

**THE STOCHASTIC DISCOUNT FACTOR AND THE GENERALIZED  
METHOD OF MOMENTS.**

**BY  
ENI KOCI**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
PROFESSIONAL MASTERS DEGREE IN FINANCIAL MATHEMATICS  
WORCESTER POLYTECHNIC INSTITUTE**

**2006**

**APPROVED BY PROF. LUIS J. ROMAN**

---

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	3
<b>INTRODUCTION</b> .....	4
<b>CHAPTER 1:</b>	
1.1 THE STOCHASTIC DISCOUNT FACTOR .....	7
1.2 SDF AND THE WEIGHTED PORTFOLIOS.....	9
1.3 SDF AND THE CONSUMPTION BASED MODEL.....	10
1.4 SDF AND THE COMPLETE MARKETS.....	14
<b>CHAPTER 2:</b>	
2.1 THE GENERALIZED METHOD OF MOMENTS .....	15
<b>CHAPTER 3:</b>	
3.1 THE LINEAR FACTOR MOMENTS .....	22
<b>CHAPTER 4:</b>	
4.1 NUMERICAL IMPLEMENTATION FOR CAPM.....	25
4.2 CONCLUSION.....	29
REFERENCES.....	30

## **ABSTRACT**

The fundamental theorem of asset pricing in finance states that the price of any asset is its expected discounted payoff. Ideally, the payoff is discounted by a factor, which depends on parameters present in the market, and it should be unique, in the sense that financial derivatives should be able to be priced using the same discount factor. In theory, risk neutral valuation implies the existence of a positive random variable, which is called the stochastic discount factor and is used to discount the payoffs of any asset. Apart from asset pricing another use of stochastic discount factor is to evaluate the performance of the managers of hedge funds. Among many methods used to evaluate the stochastic discount factor, generalized method of moments has become very popular. In this paper we will see how generalized method of moments is used to evaluate the stochastic discount factor on linear models and the calculation of stochastic discount factor using generalized method of moments for the popular model in finance, CAPM.

## Introduction

The stochastic discount factor models are used to evaluate the performance of actively managed portfolios. Hedge fund returns are mostly the result of dynamic trading strategies that are being implemented by fund managers. As the result hedge funds have time varying exposures to risk, which makes traditional approaches to performance evaluation not applicable. For an investor the problem is to choose from a large universe of investment possibilities and for this reason and other ones measuring the performance of fund managers is a very important research problem in finance. One approach to evaluate the performance of hedge fund managers is to use the stochastic discount factor. The performance of hedge funds is evaluated under the assumption that there are no arbitrage opportunities in financial markets. This assumption implies that there is a positive stochastic discount factor that can price all assets. Under such assumption the price of any asset is given by the expected value of future payoff of the asset adjusted by the stochastic discount factor and in particular the gross return of any asset will satisfy:

$$E_t[m_{t+1}R_{i,t+1}] = 1.$$

Where  $m_{t+1}$  is the stochastic discount factor at time  $t + 1$  and  $E_t$  is the expectation conditioned on the information available up to time  $t$ . The stochastic discount factor is a positive random variable that adjusts the future payoffs for passage of time and uncertainty and as we already mentioned its presence is guaranteed by the absence of arbitrage.

In order to evaluate the performance of hedge funds it is necessary to have some benchmark. Usually as benchmarks are taken portfolios of primitive assets. A conditional model of performance evaluation should assign a value of zero to any dynamic trading strategy that involves the available benchmarks as long as the strategy is based on data publicly available. If we denote by  $R_{t+1}$  the vector of gross returns on primitive assets at time  $t + 1$  and  $W$  the vector of amount invested in each asset. In what follows we will consider column vectors unless otherwise specified, and  $U'$  denotes the transpose of vector  $U$ . Then we want a portfolio  $R_{t+1}'W$  such that

$$E_t[R_{t+1}'R_{t+1}W] = \hat{1},$$

where  $\hat{1}$  is a vector of ones.

The above expression is used to find the optimal weights which is given by

$$W = E_t[R_{t+1}R_{t+1}']^{-1}\hat{1}$$

Denote the total payoff of this portfolio as  $R_{t+1}^* = R_{t+1}'W$ . Then we have

$$E_t[R_{t+1}R_{t+1}^*] = \hat{1}$$

from which it follows

$$E_t[R_{i,t+1}R_{t+1}^*] = 1 \quad (*)$$

for all  $i$ . Thus in this case the stochastic discount factor ( $R_{t+1}^*$ ) is represented by a portfolio of primitive assets, where the weights are estimated so at least the primitive assets themselves are priced by the (\*) model.

