Are parameters in Epstein-Zin asset pricing model identifiable?

Zhonghui Zhang

Abstract

This paper verifies the result of Smith (1999) that the generalized method of moment (GMM) estimators of Epstein-Zin recursive preference model are poorly identified. The Monte Carlo results suggest that the parameters of the Epstein-Zin asset pricing model are sensitive to the data. We need strong autoregressive consumption growth data to reject the null that risk aversion is not equal to the reciprocal of the elasticity of intertemporal substitution (EIS). I also apply the second approach, Sieve Minimum Distance (SMD) procedure in Chen, Favilukis, and Ludvigson (2013), to verify this conclusion. During the application process, I find some technical issues of applying GMM method on nonlinear model such as Epstein-Zin. One frequent problem is non-numerical (either infinite or undefined) conditional moment could be generated for some combination of parameters. However, there is no theoretical reason to rule out these combinations. Another finding is that the recursive model tends to collapse to the standard time separable model during the estimation process. In another word, theoretically, Epstein model can provide us more flexibility to distinguish the EIS and risk aversion. However, in practice, the GMM objective function cannot produce enough variation to identify each parameter no matter using real or simulated data.

JEL: C52, G12

Keywords: Monte Carlo simulation; GMM; Epstein-Zin preference; Small sample properties; Sieve Minimum Distance

1 Introduction

1.1 Epstein, Zin and Weil Model

In the last decades there has been a massive and growing body of literature studying the recursive preferences based asset pricing model. Among these models the Epstein and Zin (1989), and Weil (1989) (EZW hereafter) is widely used because of its higher degree of flexibility with regards attitudes towards to risk aversion and intertemporal substitution. In the standard time-separable power utility model, let $\{C_t\}_{t=0}^{\infty}$ be the consumption sequence, and $\{\mathcal{F}\}_{t=0}^{\infty}$ be the information filter, then the representative agent maximizes

$$U_t = E\left[\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} | \mathcal{F}_t\right] \tag{1}$$

In this setting, both the coefficient of relative risk aversion (CRRA) and the elasticity of intertemporal substitution are forced to be $1-\gamma$. However, some authors such as Selden (1978), and Hall (1988) claim that this presumed link should be released so that people can separate CRRA from EIS. The EZW recursive preference model satisfies this demand perfectly. In the EZW setup, the representative agent maximizes

$$V_t = \left[(1-\beta)C_t^{1-\rho} + \beta(E\left[V_{t+1}^{1-\theta}|\mathcal{F}_t\right])^{\frac{1-\rho}{1-\theta}}\right]^{\frac{1}{1-\rho}} \tag{2}$$

where V_{t+1} is the the continuation value function, θ is the CRRA, and $\frac{1}{\rho}$ is the EIS. It is easy to show that the time separable CRRA model is a special case of EZW if $\theta = \rho$.

An empirical challenge in using this model is that we can not observe the value function, V_t, which is linked to future consumption through recursion. One possible approach to solve this measurement problem, claimed in Epstein and Zin (1991), is to use the link between the continuation value function and the wealth defined as the value of the aggregate consumption stream in equilibrium. The intertemporal marginal rate of substitution (MRS) in consumption, or stochastic discount factor (SDF) is given by

$$\begin{aligned} M_{t+1} &= \left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\right)^{\frac{1-\theta}{1-\rho}} \left(\mathcal{R}_{w,t+1}\right)^{\frac{1-\theta}{1-\rho}-1} \\ \mathcal{R}_{w,t+1} &\equiv \frac{W_{t+1}}{W_t - C_t} \end{aligned} \tag{3}$$

where $\mathcal{R}_{w,t+1}$ is the return to aggregate wealth, which represents a claim to future consumption. The difficulty of this approach is that the return, $\mathcal{R}_{w,t}$, is itself unobservable. Empirically, for example, Epstein and Zin (1991) use aggregate stock market return as a proxy for $\mathcal{R}_{w,t}$. However, this proxy suffers from a severe measurement error when human capital and other nontradable assets partially constitute aggregate wealth.

