MSc Economics

Estimating the Euler Equation from the basic C-CAPM Using a Large Set of Possible Instruments

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by
Justinas Pelenis

Roman Goncharenko

1127066

Vienna, 06.06.2013

INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

CONTINUING EDUCATION CENTER

MSc Economics

Affidavit

I, Roman Goncharenko

hereby declare

that I am the sole author of the present Master's Thesis,

Estimating the Euler Equation from the basic C-CAPM Using a Large Set of Possible Instruments

53 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and that I have not prior to this date submitted this Master's Thesis as an examination paper in any form in Austria or abroad.

Vienna, 06.06.2013

Signature

Contents

List of Figures

List of Tables

Abstract

We perform the instrumental variable estimation of the Euler equation and the system of Euler equations from the basic Consumption - based Capital Asset Pricing Model (C-CAPM) using a large set of possible instruments. This large set of possible instruments is due to the Rational Expectation Hypothesis. The optimal GMM estimator, which is used in the estimation, has a finite sample bias proportional to the number of instruments. This means that there is a need of the efficient instrument dimension reduction method. The two different methods of such a reduction are compared: the FIV estimator and the optimal GMM estimator that uses preselected principal components (constructed from the large set of possible instruments) as instruments. Originally, the two methods were developed for linear models. We modify the latter method to extend it to non-linear models. The Euler equation is estimated in both nonlinear and linearized forms with different utility function specifications. Estimation results do not contradict earlier research in terms of the estimated magnitude of parameters of interest but unambiguously point on the dominance of the substitution effect in the consumers' saving decision. Moreover, we find that the optimal GMM that uses preselected principal components as instruments (the second method) performs better than the FIV estimator (the first method).

1 INTRODUCTION

This paper deals with the instrumental variable estimation of the Euler equation (as well as the system of Euler equations) from the basic Consumption-based Capital Asset Pricing Model (C-CAPM) when the set of possible instruments is very large. The large set of possible instruments arises because of the the rational expectations hypothesis which implies that the agent when making a decision uses all the information available. The estimation of the Euler equation is performed in both nonlinear and linearized forms.

There are at least two reason why one may be interested in estimating the Euler equation. First of all, the result of the estimation can be used as an inside in whether the underlying theoretical model is a good approximation to the real world. It is true that one cannot conclude easily that the theoretical model is wrong just because it is statistically rejected, yet it can be thought of as a sign a possible problem. Secondly, the estimated magnitude of the model's parameters can provide us with some understanding of economic processes. For instance, as this will be discussed later in more details, depending on the magnitude of the parameter from the CRRA type utility function one can say whether the wealth or income effect dominates in consumers' saving decision.

Mathematically, the Euler equation is nothing more than a conditional expectation (on the information available at time t) with unknown parameters and density. The fact that the expectation is conditional gives rise to a large set of possible instruments. In theory any macroeconomic series is a potential instrument. On the one hand, the fact that it is so easy to find instruments is a great news. On the other hand, the problem now is that there are too many possible instruments. It has been shown in Bai and Ng (2010) that the optimal GMM estimator has a finite sample bias proportional to N/T, where N is the number of instruments and T is the number of observations. Moreover, the consistency of the optimal GMM requires $N/T \to 0$. This means that the use of too many instruments can lead to a large bias in a finite sample. Thus, on the one side there is large set of possible instruments to estimate the Euler equation, on the other side, one cannot use them all because that would lead to aforementioned problem. Hence, there is a need for an efficient method of instrument dimension reduction. This paper compares the results of two such methods.

The first method is due to Bai and Ng (2010). In this approach a certain structure is assumed for the endogenous variable and instruments: it is assumed that the endogenous variables as well as a large set of instrumental variables share common factors. It is shown that under some additional assumptions (discussed in more details in the first section of the paper) the ideal instruments are the common factors but they are unobserved. It is proposed, instead, to use estimated common factors in the optimal GMM, which gives rise to the Factor Instrumental Variable Estimator (FIV). It is shown that the FIV has very favorable properties such as normality and efficiency (Theorem 1 and Proposition 1 in Bai and Ng (2010)).

The second method is due to Bai and Ng (2008). In this method no strict assumption is made upon the structure of the data generating processes. The method is implemented in three stages: first, the principal components are constructed from the large set of instruments, then these are pre-ordered based on the (squared) correlation with the endogenous variable, finally, using an information criterion it is decided how many of them are to be used in the estimation (a technical explanation of the method comes in later sections). This method, however, is originally designed for linear models only. I extend it to non-linear models.

Specifically, the first step in Bai and Ng's (2008) method stays the same, that is, using a large set of possible instruments I construct as many principal components as I have instruments (given that the number of instruments is strictly smaller than the number of observations). However, the second step, involving the ranking of principal components as instruments is modified. I propose to do the ranking based on the R^2 obtained from the regression of the partial derivative of the function used in the GMM estimation with respect to the parameter of interest on each orthogonal series one at a time. The problem with this step, however, is that the partial derivative is itself a function of the parameters of interest which values are unknown. To deal with this problem I use different "reasonable" parameter values to evaluate the partial derivative and to see whether the result is robust to different parameter values. In this particular application the ranking of the orthogonal series is completely insensitive to different parameter values (in reasonable range).

The selection of instruments based on the correlation between the partial derivatives and instruments is motivated by the fact that the higher correlation will decrease an asymptotic value of the variance of the optimal GMM estimator. Finally, the third step of the method, the selection of the number of the principal components to be used as instruments by information criterion, stays the same.

In general, the Euler equation from the basic C-CAPM has been estimated with the US data numerous times both in nonlinear and linearized forms. It is important to note that linearization of the Euler equation requires an additional assumption about the joint distribution of the consumption growth rate and the rate of asset return. Moreover, in the process of linearization one of the parameters of the model is lost (the discount factor). Unfortunately, this is the price that has to be paid to be able to work in a linear framework which is much better understood.

One of the earliest papers that deals with the estimation of the Euler equation from the basic C-CAPM is Hansen and Singleton (1983). They estimate the linearized Euler equation by the MLE using the post war data on the US economy. The estimated values of the coefficient of relative risk aversion for different risky assets were in the range between zero and two. Another example of the estimation of linearized Euler equation is Hall (1988). Hall uses instrumental variable estimation procedure and his instruments are twice lagged consumption growth rate and twice lagged asset return (for risky asset it is return on S&P 500 and for risk free asset it is return on T-Bills). To estimate the elasticity of intertemporal substitution (EIS) (the inverse of coefficient of relative risk aversion) the US economy data from as early as 1919 is used. The results are such that the EIS is in the range between 0.06 and 0.35. Campbell (2003) uses international data to estimate the linear Euler equation by instrumental variable estimator with lagged variables (used in estimation) as instruments. He reports 95% confidence interval for the coefficient of relative risk aversion as $[-0.73, 2.14]$. Finally, Yogo (2004) uses both the US and international data in instrumental variable estimation of the linearized Euler equation. For example, using the data from 1970:3 to 1998:4 for the US economy Yogo (2004) reports the 2SLS estimate of the coefficient of relative risk aversion around 0.53 (0.5).Yogo (2004) uses the following variables as instruments: twice lagged nominal interest rate, inflation, consumption growth, and the log dividend price ratio. The estimate of the coefficient of relative risk aversion he obtains is around 0.53 using the 2SLS.

The first nonlinear estimation of the Euler equation from the basic C-CAPM was done by Hansen and Singleton (1982). Using the US data from 1959:2 to 1978:12 they estimate the Euler equation by the GMM for a risky asset (the return on which is computed as the equally-weighted average return on all stocks listed on the NYSE) using lagged growth rate of consumption and lagged return on the risky asset as instruments. For example, using 4 lags as instruments they report the estimate of the discount factor around 0.993 (0.031) and the estimate of the coefficient of relative risk aversion around 0.78 (0.253). Using different number of lags they get similar results in terms of the magnitude of estimated parameters, though quite often the J-test rejects the model. When trying to estimate the system of equations for different risky assets the J-test rejects H_0 of correctly specified model. Finally, Hansen and Singleton (1982) estimate the system of two Euler equations, one for risky asset and the other for risk-free one. In this case they report the parameter estimate of the coefficient of relative risk aversion to be around 0.14 (0.04). Such a low value of the estimate demonstrates the equity premium puzzle. More on the equity premium puzzle is in the last section.

It is important to note that despite the fact that the rational expectations hypothesis suggests a large set of possible instruments (potentially all the macroeconomic series) the set of instruments in the aforementioned studies was restricted to either the set of lagged variables that appear in the estimated equation or the set of lagged variables that "intuitively must be important" instruments. Such an approach to preselecting of the instruments is highly undesirable because it can lead easily to the loss of important information for the identification of the parameters of interest. Observing this problem Bai and Ng (2008), using a large set of possible instruments, apply their method to replicate the result of Yogo (2004). Using the set of selected instruments they obtain the estimate of the coefficient of relative risk aversion of 0.769 (0.346).

In my paper, using a large set of 131 macroeconomic time series for the US economy, I begin by applying the FIV estimator, as the first method of efficient instrument dimension reduction, to estimating the linearized version of the Euler equation and then apply the second method of instrument dimension reduction, that is, replicate the result of Bai and Ng (2008), Section 5 Table 7a, result labeled by FIV_{ic}. Then I go further and apply both methods to the nonlinear estimation of the Euler equation with two different utility specifications (the standard CRRA and a CRRA with an external habit formation). Finally, I apply the second method of instrument dimension reduction to the system of the two Euler equations (one for a risky asset and the other is for a risk-free asset) to examine the equity premium puzzle, again using the two utility specifications.

In general I find that the second method of instrument dimension reduction performs better than the FIV estimator in terms of the size of the standard errors in this particular application. Precisely, the set of the instruments picked by the second method is either a super-set of the instruments used in the FIV (the case of a standard CRRA) or is a completely different set (the case of an external habit). One of the possible explanations is that the strict assumption upon the structure of the data generating processes made in FIV estimation is not met. Also it is argued that the estimation of the Euler equation is usually associated with the problem of weak identification (e.g., Yogo (2004) and Campbell (2003)). I do not detect any signs of weak identification while using the second method of instrument dimension reduction, that is, the different methods of the optimal GMM implementation produce very similar results.

In terms of parameter estimates when estimating the linear Euler equation for the risky asset by the FIV estimator the parameter estimate, the coefficient of relative risk aversion, is estimated to be 0.85 (0.74), and using the second method it is 0.82 (0.47). The non-linear estimation of the Euler equation with the standard CRRA utility function produces the following estimates of the coefficient of relative risk aversion 0.89 (0.74) and 0.82 (0.47) by the FIV and the second method, respectively. The same non-linear estimation but with external habit in the CRRA produces the following estimates of the coefficient of relative risk aversion 0.16 (1.3) and 0.74 (0.56) by the FIV and the second method, respectively. Finally, the result of the two-equation non-linear estimation using only the second method (the optimal GMM with selected principal components as instruments) is such that the estimates of the coefficient of relative risk aversion are 0.04 (0.02) and 0.38 (0.09) for the standard CRRA and the CRRA with an external habit cases, respectively.

The magnitude of the coefficient of the relative risk aversion has some interesting implications to certain economic processes. First of all, larger values of the coefficient would signal higher aversion towards risk. Secondly, in the basic C-CAMP with the CRRA utility function the inverse of this coefficient is the intertemporal elasticity of substitution. If this elasticity is large than unity then the wealth effect is relatively weak in the saving decision and the substitution effect is relatively strong; if it is smaller than unity then this is the other way around; finally, if this elasticity is exactly one then the both effects cancel each other. As one can see from the above literature review different researches find different result regarding the dominance of the income and substitution effects. The results of my estimation unambiguously demonstrate that the wealth effect is relatively weak in the consumers' saving decision and the substitution effect is relatively strong. More on these effects is in the next section.

The rest of the paper is structured in the following way. The Section 2 presents the basic C-CAPM and the derivation of the Euler equation, this section also discusses the GMM framework, the Large Dimensional Factor Analysis and the two methods of instrument dimension reduction such as the FIV and the optimal GMM that uses principal components as instruments. The Section 3 describes the dataset and the data preparation process. The Sections 4, 5, and 6, present the econometric models, instrument selection process, and the estimation results for the linear, non-linear single equation, and non-linear two-equation cases, respectively. Finally, the Section 7 concludes.

2 Theoretical Framework

In this section I present the theoretical macroeconomic model which lies underneath the econometric model I estimate later. I also present the main econometric techniques employed in this project. I introduce the GMM estimator and summarize its main properties, briefly explain the problem of weak identification, introduce the large dimensional static factor analysis, and finally discuss the FIV estimator.

2.1 Macroeconomic Model

Consider the basic consumption-based capital asset pricing model, for short the C-CAPM, due to Lucas (1978). The representative agent solves the following problem:

$$\max_{c_1,c_2,\ldots} \mathbb{E}_1 \left[\sum_{t=1}^{\infty} \beta^{t-1} u(c_t, X_t) \right] \tag{1}$$

s.t.

$$c_t + \sum_{j=1}^{N} p_{jt} q_{jt} \leq \sum_{j=1}^{N} (p_{jt} + d_{jt}) q_{jt-1} + w_t \tag{2}$$

where c_t is a level of consumption at time t, X_t is a variable that an agent has no direct control over (fore example, this could be an external habit), p_{jt} is a price of j-th asset at t, d_{jt} is a dividend payment on j-th asset at time t, q_{jt} is a quantity of asset j held at the end of t, and finally w_t is a real wage at t.

