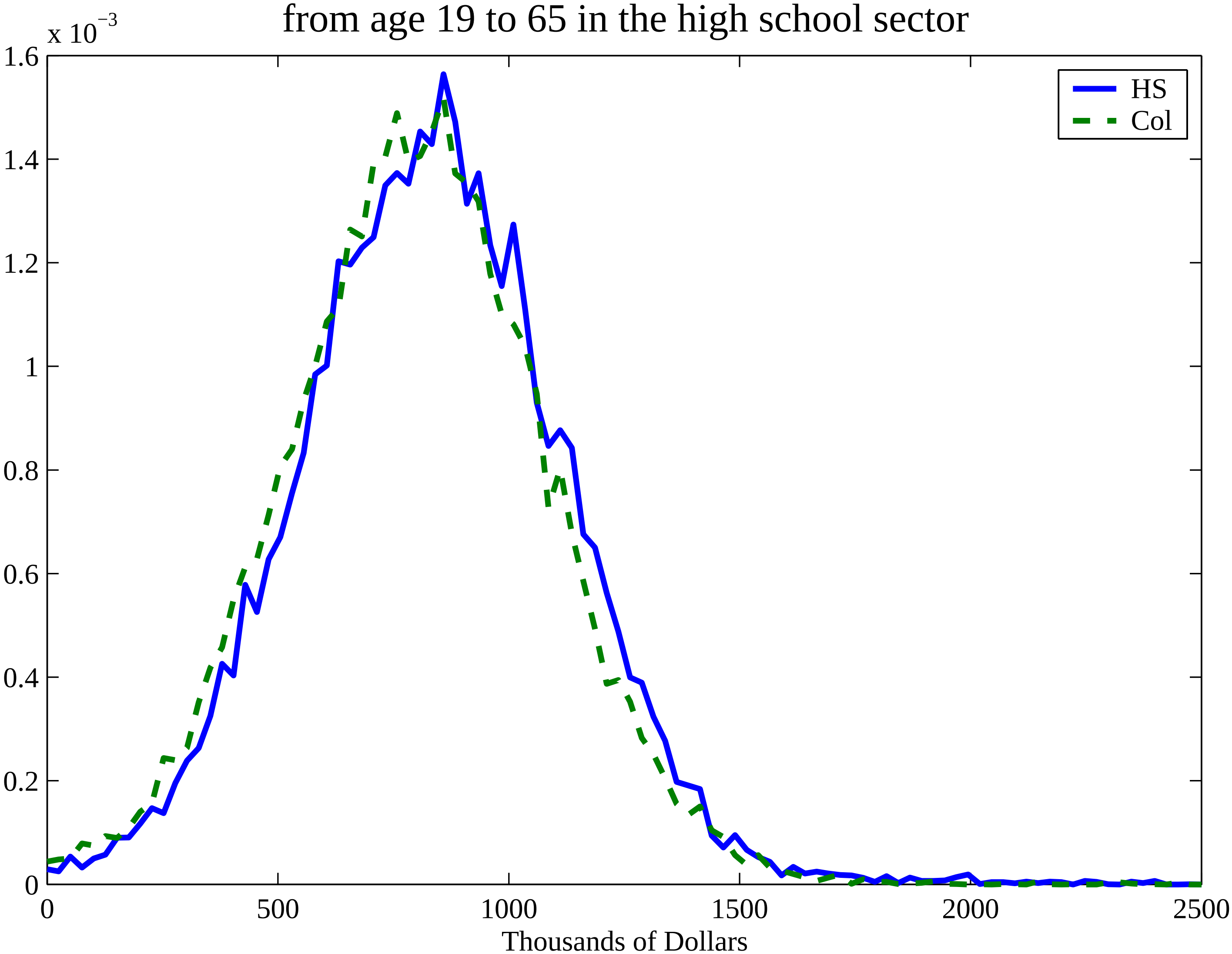
Let Y_0 denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(Y_0)$ denote its density function. The dashed line plots the predicted Y_0 density conditional on choosing high school, that is, $f(Y_0|S=0)$, while the dashed line shows the counterfactual density function of Y_0 for those agents who are actually college graduates, that is, $f(Y_0|S=1)$. This assumes that the agent chooses schooling without knowing θ_3 and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$

Figure 6.2
Densities of ex post present value of counterfactual and fitted earnings from age 19 to 65 in the college sector



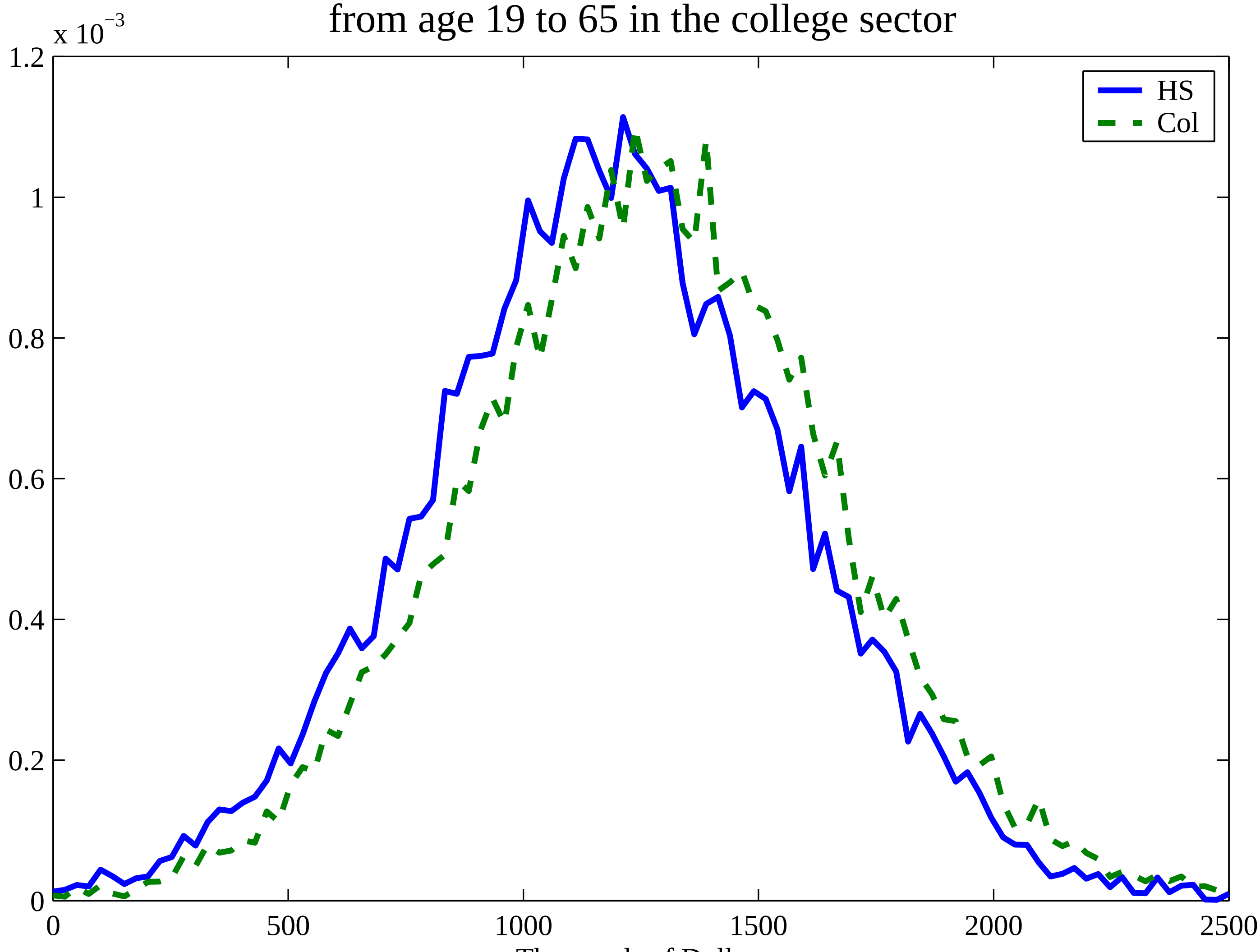
Let Y_1 denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The dashed line plots the predicted Y_1 density conditional on choosing college, that is, $f(Y_1|S=1)$, while the solid line shows the counterfactual density function of Y_1 for those agents who are actually high school graduates, that is, $f(Y_1|S=0)$. This assumes that the agent chooses schooling without knowing θ_3 and $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$

Figure 6.3
Densities of ex ante present value of counterfactual and fitted earnings from age 19 to 65 in the high school sector



