1 Structural Estimation

{sec:structural-estir

This section describes how to use the methods developed above to structurally estimate a life-cycle consumption model, following closely the work of Cagetti (2003).[1] The key idea of structural estimation is to look for the parameter values (for the time preference rate, relative risk aversion, or other parameters) which lead to the best possible match between simulated and empirical moments.

1.1 Life Cycle Model

{subsec:life-cycle-m

Realistic calibration of a life cycle model needs to take into account a few things that we omitted from the bare-bones model described above. For example, the whole point of the life cycle model is that life is finite, so we need to include a realistic treatment of life expectancy; this is done easily enough, by assuming that utility accrues only if you live, so effectively the rising mortality rate with age is treated as an extra reason for discounting the future. Similarly, we may want to capture the demographic evolution of the household (e.g., arrival and departure of kids). A common way to handle that, too, is by modifying the discount factor (arrival of a kid might increase the total utility of the household by, say, 0.2, so if the 'pure' rate of time preference were 1.0 the 'household-size-adjusted' discount factor might be 1.2. We therefore modify the model presented above to allow age-varying discount factors that capture both mortality and family-size changes (we just adopt the factors used by Cagetti (2003) directly), with the probability of remaining alive between t and $t+n$ captured by $\mathcal{L}$ and with $\hat{\beta}$ now reflecting all the age-varying discount factor adjustments (mortality, family-size, etc). Using $\beth$ (the Hebrew cognate of β) for the 'pure' time preference factor, the value function for the revised problem is

$$\mathrm{v}_t(\mathbf{p}_t, \mathbf{m}_t) = \max_{\{\mathrm{c}\}_t^T} \ \mathrm{u}(\mathbf{c}_t) + \left[\sum_{n=1}^{T-t} \beth^n \mathcal{L}_t^{t+n} \hat{\beta}_t^{t+n} \mathrm{u}(\mathbf{c}_{t+n}) \right] \tag{1}$$ {eq:lifecyclemax}

subject to the constraints

$$\begin{aligned} \mathbf{a}_t &= \mathbf{m}_t - \mathbf{c}_t \\ \mathbf{p}_{t+1} &= \mathcal{G}_{t+1} \mathbf{p}_t \psi_{t+1} \\ \mathbf{y}_{t+1} &= \mathbf{p}_{t+1} \boldsymbol{\theta}_{t+1} \\ \mathbf{m}_{t+1} &= \mathsf{R} \mathbf{a}_t + \mathbf{y}_{t+1} \end{aligned}$$

where

$$\begin{aligned} \mathcal{L}_t^{t+n} &: \text{probability to } \mathcal{L}\text{ive until age } t+n \text{ given alive at age } t \\ \hat{\beta}_t^{t+n} &: \text{age-varying discount factor between ages } t \text{ and } t+n \\ \psi_t &: \text{mean-one shock to permanent income} \\ \beth &: \text{time-invariant 'pure' discount factor} \end{aligned}$$

[1] Similar structural estimation exercises have been also performed by Palumbo (1999) and Gourinchas and Parker (2002).

and all the other variables are defined as in section 2.

Households start life at age $s = 25$ and live with probability 1 until retirement ($s = 65$). Thereafter the survival probability shrinks every year and agents are dead by $s = 91$ as assumed by Cagetti.

Transitory and permanent shocks are distributed as follows:

$$\Xi_s = \begin{cases} 0 & \text{with probability } \wp > 0 \\ \boldsymbol{\theta}_s/\wp & \text{with probability } (1-\wp)\text{, where } \log \boldsymbol{\theta}_s \sim \mathcal{N}(-\sigma_{\boldsymbol{\theta}}^2/2, \sigma_{\boldsymbol{\theta}}^2) \end{cases} \tag{2}$$
$$\log \psi_s \sim \mathcal{N}(-\sigma_{\psi}^2/2, \sigma_{\psi}^2)$$

where $\wp$ is the probability of unemployment (and unemployment shocks are turned off after retirement).

