# Empirical Asset Pricing - Midterm Exam

Maximilian Huber

March 30, 2018

## 1 Estimating Asset Pricing Models Using GMM

| | | 1st Stage | | Final Stage | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Specification | $\hat{\gamma}$ | $\hat{\psi}$ | $\hat{\gamma}$ | $\hat{\psi}$ | (std. err $\hat{\gamma}$) | (std. err $\hat{\psi}$) | $J_T$ stat | % p-val |
| | 1 | | 0.2001 | | 0.2136 | | 0.04265 | 3.7297 | 29.21 |
| | 2 | | 0.2415 | | 1.5145 | | 0.3381 | 15.645 | 0.35 |
| | 3 | | 0.1721 | | 0.1584 | | 0.0244 | 18.474 | 0.24 |
| | 4 | 7.8984 | 0.1605 | 20.914 | 0.1184 | 11.887 | 0.0171 | 18.631 | 0.09 |
| | 5 | 7.5051 | 0.1603 | 6.0217 | 0.169 | 1.8784 | 0.0222 | 29.212 | 2.25 |
| 6 | 2 step | 14.229 | 0.1318 | 13.863 | 0.1339 | 1.7408 | 0.0073 | 10.886 | 92.76 |
| 6 | $\infty$ step | 14.229 | 0.1318 | 0.2980 | 3.7015 | 0.0771 | 1.2479 | 10.843 | 92.90 |
| 7 | $\hat{\psi} > 1$ | | 1.3825 | | 1.1019 | | 0.0129 | 29.459 | 5.90 |
| 7 | $\hat{\psi} < 1$ | | 0.9134 | | 0.8656 | | 0.0143 | 52.371 | 0.00 |
| | 8 | 0.9987 | 0.9993 | 1.0503 | 1.0742 | 0.8351 | 1.312 | 36.437 | 0.93 |
| | 9 | 0.9986 | 0.9992 | 0.9999 | 0.9999 | 0.8748 | 0.5077 | 40.1946 | 15.15 |

My estimation routine evolves around $\beta$, $\psi$, and $\gamma$. Usually I choose a HAC variance estimation with 4 lags, i.e. +/- one year.

### Data Preperation

- I divided consumption by population in order to get per capita consumption.
- In order to adjust the industry returns for inflation one needs to derive an inflation measure from the PCE deflator. I assume that the deflator is presented in end-of-quarter values. Therefore I take the backward difference in order to calculate inflation within a given quarter.

### General Remarks

- From task 5 on, the HAC variance matrix of the moments has a condition number of over 1e8. This means we may loose up to 8 digits of precision! I cannot say to what extent this is a serious problem.

### 1.1 $\psi$ and Industry Stocks

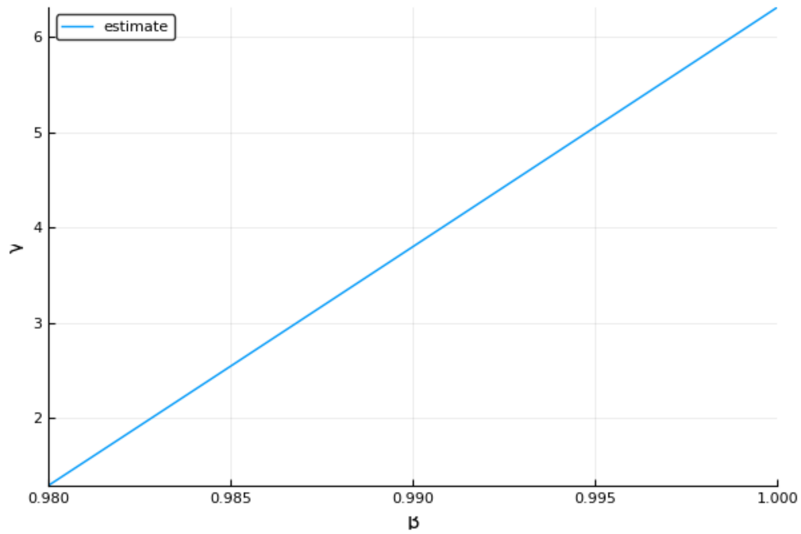
With $\beta$ fixed at 1 and CRRA utility the estimation yields a $\gamma$ around 5. This number is plausible, but using only industry portfolios might not be a good choice of test assets.