If a portfolio satisfies the above equation then a neutral performance is given to the manager, if the left hand side is greater than one the portfolio is believed to have an abnormal or positive performance and of course if the left hand side is less than one

than the performance is negative. To evaluate the performance of hedge funds we use the expression:

$$E_t \begin{bmatrix} R_{t+1}^* R_{1,t+1} \\ \dots \\ R_{t+1}^* R_{k,t+1} \\ R_{t+1}^* R_{h,t+1} \end{bmatrix} - \begin{bmatrix} 1 \\ \dots \\ 1 \\ 1 + \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \dots \\ 0 \\ 0 \end{bmatrix}$$

$R_{i,t+1}$  is the return on the primitive asset  $i$  for  $i = 1, 2, \dots, k$ ;  $R_{h,t+1}$  is the return on hedge fund index or manager and  $\alpha$  is the measure of investment risk adjusted excess return, which is to be estimated. If  **$\alpha$  is positive** the hedge fund index has outperformed investment strategies that involve dynamic strategies which use primitive assets and are based on public available information. If  **$\alpha$  is negative** we say that the performance is negative and if  **$\alpha$  is zero** the performance is said to be neutral.

However in most cases neither we know which is the exact form of the stochastic discount factor nor all the market variables involved in it. In this project we will use the generalized method of moments to find information about the stochastic discount factor.

# Chapter 1.

## Section 1.1 The Stochastic Discount Factor.

In general, the basic equation of asset pricing can be written as:

$$p_{i,t} = E_t[m_{t+1}x_{i,t+1}] \quad (1)$$

where  $p_{i,t}$  is the price of the asset  $i$  at time  $t$ ,  $E_t$  is the conditional expectation conditioned on information up to day  $t$ ,  $x_{i,t+1}$  is the random payoff on asset  $i$  at time  $t + 1$  and  $m_{t+1}$  is the stochastic discount factor at time  $t + 1$ . The stochastic discount factor is a random variable whose realized values are always positive.

If there is no uncertainty the stochastic discount factor is a constant that converts into the present value the expected payoffs. In this case the asset pricing formula can be written as:

$$p_{i,t} = \frac{1}{R^f} x_{i,t+1} ,$$

$R^f$  is the gross risk-free rate. In this case  $\frac{1}{R^f}$  is the discount factor. Riskier assets have lower prices than risk free assets and they can be valued using formula

$$p_{i,t} = \frac{1}{R^i} E_t(x_{i,t+1}) ,$$

where  $\frac{1}{R^i}$  is the risk-adjusted discount factor for asset  $i$ .

There are two important theorems that give the conditions for the existence of the stochastic discount factor. Before we give their definitions we will talk about the law of one price and absence of arbitrage. The *law of one price* states that if two

portfolios have the same payoffs in every state, then they must have the same price. The violation of this law will create the opportunity for arbitrage opportunity, as one investor could sell the expensive version and buy the cheap version of the same portfolio. *Absence of arbitrage* implies that if a payoff  $A$  is not smaller than a payoff  $B$ , and sometimes  $A$  is greater, the price of  $A$  must be greater than the price of  $B$ .

*Theorem 1:* In complete markets, no arbitrage and the law of one price imply that there exists a unique  $m_{t+1} > 0$  such that  $P_{i,t} = E_t[m_{t+1}x_{i,t+1}]$ .

*Theorem 2:* No arbitrage and the law of one price imply the existence of a strictly positive discount factor,  $m_{t+1} > 0$ ,  $P_{i,t} = E_t[m_{t+1}x_{i,t+1}]$  for every  $x$ . (For a proof of both theorems see Asset Pricing [1] chapter 4).

So the latter theorem assures the existence of the stochastic discount factor  $m > 0$ , but it does not say that  $m$  is unique therefore it does not say that every discount factor  $m$  must be positive.

However the second theorem shows that we can use stochastic discount factors without assuming the markets to be complete which is a very strong assumption.

Going back to equation (1) if  $p_{i,t}$  is not zero we can divide both sides of equation (1) by  $p_{i,t}$  and we get

$$1 = E_t[m_{t+1}R_{i,t+1}] \quad (2)$$

Where  $R_{i,t+1} = \frac{x_{i,t+1}}{p_{i,t}}$  is the gross return of asset  $i$  at time  $t+1$ . When we derive the

equation (2) we get  $p_{i,t}$  inside the conditional expectation as a constant because  $p_{i,t}$  is known to us at time  $t$ . The equation (2) for the asset pricing formula is the formula form mostly used in empirical work.



## Section 1.2. The stochastic discount factor and the weighted portfolios.

The major assumption we made to use the stochastic discount factor in asset pricing is the absence of arbitrage opportunities in the financial markets. Under this assumption the gross return of any risky asset will satisfy the equation:

$$1 = E_t[m_{t+1}R_{i,t+1}]$$

If there are no arbitrage opportunities it can be shown that a portfolio of available assets can be chosen to mimic the behavior of the stochastic discount factor (see Hansen, Richard [9]). Let  $R_{t+1}$  be the vector of gross returns of primitive assets and let  $W$  be the vector of weights (amount invested in each asset). We want a portfolio  $R'_{t+1}W$  (where ' stands for transpose) such that

$$E_t[R_{t+1}R'_{t+1}W] = \hat{1} \quad (1)$$

where  $\hat{1}$  is a vector of ones. Using equation (1) to solve for the optimal weights we get

$$W = E_t[R_{t+1}R'_{t+1}]^{-1}\hat{1}.$$

If we define the payoff of this portfolio as  $R^*_{t+1} = R'_{t+1}W$ , then the return on any portfolio of the primitive assets will satisfy equation

$$E_t[R^*_{t+1}R'_{i,t+1}W] = 1.$$

It is obvious that in this case the stochastic discount factor is of the form

$$m_{t+1} = R^*_{t+1} = R'_{t+1}W.$$

So a weighted portfolio is a special case of a stochastic discount factor.