Chen, Favilukis, and Ludvigson (2013) (CFL hereafter) introduce a nonparametric approach to estimate the unobservable value function without requiring the proxy for $\mathcal{R}_{w,t}$. The following is the brief introduction of their method. Following Hansen, Heaton, and Li (2008), we can rewrite (2) as

$$\begin{aligned} V_t &= \left[(1-\beta)C_t^{1-\rho} + \beta\mathcal{R}_t(V_{t+1})^{1-\rho}\right]^{\frac{1}{1-\rho}} \\ \mathcal{R}_t\left(V_{t+1}\right) &\equiv \left(E\left[(V_{t+1})^{1-\theta}|\mathcal{F}_t\right]\right)^{\frac{1}{1-\theta}} \end{aligned} \tag{4}$$

where $\mathcal{R}_t(V_{t+1})$ is the risk adjustment to the date $t+1$ continuation value function. To express the recursive utility function in terms of stationary variables, we rescale (3) by dividing C_t on both sides:

$$\frac{V_t}{C_t} = \left[(1-\beta) + \beta\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}} \tag{5}$$

we can show that SDF is

$$M_{t+1} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta} \tag{6}$$

Rearranging (5) and plugging it into (6), the SDF becomes

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\left(\frac{1}{\beta} \left[\left(\frac{V_t}{C_t} \right)^{1-\rho} - (1-\beta) \right] \right)^{\frac{1}{1-\rho}}} \right)^{\rho-\theta} \tag{7}$$

The only latent variable in (7) is the continuation value function-to-consumption ratio, $\frac{V_{t+1}}{C_{t+1}}$. CFL estimate $\frac{V_t}{C_t}$ by applying a profile Sieve Minimum Distance (SMD) procedure, which is to approximate latent value function to consumption ratio, $\frac{V_t}{C_t}$, by a sequence of flexible parametric functions, with the number of parameters expanding as the sample size grows. To be more specific, the $\frac{V_t}{C_t}$ is approximated as follows:

$$\frac{V_t}{C_t} \approx F_{K_T}(\cdot, \delta) = a_0(\delta) + \sum_{j=1}^{K_T} a_j(\delta) B_j \left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}} \right) \tag{8}$$

The sieve coefficients $\{a_0, a_1, ..., a_{K_T}\}$ depend on δ, but the basis functions $\{B_j(\cdot, \cdot) : j = 1, ...K_T\}$ have known functional forms independent of δ. The basis function could be, for example, polynomials or splines. To implement this approximation, the initial value of $\frac{V_0}{C_0}$ is needed. So the total parameters to be estimated are $\{\frac{V_0}{C_0}, a_0, a_1, ..., a_{K_T}\}$. Once $\frac{V_0}{C_0}, \{a_j\}_{j=1}^{K_T}, \{B_j\}_{j=1}^{K_T}$ and data of consumption growth, $\{\frac{C_t}{C_{t-1}}\}_{t=1}^{T}$ are given, one can use the approximate function F_{K_T} to recursively generate the sequence $\{\frac{V_t}{C_t}\}_{t=1}^{T}$.

1.2 Important parameters estimation

Our interest is to estimate the important parameters, θ (CRRA), and ρ (EIS). The first order condition for optimal consumption choice implies that

$$E_t \left[M_{t+1} R_{i,t+1} - 1 \right] = 0, \quad i = 1, 2, ..., N. \tag{9}$$

where M_{t+1} is the SDF, and $R_{i,t+1}$ is the return of asset i. So it is straightforward to consider GMM because GMM only requires some assumptions about the moment conditions. In finance, most data such as stock returns are characterized by heavy-tail and skewed distribution, GMM is a distribution free method which means it does not impose any restriction on the distribution of the data. There are at least two ways to build the GMM moment conditions in the model with recursive preference. For example, follow Epstein and Zin (1991), plug equation (3) into (9), the moment condition is

$$g_{i,t} \equiv E_t \left[\left(\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \right)^{\frac{1-\theta}{1-\rho}} \left(\frac{1}{\mathcal{R}_{w,t+1}} \right)^{\frac{\theta-\rho}{1-\rho}} R_{i,t+1} - 1 \right] = 0, \quad i = 1, 2, ..., N. \tag{10}$$

When we choose stock market aggregate return as the proxy for the return to aggregate wealth, $R_{w,t+1}$. The moment condition could be further simplified as

$$g_{i,t} \equiv E_t \left[\left(\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \right)^{\frac{1-\theta}{1-\rho}} (R_{i,t+1})^{\frac{1-\theta}{1-\rho}} - 1 \right] = 0, \quad i = 1, 2, ..., N. \tag{11}$$