Using the dynamic programming technique one can derive the Euler Equation, which must be satisfied in equilibrium for all assets and is given by

$$p_{jt} = \mathbb{E}_t \left[\beta \left(\frac{u'(c_{t+1})}{u'(c_t)} \right) (p_{jt+1} + d_{jt+1}) \right] , \forall j \tag{3}$$

where $\mathbb{E}_t$ is a conditional expectation on the information set I_t consisting of all information available up to time t. Rearranging some term in (3) leads to the following expression

$$\mathbb{E}_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} (1 + r_{jt+1}) - 1 \right] = 0 \forall j \tag{4}$$

where $1 + r_{jt+1} \equiv \frac{p_{jt+1} + d_{jt+1}}{p_{jt}}$ is a gross return on the asset j. The equation (4) can be used in the nonlinear GMM estimation of the parameters β and σ. In one of the later sections I explain how one can get an econometric model from (4) and perform the estimation.

I use two different utility specifications. The first one is the standard constant relative risk aversion (CRRA) utility function of the form

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

The coefficient σ has two interpretations. First of all, this is a so-called Arrow-Pratt measure of relative risk-aversion (RRA) or the coefficient of relative risk aversion which is defined as

$$R(c_t) = \frac{u''(c_t)c_t}{u'(c_t)}.$$

For the CRRA utility function $R(c_t) = \sigma$ (hence the name of the utility function). This coefficient measures the agent's attitude towards risk: the large the coefficient of relative risk aversion is the more risk averse the agent is. On the other hand, for the CRRA, the inverse of the coefficient of relative risk aversion is the coefficient of intertemporal elasticity of substitution (IES), a measure of responsiveness of the growth rate of consumption to the real interest rate which, in general, is defined as $\frac{\partial \log(c_{t+1}/c_t)}{\partial r_{t+1}}$. When $\sigma = 1$ the CRRA takes the form of logarithmic utility $\log(c_t)$. It is also important to note that when $\sigma = 1$ the income and substitution effects exactly offset each other, whereas when $\sigma > 1$ the substitution effect is relatively weak whereas the income effect is relatively strong, when $\sigma < 1$ it is the other way around. In short, the substitution effect here is the resulted decrease in saving rate due to a decrease in the interest rate on the other hand the income effect will drive the saving rate up. This also implies that a very low sigma would imply a faster convergence of the economy to the steady state (if one think of σ as the inverse of the IES).

Number of researchers has recognized that that the current consumption expenditure may be affected by its past values. Some of them, such as, Dunn and Singleton (1986) and Eichenbaum, Hansen, and Singleton (1988). Sundaresan (1989) and Constantinides(1990) have incorporated a habit formation into preferences such that a habit depends on an agent's own past consumption level (internal habit formation). Others, such as Abel (1990) and Campbell and Cochrane (1999) introduced habit in such a way that it is affected by other agentsâĂŹ decisions in the economy (e.g., by aggregate consumption), rather than by an agent's own decisions (external habit formation).

The second utility specification is the CRRA with an external habit formation. The utility is given by

$$u(c_t, X_t) = \frac{(c_t/X_t)^{1-\sigma} - 1}{1-\sigma}$$

where $X_t = c_{t-1}$ is an external habit formation. The idea behind the external habit formation is that the agent's current consumption is affected by his own consumption from the previous period (hence the name "habit") and it is not under the control of the agent (hence "external"). An external habit as opposed to an internal one, is understood in the sense of "keeping up with the Johnsons". The X_t is thought of being a past-period aggregate consumption c_{t-1} (this is why it is not under the control of the agent): if it raises then the agent is hurt, hence there is a motivation to increase agent's own consumption, to "keep up with the Johnsons". This particular specification of the CRRA with the external habit is taken from Jagannathan et al. (2002). Once again σ is equal to the coefficient of relative risk aversion.

Under the standard CRRA utility function the Euler equation (4) takes the following form

$$\mathbb{E}_t \left[\beta (\frac{c_t}{c_{t+1}})^{\sigma} (1 + r_{j\,t+1}) - 1 \right] = 0 \ , \ \forall j$$

and under the CRRA with the external habit the Euler equation (4) looks like

$$\mathbb{E}_t \left[\beta (\frac{c_t}{c_{t-1}})^{\sigma-1} (\frac{c_{t+1}}{c_t})^{-\sigma} (1 + r_{j\,t+1}) - 1 \right] = 0 \ , \ \forall j$$

Under the assumption that the asset return $r_{j\,t+1}$ and consumption c_{t+1} are conditionally on information at time t jointly log-normally distributed, and with the standard CRRA utility, the equation (4) can be linearized as follows (the details of the linearizion will be explained in next section)

$$\mathbb{E}_t \left[r_{i,t+1} \right] = \mu_{i,t} + \sigma \mathbb{E}_t \left[\Delta c_{t+1} \right] \tag{5}$$

which then can be considered as the following regression equation

$$r_{i,t+1} = \mu_i + \sigma \Delta c_{t+1} + \eta_{i,t+1} \tag{6}$$

where $r_{i,t+1} := \ln(1 + r_{i\,t+1})$ and $\Delta c_{t+1} := \ln(c_{t+1}/c_t)$ and $\mu_{i,t}$, $\eta_{i,t+1}$, and μ_i are defined in the Section 4, in which more details on the linearization are provided.

2.2 Econometric Techniques

2.2.1 GMM and Weak Identification Problem

In this section I briefly describe main properties of the GMM estimation procedure as well as the J-test. I also say a few words about weak identification since this type of problem is frequently encountered when estimating the Euler equation from the C-CAPM.

GMM Here I introduce the GMM estimator developed by Hansen (1982) and briefly summarize its properties. Suppose we have a population moment condition given by

$$g(\theta_0) = \mathbb{E}\left[f(w_t, z_t, \theta_0)\right] = 0 \tag{7}$$

This is the statement involving the data (w_t, z_t), where w_t is the structural data and z_t is instruments (the reason for this distinction will become clearer later), and the parameters of interest θ. If one could compute this expectation then it would be easy to get θ_0 (assuming that there is a unique solution). Otherwise, one could define the analogous sample moments as

$$g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f(w_t, z_t, \theta) \tag{8}$$

By setting this expression equal to zero one can derive the estimator of the parameters of interest θ. To be able to do that one needs as many equations as one has parameters. If the number of equations is exactly equal to the number of parameters then the system is said to be exactly identified and one can use the method of moments (MM) to derive the estimator of θ. If the number of equations exceeds the number of parameters then the system is said to be overidentified and one could use the Generalized Method of Moments (GMM) to derive the estimator for θ. When the number of equations exceeds the number of parameters it is impossible to set the system equal to zero exactly. Therefore, one could try to minimize the weighing sum of squares defined by the following quadratic form

$$J_T(\theta) = g_T'(\theta) W_T g_T(\theta) \tag{9}$$

where matrix W is a (symmetric and positive definite) weighting matrix (that attaches weights to the individual moments). Then the GMM estimator for the parameters of interest θ is defined as:

$$\widehat{\theta}_{GMM} \equiv \arg\min_{\theta}\{g_T'(\theta) W_T g_T(\theta)\} \tag{10}$$

Under certain conditions (such as, process (w_t, z_t) being strictly stationary ergodic, z_t are predetermined, etc.) this estimator is consistent (for any weighting matrix), and asymptotically normally distributed . The asymptotic variance-covariance matrix is given by

$$V \equiv (D'WD)^{-1} D'W\Omega WD(D'WD)^{-1} \tag{11}$$

where $D \equiv \mathbb{E}\left[\frac{\partial f(w_t,z_t,\theta)}{\partial\theta'}\right]$ and $\Omega := \lim_{T\to\infty} \mathbb{E}\left[T g_T(\theta_0) g_T'(\theta_0)\right]$. Moreover, if the weighting matrix is such that $W \propto \Omega^{-1}$, where Ω is selected to minimize (11) w.r.t. W then (10) is the most efficient estimator in the class of all asymptotically normal estimators. In such a situation one speaks about the optimal GMM estimator with the asymptotic property that

$$\sqrt{T}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} n(0, \Omega_{GMM})$$

where $\Omega_{GMM} = (D'\Omega^{-1}D)^{-1}$.

There are various ways to implement the optimal GMM estimation. It is important to note, although, there are different ways of implementing the optimal GMM, they are all asymptotically equivalent (in finite sample, on the other hand, different ways of the implementation may produce substantially different results). Most common of them are:

- two-stage optima GMM;
- iterative GMM (N times or till convergence);
- Continuously Updating GMM (CUGMM, or CUE).

For example, the two stage optimal GMM is implemented in the following way. In the first stage estimate the parameter of interest using an identity matrix as weighting matrix (or any other PD matrix). Then get a consistent estimate of Ω, for example

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} f(w_t, z_t, \hat{\theta}_1) f'(w_t, z_t, \hat{\theta}_1)$$

where $\hat{\theta}_1$ is the GMM estimator of θ from the first stage and $\text{plim}(\hat{\Omega}) = \Omega$. Then in the second stage estimate θ one more time but using $\hat{\Omega}^{-1}$ as a weighting matrix to obtain $\hat{\theta}_2$, this is the two-stage GMM estimator for θ. The other way to implement the efficient GMM estimation is to repeat the same procedure iteratively a certain number of times, say N times (N-iterative GMM), or till convergence (iterative GMM, IT GMM).

A somewhat different way of implementing the optimal GMM estimation is the CUE. The CUE was developed by Hansen et al., (1996). The CUE is a generalization of LIML (Limited Information Maximum Likelihood) to the GMM. The CUE, $\widehat{\theta}_{CUE}$, for θ is defined as

$$\widehat{\theta}_{CUE} \equiv \arg\min_{\theta}\{g_T'(\theta)\widehat{W}(\theta)g_T(\theta)\} \tag{12}$$

where $\widehat{W}(\theta) = T^{-1}\sum_{t=1}^{T}[f(w_t,z_t,\theta)-\bar{f}(w_t,z_t,\theta)][f(w_t,z_t,\theta)-\bar{f}(w_t,z_t,\theta)]'$ and $\bar{f}(w_t,z_t,\theta) = T^{-1}\sum_{t=1}^{T} f(w_t,z_t,\theta)$ If If $f(w_t,z_t,\theta)$ is serially correlated, then $\widehat{W}(\theta)$ is replaced by an estimator of the spectral density of $f(w_t,z_t,\theta)$ at frequency 0. Under the null of $\theta = \theta_0$ the corresponding J-statistic of $\widehat{\theta}_{CUE}$ is asymptotically χ^2_k, where k is a number of restrictions, whether the identification is weak or strong (Stock and Wright 2000). Moreover, robust confidence intervals for θ can be constructed by inverting the objective function, that is, $CI(\theta)_{95\%} = \{\theta \in \Theta|\widehat{J}(\theta) < \chi^2_{k,5\%}\}$, where $\widehat{J}(\theta)$ is the objective function of CUE. This is so-called S-set, that neither has to be convex nor connected. One of the advantages of the CUE is its relatively better performance in finite samples in comparison to the other optimal GMM implementations and its robustness to weak identification (Hansen et al. (1996) and Stock et al, (2002)). The CUE can only be implemented numerically.

Once the GMM estimation is completed one may want to perform the test of overidentifying moment conditions. The idea behind the test is as follows. If the number of equations (restrictions) is the same as the number of the parameters then there is no problem in setting the objective function to zero and getting the estimates. However, with additional equations (restrictions) one cannot, in general, attain zero. So, basically what the test checks is whether the value of the minimized objective function is not "too large" when additional equations are used. Formally,

- $H_0 : g(\theta_0) = 0$ (the null hypothesis that the model is "valid")
- $H_1 : g(\theta) \neq 0, \forall\theta \in \Theta$ (the alternative hypothesis that model is "in-

valid", the data do not come close to meeting the restrictions).

Under the null the J-statistic is asymptotically X^2_{k-l} (k is a number of moment conditions and l is a number of parameters estimated) and is given by

$$\widehat{J}_T(\widehat{\theta}) = Tg'_T(\widehat{\theta})\widehat{W_T}g_T(\widehat{\theta}) \tag{13}$$

it must be true that plim$\widehat{W_T} = \Omega^{-1}$, where Ω is the efficient weighting matrix. Under that alternative hypothesis plim$(\widehat{J}_T(\widehat{\theta})) = \infty$. Finally, the rejection rule is specified as follows (under α -confidence level):

- H_0 is rejected at α confidence level if $\widehat{J}_T(\widehat{\theta}) > X^2_{k-l,\alpha}$
- H_0 is not rejected at α confidence level if $\widehat{J}_T(\widehat{\theta}) \leq X^2_{k-l,\alpha}$

It is important to stress that the J-test does not test the validity of the model $per se$, particularly it does not test the underlying economic theory on correctness. What test does is it considers whether the $k-l$ over identifying restrictions are valid or not (given the identification of l moments). Unfortunately, there is no way of knowing which moment conditions the test rejects.

WEAK IDENTIFICATION Weak identification in GMM is basically a problem similar to the problem of weak instruments in IV. However, the problem of weak instruments is much better understood than the problem of weak identification (Yogo, 2004). A proper discussion about the problem of weak identification is clearly beyond the scope of this subsection, thus, I will just try to sum up main aspects of it. An interested reader can find more of weak identification, particularly in the case of nonlinear models, in Stock et al (2002).

A necessary condition for the identification of the parameters of interest θ is that the population moment condition is not equal to zero at any other value of θ but θ_0, that is, $g(\theta) \neq 0 \forall \theta \neq \theta_0$. In the linear model, weak instruments (weak identification) arise in the situation in which the population moment condition is almost zero for some $\theta \neq \theta_0$; that is, when the instruments Z_t are almost uncorrelated with the error term of the model even at untrue value of θ.