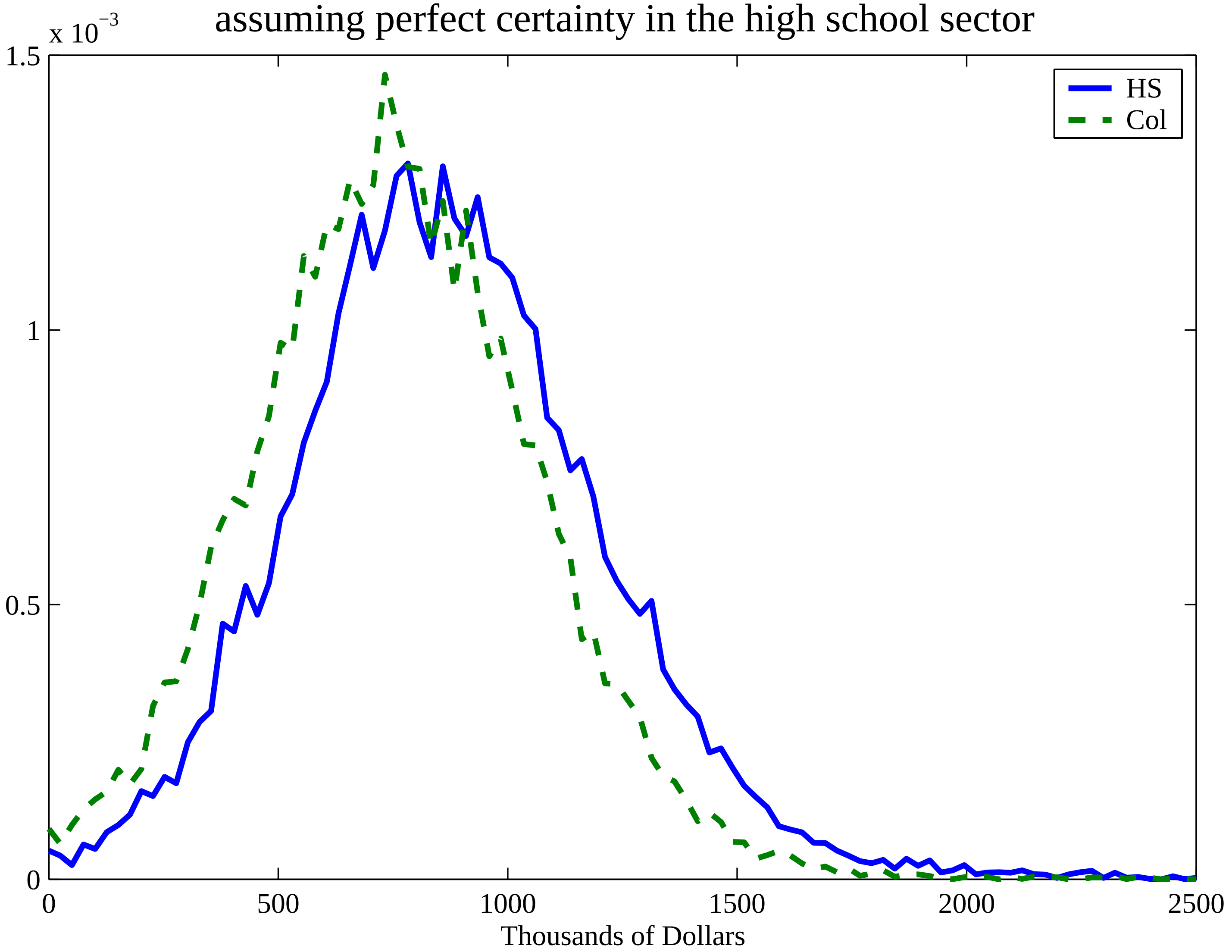
Let $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$. Let $E_{\theta_3,\varepsilon}(Y_0)$ denote the ex ante present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(E_{\theta_3,\varepsilon}(Y_0))$ denote its density function. The solid curve plots the predicted Y_0 density conditional on choosing high school, that is, $f(E_{\theta_3,\varepsilon}(Y_0)|S=0)$, while the dashed line shows the counterfactual density function of $E_{\theta_3,\varepsilon}(Y_0)$ for those agents who are actually college graduates, that is, $f(E_{\theta_3,\varepsilon}(Y_0)|S=1)$. This assumes that the agent chooses schooling without knowing θ_3 and ε.

Figure 6.4
Densities of ex ante present value of counterfactual and fitted earnings from age 19 to 65 in the college sector



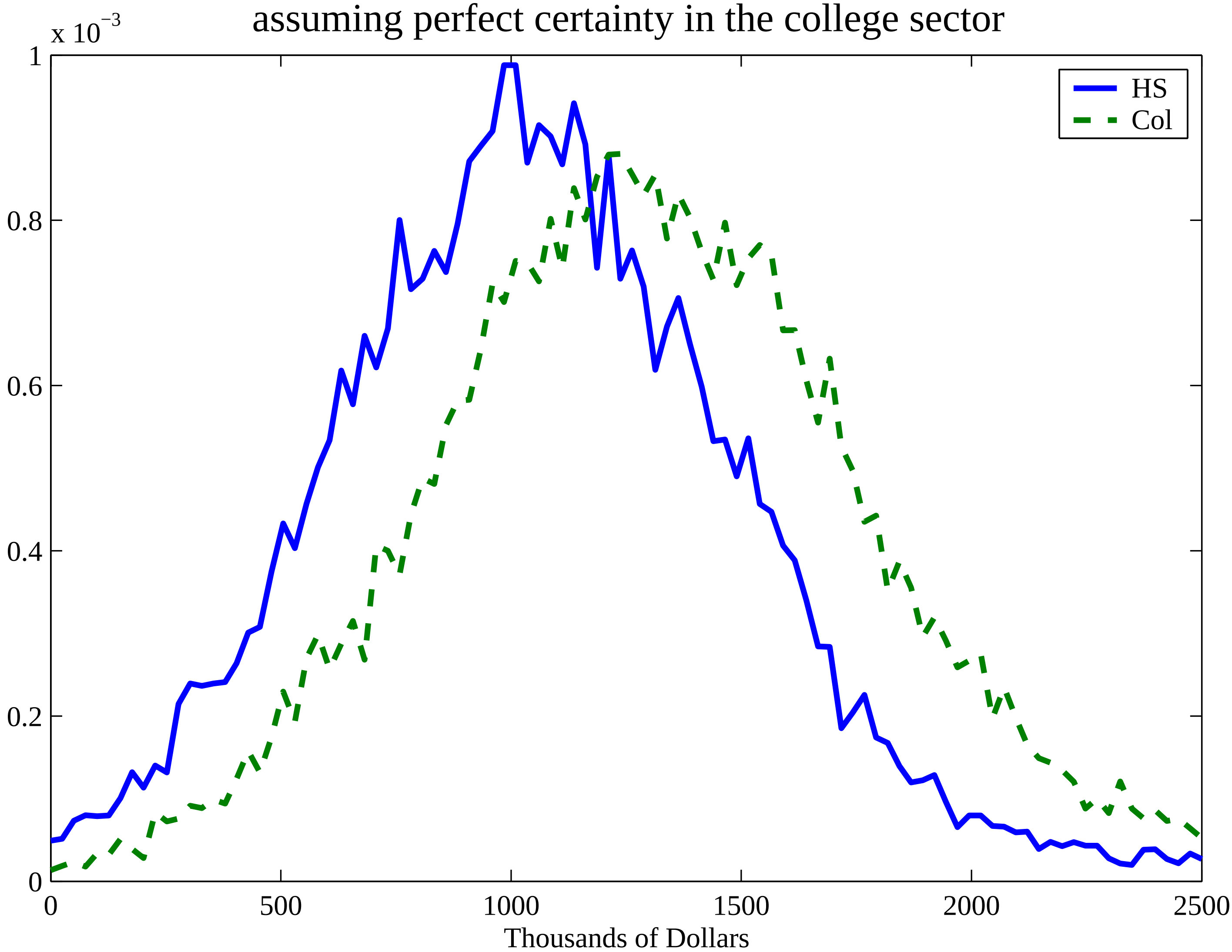
Let $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$. Let $E_{\theta_3,\varepsilon}(Y_1)$ denote the ex ante present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(E_{\theta_3,\varepsilon}(Y_1))$ denote its density function. The solid curve plots the counterfactual Y_1 density conditional on choosing high school, that is, $f(E_{\theta_3,\varepsilon}(Y_1)|S=0)$, while the dashed line shows the predicted density function of $E_{\theta_3,\varepsilon}(Y_1)$ for those agents who are actually college graduates, that is, $f(E_{\theta_3,\varepsilon}(Y_1)|S=1)$. This assumes that the agent chooses schooling without knowing θ_3 and ε.

Figure 6.5
Densities of present value of counterfactual and fitted earnings from age 19 to 65 assuming perfect certainty in the high school sector