The parameter values for the shocks are taken from Carroll (1992), $\wp = 0.5/100$, $\sigma_{\boldsymbol{\theta}} = 0.1$, and $\sigma_{\psi} = 0.1$.[2] The income growth profile $\mathcal{G}_t$ is from Carroll (1997) and the values of $\mathcal{L}_t$ and $\hat{\beta}_t$ are obtained from Cagetti (2003) (Figure 1).[3] The interest rate is assumed to equal 1.03. The model parameters are included in Table 1.

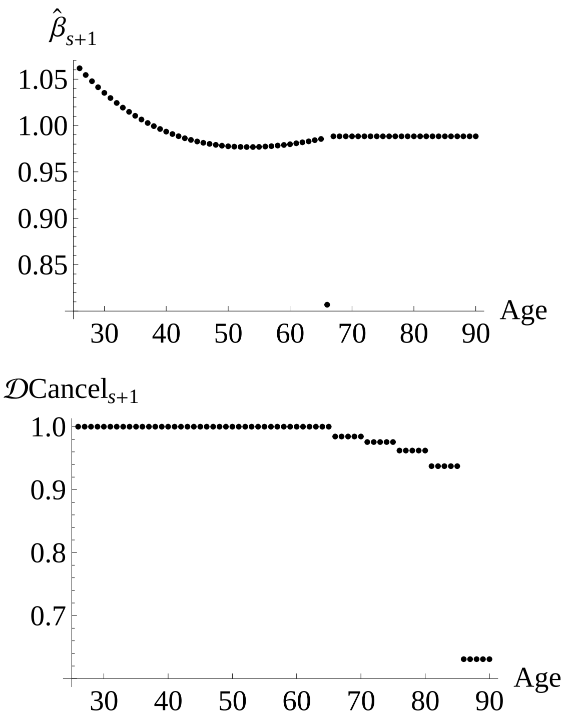

Figure 1 Time Varying Parameters

{fig:TimeVaryingPa

[2]Note that $\sigma_{\boldsymbol{\theta}} = 0.1$ is smaller than the estimate for college graduates estimated in Carroll and Samwick (1997) ($= 0.197 = \sqrt{0.039}$) which is used by Cagetti (2003). The reason for this choice is that Carroll and Samwick (1997) themselves argue that their estimate of $\sigma_{\boldsymbol{\theta}}$ is almost certainly increased by measurement error.

[3]The income growth profile is the one used by Caroll for operatives. Cagetti computes the time-varying discount factor by educational groups using the methodology proposed by Attanasio et al. (1999) and the survival probabilities from the 1995 Life Tables (National Center for Health Statistics 1998).

Table 1 Parameter Values

{table:StrEstParam

$\sigma_{\boldsymbol{\theta}}$	0.1	Carroll (1992)
σ_{ψ}	0.1	Carroll (1992)
$\wp$	0.005	Carroll (1992)
$\mathcal{G}_s$	figure 1	Carroll (1997)
$\hat{\beta}_s, \mathcal{L}_s$	figure 1	Cagetti (2003)
R	1.03	Cagetti (2003)

The structural estimation of the parameters $\beth$ and ρ is carried out using the procedure specified in the following section, which is then implemented in the `StructEstimation.py` file. This file consists of two main components. The first section defines the objects required to execute the structural estimation procedure, while the second section executes the procedure and various optional experiments with their corresponding commands. The next section elaborates on the procedure and its accompanying code implementation in greater detail.

1.2 Estimation

When economists say that they are performing "structural estimation" of a model like this, they mean that they have devised a formal procedure for searching for values for the parameters $\beth$ and ρ at which some measure of the model's outcome (like "median wealth by age") is as close as possible to an empirical measure of the same thing. Here, we choose to match the median of the wealth to permanent income ratio across 7 age groups, from age $26-30$ up to $56-60$.[4] The choice of matching the medians rather the means is motivated by the fact that the wealth distribution is much more concentrated at the top than the model is capable of explaining using a single set of parameter values. This means that in practice one must pick some portion of the population who one wants to match well; since the model has little hope of capturing the behavior of Bill Gates, but might conceivably match the behavior of Homer Simpson, we choose to match medians rather than means.