### Section 1.3 The Stochastic Discount Factor and the Consumption Based Model.

In this section we will derive the stochastic discount factor in a consumption based model. We will need to find the value at time  $t$  of a payoff  $x_{t+1}$ , that is, the stock price  $p_{t+1}$  at time  $t+1$  plus the dividend  $d_{t+1}$  issued at time  $t+1$ . So  $x_{t+1} = p_{t+1} + d_{t+1}$  and  $x_{t+1}$  is random variable. One approach to find the value of the payoff uses the utility function, which is a mathematical formalism, used to model investors over current and future values of their consumption, therefore we have

$$U(c_t, c_{t+1}) = u(c_t) + \beta E_t[u(c_{t+1})].$$

where  $c_t$  is the consumption at time  $t$ . Usually, a formula for  $u(c_t)$  is given by

$$u(c_t) = \frac{1}{1-\gamma} c_t^{1-\gamma}$$

and we can see as  $\gamma \rightarrow 1$ ,  $u(c)$  converges to  $\ln(c)$ .

The utility function captures the fundamental desire for more consumption. The period  $u(\cdot)$  utility function is increasing function reflecting the desire for more consumption, and concave, which means a decline of marginal value of additional consumption. Discounting future by  $\beta$  captures impatience, and  $\beta$  is called the subjective discount factor.

Suppose that the investor can buy or sell as much of the payoff  $x_{t+1}$  at time  $t$  as he wants. We denote by  $e$  the consumption level if the investor do not buy any assets, and we denote by  $\zeta$  the amount of the asset investor buys. It is clear that if the investor buys  $\zeta$  numbers of asset at time  $t$  then the consumption level at time  $t$  will

decrease by amount of  $p_t \xi$  and at time  $t + 1$  the consumption level will increase by  $x_{t+1} \xi$ . Investor needs to

$$\max_{\{\xi\}} u(c_t) + E_t[\beta u(c_{t+1})] \quad (2)$$

condition to:

$$c_t = e_t - p_t \xi \quad (3)$$

$$c_{t+1} = e_{t+1} + x_{t+1} \xi \quad (4)$$

Substituting restrictions (3) and (4) at equation (2) and taking the first derivative with respect to  $\xi$  and setting it to zero we get

$$p_t u'(c_t) = E_t[\beta u'(c_{t+1}) x_{t+1}], \quad (5)$$

$$p_t = E_t\left[\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1}\right] \quad (6)$$

Equation (6) is the first order condition for an optimal consumption and portfolio choice. Another way to think about the above formulas is:  $p_t u'(c)$  is the marginal loss in utility if investor buys another unit of asset and  $E_t[\beta u'(c_{t+1}) x_{t+1}]$  is the expected increase in marginal gain from buying the extra unit of asset. To get the maximum gain investor will buy or sell until marginal loss equals marginal gain (equation (5)).

Equation (6) is the central asset pricing formula. There we see that if we define

$$m_{t+1} \equiv \beta \frac{u'(c_{t+1})}{u'(c_t)}$$

then  $m_{t+1}$  plays the role of the stochastic discount factor and, the basic asset pricing formula becomes:

$$p_t = E_t(m_{t+1} x_{t+1}) \quad (7)$$

where expectation is being condition on information at time  $m_t$  is also called *kernel pricing* or *change of measure*.

In this context asset pricing formula (7) is a generalization that puts together all risk corrections by defining a single discount factor, which is the same for each asset. Of course  $m_{t+1}$  is stochastic or random because it is not known at time  $t$ . It is the correlation between the random components of stochastic discount factor  $m$  and asset-specific payoff  $x_i$  that will generate asset-specific risk corrections. Asset pricing formula is a very general formula that can be used to price any assets such as stocks, bonds, and options. For stocks one-period payoff is  $x_{t+1} = p_{t+1} + d_{t+1}$ . If we divide  $x_{t+1}$  by  $p_t$  we get the gross return,

$$R_{t+1} = \frac{x_{t+1}}{p_t}$$

and asset pricing formula can be expressed as:

$$1 = E_t[m_{t+1}R_{t+1}]$$

The latter formula for asset pricing is often used in empirical work because returns are very close to being stationary over time.