### 1.2 $\psi$, Industry Stocks and T-Bill

Including a risk-free asset depresses the risk aversion coefficient. The second stage apparently gives more weight to the moments generated from test assets with a lower excess return, likely the T-Bill.

### 1.3 $\psi$ and Value Portfolios

The relative risk aversion increases, the model is rejected by the J-test.
There appears to be a linear relation in the $\beta$ and the estimate for $\gamma$:




The relative risk aversion $\gamma$ subsumes parts of the impatience of the representative agent!

### 1.4 $\gamma$, $\psi$ and Value Portfolios

The second column of the gradient of the sample moment function is almost zero, hence the second argument, $\gamma$, is not well estimated, due to lack of slope. A different value of $\gamma$ does not change the sample moments much.

$$\nabla g_n(\theta) = \begin{bmatrix} 0.9951 & 0.0012 \\ 0.962 & 0.001 \\ 0.9659 & 0.001 \\ 0.9344 & 0.0009 \\ 0.9163 & 0.0008 \\ 0.9403 & 0.0009 \end{bmatrix}$$

### 1.5 $\gamma$, $\psi$ and Value Portfolios with lagged Consumption and Wealth Return

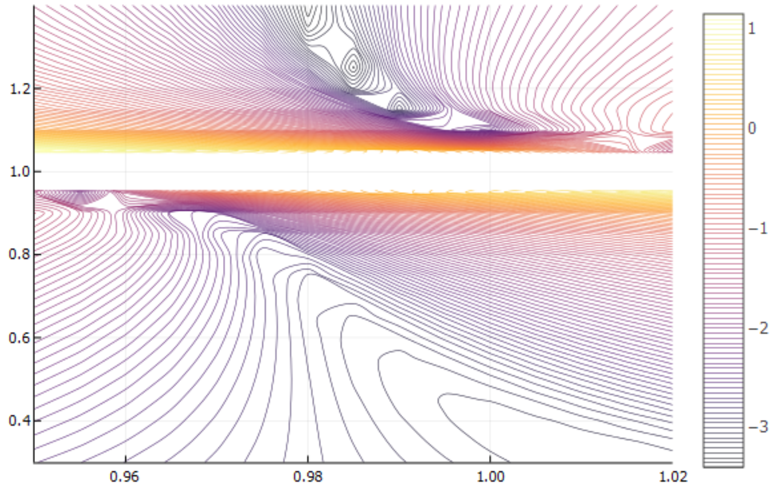
I had to restrict the GMM iteration to be a two-step procecure only. Iterating to convergence yielded an extremly high $\psi$.
Risk aversion and IES seem plausible!

### 1.6 $\gamma$, $\psi$ and Value Portfolios and T-Bill with lagged Consumption, Wealth Return

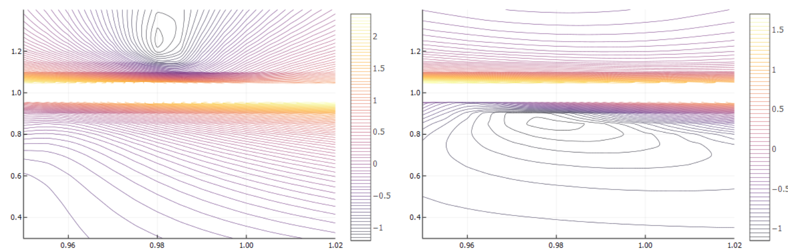
Iterating until convergence yields implausibly high IES.
The risk aversion increases and the IES drops compared to #5. I thought adding the risk-free asset would lay more focus on the risk-free rate puzzle (the inverse of the equity premium puzzle) and the IES would increase!

