The basic measurement model is the following.
Let

- i represent a person
- $$d(i)$$
  is their demographic groups (Sex×Ethnicity)
- $T_{\ell i}$ a test scores (Math SAT, Verbal SAT, ACT) so ℓ goes from 1 to three. For most people we don't observe all three but lets abstract from that (dealing with it is straight forward).
- Q_i is a measure of college quality
- θ_i is a the skill level of the person

In this version of the model we assume that

$$\begin{aligned} T_{\ell i} =& \alpha_{\ell 0} + \alpha_{\ell 1}\theta_i + \varepsilon_{\ell i} \\ Q_i =& \delta_{d(i)0} + \delta_{d(i)1}\theta_i + \nu_i \end{aligned}$$

θ_i has a discrete distribution where we estimate the points of support and probability mass. We constrain the model so that the minimum probability mass is 0.05.

Both ε_{elli} and ν_i are estimated using Gallant/Nychka.

That is their density can be written as

$$f(x;\sigma,\Gamma) = \left(\sum_{i=0}^{K} \Gamma_i \left(\frac{x}{\sigma}\right)^i\right)^2 e^{-\frac{x^2}{\sigma^2 2}} + \epsilon_0 \phi\left(\frac{x}{\sigma}\right)$$

where Γ is constrained so that this integrates to one and ϕ is the pdf of a standard normal.

I have taken $\epsilon_0 = 0.05$

Some comments

- I have have been focusing on the case H=5. It seems like the likelihood can support a higher value of H but I am not sure we want that.
- We have that the test score only depends on the actual level of ability and the noise is the same for all ethnic groups, however as a result of affirmative action and potentially different tastes for college we allow the parameters of the college quality to vary with demographic group.

We normalize the scores by standard deviation of scores we see in data so

$$
\begin{aligned}
T_{1i} \equiv & \frac{\text{Raw Math SAT} - 556.9207082371055}{100.8233233370844} \\
T_{2i} \equiv & \frac{\text{Raw Verbal SAT} - 553.9780600461894}{101.00674362034842} \\
T_{3i} \equiv & \frac{\text{Raw ACT} - 23.196212054164086}{4.592107241564743} \\
Q_i \equiv & \frac{\text{"Average" M+V SAT} - 1122.234697923378}{141.3114806648894}
\end{aligned}
$$

We estimate in three stages

1. using only the test scores to solve for the unconditional distribution of θ_i, the marginal distributions of $\varepsilon_{\ell i}$, and the α.
2. Solve for the conditional distribution of θ_i conditional on demographic group. Here we restrict the distribution of $\varepsilon_{\ell i}$ to be the same for the groups-which is an important assumption. The measurment error on the test is independent of the demographic group. Also here note that we keep the support of the distribution the same-we are just letting the probabilities of each point of theta vary across groups. I do not constrain the model so that these probabilities add up to the ones estimated unconditionally (because I am not sure how to do it and am not sure we would want to anyway)
3. Solve the college equation for each demogrphaic group, the δ and the distribution of ν_i

Stage 1

We normalize

$$
\begin{aligned}
\alpha_{10} = & 0 \\
\alpha_{11} = & 1
\end{aligned}
$$

(These are normalizations because the scale and location of θ is completely free)

Setting the number of points of θ we estimate via maximum likelihood.

The current estimates are for 8 points (I am still playing around with this-could change.)

Here are the 8 points.