When the number of moment condition is larger than the number of parameters, in our case, there are more than three assets. The parameters $\delta \equiv \{\beta, \theta, \rho\}$ are not the root of moment condition g_i, but the minimizer of the standard sandwich form

$$\delta = \underset{\delta \in \Theta}{\operatorname{argmin}} \, E[g'Wg] \tag{12}$$

where $g \equiv (g_1, g_2, ..., g_n)$ is the vector of moment conditions, and W is the corresponding weighting matrix. Empirically, based on Law of Large Numbers (LLN), we use the sample moment, $\hat{g}_{i,t}$, to replace the expected moments, $g_{i,t}$.

$$\hat{g}_{i,t} = \frac{1}{n} \sum_{t=1}^{T} \left[\left(\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\rho} \right)^{\frac{1-\theta}{1-\rho}} (R_{i,t+1})^{\frac{1-\theta}{1-\rho}} - 1 \right] \tag{13}$$

To find the optimal weighting matrix, W, Hansen (1982) introduced the two step GMM (2SGMM). The basic idea is to use identity matrix as the weighting matrix in the first step, and estimate δ^*. Once δ^* is estimated, we can compute the heteroskedasticity and auto-correlation consistent (HAC) covariance matrix proposed by, e.g., Newey and West (1987). The general form of HAC takes the form of

$$\hat{W} = \sum_{s=-(n-1)}^{n-1} k_h(s)\hat{\Gamma}_s(\delta^*) \tag{14}$$

where $k_h(s)$ is a kernel, h is the bandwidth, and $\hat{\Gamma}_s(\delta^*)$ is

$$\hat{\Gamma}_s(\delta^*) = \frac{1}{n}\sum_i g(\delta^*, x_i)g(\delta^*, x_{i+s})' \tag{15}$$

where δ^* is the estimator from first step. We can choose different kernels, for example, Quadratic spectral, Bartlett, and so different HAC. By default, the Quadratic spectral kernel is used as it was shown to be optimal by Andrews (1991) because of the minimum mean squared error. Bartlett kernel is also popular for its simplicity. The choice of HAC do not affect the asymptotic properties of GMM, but the efficiency could vary. Once the HAC is computed, δ is defined in the following way in the second step of 2SGMM.

$$\hat{\delta} = \underset{\delta\in\Theta}{\operatorname{argmin}}\, \hat{g}'[\hat{W}(\delta^*)]^{-1}\hat{g} \tag{16}$$

However, several applied studies have shown that the estimator may be strongly biased for some certain moment conditions. To solve this problem Hansen et al (1996) proposed two new methods: the iterative GMM (ITGMM) and the continuous updated GMM (CUE). ITGMM is based on 2SGMM, but iterating the procedure of 2SGMM until the estimator converges to a certain number. CUE is a highly nonlinear method, which treats HAC as a function of δ. HAC is therefore allowed to change when the optimization algorithm computes the numerical derivatives. It is therefore continuously updated as we move towards the minimum.

An alternative way to estimate interest parameters in EZW model, is to apply the SMD procedure as in CFL. Plugging the stochastic discount factor, M_{t+1} in (7) into (9), the GMM sample moment is defined as

$$\hat{g}_{i,t} = \frac{1}{n}\sum_{t=1}^{T}\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\left(\frac{1}{\beta}\left[\left(\frac{V_t}{C_t}\right)^{1-\rho}-(1-\beta)\right]\right)^{\frac{1}{1-\rho}}}\right)^{\rho-\theta}R_{i,t+1}-1\right], \quad i=1,2,...,N. \tag{17}$$

In the estimation process, the latent variable $\frac{V_t}{C_t}$ will be substituted by its approximation,

$$\frac{V_t}{C_t} \approx F_{K_T}(\cdot,\delta) = a_0(\delta) + \sum_{j=1}^{K_T} a_j(\delta) B_j\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$$

where the $B_j\cdot$ is the cardinal B-spline, which has the known form

$$B_m(y) = \frac{1}{(m-1)!}\sum_{k=0}^{m}(-1)^k \begin{pmatrix} m \\ k \end{pmatrix} [\max(0, y-k)]^{m-1} \tag{18}$$