### 1.7 $\beta$, $\psi$ and Value Portfolios and T-Bill with lagged Consumption and Wealth Return

There are two candidates for the global optimum.
The solution with $\psi > 1$ has a lower objective function value, however if I restrict $\psi < 1$ in the first stage I obtain an alternative solution. The first stage objective function with $\beta$ on the x-axis and $\psi$on the y-axis looks like:




Evaluating the HAC variance matrix and updating the objective function on the two different ($\psi > 1$ on the left) first stage candidate estimates yields:



- Restricting $\psi > 1$ yields: The estimated $\beta$ is 0.9809 in the first stage and 0.9781 after the final stage with a standard error of 0.002.
- Restricting $\psi < 1$ yields: The estimated $\beta$ is 0.9648 in the first stage and 0.9782 after the final stage with a standard error of 0.002.

## 1.8 $\beta$, $\gamma$, $\psi$ and Value Portfolios and T-Bill with more lagged Consumption

I had to switch to a gradient decent algorithm in the first stage. Then $\beta$ is 0.982 in the first stage and 0.9845 after the final stage.

## 1.9 $\beta$, $\gamma$, $\psi$ and Value Portfolios and T-Bill with more lagged Consumption and Wealth Return

$\beta$ is 0.9819 in the first stage and 0.9844 after the final stage.
In this specification there appears to be more curvature in the $\psi$ direction and less curvature in the $\gamma$ direction compared to #8.

# 2 Forecasting Excess Stock Returns

## 2.1 Regressions

**Remarks on the Methodology**

- I chose the Newey-West lag to be the number of periods that are being aggregated minus one, e.g. the horizon one variance matrix is estimated without a lag. The reason is that I want to capture any autocovariance that is generated by aggregating multiple returns.
- I calculate $R^2 = 1 - \frac{SS_{residual}}{SS_{total}}$ and $\bar{R}^2 = 1 - \frac{(1-R^2)*(T-1)}{T-k-1}$ where $T$ is the number of observations and $k$ is the number of regressors.
- $\mathbb{V}_{asympt}(\hat{\beta}) = (\frac{1}{T}X'X)^{-1}S(\frac{1}{T}X'X)^{-1}$ where S is the Newey-West estimate of the error variance matrix

The pattern of the $\bar{R}^2$ is increasing with horizon in all specifications!

#### 2.1.a *cay* only

$r_{t+h,t}$ on $cay_t$

| Horizon | coeff | t-stat | $\bar{R}^2$ |
|---|---|---|---|
| 1 | 0.6397 | 2.5384 | 0.0143 |
| 2 | 1.3309 | 2.9912 | 0.0361 |
| 3 | 2.0517 | 3.2017 | 0.0622 |
| 4 | 2.7777 | 3.3198 | 0.0903 |
| 5 | 3.4264 | 3.3825 | 0.1131 |
| 6 | 4.2060 | 3.6871 | 0.1455 |
| 7 | 5.0345 | 4.0623 | 0.1823 |
| 8 | 5.7702 | 4.4798 | 0.2194 |

*cay* appears to be a significant predictor of return in all horizons.

- Remark on endogeneity: This regression suffers somewhat from an endogeneity problem! The $y$ part of *cay* is arguably the left-hand-side of the regression, namely market return. Therefore assuming that consumption does not react much the a return shock, $\epsilon \uparrow$ implies $cay \downarrow$. Hence the coefficient is downwardly biased, hence my estimates are convervative and the argument of *cay* being a predictor goes through!

#### 2.1.b Dividend-Price Ratio

$r_{t+h,h}$ on $dp_t$

| Horizon | coeff | t-stat | $\bar{R}^2$ |
|---|---|---|---|
| 1 | 0.0265 | 2.2922 | 0.0115 |
| 2 | 0.0556 | 2.7588 | 0.0309 |
| 3 | 0.0824 | 2.7829 | 0.0490 |
| 4 | 0.1067 | 2.7191 | 0.0650 |
| 5 | 0.1303 | 2.6817 | 0.0806 |
| 6 | 0.1533 | 2.6689 | 0.0951 |
| 7 | 0.1749 | 2.6433 | 0.1086 |
| 8 | 0.1914 | 2.5952 | 0.1196 |

The Dividend-Price ratio appears to be a significant predictor of return in all horizons.

- Remark on endogeneity: This regression suffers from the Mankiw-Shapiro/Stambaugh bias. Since $dp_t$ is highly persistence and because a positive shock to $r_{t+h,t}$ also decreases $dp_{t+h+1}$ this leads to an upward bias of the coefficient on $dp_t$. This notebook illustrates the bias.