We know that the risk free rate at time  $t+1$  is known at time  $t$ . Using the asset pricing formula for the risk-free rate we get:

$$1 = E_{t+1}(m_{t+1}R_{t+1}^f) = R_{t+1}^f E(m_{t+1})$$

$R_{t+1}^f$  comes out of expectation because its value it is known at time  $t$ . So the formula for the risk free rate can be written as:

$$R^f = \frac{1}{E(m)} .$$

Now, remember that the covariance between  $m_{t+1}$  and  $x_{t+1}$ , denoted as  $\text{cov}(m_{t+1}, x_{t+1})$ , is defined by

$$\text{cov}(m_{t+1}, x_{t+1}) = E(m_{t+1}x_{t+1}) - E(m_{t+1})E(x_{t+1}),$$

therefore the asset pricing formula (7) can be written as:

$$p_t = E(m_{t+1})E(x_{t+1}) + \text{cov}(m_{t+1}, x_{t+1}),$$

replacing  $E(m_{t+1})$  with  $\frac{1}{R^f}$  we get

$$p_t = \frac{E(x_{t+1})}{R^f} + \text{cov}(m_{t+1}, x_{t+1}).$$

The first term is the asset's price in a risk neutral world. The second term is a *risk adjustment*. An asset that is positively correlated with the discount factor has its price increased and vice versa.

If  $\text{cov}(m_{t+1}, x_{t+1}) = 0$  then

$$p_t = \frac{E(x_{t+1})}{R^f}$$

no matter what the risk of  $x$  (i.e. or variance of  $x$ ) is. So if the asset is uncorrelated to the discount factor the asset receives no risk correction to its price and pays an expected return equal to the risk-free rate.

## Section 1.4 Stochastic Discount Factor and Complete Markets.

The asset pricing formula  $p_t = E_t(m_{t+1}x_{t+1})$ , does not assume that markets are complete and does not say anything about the return distributions.

A contingent claim is a security that pays one dollar in one state  $s$ . Denote  $pc(s)$  the price today of the contingent claim. A market is complete if any financial asset can be built synthetically using contingent claims. Now we will show that if markets are complete a discount factor exists and it is equal to the contingent claim price divided by probabilities. Let  $x(s)$  be the payoff of an asset at state  $s$ . Because the market is complete we can consider the asset as a bundle of contingent claims and the asset price must be equal to the value of those contingent claims,

$$p(x) = \sum_s pc(s)x(s), \quad (8)$$

where  $p(x)$  is the price of the payoff  $x$ . If we multiply both sides of equation (8) by probabilities  $\pi(s)$ , where  $\pi(s)$  is the probability that state  $s$  occurs, we get:

$$p(x) = \sum_s \pi(s) \left( \frac{pc(s)}{\pi(s)} \right) x(s) \quad (9)$$

Then we define  $m$  as the ratio of contingent claim to probability,

$$m(x) = \frac{pc(s)}{\pi(s)}.$$

We can write equation (9) as

$$p(x) = \sum_s \pi(s)m(s)x(s) = E_t(mx).$$

So we just showed that in complete market stochastic discount factor  $m$  exist and it is a set of contingent claims divided by probabilities.

## Chapter 2.

### Section 2.1 Generalized Method of Moments.

We have to solve an equation of the form

$$E_t[m_{t+1}R_{t+1}] = \widehat{1}.$$

We will use the generalized method of moments to solve the above equation.

Since Lars Hansen first introduced it in 1982, the Generalized Method of Moments (GMM) has been widely applied to analyze financial data. Generalized Method of Moments has simulated the development of a number of statistical inference techniques that are based on GMM estimators. These applications have been used in different areas of macroeconomics, finance, etc. Depending on the context of the problem GMM has been applied to time series, cross sectional and panel data.

It is natural to ask the question why GMM is being used so widely and has a great impact in such areas as macroeconomics and finance. Maximum Likelihood estimation (MLE) has been used since the beginning of the twentieth century and it is the best available estimator. However there are two problems that come from the use of MLE estimator and these have motivated the use of GMM estimators.

The first problem is the sensitivity of statistical properties to the distributional assumption. In order to use MLE estimators we need to know the probability distribution function of the population's data, but most of the time this distribution function is not known. One way to get around this is to make an educated guess and

choose distribution. However unless our guessed distribution is the same as the true distribution the result estimator is no longer optimal and, even worse it may lead to biased inferences.

The second problem is the computational burden. Maximum Likelihood estimation could be computationally very difficult.

In contrast GMM framework provides a computationally convenient method of performing inference in the models without the need to know the distribution function. GMM is based in the idea of using moment conditions for estimation.

We already have seen that the asset-pricing model gives us

$$p_t = E_t[m(data_{t+1}, parameters)x_{t+1}]. \quad (10)$$

After taking the unconditional expectations on both sides of equation (10) (using the formula  $E[E[Y|H]] = E[Y]$ ) we get

$$E[p_t] = E[m(data_{t+1}, parameters)x_{t+1}]. \quad (11)$$

In order to use Generalized Method of Moments (GMM) we need to make some statistical assumptions. The most important one is that  $m_t, p_t, x_t$  are stationary random processes which means that joint distribution of  $x_t$  and  $x_{t-j}$  depends on  $j$  not  $t$ . Sample averages must converge to population means as the size of sample increases. The latter assumption is true for stationary random processes.