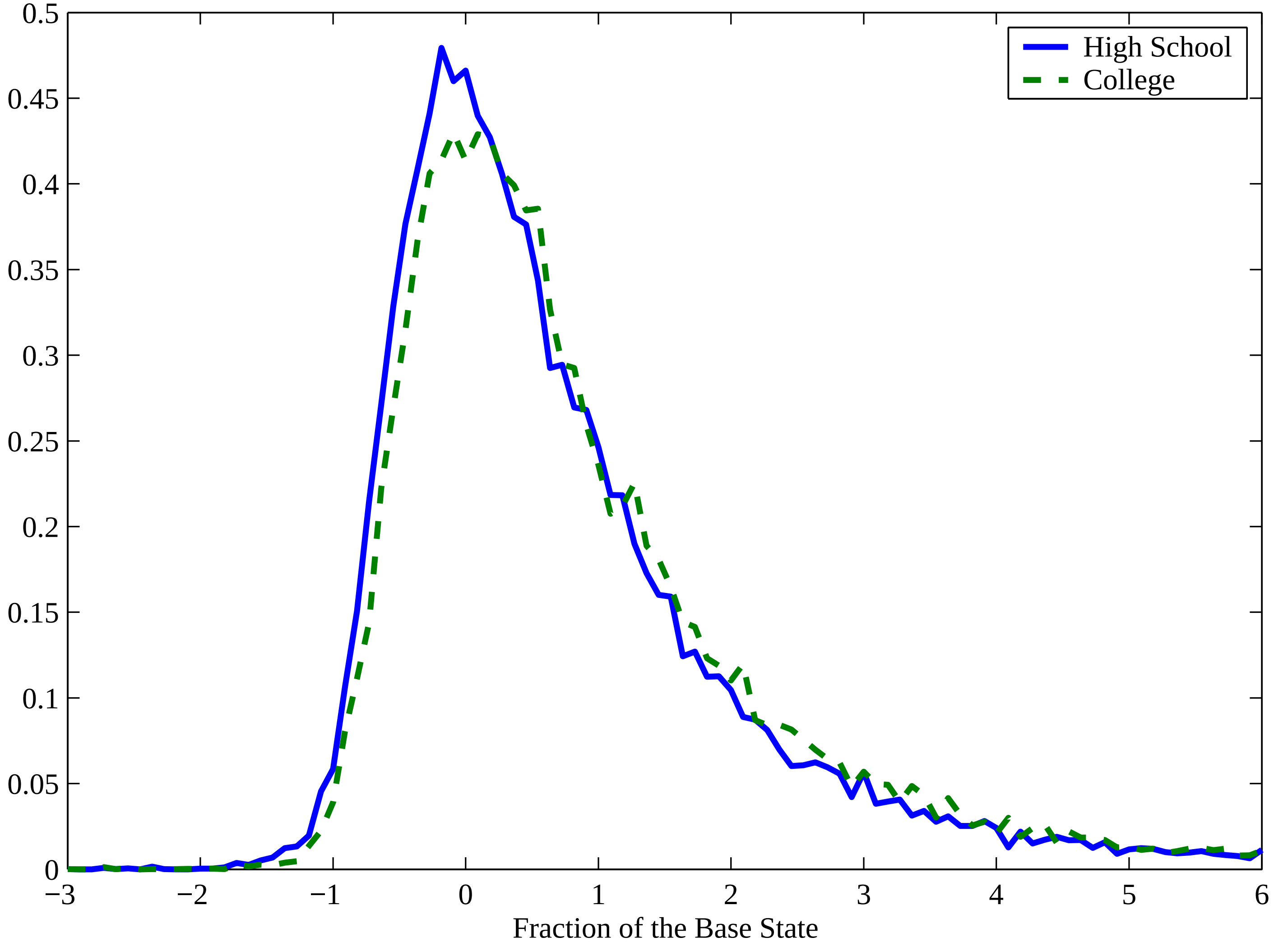
Let Y_0 denote the present value of earnings from age 19 to 65 in the high school sector (discounted at a 3% interest rate). Let $f(Y_0)$ denote its density function. The solid curve plots the predicted Y_0 density conditional on choosing high school, that is, $f(Y_0|S=0)$, while the dashed line shows the counterfactual density function of Y_0 for those agents who are actually college graduates, that is, $f(Y_0|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.

Figure 6.6
Densities of present value of counterfactual and fitted earnings from age 19 to 65 assuming perfect certainty in the college sector



Let Y_1 denote the present value of earnings from age 19 to 65 in the college sector (discounted at a 3% interest rate). Let $f(Y_1)$ denote its density function. The solid curve plots the counterfactual Y_1 density conditional on choosing high school, that is, $f(Y_1|S=0)$, while the dashed line shows the predicted density function of Y_1 for those agents who are actually college graduates, that is, $f(Y_1|S=1)$. This assumes that the agent chooses schooling with complete knowledge of future earnings.

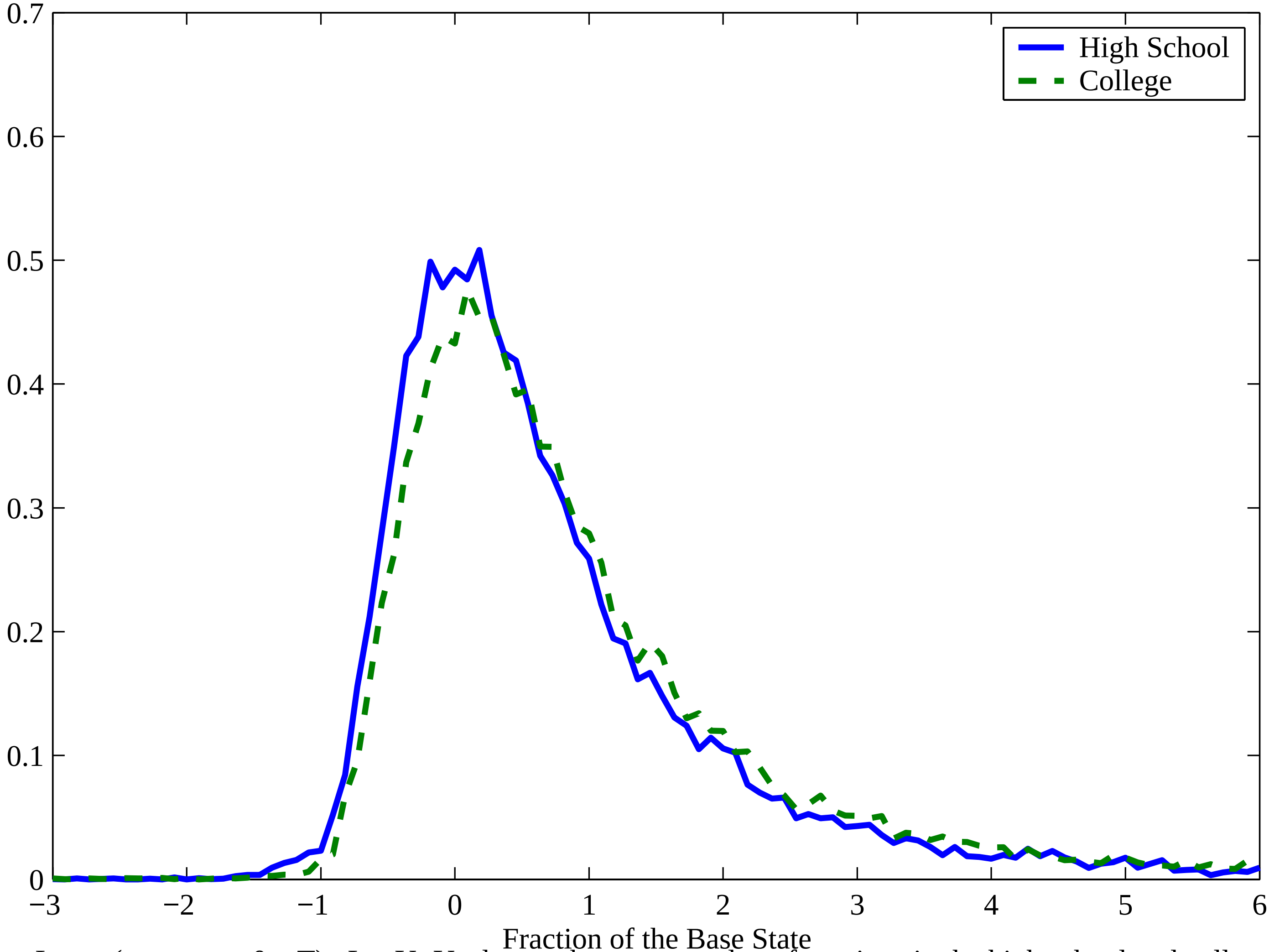
Figure 7.1
Densities of ex post returns to college by level of schooling chosen



Let Y_0, Y_1 denote the present value of earnings 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 colege 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 $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$

Figure 7.2

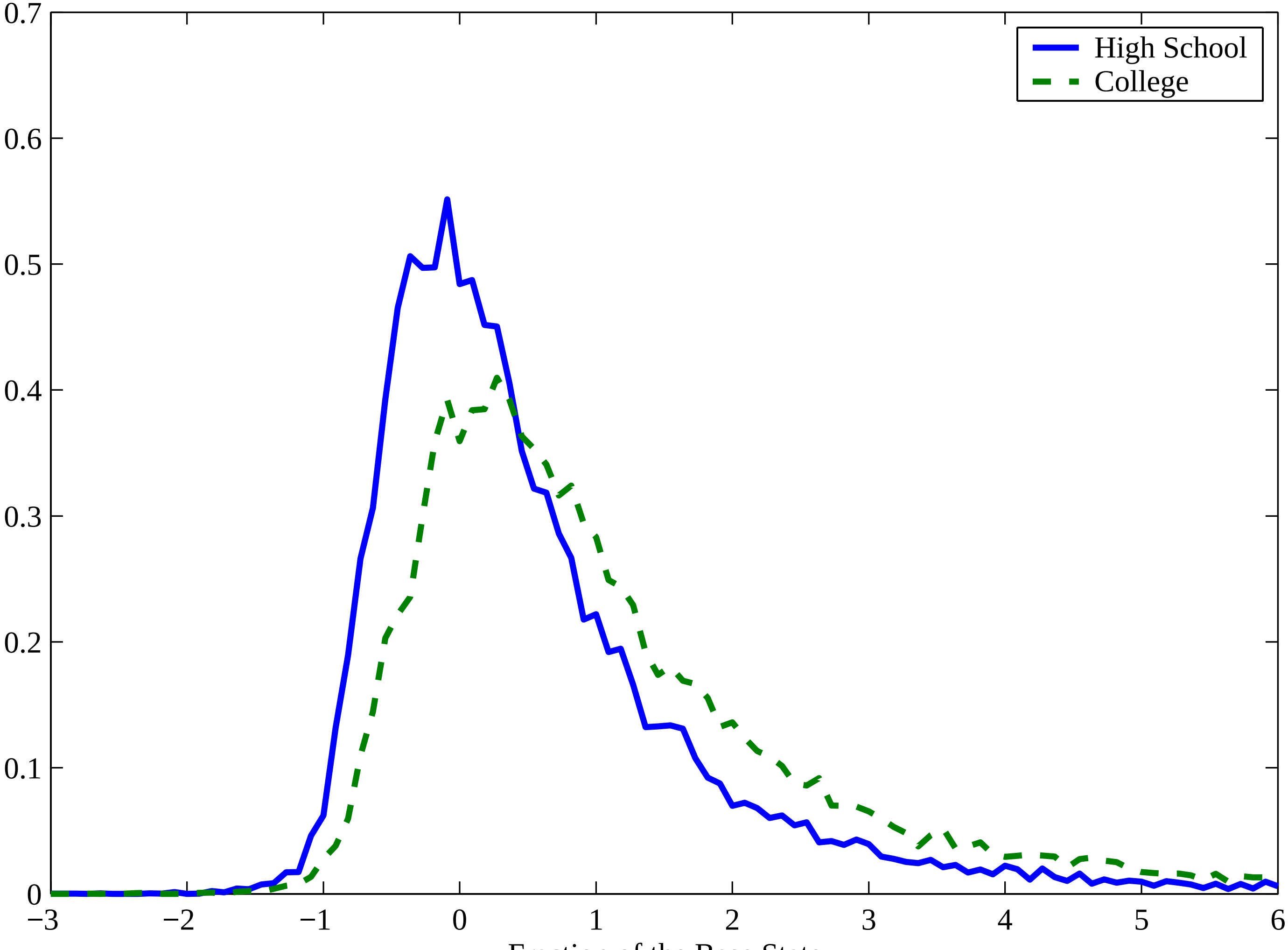
Densities of ex ante returns to college by level of schooling chosen