There are some other basis functions of $B_j(\cdot)$ (e.g. polynomials), that could be chosen. The B-splines are shape-preserving (preserving nonnegativity, monotonicity and convexity of the unknown function to be approximated). Estimation of the approximation of $\frac{V_t}{C_t}$, F_{K_T}, is based on the following criterion. For any $\delta \in \Theta$,

$$\hat{F}_T(\cdot,\delta) = \underset{F_{K_T}\in\mathcal{V}_T}{\operatorname{argmin}} \left[\hat{g}_T(\delta, F_{K_T}(\delta,\cdot))\right]' W_T \left[\hat{g}_T(\delta, F_{K_T}(\delta,\cdot))\right] \tag{19}$$

Once the approximation $\hat{F}_T^*$ is given, the estimation of finite dimensional parameters, δ, could be implemented by the standard GMM procedure. That is

$$\hat{\delta} = \underset{\delta\in\mathcal{D}}{\operatorname{argmin}} \left[\hat{g}_T(\delta, F_{K_T}(\delta,\cdot))\right]' W_T \left[\hat{g}_T(\delta, F_{K_T}(\delta,\cdot))\right] \tag{20}$$

One can iterate this two-step procedure until the $\hat{\delta}$ converge to δ.

1.3 Motivation

Theoretically, the approaches mentioned above are equivalent and both could estimate the interest parameters in the EZW model. However, the empirical results, especially for θ and ρ, vary greatly in different literature. Epstein and Zin (1991) found that the discount factor β is often considerably larger than one. θ is close to one, and ρ is less than one. These results are sensitive to the choice of the instruments. Using quarterly data on consumption growth, assets returns, and instruments, CFL show that estimation of θ ranges from 16 to 60. The estimated ρ is above one. I use the same method and data as CFL and get similar results. However, the results are sensitive to initial values. In fact, I find that not all combination of initial value of $\{\beta, \theta, \rho\}$ could lead to the convergence of GMM procedure even if I increase the number of iterations. The GMM estimation process get stuck for some initial values and return the error that singular covariance matrix produced. This implies the covariance matrix are suffering from some collinearity problems. Even if I'm lucky to find the initial values that could give me some estimators, the estimated values are usually not far from the initial values. Interestingly, the values of $\{\beta, \theta, \rho\}$ in the literature are coincidence the values that make the program run smoothly. However, these values are not unique. For example, no matter what initial values I use, either in Epstein and Zin (1991) or CFL, R can give me the corresponding estimates. The issue is that the estimates are not much different from the initial values.

An intuitive guess of this issue is that the surface of the GMM object function is too flat. So computer only looks for the minimum value around the starting point. To verify this guess, I applied the approach in CFL. Using the same data, I draw the surface of the GMM object function Q, which is a function of $\delta \equiv \{\beta, \theta, \rho\}$ takes form of

$$Q(\delta) \equiv [\hat{g}_T(\delta, F_{K_T}(\delta, \cdot))]' \, [\hat{g}_T(\delta, F_{K_T}(\delta, \cdot))] \tag{21}$$

where $\hat{g}_T(\delta, F_{K_T}(\delta, \cdot))$ is the vector of equation (17). I fixed the approximated value function $F_{K_T}(\delta, \cdot)$ and β. Because compared to CRRA θ and EIS ρ, discount factor β is much more robust. The data of consumption growth and assets returns are the same with CFL. Consumption is measured as real, per capita expenditures on nondurables and services. Asset include 3-month Treasury Bill and six value-weighted portfolios of common stocks. (Details about data are in CFL appendix.)