To my knowledge there is no a proper test for the weak identification problem in GMM. However, the main sings of weak identification are as follows:

- CUE (defined below), two-step, and iterated GMM produce noticeably different estimates (Hansen et al., (1996));

- In the case of two-step and iterated GMM the normalization matters (Stock et al, (2002));

The problem with weak identification is that it leads to non-normal distributions, even in large samples, so that conventional IV or GMM inferences are misleading and cannot be used. Fortunately, as it is argued in Stock et al. (2002), the CUE is fully robust to weak identification.

2.2.2 Large Dimensional Factor Analysis

In this subsection I would like to discuss briefly how one can construct static factors from a large data set. To be more precise, one "constructs" factors by estimating the latent factors by asymptotic principal components.

Let's first start with defining a static factor model. Let N be the number of cross-section clusters and T the number of time points in each cluster. Then the static factor model is defined as

$$x_{it} = \lambda_i' f_t + e_{it} \tag{14}$$

e_{it} is usually referred to as an idiosyncratic component, whereas $\lambda_i' f_t$ as a common component. λ_i' is a factor loading and f_t are latent factors. Even though this factor model is static, f_t is allowed to follow a dynamic process, e.g. $L(A) f_t = u_t$. Also, e_{it} is allowed to be a dynamic process and be cross-sectionally correlated.

Before explaining how the factors are estimated the following standard assumption is made:

Assumption A

1. $\mathbb{E}\|f_t\|^4 \le M < \infty$, $T^{-1}\sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \Sigma_f > 0$ is an $r \times r$ nonrandom matrix.
2. λ_i is either deterministic such hat $\|\lambda_i\| \le M < \infty$, or it is stochastic such that $\mathbb{E}\|\lambda_i\|^4 \le M < \infty$. In either case, $N^{-1}\Lambda'\Lambda \xrightarrow{p} \Sigma_\Lambda > 0$, an $r \times r$ nonrandom matrix, as $N \to \infty$.
3.
   - $\mathbb{E}[e_{it}] = 0$, $\mathbb{E}|e_{it}|^8 \le M < \infty$;
   - $\sigma_{i,j,t,s} := \mathbb{E}[e_{i,t} e_{j,s}]$, $|\sigma_{i,j,t,s}| < \bar{\sigma}_{ij}\ \forall (t,s)$ and $|\sigma_{i,j,t,s}| < \tau_{ts}\ \forall (i,j)$ such that
     $N^{-1}\sum_{i,j=1}^{N} \bar{\sigma}_{ij} \le M < \infty$, $T^{-1}\sum_{s,t=1}^{T} \tau_{ts} \le M < \infty$, and
     $(NT)^{-1}\sum_{i,j,t,s=1} |\sigma_{i,j,t,s}| \le M < \infty$;

- For every (t,s) $\mathbb{E}|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})]|^4 \leq M < \infty$.

4. $\{\lambda_i\}$, $\{f_t\}$, and $\{e_{it}\}$ are three mutually independent groups. Dependence in each group is allowed.

Basically, A.1 and A.2 imply the existence of r factors, as the largest r population eigenvalues of Σ_X will increase with N, whereas the remaining eigenvalue are bounded. The assumption that $\Sigma_\Lambda > 0$ implies that the r factors are identifiable (strong instruments). A.3 allows e_{it} to be cross-sectionally and serially correlated, but only weakly (approximate factor model). Finally, A.4 means that F_t can be serially correlated, λ_i can be correlated over i, and e_{it} can be cross-sectionally and serially correlated. In general, Assumption A, is a typical assumption that is made in the literature on static factors, for example Bai and Ng (2003, 2010).

The question now is how one estimates the latent factors f. Given the static framework, the factors can be estimated by the method of asymptotic principal components (PCA) originally invented by Connor and Korajzcyk (1986). According to this method, if one denotes $\widehat{f}$ to be an estimator for f then the $T \times r$ (r is a number of factors) matrix $\widehat{f}$ is $\sqrt{T}$ times the r eigenvectors associated with the r largest eigenvalues of the $T \times T$ matrix $XX'/(NT)$ in decreasing order with the following normalization $\widehat{f}'\widehat{f} = I_r$. The need of normalization is due to the fact that Λ, the matrix of loading coefficients, and f are not separately identifiable. In fact, what is identifiable is the span(f). For example, for any invertible matrix H one can write $f\Lambda = fH^{-1}H\Lambda = f^*\Lambda^*$. Once $\widehat{f}$ is obtained it's easy to get the estimates of the matrix of loading coefficients, which his given by $\widehat{\Lambda} = X'\widehat{f}/T$. Bai and Ng (2002) as well as Stock and Watson (2002a) demonstrated that span(f) can be consistently estimated by span$(\widehat{f})$ when $T, N \to \infty$. Finally, the number of factors can be estimated based on the information criteria developed by Bai and Ng (2002).

2.2.3 Methods of Instrument Dimension Reduction

In this section I briefly summarize the FIV estimator developed in Bai and Ng (2010). This paper analyzes a situation in which there is large set of instruments that can be used in instrumental variable estimation. It is shown that the optimal GMM has bias proportional to N/T, where N is the number of instruments and T is the length of the series.

Thus, if N is large then the bias can be quite large in finite samples. This confirms the findings of Meng et al. (2007). Moreover, unless $N/T \to 0$ the optimal GMM is inconsistent. This motivates the idea of reduction in the number of instruments by using the common factors of instruments as instruments. It is important to note that there is another paper by Kapetanios and Marcellino (2010) that also studies the use of factors as instruments (they call their estimator F-GMM). The two analyses are different in terms of assumed data generating processes of the variables: in the framework of F-GMM weak instruments are allowed whereas in FIV the standard asymptotics is adopted. Because weak instruments is a complicated topic by itself I restrict myself to the standard case.

Let's first analyze the main properties of FIV. Consider the following regression equation

$$y_t = x_t'\beta + \epsilon_t = x_{1t}'\beta_1 + x_{2t}'\beta_2 + \epsilon_t \text{ for } t = 1, 2, ..., T,$$

where x_t is $K \times 1$, x_{1t} is $K_1 \times 1$, and x_{2t} is $K_2 \times 1$. x_{2t} is assumed to be endogenous, that is $E[x_{2t}\epsilon_t] \neq 0$. Assume further

$$x_{2t} = \Psi' f_t + u_t \text{ for } t = 1, 2, ..., T,$$

where Ψ' is $K_2 \times r$, f_t is $r \times 1$ with $E[f_t\epsilon_t] = 0$ (factors are valid instruments), $r > K_2$ but is a small number, $E[\epsilon_t u_t] \neq 0$ (endogeneity problem). Finally, assume that there is "large" set of instruments $z_t = [z_{1t}, ..., z_{Nt}]$:

$$z_{it} = \lambda_i' f_t + e_{it} \text{ for } t = 1, 2, ..., T \text{ and } i = 1, ..., N,$$

where f_t is a vector of common factors, λ_i is the factor loadings, thus $\lambda_i' f_t$ is the common component of z_{it} and e_{it} is an idiosyncratic component. The two components are orthogonal. Neither the common nor an idiosyncratic components are observed. It is assumed that $E[e_{it}u_t] = E[e_{it}\epsilon_t] = 0$.

z_{it} is a valid but a "noisy" instrument. Given the assumed DGP, f_t are ideal instruments, but not observed. Hence, IV estimation with f_t is infeasible. Instead it is proposed to use use the asymptotic principal component estimates $\tilde{f}_t$. $\tilde{f} := [\tilde{f}_1, ..., \tilde{f}_T]'$ is given by a $T \times r$ matrix of r eigenvalues (multiplied by $\sqrt{T}$) associated with the r largest eigenvalues of the matrix $Z'Z/(TN)$ in decreasing order and $\tilde{\Lambda} := [\tilde{\lambda}_1, ..., \tilde{\lambda}_N] = Z\tilde{f}/T$.

The next assumption is made to derive asymptotic properties of the

FIV estimator:

Assumption B

1. $\mathbb{E}[\epsilon_t] = 0$, $\mathbb{E}|\epsilon|^{4+\delta} < \infty$ for some $\delta > 0$. The vector process $g_t(\beta^0) := F_t^+ \epsilon_t$ satisfies $\mathbb{E}[g_t(\beta^0)] = 0$ with $\mathbb{E}[g_t(\beta)] \neq 0 \ \forall \beta \neq \beta^0$. Let $\bar{g}^0 := T^{-1} \sum_{t=1}^T F_t^+ \epsilon_t$ and
$\sqrt{T}\bar{g}^0 = T^{-1/2} \sum_{t=1}^T F_t^+ \epsilon_t \xrightarrow{d} n(0, S^0)$ for some $S^0 > 0$.
2. $x_{2t} = \Psi' F_t u_t$ with $\Psi'\Psi > 0$, $\mathbb{E}[F_t u_t] = 0$, $\mathbb{E}[u_t \epsilon_t] \neq 0$, and $\mathbb{E}[F_t \epsilon_t] = 0$.
3. $\forall i, t$, $\mathbb{E}[e_{it} u_t] = 0$, and $\mathbb{E}[e_{it} \epsilon_t] = 0$.

Assume for simplicity that $K_1 = 0$. Define $\tilde{g}(\beta) := \tilde{f}_t \epsilon_t(\beta)$. Consider estimating β using the r moment restrictions $\bar{g}(\beta) := T^{-1} \sum_{t=1}^T \tilde{f}_t \epsilon_t(\beta)$. Let $S_{\tilde{f}_t x} := T^{-1} \sum_{t=1}^T \tilde{f}_t x_t'$ and $S^* := T^{-1} \sum_{t=1}^T \tilde{f}_t \tilde{f}_t' \epsilon_t^2$ Then, the efficient feasible FIV is given by

$$\hat{\beta}_{FIV} := \arg\min \bar{g}'(\beta) S^{*-1} g(\bar{\beta}) = (S'_{\tilde{f}_t x} S^{*-1} S_{\tilde{f}_t x})^{-1} S'_{\tilde{f}_t x} S^{*-1} S_{\tilde{f}_t y}.$$

Theorem 1 Under Assumptions A and B, as $N, T \to \infty$,

1. $\sqrt{T}(\hat{\beta}_{FIV} - \beta^0) \xrightarrow{d} n(0, \Omega_{FIV})$
2. $\Omega_{FIV} := \text{plim}(S'_{\tilde{f}_t x} S^{*-1} S_{\tilde{f}_t x}) = \Omega_{xf} S^{0-1} \Omega_{fx}$,
with $\Omega_{fx} := \text{plim}(T^{-1} \sum_{t=1}^T f_t x_t')$ and S^0 is as defined in Assumption B.

Additionally, Bai and Ng (2010) show that the optimal GMM that uses all the instruments is no more efficient than the FIV (even when the optimal weighting matrix is known) and requires (i) the bias correction and (ii) an additional assumption of $N/T \to 0$ (otherwise it is inconsistent). For a more technical treatment of these results one can consult Bai and Ng (2010).

Given the strict assumptions made in the FIV one can easily see how this estimator may fail. For example, suppose that $E[e_{it} u_t] \neq 0$. In that case one can contract an example in which the z_t would be better instruments than the factors f_t, or even an example in which the factors will become very weak instruments. Unfortunately, there is no way of testing the validity of the assumption $E[e_{it} u_t] = 0$.

Suppose now that the FIV is not applicable for a certain application but theres is still a need of instrument dimension reduction. What can be done instead is proposed in Bai and Ng (2008). It is proposed to construct principal components from the large set of possible instruments.

Assuming that $T > N$ one can construct N principal components, otherwise one can construct T of them. The next step is to select the instruments from the set of principal components. For instance, Bai and Ng propose different ways of selection: (i) boosting, (ii) ranking the predictive ability of the instruments one at a time, and (iii) information criteria applied to the ordered instruments. Given that all the three selection procedures have comparable properties I choose to use the latter one.

To apply the information criteria to the ordered set of instruments requires first the ordering of the set of principal components. The principal components can be ordered either based on t-statistics or R^2 from the regression of endogenous regressor on each principal component one at a time. After the ordering has been done one can apply an information criterion that selects how many instruments from the ordered set will be used. The discussion about the information criterion will be found in later section.

3 Data

In this section I describe the dataset used in the present analysis, which is basically a dataset from Ludvigson and Ng (2011). I also explain the data preparation process which must be done before the factors can be estimated.

3.1 Data Description

The dataset used in this paper is a famous macroeconomic panel used in Stock and Watson (2002b, 2005). The data set consists of 131 macroeconomic time series spanning the time period of 1960:01 to 2007:12, total of 576 observations. It is important to note that the original data set was only up to 2003:12 and had 132 series but it was later extended by Ludvigson and Ng (2011) up to 2007:12. One series, ao048, is no longer available on monthly basis after 2003. The data set comes together with MatLab code used in data reparation and construction of the static factors in Ludvigson and Ng (2011).

The data set represents a broad collection of macroeconomic time series, particularly it consists of real output and income, employment and hours, real retail, compensation and labor costs, international trade, consumer spending, housing starts, inventories and inventory sales ratios, orders and unfilled orders, manufacturing and sales data, capacity utilization measures, price indexes, interest rates and interest rate spreads, foreign exchange measures , and stock market indicators. A precise description of each series and its codding name can be found in the appendix of Ludvigson and Ng (2011).

The (monthly) data on a risky asset, r_{t+1}, that is used in estimation is constructed using S&P 500 index which is the part of the large data set (codding name "fspcom"). Particularly, this is constructed as a growth rate of S&P 500 index. The monthly return on a risk-free asset, r^f_{t+1} is constructed using the return on 3-month T-Bills. The series on per-capita consumption growth, is constructed using the data on nondurable consumption, which again the part of the large data set (codding name "ips18"), The monthly data on the US population (POPTHM), which is used to construct the per capita consumption, has been downloaded form the website of Federal Reserve Bank of St. Louis.