Let $\varepsilon=(\varepsilon_{0,t}, \varepsilon_{1,t}, t=0,...T)$. Let Y_0,Y_1 denote the present value of earnings in the high school and college sectors, respectively. Define ex ante returns to college as the ratio $E_{\theta_3,\varepsilon}(R)=E_{\theta_3,\varepsilon}((Y_1-Y_0)/Y_0)$. Let f(r) denote the density function of the random variable $E_{\theta_3,\varepsilon}(R)$. 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 ε.

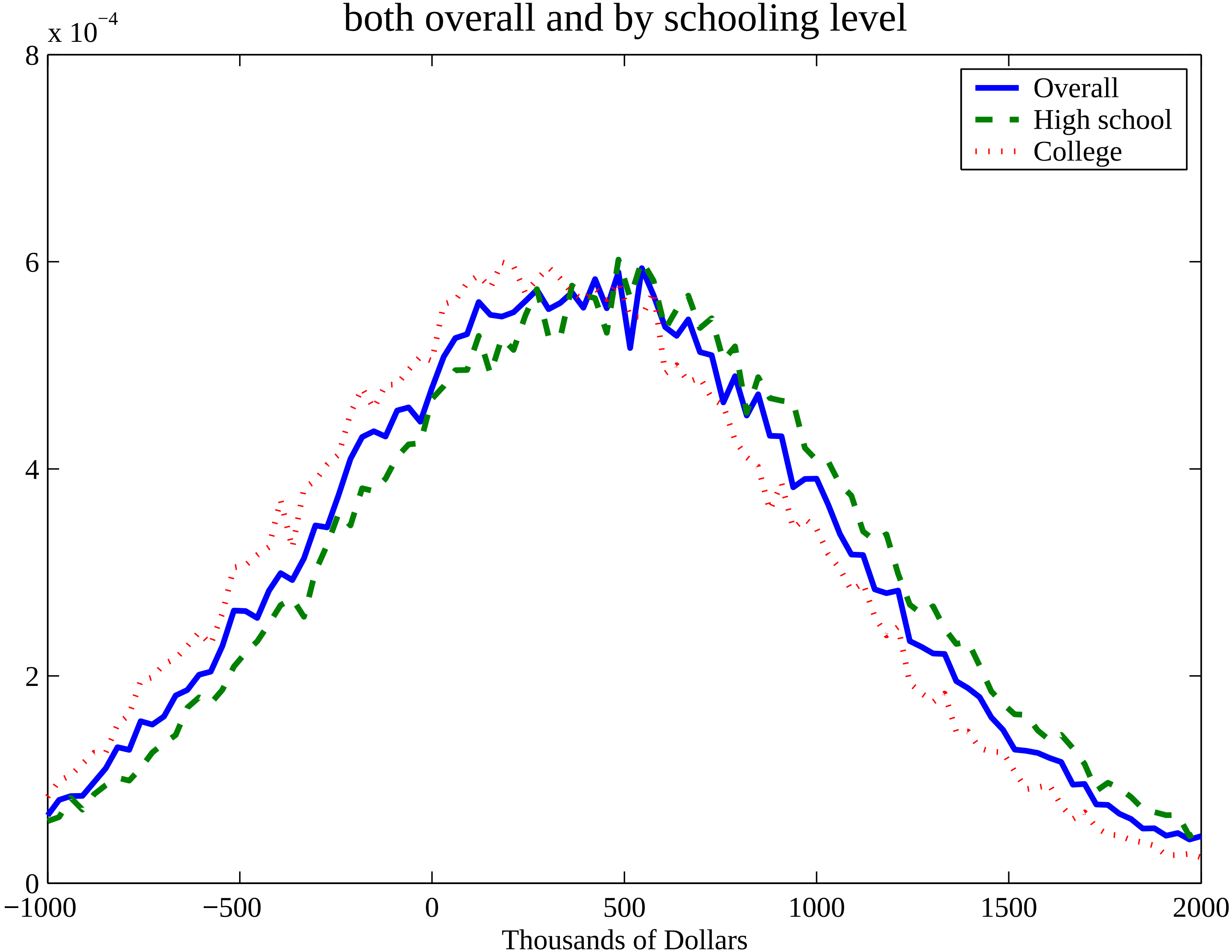
Figure 7.3

Densities of returns to college by schooling level chosen assuming perfect certainty



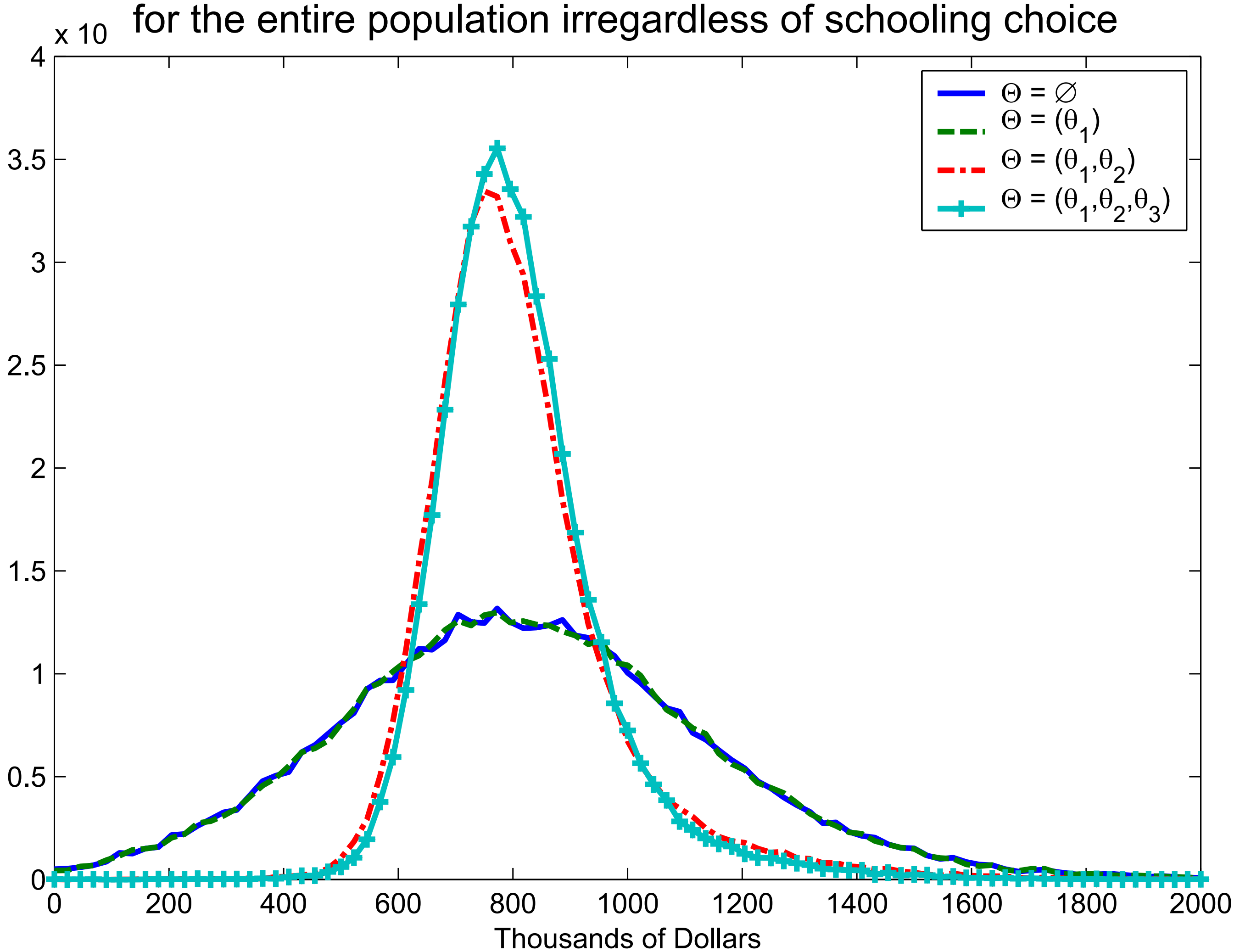
Let Y_0,Y_1 denote the present value of earnings in the high school and college sectors, respectively (discounted at a 3% interest rate). Define returns to college as the ratio $R=(Y_1-Y_0)/Y_0$. Let f(r) denote the density function of the random variable R. The solid line is the density of returns to college for high school graduates, that is f(r|S=0). The dashed line is the density of returns to college for college graduates, that is, f(r|S=1). This assumes that the agent chooses schooling with complete knowledge of future earnings.