As explained in section 3, it is convenient to work with the normalized version of the

[4] Cagetti (2003) matches wealth levels rather than wealth to income ratios. We believe it is more appropriate to match ratios both because the ratios are the state variable in the theory and because empirical moments for ratios of wealth to income are not influenced by the method used to remove the effects of inflation and productivity growth.

model which can be written in Bellman form as:

$$\begin{aligned}
\mathrm{v}_t(m_t) &= \max_{c_t} \quad \mathrm{u}(c_t) + \beth\mathcal{L}_{t+1}\hat{\beta}_{t+1}\mathbb{E}_t[(\psi_{t+1}\mathcal{G}_{t+1})^{1-\rho}\mathrm{v}_{t+1}(m_{t+1})] \\
&\text{s.t.} \\
a_t &= m_t - c_t \\
m_{t+1} &= a_t \underbrace{\left(\frac{\mathsf{R}}{\psi_{t+1}\mathcal{G}_{t+1}}\right)}_{\equiv \mathcal{R}_{t+1}} + \boldsymbol{\theta}_{t+1}
\end{aligned}$$

with the first order condition:

$$\mathrm{u}^c(c_t) = \beth\mathcal{L}_{t+1}\hat{\beta}_{t+1}\mathsf{R}\mathbb{E}_t\left[\mathrm{u}^c\left(\psi_{t+1}\mathcal{G}_{t+1}\mathrm{c}_{t+1}\left(a_t\mathcal{R}_{t+1} + \boldsymbol{\theta}_{t+1}\right)\right)\right]. \tag{3}$$ {eq:FOCLifeCycle}

The first substantive toe in this estimation procedure is to solve for the consumption functions at each age. We need to discretize the shock distribution and solve for the policy functions by backward induction using equation (3) following the procedure in sections 5 and 'Recursion.' The latter routine is slightly complicated by the fact that we are considering a life-cycle model and therefore the growth rate of permanent income, the probability of death, the time-varying discount factor and the distribution of shocks will be different across the years. We thus must ensure that at each backward iteration the right parameter values are used.

Correspondingly, the first part of the `StructEstimation.py` file begins by defining the agent type by inheriting from the baseline agent type `IndShockConsumerType`, with the modification to include time-varying discount factors. Next, an instance of this "life-cycle" consumer is created for the estimation procedure. The number of periods for the life cycle of a given agent is set and, following Cagetti, (2003), we initialize the wealth to income ratio of agents at age 25 by randomly assigning the equal probability values to 0.17, 0.50 and 0.83. In particular, we consider a population of agents at age 25 and follow their consumption and wealth accumulation dynamics as they reach the age of 60, using the appropriate age-specific consumption functions and the age-varying parameters. The simulated medians are obtained by taking the medians of the wealth to income ratio of the 7 age groups.

To complete the creation of the consumer type needed for the simulation, a history of shocks is drawn for each agent across all periods by invoking the `make_shock_history` function. This involves discretizing the shock distribution for as many points as the number of agents we want to simulate and then randomly permuting this shock vector as many times as we need to simulate the model for. In this way, we obtain a time varying shock for each agent. This is much more time efficient than drawing at each time from the shock distribution a shock for each agent, and also ensures a stable distribution of shocks across the simulation periods even for a small number of agents. (Similarly, in order to speed up the process, at each backward iteration we compute the consumption function and other variables as a vector at once.)

With the age-varying consumption functions derived from the life-cycle agent, we can proceed to generate simulated data and compute the corresponding medians. Estimating the model involves comparing these simulated medians with empirical medians, measuring the model's success by calculating the difference between the two. However, before performing the necessary steps of solving and simulating the model to generate simulated moments, it's important to note a difficulty in producing the target moments using the available data.

Specifically, defining ξ as the set of parameters to be estimated (in the current case $\xi = \{\rho, \beth\}$), we could search for the parameter values which solve

$$\min_{\xi} \sum_{\tau=1}^{7} |\varsigma^{\tau} - \mathbf{s}^{\tau}(\xi)| \tag{4}$$

{eq:naivePowell}

where ς^{τ} and $\mathbf{s}^{\tau}$ are respectively the empirical and simulated medians of the wealth to permanent income ratio for age group τ. A drawback of proceeding in this way is that it treats the empirically estimated medians as though they reflected perfect measurements of the truth. Imagine, however, that one of the age groups happened to have (in the consumer survey) four times as many data observations as another age group; then we would expect the median to be more precisely estimated for the age group with more observations; yet (4) assigns equal importance to a deviation between the model and the data for all age groups.