The GMM approach is to estimate the parameters by making sample averages in equation (11)

$$\frac{1}{T} \sum_{i=1}^T p_t \quad \text{and} \quad \frac{1}{T} \sum_{i=1}^T [m(data_{t+1}, paramaters)x_{t+1}] \quad (12)$$

as close as possible to each other.



Based on asset pricing formula (11), the GMM approach is as follows: sample averages are calculated for both sides of equation (11), so we need to calculate sample averages (12) then GMM estimates the parameters of the model by equating sample averages.

$$\text{So } E[p_t] \approx \frac{1}{T} \sum_{t=1}^T p_t \text{ and } E[m_{t+1}x_{t+1}] \approx \frac{1}{T} \sum_{t=1}^T [m(\text{data}_{t+1}, \text{parameters})x_{t+1}].$$

As we mentioned above asset pricing model implies

$$E[p_t] = E[m_{t+1}(b)x_{t+1}]$$

which can be written in the form

$$E[m_{t+1}(b)x_{t+1} - p_t] = 0 \quad (13)$$

where  $x_{t+1}$  and  $p_t$  are vectors. We need to check whether a model for  $m_{t+1}$  can price a number of assets at the same time. Equation (13) is called moment condition or orthogonal condition equation. Each component of equation (13) is the difference between predicted price  $E[m_{t+1}x_{t+1}]$  and the actual price  $E[p_t]$ . If the number of parameters we want to estimate is the same as the number of components of equations (13) then we use the method of moments. However the number of moment conditions is usually greater than the number of parameters. If we have more equations than parameters to estimate we use GMM.

If we let  $u_{t+1}(b) = m_{t+1}(b)x_{t+1} - p_t$

then  $u_t(b)$  represents the error at time  $t + 1$  and of course, the mean of this error should be zero. Parameters are chosen so that the predicted prices are as close as possible to the actual prices. Model is evaluated based on how large the errors are.

For the given values of the parameters  $b$ , we could construct a time series on  $u_t(b)$  and look at its mean.

Let  $g_t(b)$  be the sample mean of the errors  $u_t(b)$ . If the sample is of size  $T$ ,

$$g_T(b) = \frac{1}{T} \sum_{t=1}^T u_t(b) = E_T[u_t(b)] = E_T[m_{t+1}(b)x_{t+1} - p_t],$$

where we are using notation  $E_T$  to denote sample means,

$$E_T(\cdot) = \frac{1}{T} \sum_{t=1}^T (\cdot).$$

It is better to work with the asset returns so if we divide by  $p_t$  both sides of equation (13) the moment conditions are written as:

$$E_T[m_{T+1}R_{T+1} - \hat{1}] = 0.$$

$R_{T+1} = \frac{x_{T+1}}{p_T}$  is the gross return. The reason for this is that stock prices  $p_t$  and

dividends  $d$  vary over time; even more they are not stationary, whereas stock returns are close to being stationary.

The process for estimating  $b$  is a two-stage process. On the first stage we choose  $b$  that makes the pricing errors  $g_t(b)$  as small as possible by minimizing the quadratic form of the sample mean of the errors

$$\hat{b}_1 = \arg \min_{\{b\}} g_t(b)' W g_t(b)$$

$W$  is a positive definite matrix that shows us how much attention is given to each moment. Normally  $W = I$ , because GMM treats all assets symmetrically: however, we may use a matrix  $W$  different from identity matrix. We may start with a matrix  $W$  that has different values on the main diagonal. This way we will give more weight

to some assets we think are more important. It can be shown that  $\hat{b}_1$  is a consistent estimator of  $b$  and is asymptotically normal so we may choose to stop here and not go further on a second stage.

But if we go on a second stage, using the value of  $\hat{b}_1$  obtained on the first stage, we form an estimate  $\hat{S}$  of

$$S \equiv \sum_{j=-\infty}^{\infty} E[u_t(b)u'_{t-j}(b)].$$

The reason for that is that some assets may have a much higher variance than others and for those assets the sample mean is a much less accurate measurement of the population mean because the sample mean will vary from sample to sample. So it makes sense to give less weight to the assets with higher variance. We could use a diagonal matrix  $W$  with inverse variances of  $g_T = E_T[m_t R_t - \hat{1}]$  on the diagonal.

However since assets returns are correlated a good idea is to use covariance matrix of asset returns. The basic idea is to pay more attention to linear combinations of moments that contain the most of the information.