Figure 8
Densities of monetary value of psychic cost
both overall and by schooling level



Let C denote the monetary value of psychic costs. Let f(c) denote the density function of psychic costs in monetary terms. The dashed line shows the density of psychic costs for high school graduates, that is f(c|S=0). The dotted line shows the density of psychic costs for college graduates, that is, f(c|S=1). The solid line is the unconditional density of the monetary value of psychic costs, f(c).

Figure 9.1
Densities of present value of high school earnings under different information sets for the agent calculated for the entire population irregardless of schooling choice



Let Θ denote the agent's information set. Let Y_0 denote the present value of earnings in the high school sector (discounted at a 3% interest rate). Let $f(y_0|\Theta)$ denote the density of the present value of earnings in high school conditioned on the information set Θ. Then:

The solid line plots $f(y_0|\Theta)$ under no information, i.e. $\Theta=\varnothing$.

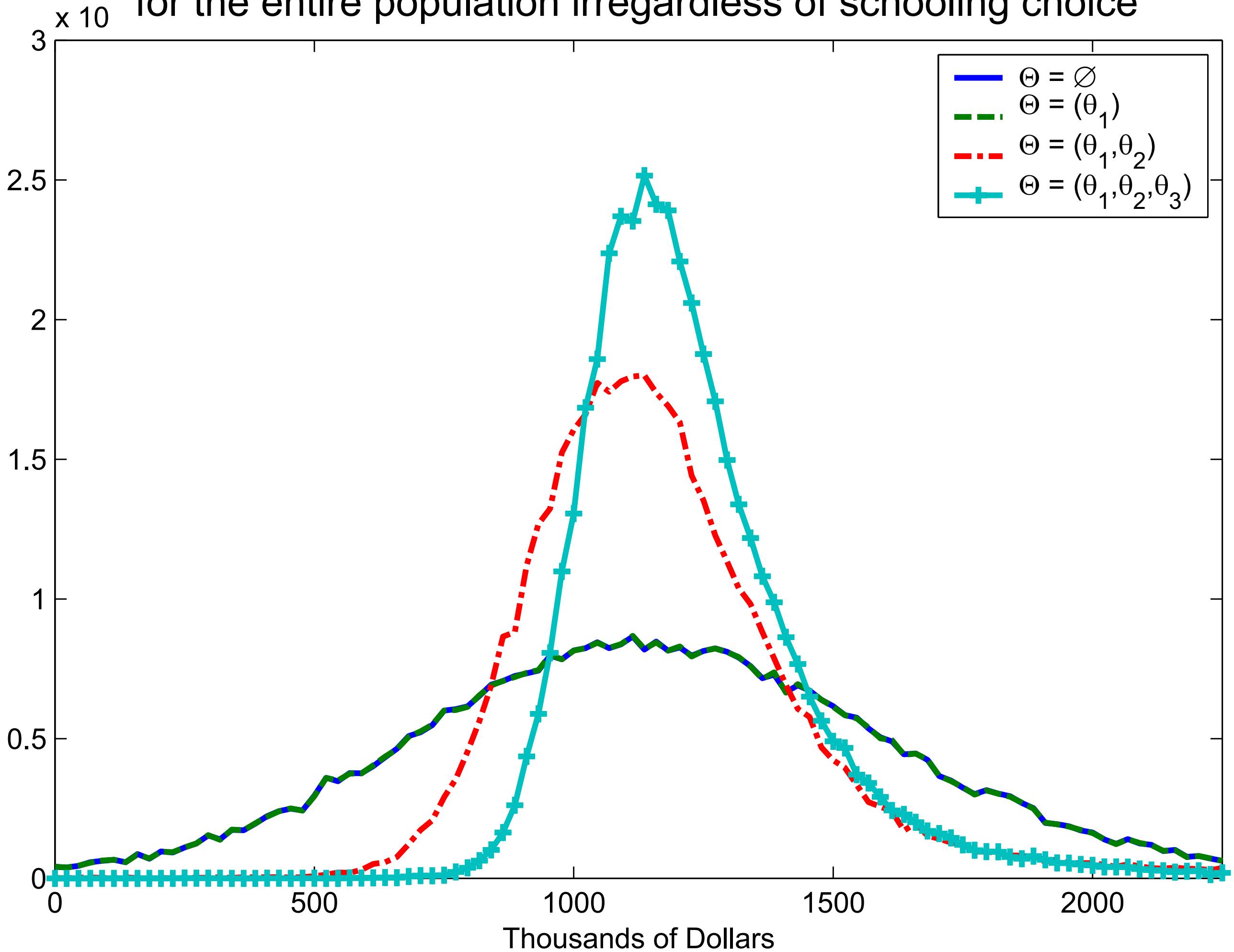
The dashed line plots $f(y_0|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.

The dashed-dotted line plots $f(y_0|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.

The crossed line plots $f(y_0|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.

The X are put at the mean and are assumed to be known. The θ, when known, are set at their mean of zero.

Figure 9.2
Densities of present value of college earnings
under different information sets for the agent calculated
for the entire population irregardless of schooling choice



Let Θ denote the agent's information set. Let Y_1 denote the present value of earnings in the college sector (discounted at a 3% interest rate). Let $f(y_1|\Theta)$ denote the density of the present value of earnings in high school conditioned on the information set Θ. Then:

The solid line plots $f(y_1|\Theta)$ under no information, i.e. $\Theta=\varnothing$.

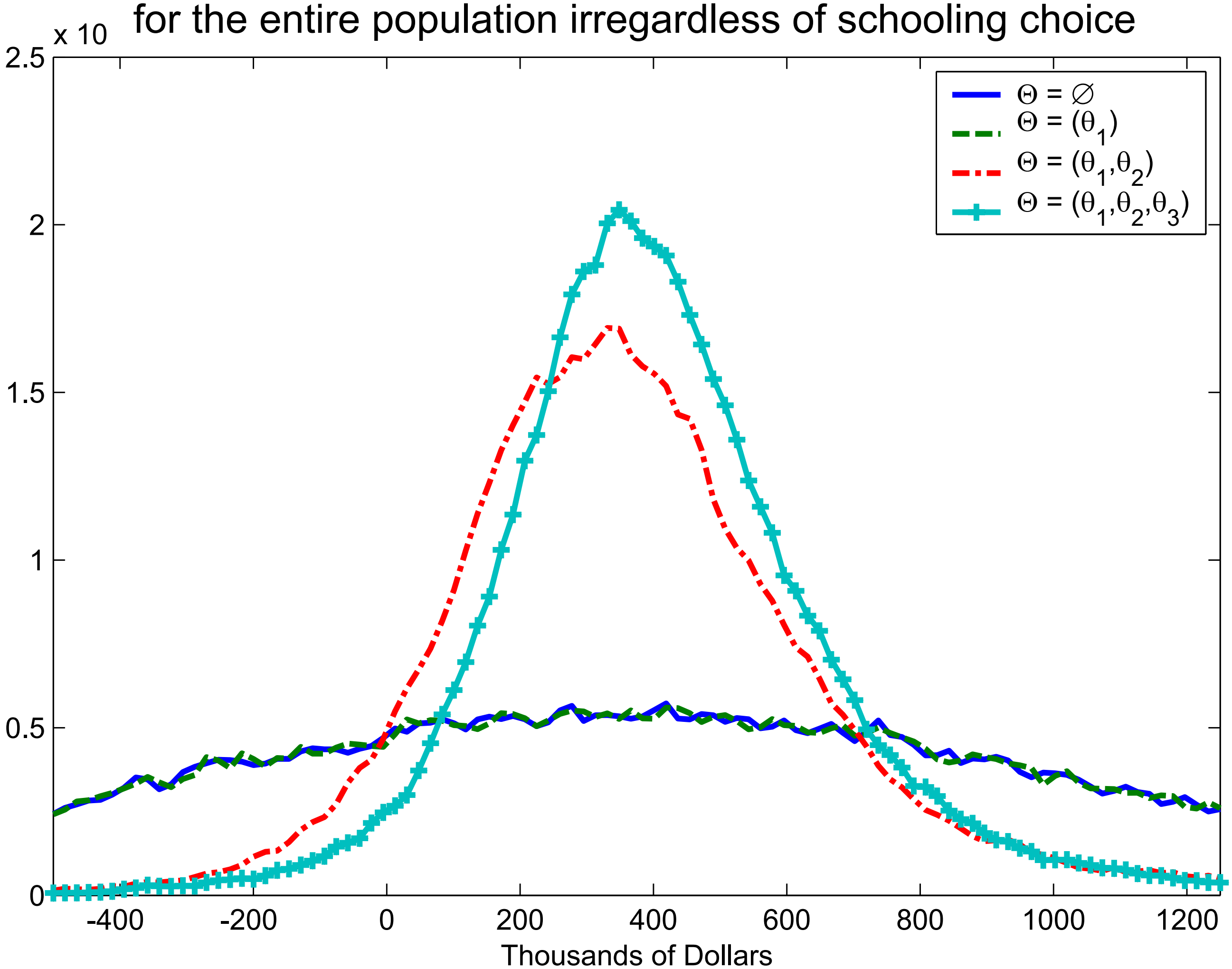
The dashed line plots $f(y_1|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.