We can get around this problem (and a variety of others) by instead minimizing a slightly more complex object:

$$\min_{\xi} \sum_{i}^{N} \omega_i \, |\varsigma_i^{\tau} - \mathbf{s}^{\tau}(\xi)| \tag{5}$$

{eq:StructEstim}

where ω_i is the weight of household i in the entire population,[5] and ς_i^{τ} is the empirical wealth to permanent income ratio of household i whose head belongs to age group τ. ω_i is needed because unequal weight is assigned to each observation in the Survey of Consumer Finances (SCF). The absolute value is used since the formula is based on the fact that the median is the value that minimizes the sum of the absolute deviations from itself.

With this in mind, we turn our attention to the computation of the weighted median wealth target moments for each age cohort using this data from the 2004 Survery of Consumer Finances on household wealth. The objects necessary to accomplish this task are `weighted_median` and `get_targeted_moments`. The actual data are taken from several waves of the SCF and the medians and means for each age category are plotted in figure 2. More details on the SCF data are included in appendix A.

We now turn our attention to the the two key functions in this section of the code file. The first, `simulate_moments`, executes the solving (`solve`) and simulation (`simulation`) steps for the defined life-cycle agent. Subsequently, the function uses the agents' tracked levels of wealth based on their optimal consumption behavior to compute and store the simulated median wealth to income ratio for each age cohort. The second function,

[5]The Survey of Consumer Finances includes many more high-wealth households than exist in the population as a whole; therefore if one wants to produce population-representative statistics, one must be careful to weight each observation by the factor that reflects its "true" weight in the population.

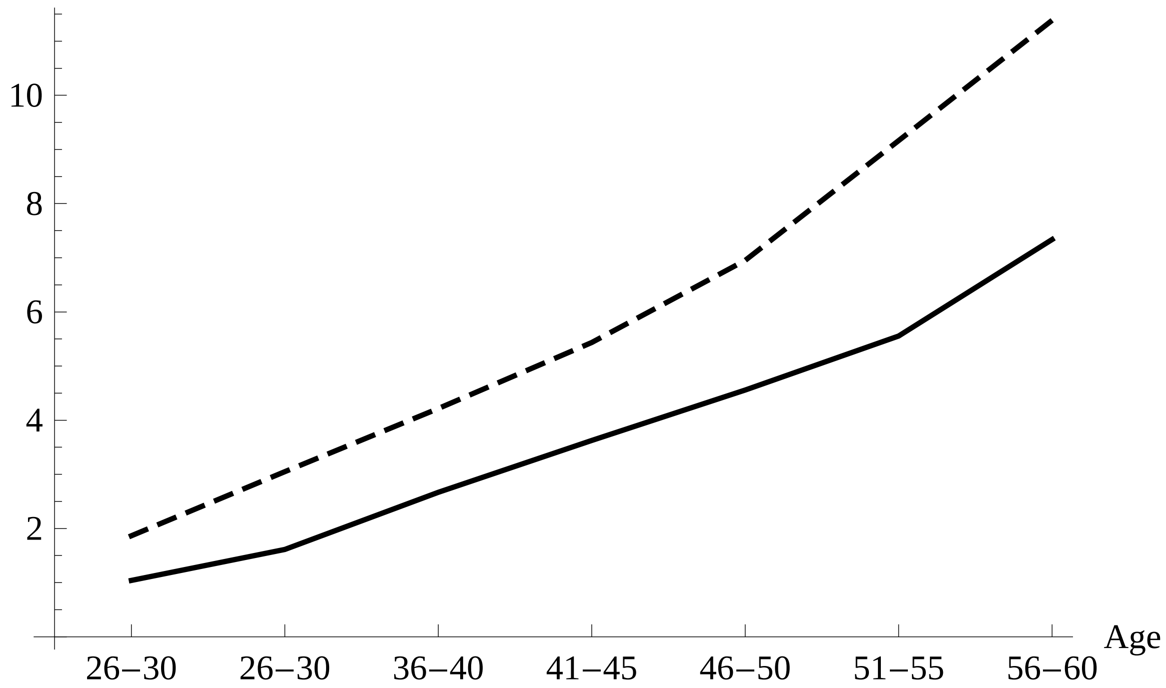

Figure 2 Wealth to Permanent Income Ratios from SCF (means (dashed) and medians (solid))