### 2.1.c $cay$, Dividend-Price Ratio, $RREL$, $DEF$ and $TRM$

| | $r_{t+h,h}$ on | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $cay_t$ | | $dp_t$ | | $RREL$ | | $DEF$ | | $TRM$ | | |
| Horizon | coeff | t-stat | coeff | t-stat | coeff | t-stat | coeff | t-stat | coeff | t-stat | $\bar{R}^2$ |
| 1 | 0.6394 | 2.3885 | 0.0273 | 1.7708 | -1.1081 | -1.5536 | 0.0050 | 0.3461 | -0.0982 | -1.3134 | 0.0392 |
| 2 | 1.3288 | 2.9105 | 0.0538 | 2.0437 | -1.6959 | -1.1158 | 0.0132 | 0.5178 | -0.0505 | -0.4729 | 0.0718 |
| 3 | 2.0331 | 3.1947 | 0.0777 | 2.0667 | -2.3036 | -1.0544 | 0.0173 | 0.5181 | -0.0138 | -0.095 | 0.1132 |
| 4 | 2.7343 | 3.3852 | 0.0995 | 2.0458 | -2.6611 | -0.9868 | 0.0165 | 0.4317 | -0.0174 | -0.1231 | 0.1505 |
| 5 | 3.3182 | 3.4313 | 0.1222 | 2.0436 | -3.6028 | -1.3385 | 0.0077 | 0.1855 | -0.0033 | -0.0233 | 0.1842 |
| 6 | 4.0528 | 3.6231 | 0.1371 | 1.9359 | -2.8338 | -1.168 | 0.0131 | 0.3007 | 0.0196 | 0.1198 | 0.2114 |
| 7 | 4.7988 | 3.8434 | 0.1470 | 1.7983 | -2.4969 | -1.163 | 0.0153 | 0.3203 | 0.1413 | 0.897 | 0.2457 |
| 8 | 5.4648 | 4.1621 | 0.1559 | 1.7269 | -2.6224 | -1.3044 | 0.0101 | 0.1951 | 0.1593 | 1.0014 | 0.28 |

While $cay$ appears to predict on all horizons equally strongly, $dp$'s strenght decays, but it is still siginificant at some horizons. $DEF$ has questionable predicitve power and $TRM$ predicts in the very short term and faces a reversal in the longer term, but the coefficients are not significant. This is interesting, an long-term interest rate hike leads to short term losses, but long term gains. $RREL$ appreas to predict, but not significantly. Short-term interest rate hikes lead to losses.

## 2.2 VAR

- Remark: In the following derivation I do not follow the slides from the lecture. My approachs seems very straight forward to me.

The return $h$ periods ahead, according to the VAR system is:

$$r_{t+h} = e^2 A^h x_t + e^2 A^{h-1} \epsilon_{t+1} + ... + e^2 \epsilon_{t+h}$$

where $e^2 = \big(0, 1, 0\big)$.
Summing over the terms $r_{t+1}, \ldots, r_{t+h}$ gives rise to the multi-period return in the next $h$ periods starting from $t$:

$$r_{t+h,t} = e^2\big((A^h + \cdots + A^1)x_t\big) + e^2\big(A^{h-1}\epsilon_{t+1} + \cdots + A^0(\epsilon_{t+h} + \cdots + \epsilon_{t+1})\big)$$

Mind that $x_t$ is known and the errors $\epsilon$ have mean zero.
The R-squared statistic is defines as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

where

$$\begin{aligned} SS_{tot} &= \sum_{t=1}^{T-h} (r_{t+h,t} - \bar{r}_h)^2 \\ &= \sum_{t=1}^{T-h} \Big(e^2(A^h + \cdots + A^1)x_t + e^2\big(A^{h-1}\epsilon_{t+1} + \cdots + A^0(\epsilon_{t+h} + \cdots + \epsilon_{t+1})\big) - e^2(A^h + \cdots + A^1)(x_t - \bar{x}_t)\Big)^2 \\ &= \sum_{t=1}^{T-h} \Big(e^2\big(A^{h-1}\epsilon_{t+1} + \cdots + A^0(\epsilon_{t+h} + \cdots + \epsilon_{t+1})\big) + e^2(A^h + \cdots + A^1)\bar{x}_t\Big)^2 \end{aligned}$$