The dashed-dotted line plots $f(y_1|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.

The crossed line plots $f(y_1|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.

The X are put at the mean and are assumed to be known. The θ, when known, are set at their mean of zero.

Figure 9.3
Densities of returns college vs high school
under different information sets for the agent calculated
for the entire population irregardless of schooling choice



Let Θ denote the agent's information set. Let Y_0, Y_1 denote the present value of earnings in the high school and college sectors, respectively (discounted at a 3% interest rate). Let $D=Y_0-Y_1$ be the difference of the present value of earnings in the college and high school sector. $f(d|\Theta)$ denote the density of the difference of present value of earnings conditioned on the information set Θ. Then:

The solid line plots $f(d|\Theta)$ under no information, i.e. $\Theta=\varnothing$.
The dashed line plots $f(d|\Theta)$ when only factor 1 is in the information set, i.e. $\Theta=(\theta_1)$.
The dashed-dotted line plots $f(d|\Theta)$ when factors 1 and 2 are in the information set, i.e. $\Theta=(\theta_1,\theta_2)$.
The crossed line plots $f(d|\Theta)$ when all factors are in the information set, i.e. $\Theta=(\theta_1,\theta_2,\theta_3)$.
The X are put at the mean and are assumed to be known The θ when known are set at their mean of zero

Table 1

Estimated Ex Post Returns to Schooling on Schooling and Their Effects on Schooling Choice using OLS and IV To Estimate The Ex Post Returns To Schooling

Log Earnings Regression[1]

Variable	Coefficient	Std. Error
	OLS	
School (High School vs. College)	0.2735	0.0344
School*ASVAB	0.0279	0.0063
	Instrumental Variables[2]	
School	0.2573	0.0451
School*ASVAB	0.0153	0.0083

Schooling Choice Probit Equation[3]

Variable	Coefficient	Std. Error
	Using OLS Results	
$b_{School} + b_{School*ASVAB}$*ASVAB	12.6244	0.7284
Marginal Effect	4.8333	0.2654
	Using IV Coefficients	
$b_{School} + b_{School*ASVAB}$*ASVAB	22.9150	1.3221
Marginal Effect	8.7731	0.4817

1 Includes controls for Mincer experience (age - years of schooling - 6), experience squared, cohort dummies, and ASVAB scores.

2 We use parental education, family income, broken home, number of siblings, distance to college, local tuition, cohort dummies, South at age 14 and urban at age 14 to instrument for schooling. We then interact the instrumente value of school with ASVAB scores.

3 We use the predicted return to school to test whether future earnings affect current schooling choices. We include controls for family background, cohort dummies, distance to college, and local tuition.

Table 2.1
Descriptive Statistics from the Pooled NLSY/1979 and PSID (white males)

	Full Sample					High School Sample					College Sample				
Variable Name	Obs	Mean	Std. Dev	Min	Max	Obs	Mean	Std. Dev	Min	Max	Obs	Mean	Std. Dev	Min	Max
Asvab AR*	1362	0.72	0.95	-1.78	1.96	747	0.26	0.89	-1.78	1.96	615	1.27	0.70	-1.36	1.96
Asvab PC*	1362	0.42	0.80	-2.68	1.36	747	0.07	0.86	-2.68	1.36	615	0.84	0.44	-1.06	1.36
Asvab WK*	1362	0.52	0.72	-2.29	1.34	747	0.20	0.76	-2.29	1.34	615	0.92	0.41	-1.36	1.34
Asvab MK*	1362	0.62	1.03	-1.62	2.11	747	0.00	0.81	-1.62	2.11	615	1.38	0.73	-1.46	2.11
Asvab CS*	1362	0.21	0.85	-2.52	2.49	747	-0.08	0.79	-2.52	2.08	615	0.56	0.77	-2.52	2.49
Urban at age 14	3695	0.79	0.40	0.00	1.00	1953	0.75	0.44	0.00	1.00	1742	0.85	0.36	0.00	1.00
Parents Divorced	3695	0.15	0.36	0.00	1.00	1953	0.18	0.38	0.00	1.00	1742	0.13	0.34	0.00	1.00
Number of Siblings	3695	2.86	1.96	0.00	17.00	1953	3.19	2.08	0.00	14.00	1742	2.49	1.74	0.00	17.00
Father's Education	3695	4.31	1.94	1.00	8.00	1953	3.56	1.51	1.00	8.00	1742	5.15	2.03	1.00	8.00
Mother's Education	3695	4.21	1.55	1.00	8.00	1953	3.68	1.26	1.00	8.00	1742	4.79	1.63	1.00	8.00
Born between 1906 and 1915	3695	0.01	0.10	0.00	1.00	1953	0.01	0.12	0.00	1.00	1742	0.00	0.06	0.00	1.00
Born between 1916 and 1925	3695	0.04	0.19	0.00	1.00	1953	0.04	0.21	0.00	1.00	1742	0.03	0.18	0.00	1.00
Born between 1926 and 1935	3695	0.07	0.25	0.00	1.00	1953	0.07	0.26	0.00	1.00	1742	0.06	0.24	0.00	1.00
Born between 1936 and 1945	3695	0.09	0.29	0.00	1.00	1953	0.07	0.26	0.00	1.00	1742	0.11	0.31	0.00	1.00
Born between 1946 and 1955	3695	0.20	0.40	0.00	1.00	1953	0.17	0.37	0.00	1.00	1742	0.24	0.43	0.00	1.00
Born between 1956 and 1965	3695	0.55	0.50	0.00	1.00	1953	0.56	0.50	0.00	1.00	1742	0.53	0.50	0.00	1.00
Born between 1966 and 1975	3695	0.04	0.21	0.00	1.00	1953	0.07	0.25	0.00	1.00	1742	0.02	0.14	0.00	1.00
Education	3695	1.47	0.50	1.00	2.00	1953	1.00	0.00	1.00	1.00	1742	2.00	0.00	2.00	2.00
Age in 1980	3695	26.87	12.32	5.00	68.00	1953	26.53	13.10	5.00	68.00	1742	27.25	11.39	9.00	68.00
Grade Completed 1980	1362	12.06	1.66	8.00	18.00	747	11.44	0.92	8.00	12.00	615	12.80	2.03	9.00	18.00
Enrolled in 1980	1362	0.57	0.50	0.00	1.00	747	0.33	0.47	0.00	1.00	615	0.86	0.35	0.00	1.00
PV of Earnings**	7152	2.38	1.64	0.00	18.59	3708	1.95	1.14	0.00	11.52	3444	2.83	1.95	0.00	18.59
Tuition at age 17	3695	1.80	0.72	0.00	5.55	1953	1.82	0.74	0.00	5.55	1742	1.76	0.70	0.00	5.55

*Note:
AR=Arithmetic Reasoning
PC=Paragraph Composition
WK= Word Knowledge
MK=Math Knowledge
CS=Coding Speed
**In thousands of Dollars

Table 2.2
List of Variables Included and Excluded in Each System

Variable Name	Cost Function	Test System	Earnings
Urban at age 14	Yes	Yes	No
Parents Divorced	Yes	Yes	No
Number of Siblings	Yes	Yes	No
Father's Education	Yes	Yes	No
Mother's Education	Yes	Yes	No
Born between 1906 and 1915	Yes	No	Yes
Born between 1916 and 1925	Yes	No	Yes
Born between 1926 and 1935	Yes	No	Yes
Born between 1936 and 1945	Yes	No	Yes
Born between 1946 and 1955	Yes	No	Yes
Born between 1956 and 1965	Yes	No	Yes
Born between 1966 and 1975	Yes	No	Yes
Age in 1980	No	Yes	No
Grade Completed 1980	No	Yes	No
Enrolled in 1980	No	Yes	No
Tuition at age 17	Yes	No	No

Table 2.3
Estimated Coefficients in Schooling Choice Equation

Coefficients	Mean	Standard Deviation
Constant	-2.2504	0.3587
Mother's Education	0.2250	0.0274
Father's Education	0.3386	0.0246
Parents Divorced	-0.1976	0.0845
Number of Siblings	-0.1012	0.0163
Urban Residence at age 14	0.1998	0.0755
Dummy birth 1916-1925	0.6076	0.3582
Dummy birth 1926-1935	0.5553	0.3471
Dummy birth 1936-1945	0.7050	0.3417
Dummy birth 1946-1955	0.4160	0.3355
Dummy birth 1956-1965	-0.2064	0.3346
Dummy birth 1966-1975	-1.4159	0.3703
Tuition at 4-year college	-0.0953	0.0447
Loading Factor 1	1.3523	0.1315
Loading Factor 2	0.4785	0.1335
Loading Factor 3	-0.0624	0.1274