{fig:MeanMedianSC

`smmObjectiveFxn`, calls the `simulate_moments` function to create the objective function described in (5), which is necessary to perform the SMM estimation.

Thus, for a given pair of the parameters to be estimated, the single call to the function `smmObjectiveFxn` executes the following:

1. solves for the consumption functions for the life-cycle agent
2. simulates the data and computes the simulated medians
3. returns the value of equation (5)

We delegate the task of finding the coefficients that minimize the `smmObjectiveFxn` function to the `minimize_nelder_mead` function, which is defined elsewhere and called in the second part of this file. This task can be quite slow and rather problematic if the `smmObjectiveFxn` function has very flat regions or sharp features. It is thus wise to verify the accuracy of the solution, for example by experimenting with a variety of alternative starting values for the parameter search.

The final object defined in this first part of the `StructEstimation.py` file is `calculateStandardErrorsByBootstrap`. As the name suggsts, the purpose of this function is to compute the standard errors by bootstrap.[6] This involves:

1. drawing new shocks for the simulation

[6]For a treatment of the advantages of the bootstrap see Horowitz (2001)

2. drawing a random sample (with replacement) of actual data from the SCF

3. obtaining new estimates for ρ and $\beth$

We repeat the above procedure several times (`Bootstrap`) and take the standard deviation for each of the estimated parameters across the various bootstrap iterations.

1.2.1 An Aside to Computing Sensitivity Measures

{subsubsec:sensmea

A common drawback in commonly used structural estimation procedures is a lack of transparency in its estimates. As Andrews, Gentzkow, and Shapiro (2017) notes, a researcher employing such structural empirical methods may be interested in how alternative assumptions (such as misspecification or measurement bias in the data) would "change the moments of the data that the estimator uses as inputs, and how changes in these moments affect the estimates." The authors provide a measure of sensitivity for given estimator that makes it easy to map the effects of different assumptions on the moments into predictable bias in the estimates for non-linear models.

In the language of Andrews, Gentzkow, and Shapiro (2017), section 1 is aimed at providing an estimator $\xi = \{\rho, \beth\}$ that has some true value ξ_0 by assumption. Under the assumption a_0 of the researcher, the empirical targets computed from the SCF is measured accurately. These moments of the data are precisely what determine our estimate $\hat{\xi}$, which minimizes (5). Under alternative assumptions a, such that a given cohort is mismeasured in the survey, a different estimate is computed. Using the plug-in estimate provided by the authors, we can see quantitatively how our estimate changes under these alternative assumptions a which correspond to mismeasurement in the median wealth to income ratio for a given age cohort.

1.3 Results

The second part of the file `StructEstimation.py` defines a function `main` which produces our ρ and $\beth$ estimates with standard errors using 10,000 simulated agents by setting the positional arguments `estimate_model` and `compute_standard_errors` to true.[7] Results are reported in Table 2.[8]

The literature on consumption and saving behavior over the lifecycle in the presenece of labor income uncertainty[9] warns us to be careful in disentangling the effect of time preference and risk aversion when describing the optimal behavior of households in this setting. Since the precautionary saving motive dominates in the early stages of life, the coefficient of relative risk aversion (as well as expected labor income growth) has

[7]The procedure is: First we calculate the ρ and $\beth$ estimates as the minimizer of equation (5) using the actual SCF data. Then, we apply the `Bootstrap` function several times to obtain the standard error of our estimates.

[8]Differently from Cagetti (2003) who estimates a different set of parameters for college graduates, high school graduates and high school dropouts graduates, we perform the structural estimation on the full population.