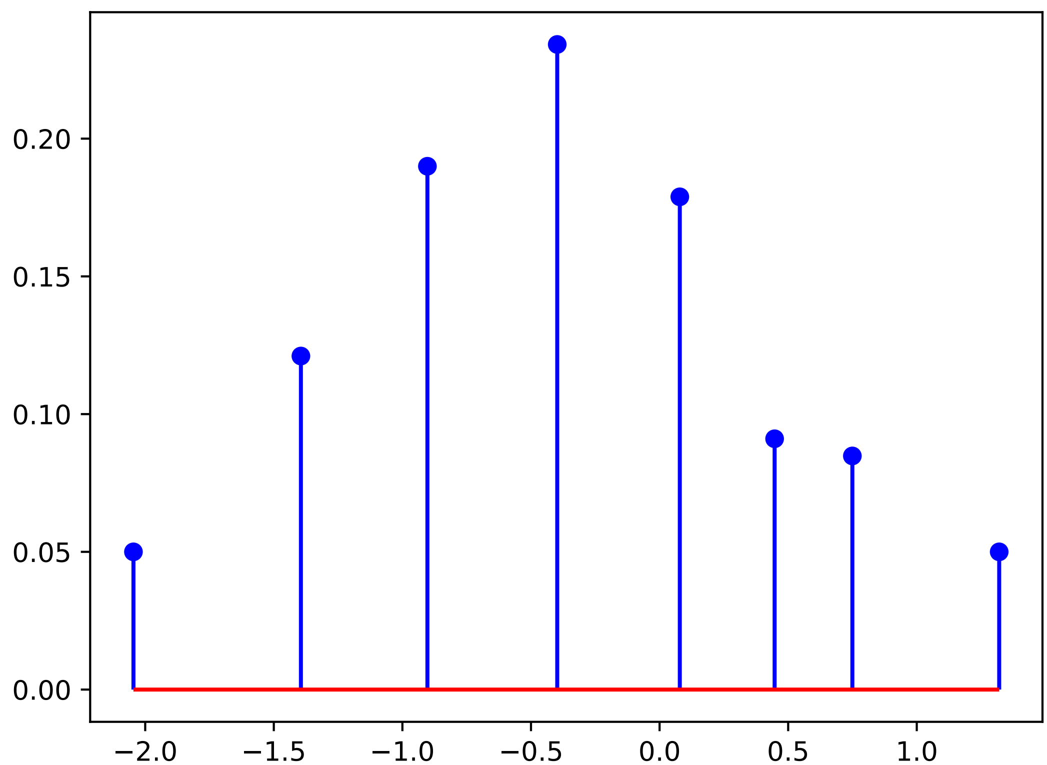
Given these parameters the second stage is easy. We just estimate δ and $\sigma^2_{\nu d(i)}$ for each demographic group.

I will focus on two cases $K^* = 10$ and $K^* = 20$

Here are the 7 points

Probability	θ
0.050	-2.046
0.121	-1.395
0.190	-0.903
0.234	-0.398
0.179	0.079
0.091	0.447
0.085	0.750
0.050	1.320

And as a histogran



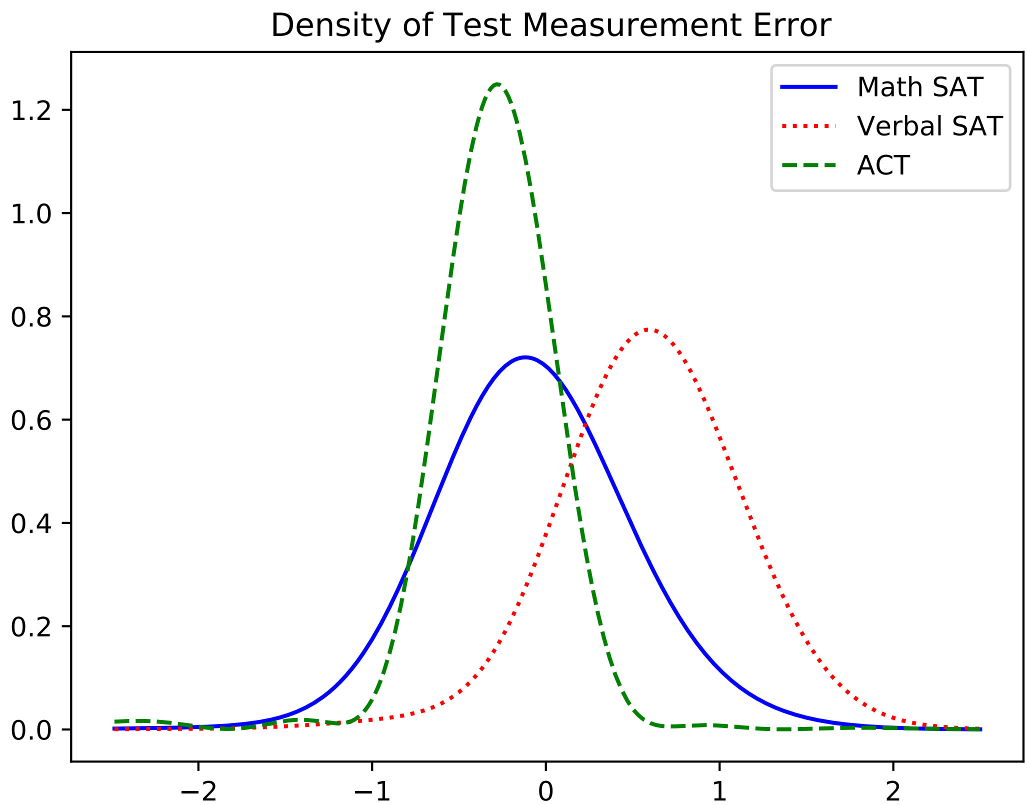
And here is the distribution of the α

Parameter	SAT Math	SAT Verbal	ACT
Intercept	0.000	-0.661	0.417
θ	1.000	1.029	1.156

And the distribution of the measument error.