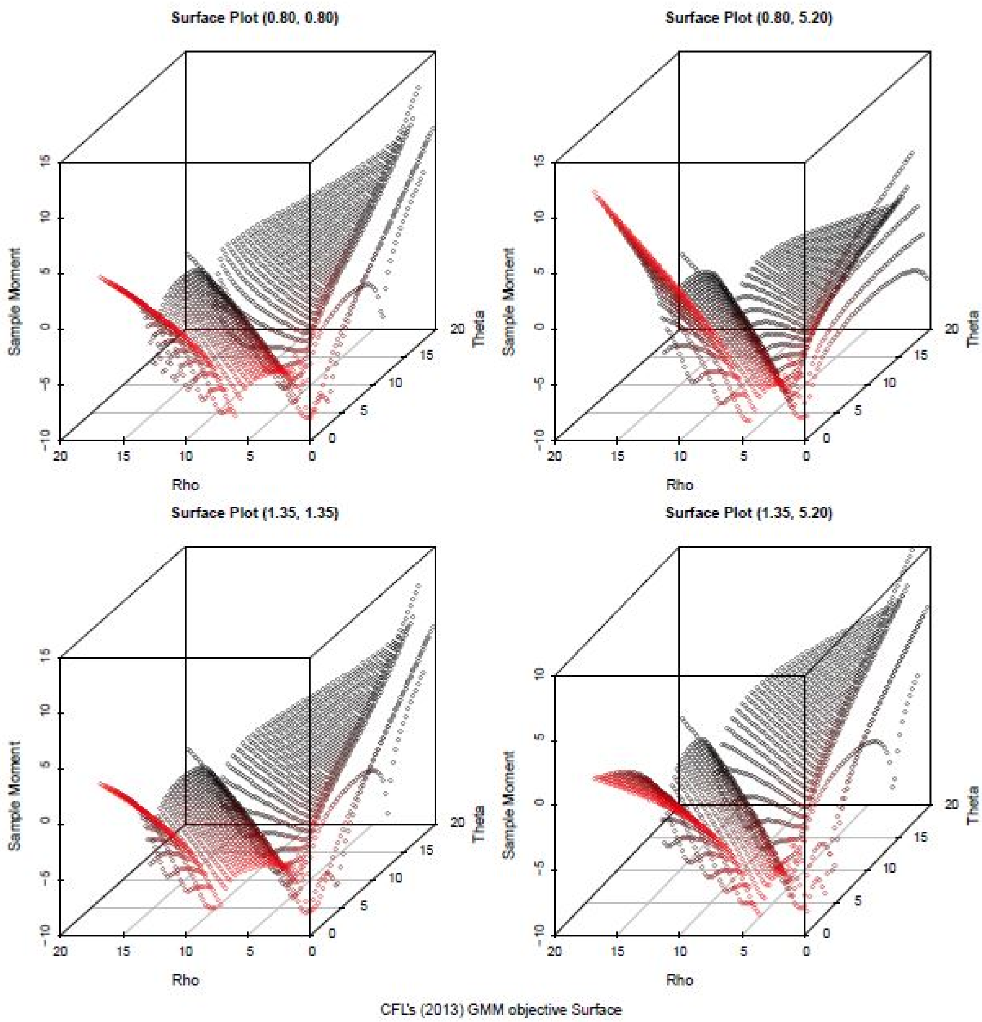

Figure 1: Surface of GMM object function

Figure 1 is the surface of the GMM objective function. Panel (A), (B), (C), and (D) are surfaces with initial values of ($\theta = 0.8, \rho = 0.8$), ($\theta = 0.8, \rho = 5.2$), ($\theta = 1.35, \rho = 1.35$),

($\theta = 1.35, \rho = 5.2$), respectively (Different initials affect the approximated value function). β is fixed at 0.98. From figure 1, we can have a rough impression that the surface is not globally concave. It's hardly to find a unique minimizer. The surface is comparably flat along 45 degree line. This indicates that θ and ρ are not well identified and many combination of (θ, ρ) will result in nearly the same value of the objective function. One possible reason is that the moment condition $\hat{g}_T$ is not a stationary and ergodic sequence. In this paper, I use Monte Carlo simulation to generate a stationary data, and two methods to verify the identification of parameters in EZW asset pricing model. The paper is organized as follows. Section 2 describes the data generating process and the Monte Carlo simulation methodology of two methods. One is following Smith (1999) and another one is to apply CFL. Section 3 will concludes the paper.

2 Background for experiments

2.1 Tauchen and Hussey's (1991) method

To find the counterfactuals of all possible returns of assets, I adopt Tauchen and Hussey's (1991) method to calibrate a Markov chain, with a discrete state space, whose probability distribution closely approximates the distribution of a given time series. This approach (Tauchen and Hussey (1991) call it Quadrature) is centrally trying to use the sum of a set of N abscissa, or state values, $u_k \in \mathbb{R}^M$, and weights, or stationary probabilities, $w_k \in \mathbb{R}^M$ to approximate the integration of functions $g(u)$, $u \in \mathbb{R}^M$ against a density $\omega(u)$, such that