The assumption is that  $E[u_t(b)] = 0$  and  $u_t(b)$  is stationary. With that in mind we get

$$\text{var}(g_T) = \text{var}\left(\frac{1}{T} \sum_{t=1}^T u_{t+1}\right) = \frac{1}{T^2} [TE(u_t u'_t) + (T-1)E(u_t u'_{t-1}) + E(u_t u'_{t-1}) + \dots]$$

$$\text{and as } T \rightarrow \infty, \frac{T-j}{T} \rightarrow 1, \text{ var}(g_t) \approx \frac{1}{T} \sum_{j=-\infty}^{\infty} E(u_t u'_{t-j}) = \frac{1}{T} S \text{ for large values of } T.$$

So a good weighting matrix is the inverse of  $S$ . It can be shown that  $W = S^{-1}$  is the optimal weighing matrix that gives us the estimates with the smallest asymptotic variance (Hansen [10]).

Using the value we found for matrix  $\hat{S}$  ( $\hat{S}$  is an estimate for  $S = \sum_{j=-\infty}^{\infty} E(u_t u_{t-j}')$ ), we

will talk more about this later at linear models), we calculate the estimate  $\hat{b}_2$

$$\hat{b}_2 = \arg \min_{(b)} g_t(b)' \hat{S}^{-1} g_t(b).$$

$\hat{b}_2$  is not only consistent and asymptotically normal but it is also asymptotically efficient estimate of the parameter vector  $b$ . By efficient we mean that  $\hat{b}_2$  has the smallest variance-covariance matrix among all estimators that make  $g_t(b)' \hat{S}^{-1} g_t(b)$  equal to zero for different choices of weighted matrix  $W$ . Using delta method which says that the asymptotic variance of  $f(x)$  is  $f'(x)^2 \text{var}(x)$  (for more on delta method see Casella and Berger [5] section 5.5.4) the variance-covariance matrix of  $\hat{b}_2$  is

$$\text{Var}(\hat{b}_2) = \frac{1}{T} (d' S^{-1} d)^{-1},$$

where  $d = \frac{\partial g_t(b)}{\partial b}$ .

The test of *overidentifying restrictions* is done to check the overall fit of the model. It can be shown that  $T$  which is the sample size, times the minimized value of  $g_t(b)' \hat{S}^{-1} g_t(b)$  calculated on second stage is  $\chi^2$  distributed with degrees of freedom equal to number of moments less the number of parameters.

$$TJ_T = T \min[g_t(b)' S^{-1} g_t(b)] \sim \chi^2 (\#moments - \#parameters)$$

The  $J_T$  test basically evaluates the model by looking at the sum of squared pricing errors and evaluates how big they are. The  $J_T$  test asks whether errors are big by

statistical standards and how often we should see a weighted sum of squared pricing errors this big.

## Chapter 3.

### Section 3.1 The Linear Factor Models.

We start with basic formula of asset pricing

$$p_t = E[m_{t+1}x_{t+1}] . \quad (14)$$

Let the discount factor be of the form  $m = b' f$  where  $p, x$  are  $N \times 1$  vectors of asset pricing and pay offs,  $f$  is a  $K \times 1$  vector of factors,  $b$  is a  $K \times 1$  vector of parameters that we want to estimate. As before, to simplify notation we will drop the sub indices in  $m, p$  and  $x$  unless otherwise specified.

After taking the unconditional expectations on both sides of equation (14) (and using the formula  $E[E[Y|H]] = E[Y]$ ) we get

$$E[p] = E[mx]$$

Substituting  $m$  with  $b'f$  we get

$$E[p] = E[b'fx] = E[xf'] b$$

To implement GMM we need to choose a set of moments. The obvious ones are the pricing errors:

$$g_T(b) = E_T[xf' - p]$$

Our goal is to find parameter  $b$  such that makes the sum of squared of pricing errors as small as possible. The GMM estimator of  $b$  is defined as

$$\hat{b} = \arg \min_b g_T(b)'Wg_T(b)$$

where  $W$  is an  $N \times N$  positive definite weighting matrix.  $W$  is chosen in a way that the more volatile assets get less weight than less volatile assets.

Start with step one by choosing  $W = I$ , then we calculate  $\hat{b}_1$  by setting

$$\frac{\partial g_t(\hat{b}_1)' W g_t(\hat{b}_1)}{\partial b} = 0$$

$$2d' W g_t(\hat{b}_1) = 0$$

$$d' W \left( \frac{1}{T} \sum_{j=1}^T (x f' \hat{b}_1 - p) \right) = 0$$

$$d' W \left( \frac{1}{T} \hat{b}_1 \sum_{j=1}^T (x f' - p) \right) = 0$$

$$d' W d \hat{b}_1 - d' W \bar{p} = 0$$

where  $\bar{p} = \frac{1}{T} \sum_{t=1}^T p_t$  and  $d = \frac{\partial g_t(\hat{b}_1)}{\partial b} = E_T[x f'] = \frac{1}{T} \sum_{t=1}^T x f'$