[9]For example, see Gourinchas and Parker (2002) for an exposition of this.

Table 2 Estimation Results

{tab:EstResults}

ρ	$\beth$
3.69	0.88
(0.047)	(0.002)

a larger effect on optimal consumption and saving behavior through their magnitude relative to the interest rate. Over time, life-cycle considerations (such as saving for retirement) become more important and the time preference factor plays a larger role in determining optimal behavior for this cohort.

Using the positional argument `compute_sensitivity`, Figure 3 provides a plot of the plug-in estimate of the sensitivity measure described in 1.2.1. As you can see from the figure the inverse relationship between ρ and $\beth$ over the life-cycle is retained by the sensitivity measure. Specifically, under the alternative assumption that *a particular cohort is mismeasured in the SCF dataset*, we see that the y-axis suggests that our estimate of ρ and $\beth$ change in a predictable way.

Suppose that there are not enough observations of the oldest cohort of households in the sample. Suppose further that the researcher predicts that adding more observations of these households to correct this mismeasurement would correspond to a higher median wealth to income ratio for this cohort. In this case, our estimate of the time preference factor should increase: the behavior of these older households is driven by their time preference, so a higher value of $\beth$ is required to match the affected wealth to income targets under this alternative assumption. Since risk aversion is less important in explaining the behavior of this cohort, a lower value of ρ is required to match the affected empirical moments.

To recap, the sensitivity measure not only matches our intuition about the inverse relationship between ρ and $\beth$ over the life-cycle, but provides a quantitative estimate of what would happen to our estimates of these parameters under the alternative assumption that the data is mismeasured in some way.

By setting the positional argument `make_contour_plot` to true, Figure 4 shows the contour plot of the `smmObjectiveFxn` function and the parameter estimates. The contour plot shows equally spaced isoquants of the `smmObjectiveFxn` function, i.e. the pairs of ρ and $\beth$ which lead to the same deviations between simulated and empirical medians (equivalent values of equation (5)). Interestingly, there is a large rather flat region; or, more formally speaking, there exists a broad set of parameter pairs which leads to similar simulated wealth to income ratios. Intuitively, the flatter and larger is this region, the harder it is for the structural estimation procedure to precisely identify the parameters.