and

$$\begin{aligned} SS_{res}(A) &= \sum_{t=1}^{T-h} (r_{t+h,t} - \hat{r}_{t+h,t})^2 \\ &= \sum_{t=1}^{T-h} \Big(e^2\big(A^{h-1}\epsilon_{t+1} + \cdots + A^0(\epsilon_{t+h} + \cdots + \epsilon_{t+1})\big)\Big)^2 \end{aligned}$$

.
I coded up $R^2$ as a function of the vectorized $A$ matrix. Then I used automatic differentiation to calculate its derivative and applied the Delta Method.
The estimated $A$ matrix with standard errors and the imputed R-square statistic is:

| $A$ matrix | | |
|---|---|---|
| | $r_t$ | $cay_t$ |
| $r_{t+1}$ | 0.0871068<br>(0.0621) | 0.654659<br>(0.0060) |
| $cay_{t+1}$ | 0.00618562<br>(0.2672) | 0.912555<br>(0.0260) |

| $R^2$ | | |
|---|---|---|
| Horizon | coeff | (std. err) |
| 2 | 0.0524 | 0.0014 |
| 4 | 0.1090 | 0.0055 |
| 8 | 0.2364 | 0.0189 |
| 12 | 0.3500 | 0.0320 |
| 16 | 0.4330 | 0.0426 |

## 3 Summary Statistics

The Campbell Shiller decomposition starts out from the defininition of the return on a stock: $R_{t+1} = \frac{P_{t+1}+D_{t+1}}{P_t}$. Taking logs:

$$r_{t+1} = p_{t+1} - p_t + \log\Big(1 + \exp(\underbrace{d_{t+1} - p_{t+1}}_{dp_{t+1}})\Big)$$

The Taylor approximation of the last term around $\bar{dp}$ is:

$$\log\Big(1 + \exp(dp_{t+1})\Big) \approx \underbrace{\log\Big(1 + \exp(\bar{dp})\Big)}_{\kappa} + \underbrace{\frac{\exp(\bar{dp})}{1 + \exp(\bar{dp})}}_{1-\rho}\Big(dp_{t+1} - \bar{dp}\Big) + o\Big(dp_{t+1} - \bar{dp}\Big)^2$$

where $\tilde{dp}$ is between $dp_{t+1}$ and $\bar{dp}$ (this is the Lagrange form of the remainder).
Hence,

$$pd_t = \rho pd_{t+1} + \kappa - (1-\rho)\bar{dp} + \Delta d_{t+1} - r_{t+1}$$

subsituting forward and using the no-bubbles assumption yields:

$$pd_t = \frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \Big(\Delta d_{t+j+1} - r_{t+j+1}\Big)$$

or taking conditional expectations:

$$pd_t = \frac{\kappa}{1-\rho} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j \Big(\Delta d_{t+j+1} - r_{t+j+1}\Big)$$

This equation allows me to do the variance decomposition. Using the unpredictability of dividend growth by the price-dividend ration (and the law of iterated expectations):

$$\mathbb{V}\Big[pd_t\Big] = \underbrace{\mathbb{C}\Big[\sum_{s=1}^{\infty} \rho^{s-1}\mathbb{E}_t\big[\Delta d_{t+s}\big], pd_t\Big]}_{=0} + \mathbb{C}\Big[-\sum_{s=1}^{\infty} \rho^{s-1}\mathbb{E}_t\big[r_{t+s}\big], pd_t\Big]$$

If the price-dividend ration has positive variance (which it has in the data), then there is predictability of future returns by the price-dividend ratio.
Now to answer the question, while $pd_t$ does forecast a geometric sum of returns, it does not give specific information on the time series of the returns. For example, $\mathbb{E}_t\big[r_{t+s}\big]$ could be constant in $s$, or decreasing, or hump-shaped. Therefore, the price-dividend ratio is a summary statistic, but is by no means sufficient.