So  $d$  is a  $N \times K$  matrix and

$$\hat{b}_1 = (d' W d)^{-1} d' W \bar{p}$$

But  $W = I$  so

$$\hat{b}_1 = (d' d)^{-1} d' \bar{p}. \quad (15)$$

Second step: using the value of  $\hat{b}_1$  we found on the first step, form an estimate  $\hat{S}$  of

$$S \equiv \sum_{j=-\infty}^{\infty} E[u_t(\hat{b}_1) u_{t-j}'(\hat{b}_1)].$$

$$\hat{S} = \hat{S}_0 + \sum_{j=1}^k (\hat{S}_j + \hat{S}_j')$$

where 
$$\hat{S}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}' ,$$

and  $\hat{u}_t = (x_{t+1}f'_{t+1}\hat{b}_1 - p_t)$ ,  $j = 0, 1, 2, \dots, k$ , where  $k$  is equal to maximum lag value selected. Choosing  $W = \hat{S}^{-1}$  will give us the optimal value for the weighting matrix. (Optimal estimators here means estimator with the smallest variance. For more on that see Hansen [10]).

However, in practice longer lags get less weight than the shorter ones. One example of that is the Bartlett kernel where

$$\hat{S} = \hat{S}_0 + \sum_{j=1}^k \left(1 - \frac{j}{k+1}\right) (\hat{S}_j + \hat{S}'_j)$$

(For more on Bartlett kernel see Newey and West [11]).

Then we precede the same way as on the first step but this time  $W = \hat{S}^{-1}$ .

The solution we get for  $\hat{b}_2$  is

$$\hat{b}_2 = (d' \hat{S}^{-1} d) d' \hat{S}^{-1} \bar{p}.$$

We already have shown that  $\text{cov}(\hat{b}_2) = \frac{1}{T} (d' S^{-1} d)^{-1}$ .

Let  $J_T = g_T(b)' \hat{S}^{-1} g_T(b)$ .  $J_T$  is the minimum criterion J-static that is used to test for the overidentifying conditions. Under null hypothesis moment conditions are zero and  $TJ_T \sim \chi^2(\# \text{moments} - \# \text{parameters})$ .



## Chapter 4.

### Section 4.1 Numerical Implementation for CAPM.

First let see the relation between stochastic discount factor and betas.

Using the asset pricing formula for the returns we get:

$$1 = E(mR_i) = E(m)E(R_i) + \text{cov}(m, R_i)$$

from the above formula we get:

$$E(R_i) = \frac{1}{E(m)} - \frac{\text{cov}(m, R_i)}{E(m)} \quad (16)$$

If we multiply and divide both sides of equation (16) by  $\text{var}(m)$  we get:

$$E(R_i) = \gamma + \left(\frac{\text{cov}(m, R_i)}{\text{var}(m)}\right)\left(-\frac{\text{var}(m)}{E(m)}\right)$$

where  $\gamma = \frac{1}{E(m)}$ .

As we see we have a single-beta representation. So  $p = E(mx)$  implies

$E(R^i) = \gamma + \beta_{i,m} \lambda_m$  which is the beta representation model. CAPM, APT are

expected return-beta models and can be shown that beta pricing models are

equivalent to linear models for the discount factor  $m$ ,

$$E(R^i) = \gamma + \beta_{i,m} \lambda_m \Leftrightarrow m = a + b' f$$

(For the proof of the above result see Asset Pricing [1] chap.6)

A special case of the above result is the CAPM model. For the CAPM model

$$m = a - b' f.$$

CAPM implied stochastic discount factor is:

$$m_t = a - bR_{m,t}^e \text{ where } R_{m,t}^e = R_{m,t} - R_{f,t}$$

$R_{m,t}^e$  is the excess return of the market,  $R_{m,t}$  is the gross return of the market and

$R_{f,t}$  is the risk free return.

From the asset pricing formula we have

$$E[mR_t] = 1 \text{ and } E[mR_f] = 1.$$

Combining the two formulas we get:

$$E[mR_t] - E[mR_f] = 0. \text{ That can be written as:}$$

$$E[m(R_t - R_f)] = 0 \text{ or } E[mR_t^e] = 0 \text{ where } R_t^e \text{ is the asset excess return.}$$

The problem with the model  $m_t = a - bR_{m,t}^e$  is that we cannot separately identify  $a$

and  $b$  so we have to choose some normalization. The reason for that is that if

$$E[mR_t^e] = 0 \text{ then } E[(2m)R_t^e] = 0.$$

We write the equation  $E[mR_t^e] = 0$  as  $E[R_t^e(a - bR_{m,t}^e)] = 0$ .