3.2 Data Preparation

To construct the static factors using asymptotic principal components from the large data set the series must firstly be stationarized, demeaned and standardized. How the data is transformed to make sure it is stationary depends on a particular series, the important part is that it is stationary. As it is explained in Ludvigson and Ng (2011), most of the series are expressed in the growth rates (e.g. GDP, exchange rates, stock market indexes, etc), some series, such as nominal interest rates (or any that are understood to have a unit root), are first-differenced, whereas others, such as prices, are second-log-differenced. However, the detailed information regarding the stationarization of each series can be found in the appendix of Ludvigson and Ng (2011).

In short, all the series after being transformed must be $I(0)$: all real series are in growth rates and as far as nominal series are concerned one can use either "light" transformation, which means that prices are in growth rates and interest rates are in levels or one can use "heavy" transformation, which means that the growth rate of prices as well as interest rate are in first differences.

4 Estimation: Linear Case

In this section I present the linear econometric model to be estimated, explain the problems related to the estimation of the model and motivate the choice of estimation methods. I further explain the process of instruments selection. Finally, I present the results of the estimation of the econometric model using the two methods of instrument dimension reduction and compare these results. Note that the linear estimation is done only for the standard CRRA case because the linearization of the Euler equation requires certain distributional assumptions.

4.1 Econometric Model

Recall that the Euler equation from C-CAPM (4), with the standard CRRA utility function, under the assumption that the asset return r_{jt+1} and consumption C_{t+1} are conditionally on information at time t jointly log-normally distributed, the equation (4) can be linearized as follows

$$\mathbb{E}_t[r_{i,t+1}] = \mu_{i,t} + \sigma \mathbb{E}_t[\Delta c_{t+1}] \tag{15}$$

following Yogo (2004) and letting

$$\mu_{i,t} := \mu_f - 0.5\text{Var}_t(r_{i,t+1} - \mathbb{E}[r_{i,t+1}]) + \sigma \text{Cov}_t(r_{i,t+1} - \mathbb{E}[r_{i,t+1}], \Delta c_{t+1} - \mathbb{E}[\Delta c_{t+1}]) \tag{16}$$

and for risk-free asset

$$\mu_{f,t} := -\log(\beta) - 0.5\sigma^2 \text{Var}_t(\Delta c_{t+1} - \mathbb{E}_t[\Delta c_{t+1}]) \tag{17}$$

where μ_f is defined as in (17) but the variance is not conditional on time t.

Given (16) one can rewrite (15) as a regression equation

$$r_{i,t+1} = \mu_i + \sigma \Delta c_{t+1} + \eta_{i,t+1} \tag{18}$$

where

$$\eta_{i,t+1} := r_{i,t+1} - \mathbb{E}_t[r_{i,t+1}] - \sigma(\Delta c_{t+1} - \mathbb{E}_t[\Delta c_{t+1}]) - (\mu_{i,t} - \mu_i) \tag{19}$$

where μ_i is defined the same way as (16) but the variance is not conditional on time t, thus $(\mu_{i,t} - \mu_i)$ can be thought of being innovations to the conditional variance of consumption.

One can already see several problems with the regression equation (18). First of all, there is endogeneity problem. Clearly, from (19)

$$\mathrm{Cov}_t(\Delta c_{t+1}, \eta_{i,t+1}) \neq 0.$$

Moreover, since asset returns and consumption return are conditionally heteroscedastic so is $\eta_{i,t+1}$. In general, the endogenaity problem requires an instrumental variable estimation and the conditionally heteroscedasticity of the error term, specifically, requires the GMM estimation due to efficiency reasons.

4.2 INSTRUMENTS

Acknowledging the endogeneity problem in (18) one should decide about the instruments which are to be used in the estimation. Recall that a valid instrument is such a variable that is correlated with the endogenous regressor and the same time is uncorrelated with the error term. Consider the linearized version of (4) written in the form of the econometric model (18). As it is discussed in Yogo (2004), under the assumption of conditional heteroscedasticity, the parameter of interest σ is identified by the moment restriction $\mathbb{E}[Z_t \eta_{it+1}] = 0$ only if $\mathbb{E}[Z_t(\mu_{i,t} - \mu_i)] = 0$. So, it could happen that, for example, some instrument z_{it} is correlated with the innovation to the conditional variance of consumption. This in turn suggest that for any z_{t-1} it holds that $\mathbb{E}[z_{t-1}(\mu_{i,t} - \mu_i)] = 0$. Hence, I use twice lagged instruments. For more details one can consult Yogo (2004).

Clearly, the rational expectations hypothesis suggest that all the information that is available for an economic agent at the time he is making a decision should be used in estimating (18). However, this implies the use of very large set of instruments, possibly all the macroeconomic series, and as it known from the work of Marimune, (1983) and Bekker, (1994), the IV estimators perform rather badly when the number of instruments tends to infinity. Moreover, as this is known from the work of Bai and Ng (2010) the efficient GMM estimator has a finite sample bias proportional to N/T, where N is the number of instrument and T is the number of observations, and requires $N/T \to 0$ for consistency. On the other hand, restricting the number of instruments to some subset of them, even to those that are intuitively seemed to be very "important", can be very misleading and can lead the loss of information. Therefore,there is a need of an efficient instrument dimension reduction method. The first such a method is the FIV estimator explained in the Section 2. So, it is proposed to use static factors contracted from the large data set of macroeconomic variables as instruments in the GMM estimation of (18). Construction of the factors is an efficient data reduction technique that enables one to overcome the problem of having too many instruments; the constructed factors are then to be used as instruments in the GMM estimation.

Given the data set of 131 series macroeconomic series covering the time period 1964:01-2007:12 and using the asymptotic principal component technique I construct 8 static factors. The number of static factors is the same as in Ludvigson and Ng (2011) since I am using the same data set as they do. This number in turn was selected by the information criteria developed in Bai and Ng (2002). I do not use all 8 factors as instruments in estimation of (18) because some of them have very low correlation with the regressor. To decide which out of the 8 factors are to be used in the estimation I look at at the R^2 from regressing Δc_{t+1} on $\hat{f}_{t-1}$ one at a time.The justification for the selection of instruments based on R^2 is that the factors are orthogonal by construction. In the Figure 1 I present the squared coefficient of correlation between the factors and the regressor (from now on to simplify notation I use f_{it} for $\hat{f}_{it}$). First of all, one can see immediately that the the correlation is rather low which is not a surprise since it is hard to predict the growth of consumption in general. Secondly, some factors are virtually uncorrelated with the regressor. Specifically, factors 3, 4, 6 and 7 have the squared coefficient of correlation equal to 0.000. On the other hand, factors 1, 2, 5, and 8 have nonzero squared correlation of similar magnitude larger than 0.011. Thus, it would be natural to use only the latter four factors as instruments in estimation of (18).

The other method of instrument dimension reduction is presented in Bai and Ng (2008). In that paper it is argued that the common factors of the large set of instrumental variables for a certain endogenous regressor do not have to be the best instruments for that regressor. Actually this would depend on the underlying DGP for endogenous variable and its instruments. It can well happen that some of the series have better instrument quality than the common factors. This means that ideally one should try all the possible combinations of instruments, which

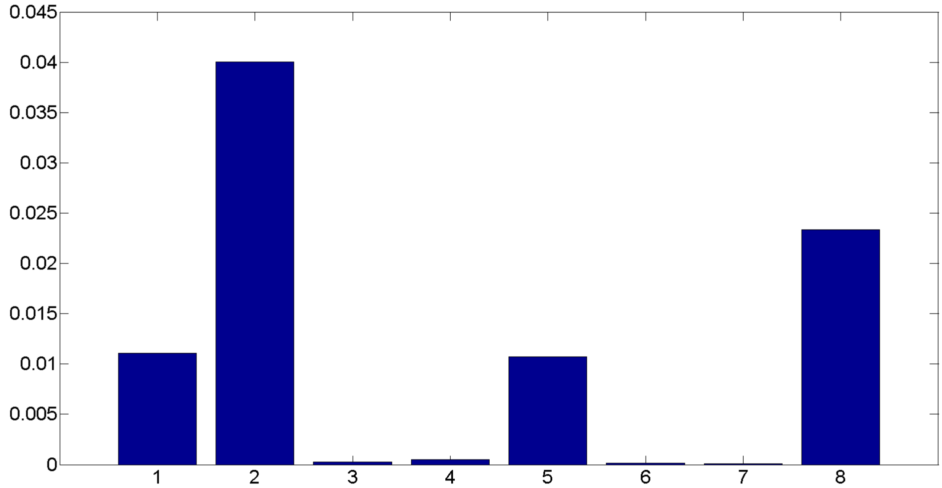

Figure 1: R^2 from regressing Δc_{t+1} on $f_{i,t-1} \in F_{t-1}$.

means that one should try $2^{|Z_t|}$ possibilities, which sometimes is simply not feasible. For instance, given 131 instruments this means one should try $2.7222589e^{39}$ combinations. What can be done in that case is, Bai and Ng (2008) argue, one can construct as many principal components as one has series (assuming $T > N$, otherwise one can construct as many principal components as one has observations), denote this collection of the principal components by $\tilde{F}$ and then search for the instruments in this set of orthogonalized series. After the principal components have been constructed one orders them based on the R^2 from the regression of the endogenous regressor on a principal component one at a time. Let $\tilde{F}_{\text{ord}}$ denote the ordered set of principal components. Next it is to be decide how many of these principal components are to be used as instruments. For that one can use some information criteria, such as BIC, to select the optimal number of instruments. Let Q be the first n elements of $\tilde{F}_{\text{ord}}$ that will be used as instruments in the estimation. Then the following procedure selects the number n

$$Q := \arg\min_n \{\log(\hat{\sigma}_n^2) + n\log(T)/T\} \tag{20}$$

where $\hat{\sigma}_n^2 := T^{-1}\sum_{t=1}^{t} \hat{e}_{t,n}^2$ and $\hat{e}_{t,n}^2$ denote the squared residual from regressing the endogenous variable of the first n ordered principal components.

Following this approach I construct 131 principal components from 131 series. I give a plot of R^2's from the regression of Δc_{t+1} on $f_{i,t-1}$ for

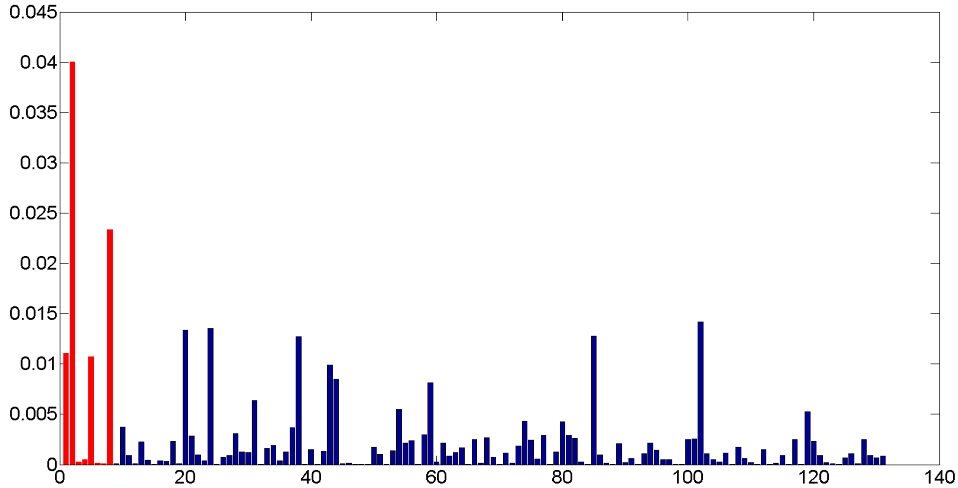

Figure 2: R^2 from regressing Δc_{t+1} on $\tilde{f}_{i,t-1} \in \tilde{F}_{t-1}$ for $i = 1, ..., 131$.

$i = 1, ..., 131$ in the Figure 2 (the ones in red are the R^2 from regressing $f_i \in F$ for $i = 1, ..., 8$). Then I order the principal components based on the R^2 and perform the minimization of (20) to select the principal components that are to be used as instruments. The value of the criterion, as a function of n, is depicted in the Figure 3. Clearly, the number of instruments selected is either first 9 or 10 principal components from $\tilde{F}_{\text{ord}}$. The difference between the value of the minimized criterion at $n = 9$ and $n = 10$ is virtually zero so I select the more parsimonious option, that is the first 9 principal components are to be used as instruments. The first 9 elements from $\tilde{F}_{\text{ord}}$ are actually 2nd, 8th, 102nd, 24th, 20th, 85th, 38th, 1st and 5th elements of $\tilde{F}$. This means that if one denotes the set of instruments in which the latter are selected from the common factors F_t by Z_t and the set of instruments in which the instruments are selected from the set of all principal components $\tilde{F}_t$ by Q_t, then it holds, in this particular case, that $Z_t \subset Q_t$.