Table 2.4
Estimated Coefficients for High School Earnings Equation

	Period Zero		Period One		Period Two		Period Three		Period 4	
Coefficients	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Dummy birth 1916-1925	-	-	-	-	-	-	-	-	-0.1054	0.0829
Dummy birth 1926-1935	-	-	-	-	-	-	-0.0225	0.0974	-0.1443	0.0809
Dummy birth 1936-1945	-	-	-	-	-0.1105	0.1034	-0.0201	0.0989	0.0616	0.1276
Dummy birth 1946-1955	-	-	-0.1779	0.0987	-0.2636	0.0917	0.1657	0.1973	-	-
Dummy birth 1956-1965	-0.7107	0.0637	-0.2936	0.0883	-0.0757	0.1385	-	-	-	-
Dummy birth 1966-1975	-0.6730	0.0960	-0.2360	0.2267	-	-	-	-	-	-
Constant	2.6276	0.0658	2.4021	0.0935	1.8880	0.0870	1.2819	0.0870	0.6147	0.0746
Loading Factor 1	0.1636	0.0433	0.1059	0.0485	0.0164	0.0949	0.0466	0.1122	-0.0077	0.0775
Loading Factor 2	-1.2138	0.0903	-1.6282	0.1142	-1.4415	0.1172	-1.1225	0.1056	-0.3924	0.0763
Loading Factor 3	0.0000	0.0000	0.0000	0.0000	0.2428	0.1684	0.2791	0.1510	0.1327	0.1013

Estimated Coefficients for College Earnings Equation

	Period Zero		Period One		Period Two		Period Three		Period 4	
Coefficients	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Dummy birth 1916-1925	-	-	-	-	-	-	-	-	-0.2976	0.3218
Dummy birth 1926-1935	-	-	-	-	-	-	-0.0881	0.1846	-0.3743	0.3147
Dummy birth 1936-1945	-	-	-	-	-0.0059	0.1710	0.0384	0.1696	-0.2256	0.3457
Dummy birth 1946-1955	-	-	-0.1944	0.1262	-0.0512	0.1568	0.2122	0.2238	-	-
Dummy birth 1956-1965	-0.7375	0.0686	-0.2340	0.1182	-0.1081	0.2910	-	-	-	-
Dummy birth 1966-1975	-0.3459	0.1736	1.3144	0.7365	-	-	-	-	-	-
Constant	2.2802	0.0670	3.5270	0.1191	3.1859	0.1720	2.4843	0.1914	1.3632	0.3367
Loading Factor 1	0.2225	0.0853	0.3137	0.1296	-0.2870	0.2415	-0.2676	0.2656	-0.0144	0.2300
Loading Factor 2	1.0000	0.0000	2.3887	0.1573	2.3194	0.1715	1.7102	0.1806	0.7481	0.1231
Loading Factor 3	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	1.5354	0.1627	0.8876	0.1665

Table 2.5
Estimated Coefficients of Test Equations

	Arithmetic Reasoning		Paragraph Composition		Word Knowledge		Math Knowledge		Coding Speed	
Coefficients	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Constant	-1.1198	0.2256	-1.0262	0.1719	-0.5180	0.2032	-1.4751	0.2265	-1.2706	0.2281
Mother's Education	0.0735	0.0177	0.0529	0.0136	0.0614	0.0158	0.0469	0.0178	0.0561	0.0175
Father's Education	0.0494	0.0136	0.0593	0.0105	0.0461	0.0121	0.0168	0.0139	0.0870	0.0135
Family Income in 1979	0.0008	0.0015	0.0009	0.0012	0.0000	0.0014	0.0038	0.0016	0.0021	0.0015
Parents Divorced	-0.0584	0.0564	-0.0514	0.0440	-0.0947	0.0508	0.0458	0.0569	-0.0138	0.0560
Number of Siblings	-0.0193	0.0111	-0.0397	0.0086	-0.0143	0.0099	-0.0273	0.0115	-0.0313	0.0110
South Residence at age 14	-0.1278	0.0463	-0.0906	0.0358	-0.0064	0.0423	-0.1418	0.0475	-0.1365	0.0464
Urban Residence at age 14	0.0640	0.0461	-0.0243	0.0361	0.0117	0.0422	0.0258	0.0468	0.0529	0.0466
Enrolled at School at Test Date	0.0646	0.0528	-0.0036	0.0403	-0.0515	0.0471	0.0074	0.0527	0.3122	0.0529
Age at Test Date	0.0096	0.0164	0.0237	0.0128	-0.0170	0.0148	0.0048	0.0165	-0.0510	0.0166
Highest Grade Completed at Test Date	0.0911	0.0198	0.0604	0.0155	0.0721	0.0179	0.1082	0.0201	0.1732	0.0198
Loading Factor 1	1.0000	0.0000	0.6801	0.0321	0.8069	0.0377	0.5648	0.0319	0.9562	0.0293
Loading Factor 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Loading Factor 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 3a
Goodness of Fit Tests: Predicted Earnings Densities vs. Actual Densities
The Three-Factor Model

		High School	College	Overall
Period 1	χ^2 Statistic	91.9681	74.2503	204.3823
	Critical Value*	107.5217	82.5287	178.4854
Period 2	χ^2 Statistic	86.6649	107.6417	207.6152
	Critical Value*	116.5110	116.5110	218.8205
Period 3	χ^2 Statistic	26.2658	45.5301	106.5721
	Critical Value*	43.7730	55.7585	91.6702
Period 4	χ^2 Statistic	35.3846	29.7218	55.5758
	Critical Value*	31.4104	30.1435	55.7585
Period 5	χ^2 Statistic	23.2193	14.9131	41.8657
	Critical Value*	23.6848	16.9190	35.1725

* 95% Confidence, equiprobable bins with aprox. 15 people per bin

Table 3b

Goodness of Fit Tests: Predicted Earnings Densities vs for Actual Earnings Densities

The Two-Factor Model

		High School	College	Overall
Period 1	χ^2 Statistic	109.5702	132.3027	267.4894
	Critical Value*	107.5217	82.5287	178.4854
Period 2	χ^2 Statistic	104.1649	150.5556	247.6732
	Critical Value*	116.5110	116.5110	218.8205
Period 3	χ^2 Statistic	40.7028	61.7322	114.1692
	Critical Value*	43.7730	55.7585	91.6702
Period 4	χ^2 Statistic	39.7253	47.5559	64.2503
	Critical Value*	31.4104	30.1435	55.7585
Period 5	χ^2 Statistic	18.3217	26.5855	40.4078
	Critical Value*	23.6848	16.9190	35.1725

* 95% Confidence, equiprobable bins with aprox. 15 people per bin

Table 4.1: Ex-Post Conditional Distributions (College Earnings Conditional on High School Earnings)
$Pr(d_i<Yc<d_i+1 \mid d_j<Yh<d_j+1)$ where d_i is the ith decile of the College Lifetime Ex-Post Earnings Distribution and d_j is the jth decile of the High School Ex-Post Lifetime Earnings Distribution
Corrrelation(Y_C,Y_H) = -0.3899

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0035	0.0109	0.0240	0.0326	0.0524	0.7538	0.1137	0.1557	0.2511	0.2808
2	0.0098	0.0244	0.0419	0.0631	0.0894	0.1122	0.1391	0.1747	0.2048	0.1407
3	0.0160	0.0466	0.0741	0.0877	0.1041	0.1213	0.1441	0.1549	0.1581	0.0931
4	0.0236	0.0603	0.0911	0.1062	0.1220	0.1298	0.1348	0.1372	0.1266	0.0683
5	0.0439	0.0848	0.1108	0.1227	0.1303	0.1309	0.1211	0.1139	0.0928	0.0489
6	0.0627	0.1074	0.1214	0.1304	0.1330	0.1218	0.1168	0.0954	0.0695	0.0415
7	0.0963	0.1256	0.1340	0.1334	0.1200	0.1200	0.0937	0.0784	0.0554	0.0433
8	0.1378	0.1659	0.1529	0.1396	0.1114	0.0925	0.0740	0.0561	0.0296	0.0402
9	0.1939	0.1970	0.1498	0.1180	0.1002	0.0771	0.0534	0.0362	0.0200	0.0543
10	0.3354	0.1983	0.1167	0.0812	0.0515	0.0351	0.0266	0.0152	0.0130	0.1271

Table 4.2: Ex-Ante Conditional Distribution (College Earnings Conditional on High School Earnings)
$Pr(d_i<Yc<d_i+1 \mid d_j<Yh<d_j+1)$ 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
Individual expects out θ_3 and $\varepsilon_{s,t}$ for t=0, … , 4, which are unknown by the agent at the time of the schooling choice.
Corrrelation(Y_C,Y_H) = -0.69936591

	College									
High School	1	2	3	4	5	6	7	8	9	10
1	0.0002	0.0079	0.0108	0.0226	0.0421	0.0594	0.0909	0.1447	0.2236	0.3978
2	0.0044	0.0180	0.0286	0.0530	0.0720	0.1010	0.1362	0.1686	0.2114	0.2068
3	0.0106	0.0362	0.0578	0.0786	0.1062	0.1152	0.1498	0.1618	0.1692	0.1146
4	0.0200	0.0546	0.0786	0.1024	0.1204	0.1266	0.1376	0.1406	0.1290	0.0902
5	0.0390	0.0740	0.1004	0.1130	0.1291	0.1387	0.1295	0.1206	0.1010	0.0546
6	0.0454	0.1017	0.1253	0.1353	0.1333	0.1323	0.1189	0.1011	0.0754	0.0314
7	0.0873	0.1299	0.1437	0.1451	0.1299	0.1199	0.0965	0.0777	0.0519	0.0180
8	0.1336	0.1603	0.1613	0.1431	0.1160	0.0974	0.0793	0.0589	0.0389	0.0112
9	0.2063	0.2016	0.1651	0.1293	0.1056	0.0840	0.0540	0.0317	0.0155	0.0068
10	0.4123	0.2318	0.1393	0.0868	0.0556	0.0365	0.0210	0.0110	0.0049	0.0006

Table 5.1
Average present value of earnings[1] for high school graduates
Fitted and Counterfactual[2]

White males from NLSY79

	High School (Fitted)	College (counterfactual)
Average Present Value of Earnings	605.92	969.34
Std. Err.	*13.719*	*67.164*
Random[3]	573.53	987.21
Std. Err.	*15.799*	*47.132*
Average returns[4] to college for high school graduates		
Average returns	1.17	
Std. Err.	*0.1350*	

[1] Thousands of dollars. Discounted using a 3% interest rate.