Parameter	SAT Math	SAT Verbal	ACT
σ	0.611	0.623	0.548
x^0	0.616	0.241	0.805
x^1	0.165	0.369	0.154
x^2	-0.040	0.166	-0.376
x^3	-0.004	0.018	-0.028
x^4	0.006	-0.004	0.034

And as a picture

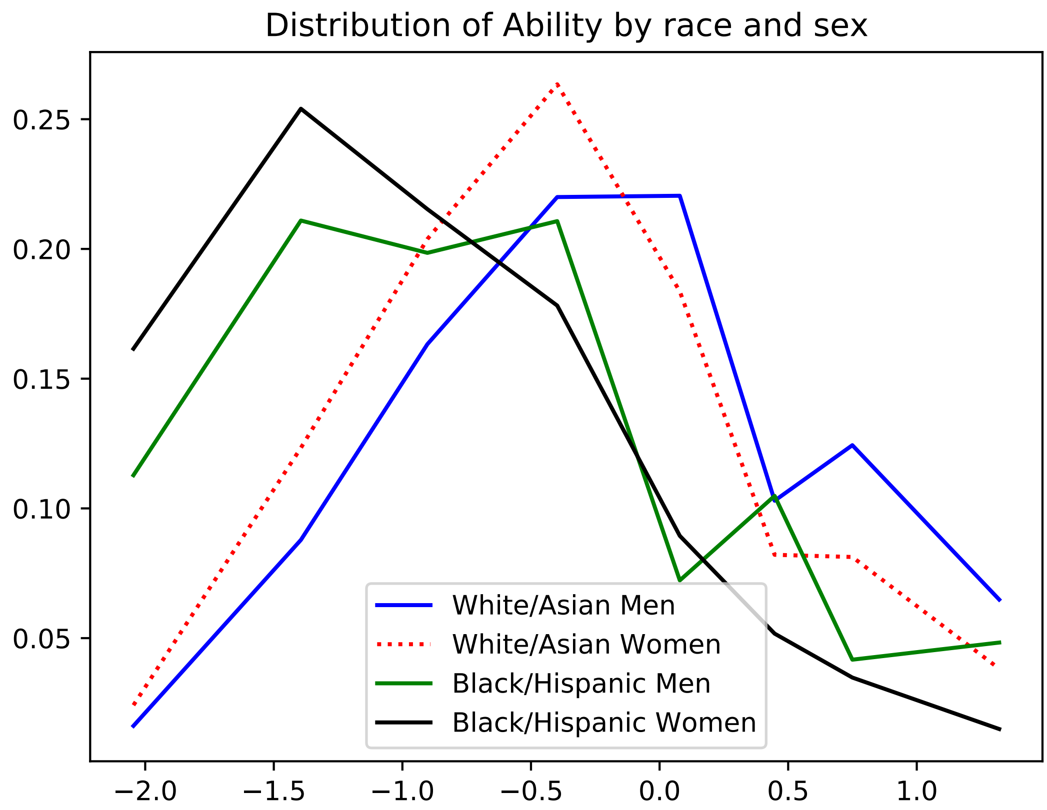


Stage 2

This is just a re-estimation of the mass probabilities for each of the four groups:

θ	While Male	White Female	Minority Male	Minority Female
-2.046	0.016	0.024	0.113	0.162
-1.395	0.088	0.123	0.211	0.254
-0.903	0.163	0.204	0.198	0.215
-0.398	0.220	0.263	0.211	0.178
0.079	0.220	0.184	0.072	0.089
0.447	0.103	0.082	0.105	0.052
0.750	0.124	0.081	0.042	0.035
1.320	0.065	0.038	0.048	0.015

and graphically



Stage 3

Second Stage Estimates

Parameter	While Male	White Female	Minority Male	Minority Female
Constant	1.313	1.234	0.365	0.403
θ	0.626	0.544	0.706	0.643
σ	0.693	0.694	0.583	0.628
x^0	0.351	0.341	0.488	0.492
x^1	-0.128	-0.128	-0.116	-0.092
x^2	0.067	0.083	0.022	0.037
x^3	-0.063	-0.062	0.022	0.004
x^4	0.013	0.011	0.020	0.016

And a plot of the densities