If we divide both sides of the last equation by  $a$ , and replace  $\frac{b}{a} = \beta$  we get:

$$E[R_t^e(1 - \beta R_{m,t}^e)] = 0.$$

Following the same procedure we did for linear factor models we get

$$g_T(\beta) = E_T[R_t^e(1 - \beta R_{m,t}^e)] = \frac{1}{T} \sum_{t=1}^T R_t^e(1 - \beta R_{m,t}^e)$$

We need to find the value of  $\beta$  that minimizes  $g_T(\beta)'Wg_T(\beta)$ . For that we take the

first derivative of  $g_T(\beta)'Wg_T(\beta)$  with respect to  $\beta$  and equate it to zero.

$$\frac{\partial}{\partial \beta} [g_T(\beta)'Wg_T(\beta)] = 0. \text{ Taking this derivative we get:}$$

$$2d'Wg_T(\beta) = 0, \quad (17)$$

where  $d$  is a  $N \times 1$  matrix and  $N$  is the number of assets we are using in our model.

$$d = \frac{\partial g_T(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \left[ \frac{1}{T} \sum_{t=1}^T R_t^e (1 - \beta R_{m,t}^e) \right] = \frac{1}{T} \sum_{t=1}^T -R_t^e R_{m,t}^e.$$

We can rewrite the formula for  $g_T(\beta)$  as

$$g_T(\beta) = \frac{1}{T} \sum_{t=1}^T R_t^e - \frac{1}{T} \beta \sum_{t=1}^T R_t^e R_{m,t}^e = \bar{R}_T^e + \beta d$$

where  $\bar{R}_T^e = \frac{1}{T} \sum_{t=1}^T R_t^e$ .

Substituting  $g_T(\beta)$  at equation (17) we get:

$$d'W(\bar{R}_T^e + \beta d) = 0$$

$$d'W\bar{R}_T^e + \beta d'Wd = 0$$

$$\beta d'Wd = -d'W\bar{R}_T^e$$

And the solution is  $\hat{\beta}_1 = -(d'Wd)^{-1} d'W\bar{R}_T^e = -(d'd)^{-1} d'\bar{R}_T^e$  because  $W = I$  on the

first step. We also get:  $\text{cov}(\hat{\beta}_1) = \frac{1}{T} (d'd)^{-1} d'S(d'd)^{-1}$  and

$$\text{cov}[g_T(\beta)] = \frac{1}{T} (I - d(d'd)^{-1} d') S (I - d(d'd)^{-1} d')$$

where  $S \equiv \sum_{j=-\infty}^{\infty} E[u_t(\beta) u_{t-j}(\beta)]$

On the second step we get  $\hat{\beta}_2 = -(d'S^{-1}d)^{-1} d'S^{-1}\bar{R}_T^e$  and hence

$$\text{cov}(\hat{\beta}_2) = \frac{1}{T} (d'S^{-1}d)^{-1} \text{ and } \text{cov}[g_T(\hat{\beta}_2)] = \frac{1}{T} (S - d(d'S^{-1}d)^{-1} d').$$

We have chosen to use eight stocks for numerical implementation of CAPM model.

We have chosen the following stocks taken from S&P500 index: Adobe Systems Inc.

(ADBE), Citigroup Inc. (C), Chevron Corp. (CVX), Duke Energy Corp. (DUK), Halliburton Co. (HAL), Lowe's Companies (LOW), Microsoft Corp. (MSFT) and Pfizer Inc. (PFE) with market factor SP500 as the single factor. Data used in our computation are weekly excess returns from April 4, 1996 to April 4, 2006, so sample size is  $T = 520$ .

We are using  $R_f = 1.01\%$  for the risk free rate. After the computation we get

$\beta = 5.936$ ,  $J_T = 0.0030$ , and the p-value  $P(TJ_T > 1.56) = 0.97$ .

The code was implemented in Matlab.

## **Section 4.2 Conclusion**

The stochastic discount factor is a generalization of a more intuitive and practical concept, which arises, in more specific problems or models such as weighted portfolios, consumption based models, and CAPM. The GMM is a technique that provides means to estimate certain forms of the stochastic discount factor.

## References

- [1] John H. Cochrane, “*Asset Pricing*” Princeton University Press January 2005.
- [2] Alastair R. Hall, “*Generalized Method of Moments*” Oxford University Press February 2005.
- [3] Ashish Sen and Muni Srivastava, “*Regression Analysis*” Springer April 1997.
- [4] James Douglas Hamilton, “*Time series Analysis*” Princeton University Press January 1994.
- [5] George Casella and Roger L. Berger, “*Statistical Inference*” Duxbury Press June 2001.
- [6] William H. Greene “*Econometric Analysis*” Prentice Hall March 1998
- [7] Heber Farnsworth, Wayne Ferson, David Jackson, Steven Todd “*Performance Evaluation with Stochastic Discount Factors*” October 2000.
- [8] Hossein Kazemi and Thomas Schneeweis, “*Conditional Performance of Hedge Funds*” March 2003.
- [9] Hansen L. P. and S. F. Richard 1987, “The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models” *Econometrica*, Vol 55.
- [10] Hansen L. P 1982, “Large sample properties of generalized method of moments estimators”. *Econometrica*, Vol 50.
- [11] Newey Whitney K. and Kenneth D. West 1987, “A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix”. *Econometrica*, Vol 55.
- [12] Seppo Pynnonen, [www.uwasa.fi/~sjp/](http://www.uwasa.fi/~sjp/).
- [13] Yahoo finance, <http://finance.yahoo.com/>.