[2] The counterfactual is constructed using the estimated college outcome equation applied to the population of persons selecting high school

[3] It defines the result of taking a person at random from the population regardless of his schooling choice.

[4] As a fraction of the base state, ie (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Table 5.2
Average present value of earnings[1] for college graduates
Fitted and Counterfactual[2]

White males from NLSY79

	High School (Counterfactual)	College (fitted)
Average Present Value of Earnings	536.43	1007.64
Std. Err.	*26.187*	*35.113*
Random[3]	573.53	987.21
Std. Err.	*15.799*	*47.132*

Average returns[4] to college for college graduates

Average returns	1.33
Std. Err.	*0.0958*

[1] Thousands of dollars. Discounted using a 3% interest rate.

[2] The counterfactual is constructed using the estimated high school outcome equation applied to the population of persons selecting college

[3] It defines the result of taking a person at random from the population regardless of his schooling choice.

[4] As a fraction of the base state, ie (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Table 5.3

Average present value of earnings[1] for population of persons indifferent between high school and college

Conditional on education level

White males from NLSY79

	High School	College
Average Present Value of Earnings	571.33	975.16
Std. Err.	*37.066*	*70.557*

Average returns[3] to college for people indifferent between high school and college

	High School vs Some College
Average returns	1.26
Std. Err.	*0.3691*

[1] Thousands of dollars. Discounted using a 3% interest rate.

[2] It defines the result of taking a person at random from the population regardless of his schooling choice.

[3] As a fraction of the base state, ie (PVearnings(Col)-PVearnings(HS))/PVearnings(HS).

Table 5.4

Average ex-post, ex-ante and perfect certainty returns[1]

White males from NLSY79

	ex-post[2]	ex-ante[3]	perfect certainty[4]
For People who choose High School			
Average	1.1594	1.1594	0.9337
Std. Err.	*0.1362*	*0.1362*	*0.1154*
For People who choose College			
Average	1.3398	1.3398	1.6121
Std. Err.	*0.1083*	*0.1083*	*0.1082*
For People Indifferent Between High School and College			
Average	1.2585	1.2585	1.2418
Std. Err.	*0.3868*	*0.3868*	*0.1067*

[1] Let Y_0, Y_1 denote the present value of earnings in high school and college, respectively. The return to college R is defined as

$$R = \left(\frac{Y_1 - Y_0}{Y_0} \right)$$

[2] Let I denote the schooling choice index. Let Θ_0 denote the information set of the agent at the time of the schooling choice. Let R denote the return to college. The ex-post mean return to college for a high-school graduate is $E\left(R \mid E_0\left(I\right) < 0\right)$, where $E_0\left(I\right) = E\left(I \mid \Theta_0\right)$. Similarly, the ex-post mean return to college for a college graduate is $E\left(R \mid E_0\left(I\right) \geq 0\right)$. The ex-post mean return to an agent just indifferent between college and high-school is $E\left(R \mid E_0\left(I\right) = 0\right)$.

[3] Let I denote the schooling index. Let Θ_0 denote the information set of the agent at the time of the schooling choice. Let R denote the return to college. The ex-ante mean return to college for a high-school graduate is $E\left(E_0\left(R\right) \mid E_0\left(I\right) < 0\right)$. Similarly, the ex-ante mean return to college for a college graduate is $E\left(E_0\left(R\right) \mid E_0\left(I\right) \geq 0\right)$. The ex-ante mean return to an agent just indifferent between college and high-school is $E\left(E_0\left(R\right) \mid E_0\left(I\right) = 0\right)$. By a property of means, the mean ex-ante and the mean ex-post returns must be equal for the same conditioning set, i.e. $E\left(E_0\left(R\right) \mid E_0\left(I\right) \geq 0\right) = E\left(R \mid E_0\left(I\right) \geq 0\right)$.

[4] Let I denote the schooling index. Let R denote the return to college. The return to college under perfect certainty for a high-school graduate is $E\left(R \mid I < 0\right)$. Note that now the agent makes his schooling choice under perfect certainty (that is why we condition on I). Similarly, the return to college under perfect certainty for a college graduate is $E\left(R \mid I \geq 0\right)$. The return to college under perfect certainty for an agent just indifferent between college and high-school is $E\left(R \mid I = 0\right)$.

Table 6.1

Agent's Forecast Variance of Present Value of Earnings*
Under Different Information Sets

(fraction of the variance explained by Θ)**

The Calculation is for the Entire Population Irregardless of Schooling Choice.

	Var(Y_c)	Var (Y_h)	Var(Y_c-Y_h)
For lifetime:+			
Variance when $\Theta = \varnothing$	156402.14	73827.89	267796.38
$\Theta = \{\theta_1\}$	0.95%	0.27%	0.44%
$\Theta = \{\theta_1, \theta_2\}$	29.10%	29.43%	47.42%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	68.03%	32.27%	62.65%

*We use an interest rate of 3% to calculate the present value of earnings.

+Variance of the unpredictable component of earnings between age 19 and 65 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0095)*156402.14

Table 6.2
Agent's Forecast Variance of Period Zero Earnings*
Under Different Information Sets
(fraction of the variance explained by Θ)**

The Calculation is for the Entire Population Irregardless of Schooling Choice.

	Var(Y_c)	Var (Y_h)	Var(Y_c-Y_h)
For lifetime:+			
Variance when $\Theta = \varnothing$	13086.24	14303.35	33910.17
$\Theta = \{\theta_1\}$	1.90%	0.91%	0.05%
$\Theta = \{\theta_1, \theta_2\}$	23.58%	30.08%	41.02%

*We use an interest rate of 3% to calculate the present value of earnings.

+Variance of the unpredictable component of earnings between age 19 and 28 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0190)*13086.24

Table 6.3
Agent's Forecast Variance of Period One Earnings*
Under Different Information Sets
(fraction of the variance explained by Θ)**

The Calculation is for the Entire Population Irregardless of Schooling Choice.

	Var(Y_c)	Var (Y_h)	Var(Y_c-Y_h)
For lifetime:+			
Variance when $\Theta = \varnothing$	26618.64	17545.90	65804.89
$\Theta = \{\theta_1\}$	1.90%	0.31%	0.34%
$\Theta = \{\theta_1, \theta_2\}$	62.43%	43.00%	69.60%

*We use an interest rate of 3% to calculate the present value of earnings.

+Variance of the unpredictable component of earnings between age 29 and 38 as predicted at age 19.

**So we would say that the variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0190)*26618.64

Table 6.4
Agent's Forecast Variance of Period Two Earnings*
Under Different Information Sets

(fraction of the variance explained by Θ)**

The Calculation is for the Entire Population Irregardless of Schooling Choice.

	Var(Y_c)	Var (Y_h)	Var(Y_c-Y_h)
For lifetime:+			
Variance when $\Theta = \varnothing$	40406.20	16716.50	68918.36
$\Theta = \{\theta_1\}$	0.95%	0.00%	0.63%
$\Theta = \{\theta_1, \theta_2\}$	38.66%	35.02%	58.63%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	75.25%	40.17%	70.98%

*We use an interest rate of 3% to calculate the present value of earnings.

+Variance of the unpredictable component of earnings between age 39 and 48 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0095)*40406.20

Table 6.5

Agent's Forecast Variance of Period Three Earnings*
Under Different Information Sets

(fraction of the variance explained by Θ)**

The Calculation is for the Entire Population Irregardless of Schooling Choice.

	$Var(Y_c)$	$Var(Y_h)$	$Var(Y_c - Y_h)$
For lifetime:+			
Variance when $\Theta = \varnothing$	53194.23	14605.29	66926.12
$\Theta = \{\theta_1\}$	0.65%	0.08%	0.73%
$\Theta = \{\theta_1, \theta_2\}$	16.18%	24.55%	34.65%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	81.20%	31.53%	70.11%

*We use an interest rate of 3% to calculate the present value of earnings.

+Variance of the unpredictable component of earnings between age 49 and 58 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.0065)*53194.23

Table 6.6

Agent's Forecast Variance of Period Four of Earnings*
Under Different Information Sets

(fraction of the variance explained by Θ)**

The Calculation is for the Entire Population Irregardless of Schooling Choice.

	Var(Y_c)	Var (Y_h)	Var(Y_c-Y_h)
For lifetime:+			
Variance when $\Theta = \varnothing$	23096.81	10656.83	32236.82
$\Theta = \{\theta_1\}$	0.00%	0.00%	0.00%
$\Theta = \{\theta_1, \theta_2\}$	6.84%	4.10%	11.41%
$\Theta = \{\theta_1, \theta_2, \theta_3\}$	56.70%	6.16%	37.95%

*We use an interest rate of 3% to calculate the present value of earnings.

+Variance of the unpredictable component of earnings between age 59 and 65 as predicted at age 19.

**The variance of the unpredictable component of period 1 college earnings $\Theta = \{\theta_1\}$ is (1-0.00)*23096.81