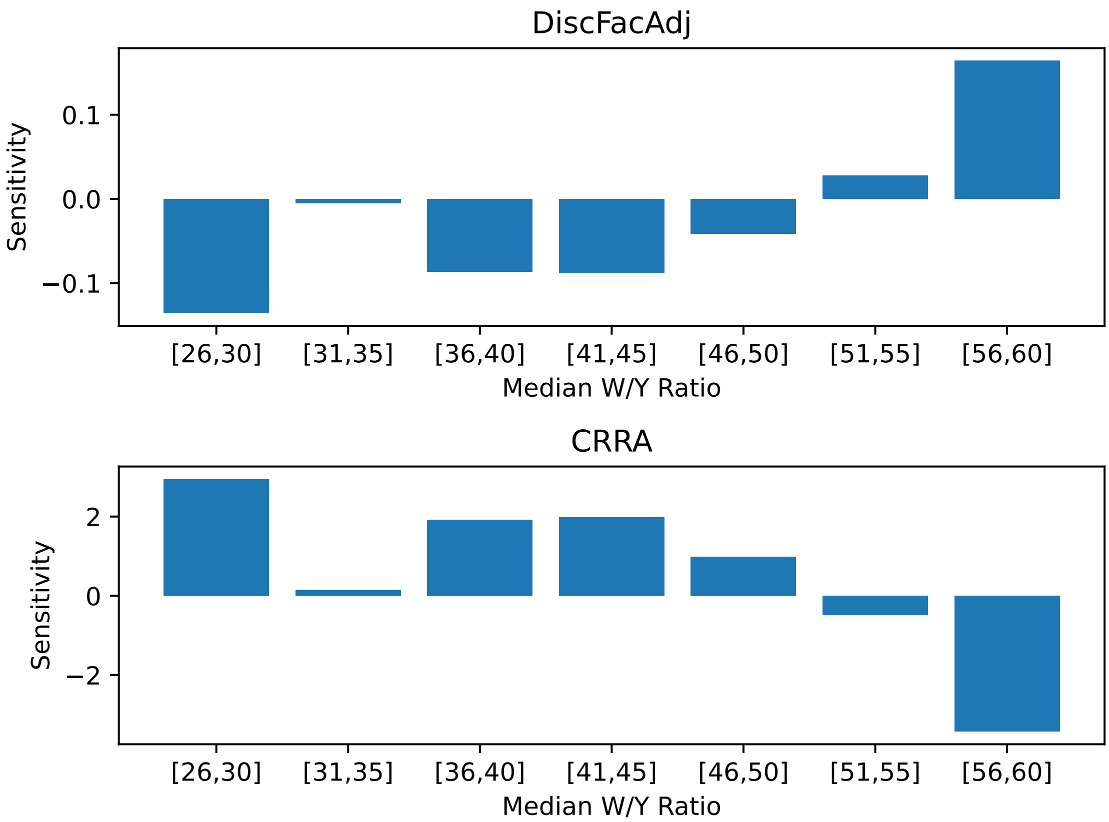

Figure 3 Sensitivty of Estimates $\{\rho, \beth\}$ regarding Alternative Mismeasurement Assumptions.

{fig:PlotSensitivityI

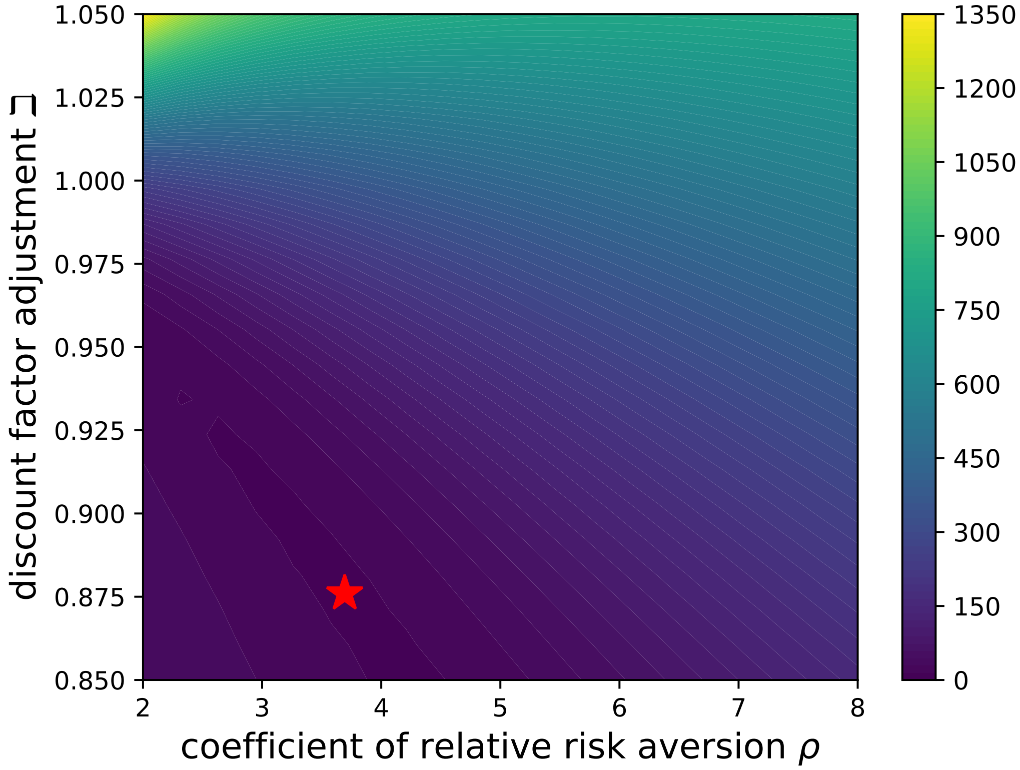

Figure 4 Contour Plot (larger values are shown lighter) with $\{\rho, \beth\}$ Estimates (red dot).

{fig:PlotContourMe