A common critic of the use of factors is that it is hard to interpret them. One of the ways to give an interpretation to a factor or a principal component is to consider the regression in which each individual series from which a factor was constructed is regressed on the factor and then the R^2's from this regressions is used to give interpretation (this method of interpreting factors is used in Bai and Ng (2008) and Ludvigson and Ng (2011)). The interpretation of the common factors is taken from Ludvigson and Ng (2011). Particularly, the common factors that are used as instruments $Z_t = \{f_{1,t}, f_{2,t}, f_{5,t}, f_{8,t}\}$ represent real activity factor, in-

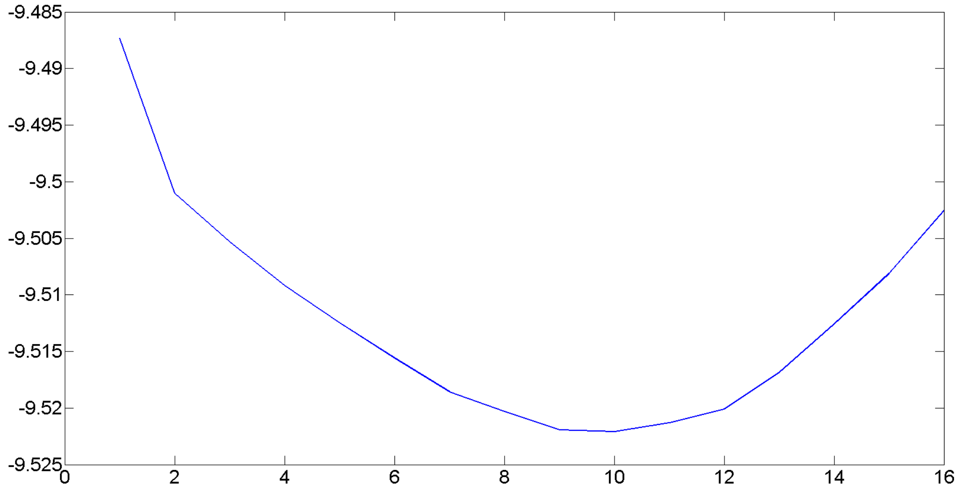

Figure 3: The value of the minimized criterion in (20) as a function of n.

terest rate spreads, interest rates, and stock market, respectively. The interpretation of the instruments selected from the set of 131 principal components $Q_t = \{f_{1,t}, f_{2,t}, f_{5,t}, f_{8,t}, f_{20,t}, f_{24,t}, f_{38,t}, f_{85,t}, f_{102,t}\}$ as follows: the first four are the same, then starting from the fifth one "money stock: currency held by the public" (72nd series), "commercial and industrial loans outstanding" (77th series), "employees on nonfarm payrolls - mining" (35th series), "CPI-U: all items less medical care" (123rd series), and finally "civilian labor force: total employed" (23rd series).

To finish this subsection I would like to illustrate using a simple example why it is wise to select instruments based on their correlation with the endogenous variable. To illustrate how exactly a high correlation between an instrument and the endogenous variable will lead to a more efficient estimation let's consider a simple example. Suppose, we have a simple regression

$$y_t = \alpha + \beta x_t + \epsilon_t, \text{ for } t = 1, ..., T$$

where x_t is an endogenous regressor. Suppose also, that we have two mutually orthogonal instruments $z_t = (z_{1t}, z_{2t})'$ and ϵ_t is conditionally heteroskedastic. Recall, that for the linear model the asymptotic variance covariance matrix of the optima GMM is given by

$$\text{var}(\hat{\theta}^{GMM}) =: \Omega := \left(\mathbb{E}[\tilde{x}_t z_t']\mathbb{E}[\epsilon_t^2 z_t z_t']^{-1}\mathbb{E}[z_t \tilde{x}'_t]\right)^{-1},$$

where

$$\hat{\theta}^{GMM} := (\hat{\alpha}^{GMM}, \hat{\beta}^{GMM})'.$$

Let's also assume for the simplicity of algebraic expressions that $\text{var}(x_t) = \text{var}(z_{it}) = \text{var}(\epsilon) = 1$ and $\mathbb{E}[z_{it}] = 0$ for $i = 1,2$. Then the off-diagonal elements of Ω are all zero and $\text{var}(\hat{\alpha}^{GMM}) = 1$ and $\text{var}(\hat{\beta}^{GMM}) = \frac{1}{r_1^2 + r_2^2}$, where r_i^2 is a squared correlation between the endogenous regressor x_t and an instrument z_{it} for $i = 1,2$. From this simple example one can see that the larger the correlation between the endogenous regressor and the instrument is the smaller the asymptotic variance of the estimator is and vice versa.

4.3 Estimation Results

In this subsection I present the result of the estimation of (18) using the two sets of instruments. The one set of instruments $Z_t \subset F_t$ has been selected from the common factors of the large data set of macroeconomic series. The other set of instruments $Q_t \subset \tilde{F}_t$ has been selected from all principal components of the same large data set. As it was argued in the previous subsection it happens to be for this particular situation that $Z_t \subset Q_t$. Additionally, in all estimations the vector of ones is included in the set of instruments.

I use CUE as a method of implementing the optimal GMM estimation, because it is fully robust to weak identification problem and MC studies show that its J-statistic is more reliable in finite samples as it is shown in Hansen et. al. (1996). Additionally, because I deal with time series I use the HAC estimator of the weighting matrix developed by Newey and West (1987).

The result of the estimation of (18) using Z_t is presented in the Table 1. The estimate of the parameter of interest σ is not significant at even the15% level but its magnitude is reasonable. Some studies on micro level, using panel data, suggest that the median of the CRRA is around 1.7, for example see Chiappori and Paiella (2011), the others such as Hartley et al. (2005) obtain an estimate of the CRRA around 1.The J-statistic is small with associated p-value of roughly 0.15 implying that the H_0 of correctly specified model cannot be rejected at the 15% confidence level. This is not a very tremendous result. The 90% confidence interval for σ is given by $[-0.381, 2.085]$. Which is again a more or less satisfactory, although the fact that it includes negative values is disturb-

	Coefficient	Std. Error	t-Statistic	Prob.
$\hat{\mu}$	0.004020	0.002043	1.968014	0.0495
$\hat{\sigma}$	0.852018	0.748163	1.138813	0.2553
		J-statistic		5.370839
Instrument rank	5	Prob(J-statistic)		0.146571

Table 1: The result of the GMM estimation of (18) using Z_t as instruments.

ing. Let's now turn to the estimation result when Q_t is used as instruments.

The result of the estimation of (18) using Q_t is presented in the Table 2. This is clearly an example of a successful estimation. The estimate of σ has somewhat small standard errors, that results in a tighter confidence interval, and has a reasonable magnitude. The 90% confidence interval for σ is given by $[0.032, 1.609]$. Moreover, the J-statistic is low with the associated p-value of roughly 0.64 implying that the H_0 of correctly specified model cannot be rejected at any reasonable significance level. In fact, the result in the Table 2 is a replication of Bai and Ng's (2008) result presented in their Table 7a and labeled under FIV_{ic}

It is important to note that with this careful selection of instruments the resulted estimation is quite robust to the implementation method of the optimal GMM. The results presented in Table 2 and 3 are very similar when instead of CUE-GMM the iterative or 2 stage GMM is used. This seems to suggest that the problem of weak identification may not be present, in contrast to what is argued in Yogo (2004). In general, no obvious signs of weak identification have been detected.

Comparing the results of the two estimations using different sets of instruments there is little doubt that the estimation in which the set of instruments is selected from the principal components over-performs the one that uses factors as instruments in terms of the estimation pre-

	Coefficient	Std. Error	t-Statistic	Prob.
$\hat{\mu}$	0.003523	0.001979	1.780265	0.0756
$\hat{\sigma}$	0.820417	0.478854	1.713295	0.0872
		J-statistic		6.077228
Instrument rank	10	Prob(J-statistic)		0.638581

Table 2: The result of the GMM estimation of (18) using Q_t as instruments.

cision (smaller standard errors, narrower the confidence intervals). This result supports the reasoning of Bai and Ng (2008) that the common factors of a large set of instruments (for some endogenous variable) need not be the best instruments (for this variable). It can well happen that a combination of some instruments will do a better job than a subset of common factors. To somehow justify this point consider the following two DGP.

DGP 1: THE ENDOGENOUS VARIABLES AND THE INSTRUMENTS HAVE THE SAME COMMON FACTORS.

$$y_t = \alpha + \beta' x_t + e_t, \forall t = 1, \ldots T \tag{21}$$

$$x_t = \Lambda F_t + \nu_t, \forall t = 1, \ldots T \tag{22}$$

$$z_t = \Psi F_t + \epsilon_t, \forall t = 1, \ldots T \tag{23}$$

also assume that $\mathbb{E}_t[e_t \nu_t] \neq 0$ (endogeniety problem), $\mathbb{E}_t[e_t \epsilon_t] = 0$ (z_t a valid instrument). This is the DGP assumed in Bai and Ng (2010). In this case it is always the case, at least asymptotically, that the factors are better instruments (even estimated factors, as it is proved in Bain and Ng (2010)). On the other hand, if $\mathbb{E}_t[e_t \epsilon_t] \neq 0$ then it is no longer obvious whether the factors are better instruments than a combination of some z_{it}, even asymptotically. For example, one can construct an example in which with $\mathbb{E}_t[\nu_t \epsilon_t] \neq 0$ z_t would be better instruments (keeping N fixed).

DGP 2: HIERARCHICAL STRUCTURE.

$$y_t = \alpha + \beta' x_t + e_t, \forall t = 1, \ldots T \tag{24}$$

$$x_t = B z_t + \psi_t, \forall t = 1, \ldots T \tag{25}$$

$$z_t = \Phi F_t + \xi_t, \forall t = 1, \ldots T \tag{26}$$

assuming that $\mathbb{E}_t[\psi_t \xi_t] = \mathbb{E}_t[e_t \xi_t] = 0$.Clearly, in this case z_t are better instruments than F_t (assuming that the number of instruments, whatever large, stays constant). Therefore, which instruments are better, at least asymptotically, depends crucially on the assumed DGP of explained, endogenous explanatory and instrumental variables.

5 Estimation: Nonlinear Case - Single Equation

In this section I perform the estimation of the Euler Equation (4) for the risky asset in its nonlinear form. I use two different specifications for the utility function one is the standard CRRA, $u(c_t) = \frac{c_t^{1-\sigma}-1}{1-\sigma}$, the other is the CRRA with external habit $u(c_t, X_t) = \frac{(c_t/X_t)^{1-\sigma}-1}{1-\sigma}$, where $X_t = c_{t-1}$ stands for external habit. These two different utility function specifications lead to two different Euler equations. I am not aware of any other work that deals with the nonlinear Euler equation estimation using a large set of possible instruments. The novelty of what I am doing is in attempt to resolve the problem of having too many instruments in the non-linear Euler equation estimation. This estimation is a lot trickier because of two reasons: to my knowledge there is no proper methodology of ranking and selecting instruments from a large set of instruments in nonlinear GMM estimation and secondly, the FIV methodology has been developed for linear models only. To overcome the second problem I extend Bai and Ng's (2008) method of instrument selection to nonlinear case. I also apply the FIV to the non-linear set up, thought, it must be mentioned that the use of the FIV estimator in the non-linear set up is rather experimental.

5.1 Econometric Model

Recall that the Euler Equation for an asset i, with the standard CRRA utility function specification, is given by

$$\mathbb{E}[\beta(\frac{c_t}{c_{t+1}})^{\sigma}(1+r_{it+1})|I_t] = 1 \tag{27}$$

and with external habit formation it is given by

$$\mathbb{E}[\beta(\frac{c_t}{c_{t-1}})^{\sigma-1}(\frac{c_{t+1}}{c_t})^{-\sigma}(1+r_{it+1})|I_t] = 1 \tag{28}$$

Define

$$h_t(\beta,\sigma) := h(c_t, c_{t+1}, r_{t+1}, \beta, \sigma) := \beta(\frac{c_t}{c_{t+1}})^{\sigma}(1+r_{t+1}) - 1$$

and

$$q_t(\beta,\sigma) := q(c_{t-1}, c_t, c_{t+1}, r_{t+1}, \beta, \sigma) := \beta(\frac{c_t}{c_{t-1}})^{\sigma-1}(\frac{c_{t+1}}{c_t})^{-\sigma}(1 + r_{it+1}) - 1.$$

Given some vector of instrumental variables $z_t \in I_t$ one can write (27) and (28) as $\mathbb{E}[h_t(\beta,\sigma)z_t|I_t] = 0$ and $\mathbb{E}[q_t(\beta,\sigma)z_t|I_t] = 0$; applying the law of iterated expectations one gets the following unconditional population moment restriction

$$g(\beta^0, \sigma^0) := \mathbb{E}[h_t(\beta^0, \sigma^0)z_t] = 0 \tag{29}$$

$$q(\beta^0, \sigma^0) := \mathbb{E}[q_t(\beta^0, \sigma^0)z_t] = 0 \tag{30}$$

Let $g_{hT}(\beta,\sigma)$ and $g_{qT}(\beta,\sigma)$ denote the sample equivalence of (29) and (30), respectively, that is

$$g_{hT}(\beta,\sigma) := T^{-1}\sum_{t=1}^{T} h_t(\beta,\sigma)z_t$$

$$g_{qT}(\beta,\sigma) := T^{-1}\sum_{t=1}^{T} q_t(\beta,\sigma)z_t.$$

Given $g_{hT}((\beta,\sigma))$ and $g_{qT}((\beta,\sigma))$ one can now perform the optimal GMM estimation the way it was presented earlier, that is by minimizing the following objective function for the standard case

$$J_{hT}(\theta) = g'_{hT}(\theta)W_T(\theta)g_{hT}(\theta)$$

and for external habit formation

$$J_{qT}(\theta) = g'_{qT}(\theta)W_T(\theta)g_{qT}(\theta)$$

where $\theta := (\beta,\sigma)'$, and $W_T(\theta)$ is an estimator of the optimal weighting matrix using HAC because I am operating in time series environment. The implementation method of the optimal GMM which I use is CUE due to its good performance in finite sample mentioned earlier.

5.2 Instruments

In this subsection I discussed the process of selection instruments for estimating (27) and (28). Again as before let's denote the set of instruments selected from the common factors by $Z_t^h \subset F_t$ for the standard case and $Z_t^q \subset F_t$ for the case with an external habit, further, let denote the set of instruments selected from the principal components by $Q_t^h \subset \tilde{F}_t$ for the standard case and $Q_t^q \subset \tilde{F}_t$ for the case with an external habit.

One cannot directly implement the Bai and Ng's (2008) method of instrument dimension reduction because the latter one has been designed for the linear set up only. To overcome this problem I extend the method to non-linear models.

Specifically, the first step in Bai and Ng's (2008) method stays the same, that is,using a large set of possible instruments I construct as many principal components as I have instruments (given that the number of instruments is strictly smaller than the number of observations). However, the second step, involving the ranking of principal components as instruments is modified. The ranking now is based on the R^2 obtained from the regression of the partial derivative of $h_t(\beta,\sigma)$ w.r.t. σ on each orthogonal series one at a time for the standard case and the regression of the partial derivative of $q_t(\beta,\sigma)$ w.r.t. σ on each orthogonal series one at a time. The problem with this step, however, is that the partial is a function of the parameters of interest which values are unknown. To deal with this problem I set $\beta = 0.995$ this is the value one obtains nearly always when estimating (29) or (30), also this is about the value that one would expect to be true. At the same time I try different value of σ. The result reported below is the result obtained with $\sigma = 0.8$. It is important to note though, that the following analysis is absolutely robust to different value of σ: I tried different values of σ on the interval of $[0,100]$ and there was virtually no difference (the reason why the value of σ had no effect on the instrument ranking and selection will become clear later).

The motivation of looking at the correlation between the gradient and instruments, as the criterion for instrument ranking, in nonlinear setup is basically the same as the motivation of looking at the correlation between an instrument and the endogenous variable in the linear set up. Specifically, recall that the asymptotic variance of the optimal GMM is given by $\Omega_{GMM} = (D'\Omega^{-1}D)^{-1}$ where $D \equiv \mathbb{E}[\frac{\partial[h_t(\theta^0)z_t]}{\partial\theta'}]$. Therefore, by selecting the instruments that are highly correlated with the gradient one can reduce the asymptotic variance. However, I only look at the derivative with respect to parameter σ because there is little problem with estimating the discount factor β.

Finally, the third step of the method, the selection of the number of

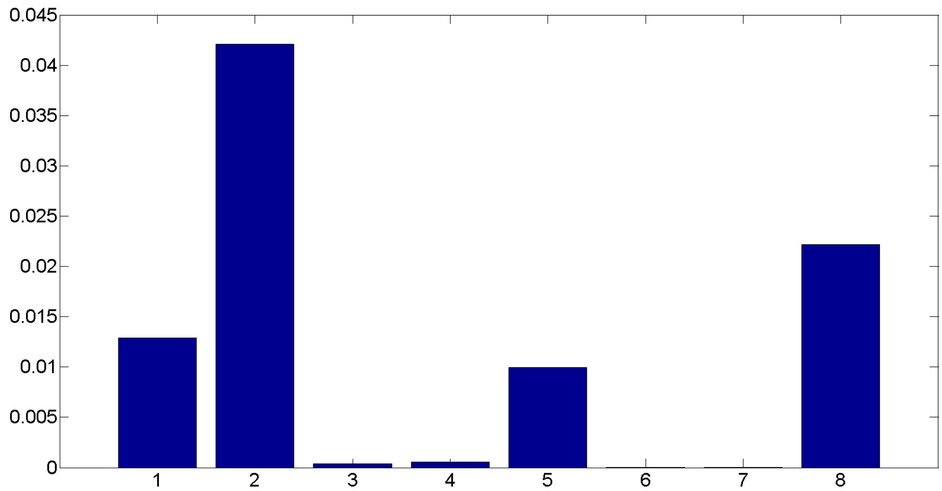

Figure 4: R^2 from regressing $d^h_{t+1}(\tilde{\beta},\tilde{\sigma})$ on $f_{i,t-1} \in F_{t-1}$

the principal components to be used as instruments by information criterion, stays the same. Below I present in details the implementation of the modified version of the method.

The partial derivative of $h_t(\beta^0,\sigma^0)$ w.r.t. σ is given by

$$d^h_{t+1}(\beta,\sigma) := \frac{\partial h_t}{\partial\sigma}(\beta,\sigma) = \left[h_t(\beta,\sigma)+1\right]\ln(c_t/c_{t+1}),$$

and the partial derivative of $q_t(\beta^0,\sigma^0)$ w.r.t. σ is given by

$$d^q_{t+1}(\beta,\sigma) := \frac{\partial q_t}{\partial\sigma}(\beta,\sigma) = \left[q_t(\beta,\sigma)+1\right](\ln(c_t/c_{t-1}) - \ln(c_{t+1}/c_t)).$$

The Figure 4 and 5 present the R^2 from the regression of $d^h_{t+1}(\beta,\sigma)$ and $d^q_{t+1}(\beta,\sigma)$ on each component of F_{t-1}, respectively. That is, the Figure 4 presents R^2's from the following regressions

$$d^h_{t+1}(\tilde{\beta},\tilde{\sigma}) = \alpha^h_0 + \alpha^h_1 f_{i,t-1} + \nu^h_{t+1}\ ,\ f_{i,t-1} \in F_{t-1} \forall i = 1,...8$$

and the Figure 5 presents the R^2's from the regressions

$$d^q_{t+1}(\tilde{\beta},\tilde{\sigma}) = \alpha^q_0 + \alpha^q_1 f_{i,t-1} + \nu^q_{t+1}\ ,\ f_{i,t-1} \in F_{t-1} \forall i = 1,...8.$$

Where $(\tilde{\beta},\tilde{\sigma}) = (0.8, 0.995)$ and as it was already discussed different values for the parameters were tried as well but they did not change any result.

One can see that the only candidates for instruments, in the standard case, are the factors 1,2, 5, and 8, and for the case with an external habit are the factors 1, 2, 3, 5, 6, 8. Thus, $Z_t^h := \{f_{1,t}, f_{2,t}, f_{5,t}, f_{8,t}\}$ and $Z_t^q := \{f_{1,t}, f_{2,t}, f_{3,t}, f_{5,t}, f_{6,t}, f_{8,t}\}$. One can notice that the Z_t from the previous section concise with Z_t^h.

In the Figure 6 and 7 I present the R^2 from the regression of $d^h_{t+1}(\tilde{\beta}, \tilde{\sigma})$ and $d^q_{t+1}(\tilde{\beta}, \tilde{\sigma})$ on each component of $\tilde{F}_{t-1}$, respectively. That is, the Figure 6 presents the R^2's from the following regressions

$$d^h_{t+1}(\beta, \sigma) = \tilde{\alpha}^h_0 + \tilde{\alpha}^h_1 \tilde{f}_{i,t-1} + \tilde{v}^h_{t+1}\ ,\ \tilde{f}_{i,t-1} \in \tilde{F}_{t-1} \forall i = 1, ...131$$

and the Figure 7 presents the R^2's from the regressions

$$d^q_{t+1}(\beta, \sigma) = \tilde{\alpha}^q_0 + \tilde{\alpha}^q_1 \tilde{f}_{i,t-1} + \tilde{v}^q_{t+1}\ ,\ \tilde{f}_{i,t-1} \in \tilde{F}_{t-1} \forall i = 1, ...131.$$

In green color are the principal components that were selected as instruments. The selection procedure I use here is the same as the one in explained in the linear case, however, this time instead of an endogenous variable I used the partial derivative $d^h_{t+1}(\beta, \sigma)$ and $d^q_{t+1}(\beta, \sigma)$ minimizing (20).

One can easily see that the Figure 6 is nearly identical to the Figure 2. Moreover, the set of instruments selected from the set of ranked (based on the R^2) principal components $\tilde{F}_t$ by minimization of (20) in the standard case is again the same as in the linear case, that is, $Q^h_t = Q_t$. On the other hand, minimizing the criterion in (20) for the case with an ex-

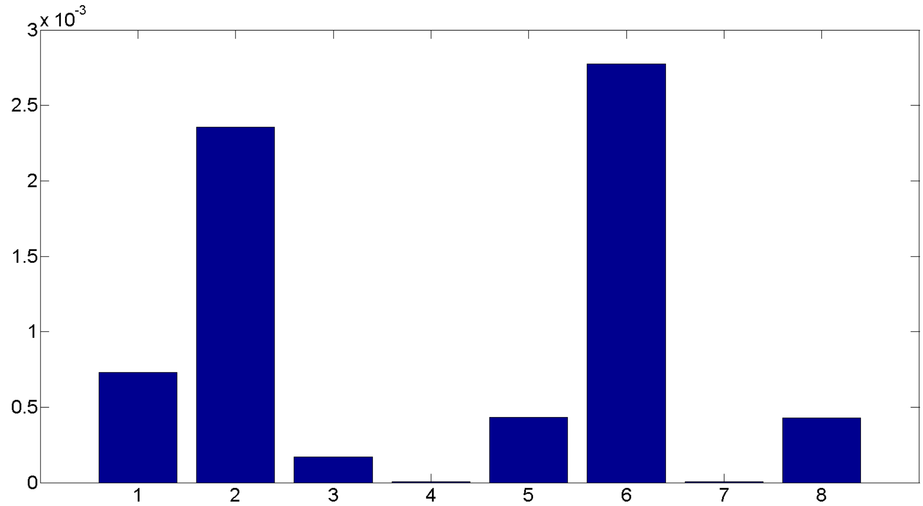

Figure 5: R^2 from regressing $d^q_{t+1}(\tilde{\beta}, \tilde{\sigma})$ on $f_{i,t-1} \in F_{t-1}$

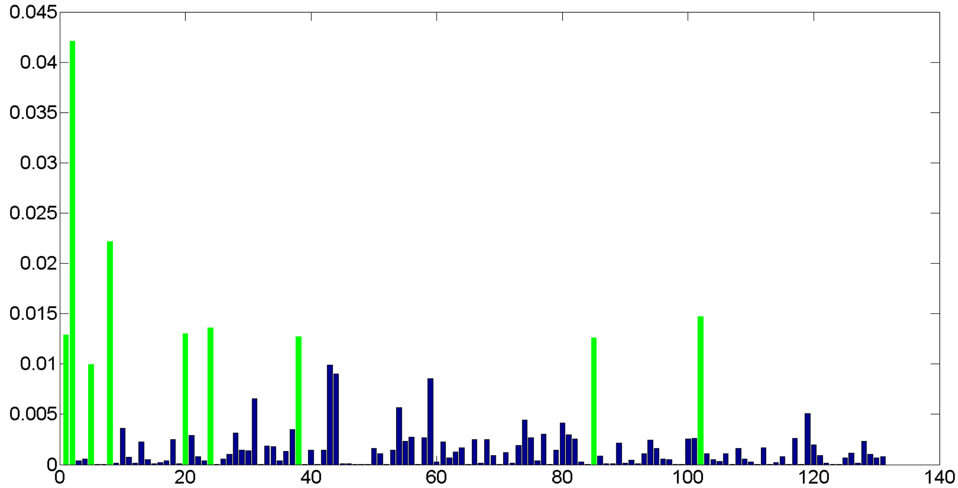

Figure 6: R^2 from regressing $d^h_{t+1}(\tilde{\beta},\tilde{\sigma})$ on $\tilde{f}_{i,t-1} \in \tilde{F}_{t-1}$

ternal habit selects only three instruments, those are the principal components 44, 85, and 102. Thus, $Q^q_t = \{f_{44,t}, f_{85,t}, f_{102,t}\}$ The Figure 8 and 9 show the minimized value of the criterion in (20) as a function of the first n principal components from $\tilde{F}_{t,ord}$ for the standard case and the case with an external habit, receptively.

To give some interpretation of the instruments Q^q_t I use the same method as in the previous section, that is, I regress each series on each instrument and then I look at the R^2 from these regressions. As a result I find that $f_{44,t}$ has the highest correlation with 125th series (price deflater),

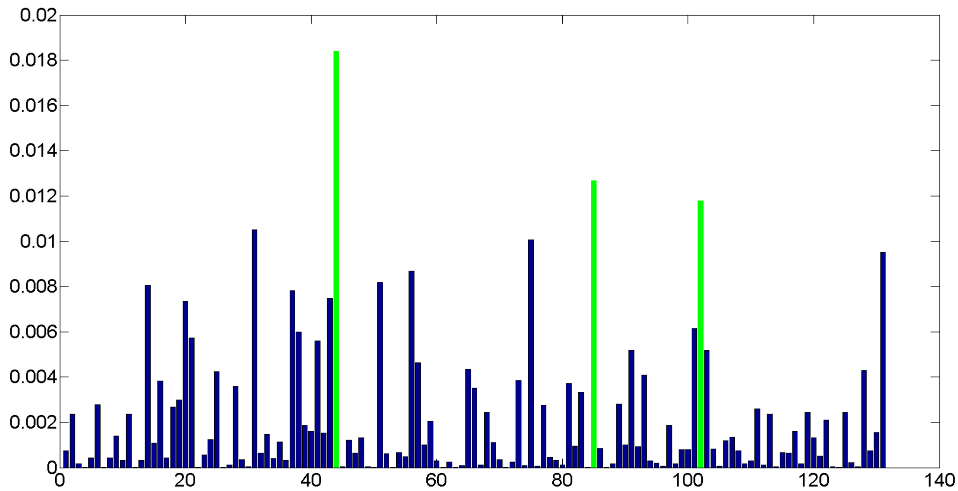

Figure 7: R^2 from regressing $d^q_{t+1}(\tilde{\beta},\tilde{\sigma})$ on $\tilde{f}_{i,t-1} \in \tilde{F}_{t-1}$

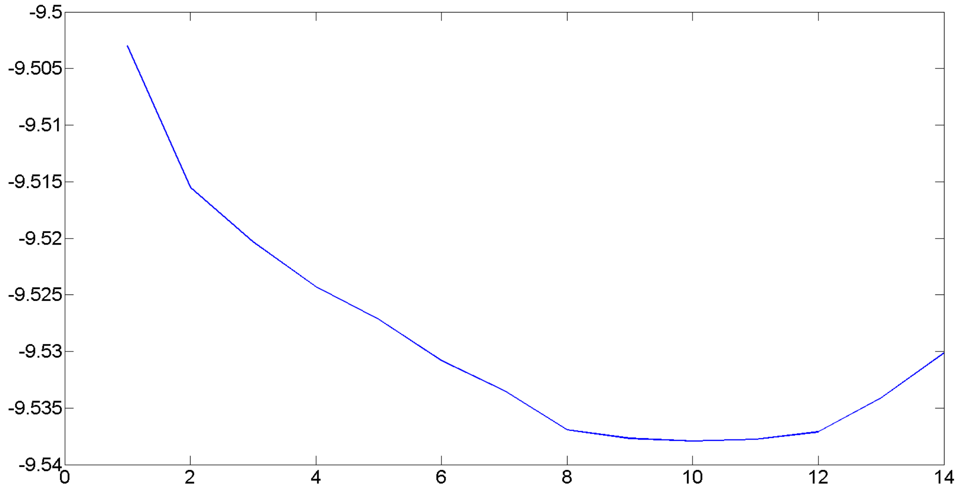

Figure 8: The value of the minimized criterion in (20), the standard case, as a function of n.

$f_{85,t}$ has the highest correlation with 123rd series (CPI less medical expenditures), and finally $f_{102,t}$ has already been given the interpretation, this one is mostly correlated with the civilian labor force.

The actual reason why Z_t^h and Q_t^h are equal to their counterparts in linear case is because $d_{t+1}^h(\beta,\sigma) \approx -\Delta c_{t+1} := -\ln(c_{t+1}/c_t)$. To see this, consider the definition of $d_{t+1}^h(\beta,\sigma)$. It is defined as $h_t(\beta,\sigma)\ln(c_t/c_{t+1})$. But one can see, that because β is close to 1, $[c_t/c_{t+1}]^\sigma$ (monthly data) is close to one, and finally the gross return on the risky asset is close to one

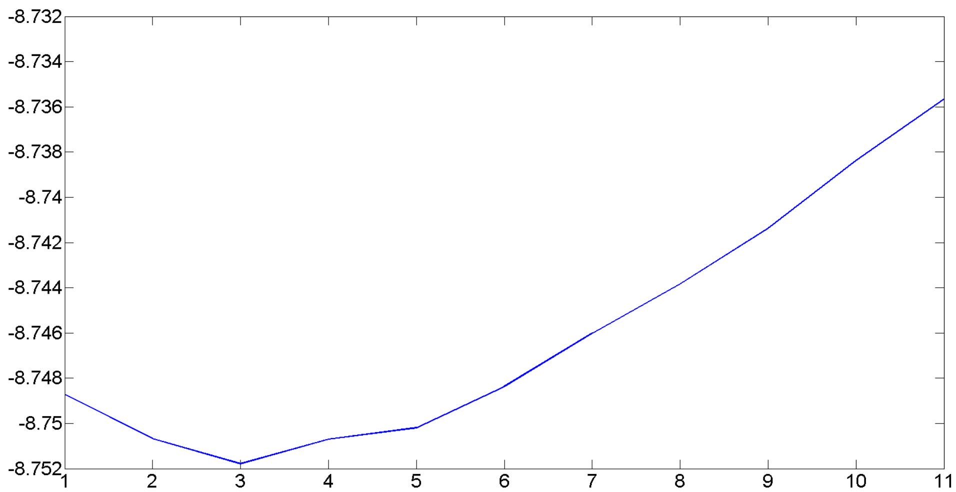

Figure 9: The value of the minimized criterion in (20), in the case of an external habit, as a function of n.

(monthly data) $h_t(\beta,\sigma)$ itself close to one. Hence, $d_{t+1}(\beta,\sigma) \approx -\Delta c_{t+1}$ and the resulted instruments are identical. Finally, since the instruments are the same their interpretation stays the same as well. On the other hand, in the case with an external habit the partial $d^q_{t+1}(\beta,\sigma) \approx (\Delta c_t - \Delta c_{t+1})$. So, there is no surprise why $Q^q_t \neq Q^h_t$.

5.3 Estimation Results

The result of the estimation of (27) and (28) using Z^h_t and Z^q_t, respectively, are presented in the Table 3 and 4. The magnitude of the estimated σ, using the standard utility specification, is in the reasonable range, however the standard error associated with a parameter estimate σ is quite large which will result in a wide confidence interval; β is estimated well with an expected magnitude and a very small standard error. The 90% confidence interval for σ is $[-0.334, 2.120]$. The J-statistic is small with associated p-value of nearly 0.15 implying that the H_0 of correctly specified model cannot be rejected at around the 15% confidence level.

The magnitude of the estimated σ, in the case of an external habit formation, is somewhat low, moreover, the standard error associated with a parameter estimate σ is very large which results in an extremely wide confidence interval; β is estimated well with an expected magnitude and a very small standard error. The 90% confidence interval for σ is $[-1.939, 2.257]$. The J-statistic is small with associated p-value of nearly 0.24 implying that the H_0 of correctly specified model cannot be rejected at around the 20% confidence level. In general the result of this estimation is rather poor. Let's now turn to the estimation result when Q^h_t and Q^q_t are used as instruments.

The result of the estimation of (27) and (28) using Q^h_t and Q^q_t, respectively, are presented in the Table 5 and 6. The results of these estima-

		Coefficient	Std. Error	t-Statistic	Prob.
$\hat{\beta}$		0.996048	0.001993	499.6905	0.0000
$\hat{\sigma}$		0.892867	0.744662	1.199024	0.2310
			J-statistic		5.348408
Instrument rank		5	Prob(J-statistic)		0.147992

Table 3: The result of the GMM estimation of (27) using Z^h_t as instruments.

	Coefficient	Std. Error	t-Statistic	Prob.
$\widehat{\beta}$	0.995903	0.001345	740.3803	0.0000
$\widehat{\sigma}$	0.159009	1.273443	0.124866	0.9007
		J-statistic		6.811345
Instrument rank	7	Prob(J-statistic)		0.235053

Table 4: The result of the GMM estimation of (28) using Z_t^q as instruments.

tions are rather successful. The magnitude of the estimated σ, using the standard utility specification, is in the reasonable range, also the standard error associated with a parameter estimate σ is rather small. Once again β is estimated well with a desirable magnitude. The 90% confidence interval for σ is $[0.031, 1.606]$. Moreover, the J-statistic is low with the associated p-value of 0.64 implying that the H_0 of correctly specified model cannot be rejected at any reasonable significance level.

The magnitude of the estimated σ, in the case of an external habit formation, is similar to most of the previous results, unfortunately, the standard error associated with a parameter estimate σ is a little bit large; β is estimated well with an expected magnitude and a very small standard error. The 90% confidence interval for σ is $[-0.185, 1.669]$. The J-statistic is small with associated p-value of more than 0.8 implying that the H_0 of correctly specified model cannot be rejected at any reasonable confidence level. In general the result of this estimation is quite good.

Let's now sum up the results of this subsection. First of all, it is important to note that, except the estimation results of (28) using Z_t^q, the results are quantitatively similar in terms of parameter estimates as well as confidence intervals for the coefficient of relative risk aversion. So, not only the results of estimating (27) are similar to the results obtained estimating the linearized version of the same equation but also these re-

	Coefficient	Std. Error	t-Statistic	Prob.
$\widehat{\beta}$	0.996466	0.001952	510.3692	0.0000
$\widehat{\sigma}$	0.818520	0.478147	1.711859	0.0875
		J-statistic		6.054008
Instrument rank	10	Prob(J-statistic)		0.641182

Table 5: The result of the GMM estimation of (27) using Q_t^h as instruments.

	Coefficient	Std. Error	t-Statistic	Prob.
$\widehat{\beta}$	0.995387	0.001870	532.2665	0.0000
$\widehat{\sigma}$	0.741885	0.562563	1.318760	0.1878
		J-statistic		0.411073
Instrument rank	4	Prob(J-statistic)		0.814211

Table 6: The result of the GMM estimation of (28) using Q_t^q as instruments.

sults are similar to the ones obtained using different utility specification (external habit formation). One of the reason the coefficient of relative risk aversion in (28) is estimated poorly using Z_t^q is that the correlations between the partial derivative d_{t+1}^q and the factors are very low. That is, factors, in this situation, are very weak instruments. For example, if instead of the CUE I use, say, the two step efficient GMM, as a way of implementing the optimal GMM, then the parameter estimate of σ in (28) using Z_t^q is 1.435 (compare to 0.16 with the CUE). This seem to indicate that the use of Z_t^q in estimating (28) leads to a severe weak identification problem. Hence, the result of this particular estimation ought to be disregarded.

6 Estimation: Nonlinear Case - Two Equations

In this section I turn to the multiple equation GMM estimation. Recall that the Euler equation must hold for all assets i. As it was shown first by Mehra and Prescott (1985) in their seminal work, trying to reconcile an observed equity premium (the difference between returns on risky asset and risk-free asset) in the U.S. economy using the framework of the C-CAPM one needs either an unrealistically large or very small (virtually zero) value of the coefficient of relative risk aversion (the equity premium puzzle). On the one hand, too large coefficient of relative risk aversion would contradict the findings on micro level, on the other hand, the coefficient of relative risk aversion which is very close to zero implies risk neutrality. If the latter was true then it would be puzzling why the risky asset pays so much more than the risk-free one. Using the GMM I estimate the system of two Euler equations: one is for risky asset (return on S&P500) and the other is for risk free asset (3-month Treasury bill), Once again I use two specifications for the utility function: (i) the standard CRRA, (ii) a CRRA with an external habit.

6.1 Econometric Model

Let's first specify the econometric model for the case with standard utility specification. Let

$$h_t(\beta,\sigma) := [\beta(\frac{c_t}{c_{t+1}})^{\sigma}(1+r_{t+1})-1, \beta(\frac{c_t}{c_{t+1}})^{\sigma}(1+r^f_{t+1})-1]' \tag{31}$$

where r_{t+1} and r^f_{t+1} stand for the rate or return on risky and risk-free assets, respectively. Then thy system to be estimated is given by

$$\mathbb{E}[h_t(\beta^0,\sigma^0)|I_t] = 0 \tag{32}$$

Using the same arguments as in the previous section and given a vector of instruments $z_t \in I_t$ one can write

$$\mathbb{E}[h_t(\beta^0,\sigma^0) \otimes z_t | I_t] = 0$$

and applying the law of iterated expectations one obtains the unconditional expectation

$$g(\beta^0.\sigma^0) := \mathbb{E}[h_t(\beta^0,\sigma^0) \otimes z_t] = 0 \tag{33}$$

Let $g_T(\beta,\sigma)$ denote the sample equivalence of (33), that is

$$g_T(\beta,\sigma) := T^{-1}\sum_{t=1}^{T} h_t(\beta,\sigma) \otimes z_t$$

Given $g_T(\beta,\sigma)$ one can now perform the optimal GMM estimation the way it was presented earlier, that is by minimizing the following objective function

$$J_T(\theta) = g'_T(\theta)W_T(\theta)g_T(\theta) \tag{34}$$

where $\theta := (\beta,\sigma)$, and $W_T(\theta)$ is an estimator of the optimal weighting matrix using HAC because I am operating in time series environment.

For the case with an external habit (31) takes a different form

$$k_t(\beta,\sigma) := [\beta(\frac{c_t}{c_{t-1}})^{\sigma-1}(\frac{c_{t+1}}{c_t})^{-\sigma}(1+r_{t+1})-1, \beta(\frac{c_t}{c_{t-1}})^{\sigma-1}(\frac{c_{t+1}}{c_t})^{-\sigma}(1+r^f_{t+1})-1]' \tag{35}$$

then the system of equations is

$$\mathbb{E}[k_t(\beta^0,\sigma^0)|I_t] = 0 \tag{36}$$

Given a vector of instrument $z_t \in I_t$ and applying the law of iterated expectations one obtains

$$w(\beta^0.\sigma^0) := \mathbb{E}[k_t(\beta^0,\sigma^0) \otimes z_t] = 0 \tag{37}$$

defining the sample counterpart of (37) as

$$w_T(\beta,\sigma) := T^{-1}\sum_{t=1}^{T} h_t(\beta,\sigma) \otimes z_t$$

one can proceed with the GMM estimation of $\theta = (\beta,\sigma)'$ by minimizing the following quadratic form

$$G_T(\theta) = w'_T(\theta)W_T(\theta)w_T(\theta) \tag{38}$$

where $W_T(\theta)$ is an estimator of the optimal weighting matrix using HAC because I am operating in time series environment. The implementation method of the optimal GMM which I use for both cases is the CUE due to its good performance in finite sample mentioned earlier.

6.2 Instruments

Given a rather poor performance of the FIV estimator in both (linear and nonlinear) cases when estimating a single equation I decide to use only instruments selected from the set of the principal components. The set of instruments used in two-equation estimation is identical to the set of instruments used in a single-equation estimation. The reason is simple, recall that the partial derivatives $d^h_{t+1}(\beta,\sigma)$ and $d^q_{t+1}(\beta,\sigma)$ were proportional to $\ln(c_t/c_{t+1})$ and $(\ln(c_t/c_{t-1})-\ln(c_{t+1}/c_t))$, respectively. This result is partially due to the fact that $(1+r_{t+1})$ is very close to 1. However, this is also true for the return on a risk-free asset $(1+r^f_{t+1})$. This motivates the fact that the set of instruments selected from the set of principal components in the case of a single-equation estimation stays the same for the case of two-equation estimation. So, lets Q^1_t denote the set of instruments used to estimate (32) and Q^2_t the set of instruments used to estimate (36) then $Q^1_t = \{f_{1,t}, f_{2,t}, f_{5,t}, f_{8,t}, f_{20,t}, f_{24,t}, f_{38,t}, f_{85,t}, f_{102,t}\}$ and $Q^2_t = \{f_{44,t}, f_{85,t}, f_{102,t}\}$.

6.3 Estimation Results

Let's firstly consider the result of the two-equation estimation with a standard CRRA utility specification (32). The system is estimated using the set of instruments Q^1_t with a constant added. The Table 7 presents the result of the estimation. One can immediately see that there are several problems with this result. First of all, the estimate of the coefficient or relative risk aversion is unrealistically low, nearly zero (this is what Mehra and Prescott (1985) argued about). In the Figure 12 I present the 90% confidence ellipse for the coefficients of interest. Also, based on the J-test the model is barely correctly specified at 5% confidence level. It is important to note that in the estimation process the initial value of σ was set to 2. If instead one sets this initial value to be 100, then the resulted estimate of σ will be way large than 100 (I do not report this results). To sum up, the estimation of (32) using Q^1_t produces unrealistically low estimate of the coefficient of relative risk aversion and, in general, there is concern about correct specification of the model.

	Coefficient	Std. Error	t-Statistic	Prob.
$\widehat{\beta}$	0.995941	0.000155	6422.778	0.0000
$\widehat{\sigma}$	0.043436	0.018117	2.397613	0.0167
J-statistic				15.4887
Prob(J-statistic)				0.0503

Table 7: The result of the GMM estimation of (32) using Q_t^1 as instruments.

The result of the estimation of (36) using Q_t^2 is presented in the Table 8. One can see that the change in the utility specification leads to a big improvement in terms of the estimation result. First of all, the J-test does not reject the null of correctly specified model at any confidence level. Secondly, the parameter estimate of σ is of possible magnitude

	Coefficient	Std. Error	t-Statistic	Prob.
$\widehat{\beta}$	0.996572	0.000433	2300.156	0.0000
$\widehat{\sigma}$	0.388158	0.086250	4.500376	0.0000
J-statistic				2.4456
Prob(J-statistic)				0.6544

Table 8: The result of the GMM estimation of (36) using Q_t^2 as instruments.

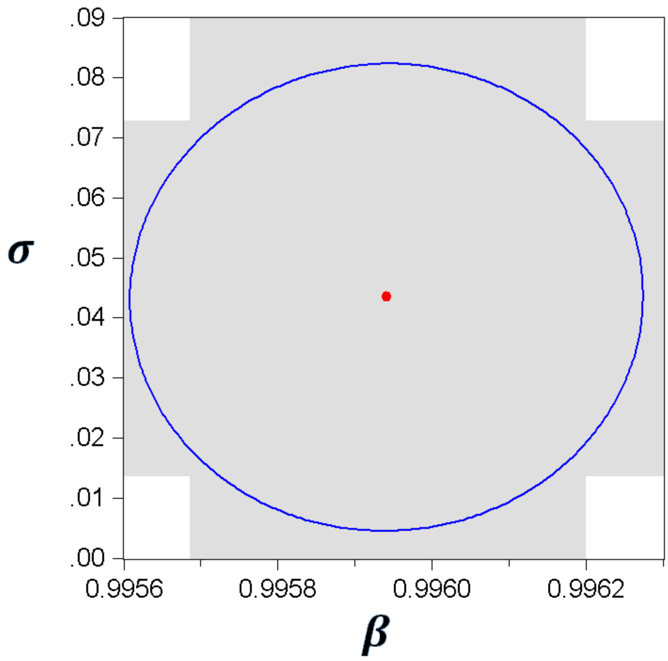

Figure 10: The 90% confidence ellipse for (β, σ); standard case.

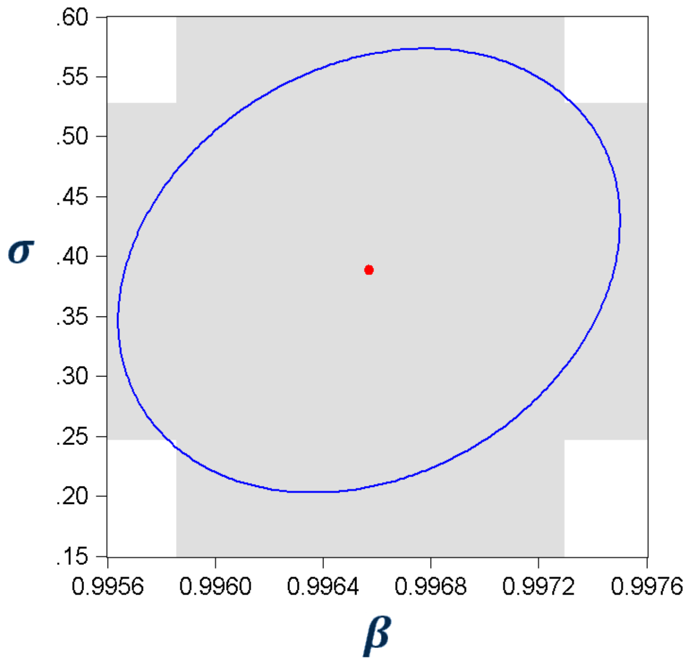

Figure 11: The 90% confidence ellipse for (β, σ); external habit formation case.

even though is somewhat low. I provide the 90% confidence ellipse for the coefficients in the Figure 11. It is important to note that this result is independent of the initial value specified for σ in the estimation. To sum up, the result of the estimation of (36) using Q_t^2 produces good result in terms of the J-test and the magnitude of the estimates. In general,the estimation of the system of two Euler equations, one for risky asset and the other one for risk-free asset, obtained from the C-CAPM with external habit formation would seem to suggest that there is no presence of the risk-premium puzzle.

Also, one can notice a huge drop in the size of the standard error in the two-equation estimation compare to the result of one equation estimation. This drop can be explained by a small variation in the risk-free interest rate (relative to risky asset return). Moreover, the drop in the estimate of the coefficient of relative risk aversion is also due to the fact that when estimating the Euler equation for a risk-free asset alone the resulted parameter estimate is small. Thus, in some sense, the estimate of the coefficient of relative risk aversion in the system of two equations is a weighted average of the estimates of the same coefficient produced in single-equation estimations.

7 CONCLUSION

In conclusion I would like to sum up the main results of the present paper. First of all, I find that in all estimations performed the second method of the instrument dimension reduction, that is the method involving the selection of the subset of the set of the pre-ordered principal components, performs better than the FIV estimator. The possible explanation is that the restrictive assumption upon the structure of the data generating processes made in FIV is not met in this particular application.

In general, the estimates of the coefficient of relative risk aversion using the linearized version of the Euler equation are in the range of 0.7-0.9. This does not contradict the findings of the earlier research, however it unambiguously point on the dominance of the substitution effect in consumers' saving decision. On the other hand, when using the second method of instrument dimension reduction the p-value of the J-statistic was much higher, the confidence intervals much tighter, and moreover, no signs of weak identification problem were detected.

The nonlinear estimation of the Euler equation using the two different utility function specifications also produced interesting results. For example, with the standard CRRA utility function the estimated coefficient of relative risk aversion had very similar magnitude as in the linear estimation. However, even when the CRRA with external habit was used the resulted point estimate of the coefficient of relative risk aversion and its confidence interval were very similar.

In the estimation of the system of the two Euler equations, when the utility function was specified as the standard CRRA, the model was basically rejected by J-test. The point estimate of the coefficient of relative risk aversion signaled the equity premium puzzle problem. On the other hand, using the CRRA with external habit and instruments selected by the second method produced somewhat satisfactory results: a possible magnitude of the coefficient of relative risk aversion and no model rejection by the J-test.

Given that the second method of the instrument dimension reduction has performed so well one may ask why not to go for the second method directly avoiding the use of the first one. My position here is as follows. If the assumptions of the FIV estimator are met then the optimal GMM estimator that uses all instruments or only subset of them is no more efficient than the former estimator. Thus, in some special cases the first method is better than the second one. However, the problem is that it may be hard to see whether you are in that special case. For instance, in the application considered here it is rather clear that the FIV estimator assumptions are not met, but in others this may not be so obvious. In that sense it would be great to have some testing methodology that would allow to test the assumptions of the FIV estimator. Secondly, if one wants to give some interpretation to instruments then it is much easier to do with factors than with principal components.

Finally, in the future research it would be interesting to extend the theory of the FIV estimator to non-linear models. It could also be interesting to develop a testing procedure to test the validity of the assumptions made in the FIV.

REFERENCES

[1] Abel, A. (1990) Asset Prices Under Habit Formation and Catching Up With the Joneses. American Economic Review, 80, 38-42.

[2] Bai, J. & Ng, S. (2002) Determining the Number of Factors in Approximate Factor Models. Econometrica, 70:191–221.

[3] Bai, J. (2003) Inferential theory for factor models of large dimensions. Econometrica 71, 135–172.

[4] Bai, J. & S. Ng (2008) Selecting instrumental variables in a data rich environment. Journal of Time Series Econometrics 1(1), article 4.

[5] Bai, J. & S. Ng (2010) Instrumental Variable Estimation in a Data Rich Environment. Econometric Theory, 26, 1577-1606.

[6] Bekker , P. A. (1994) Alternative Approximations to the Distributions of Instrumental Variable Estimation. Econometrica, 62:657–681.

[7] Campbell, J. Y. & Mankiw, G. N. (1989) Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence NBER Macroeconomics Annual, MIT Press, 185–216.

[8] Campbell, J. Y., & Cochrane, J. H. (1999) By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. Journal of Political Economy, 107, 205-251.

[9] Campbell, J. Y. (2003) Consumption-Based Asset Pricing. Handbook of the Economics of Finance, Elsivier, Vol. 1B:801–885.

[10] Chiappori, P.A. & M. Paiella, (2011) Relative Risk Aversion Is Constant: Evidence From Panel Data. Journal of the European Economic Association, European Economic Association, vol. 9(6), 1021-1052.

[11] Constantinides, G. (1990) Habit Formation: A Resolution of the Equity Premium Puzzle," Journal of Political Economy, 98, 519-543.

[12] Dunn, K. B., & Singleton, K. J. (1986) Modeling the Term Structure of Interest Rates Under Nonseparable Utility and Durability of Goods. Journal of Financial Economics, 17, 27-55.

[13] Eichenbaum, M., Hansen, L. P., & Singleton, K. J. (1988) A Time-Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty. Quarterly Journal of Economics, 103, 51-78.

[14] Hall , R. E. (1988) Intertemporal Substitution in Consumption. Journal of Political Economy, 96:339–357.

[15] Hansen, L.P. & K. J. Singleton (1982) Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. Econometrica, 50, 1269-1286.

[16] Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. Econometrica 50, 1029-1054.

[17] Hansen, L. P. & Singleton, K. J. (1983) Stochastic Consumption, Risk Aversion, and The Temporal Behavior of Asset Return. Journal of Political Economy, 91:249–265.

[18] Hansen, L.P., Heaton, J. & A. Yaron (1996) Finite Sample Properties of Some Alternative GMM Estimators. Journal of Business and Economic Statistics, 14, 262-280.

[19] Hartley, R, Lanet G., & I. Walker (2005) Who Really Wants to Be a Millionaire? Estimates of Risk Aversion from Gameshow Data. Warwick Economic Research Papers. No 719.

[20] Jagannathan, R & G. Skoulakis & Z. Wang (2002) Generalized Method of Moments: Applications in Finance. Journal of Business & Economic Statistics, American Statistical Association 20(4), 470-81.

[21] Kapetanios, G. & M. Marcellino (2010) Factor-GMM Estimation with Large Sets of Possibly Weak Instruments. Computational Statistics & Data Analysis, Elsevier, vol. 54(11), 2655-2675.

[22] Lucas, R.E. (1978) Asset Prices in an Exchange Economy. Econometrica, Vol. 46, No. 6, 1429-1445.

[23] Ludvigson, S. C & Ng, S. (2011) A Factor Analysis of Bond Risk Premium. Handbook of Empirical Economics and Finance, Chapman and Hall, A. Ulah and D. Giles Ed. 313-372.

[24] Marimune ,M. (1983) Approximate Distributions of k-class Estimators when the Degree of Overidentification is Large Compared with the Sample Size. Econometrica, 51:821–842.

[25] Mehra, R & E. C. Prescott (1985) The Equity Premium: A Puzzle. Journal of Monetary Economics 15 (2): 145–161.

[26] Meng, G., G. Hu, & J. Bai (2007) A Simple method for Estimating Betas when Factors are Measured with Error. Mimeo, Boston College.

[27] Newey, W. K. & K.D. West (1987) A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55 (3), 703–708.

[28] Stock, J.H. & J.H. Wright (2000) GMM with Weak Identification. Econometrica, 68, 1055-1096.

[29] Stock, J.H. & M.W. Watson (2002a) Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

[30] Stock, J.H. & M.W. Watson (2002b): Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics, 20:2, 147–162.

[31] Stock, J., J. Wright, & M. Yogo (2002) A Survey of weak instruments and weak identification in Generalized Method of Moments. Journal of the American Statistical Association, 20(4), 518-29.

[32] Stock, J.H. & M.W. Watson (2005) Implications of Dynamic Factor Models for VAR analysis. NBER WP 11467.

[33] Sundaresan, S. (1989) Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth. Review of Financial Studies, 2, 73-88.

[34] Yogo, M. (2004) Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak. Review of Economics and Statistics, 86:797–810.