$$\int g(u)\omega(u)du \approx \sum_{k=1}^{N} g(u_k)w_k \tag{22}$$

where the abscissa u_k and weight w_k depend only on the density ω, and not directly on the function g. The quality of approximation could be improved as the discrete state space becomes finer, i.e., as $N \to \infty$. There are many good quadrature rule could be chosen. For example, for an N-point Gauss rule along the real line, $u \in \mathbb{R}^1$, the abscissa u_k and weights w_k are determined by forcing the rule to be exact for all polynomials of degree less than or equal to $2N-1$ (Davis and Rabinowitz (1975)). A quadrature rule can be considered as a discrete probability model that approximates the density ω. In particular, the Gauss rules are discrete approximations to ω determined by the method of moments using moments up through $2N-1$. Gauss rules are close to minimum norm rules and possess several optimum properties (Davis and Rabinowitz (1975)). This is the best that can be done with N points using moments as a criterion since if two probability distributions have the same moments up to $2N$, and if one of the distributions is a discrete distribution concentrated on N points, then the two distributions must coincide (Norton and Arnold (1985)).

Consider a Lucas-style representative agent asset pricing model with Euler equation such as

$$p_{it} = E_t[(p_{i,t+1} + d_{t,t+1})m_{t+1}(c_t, c_{t+1})] \tag{23}$$

where p_{it} is the ex-dividend price in period t of the ith asset, $d_{i,t+1}$ is the dividend, m is the agent's marginal rate of substitution between consumption in periods t and $t+1$, and $E[\cdot]$ is the conditional expectations operator given information available to the agent at period t. By dividing $d_{i,t}$ on both sides of equation (23), and letting $\lambda_{t+1} = \frac{c_{t+1}}{c_t}$, and $\xi_{t+1} = \frac{d_{t+1}}{d_t}$, (23) could be rewritten as

$$v_{it} = E_t[(1 + v_{i,t+1})\xi_{t+1} m_t(\lambda_t)] \tag{24}$$

where v_{it} is the ith asset's price dividend ratio. The integral form for the equation (24) is

$$v_i(x) = \int (1 + v_i(y,x))\xi(y) m_t(y,x) f(y|x) dy \tag{25}$$

To simplify the notation, denote $\psi(y,x) = (1 + v_i(y,x))\xi(y)m_t(y,x)$. Also define $I[\psi]$ to be the integral operator such that

$$I[\psi](x) = \int \psi(y,x) f(y|x) dy \tag{26}$$

Now write equation(26) as

$$I[\psi](x) = \int \psi(y,x) \frac{f(y|x)}{\omega(y)} \omega(y) dy \tag{27}$$

where $\omega(y)$ is some strictly positive weighting function. Let $\bar{y}_k$ and w_k, $k = 1, 2, ..., N$, denote the state values and stationary probabilities for an N-point quadrature rule for the density $\omega(y)$. Based on the quadrature rule, we can use $\sum_k [\cdot] w_k$ to approximate the integral $\int [\cdot] \omega dy$ such as

$$\begin{aligned} I[\psi](x) \approx I_N[\psi](x) &\equiv \sum_{k=1}^{N} \psi(\bar{y}_k, x) \pi_k^N(x) \\ \pi_k^N(x) &= \frac{f(\bar{y}_k|x)}{s(x)\omega(\bar{y}_k)} w_k \quad (k = 1, 2, ..., N) \\ s(x) &= \sum_{i=1}^{N} \frac{f(\bar{y}_i|x)}{\omega(\bar{y}_i)} w_i \end{aligned} \tag{28}$$

We can obtain the Markov chain π^N by replacing the integration against $\omega(y)$ with the summation. The $s(x)$ in equation (28) is used for normalization so that the Markov chain sums up to one. There is a large number of options of the weighting function $\omega(y)$. Tauchen and Hussey (1991) mention in their paper that one reasonable choice is $\omega(y) = f(y|0)$, i.e., the conditional density given that the process is at the unconditional mean, which is taken to be zero without loss of generality.[1] Another choice is $\omega = f_s(y)$, where f_s is the unconditional or stationary density of the process. Compare to $f_s(y)$, $f(y|0)$ places relatively more weight in the central part of the distribution and less weight in the tails than does $f_s(y)$. In practice,

[1]Tauchen and Hussey didn't mention the form of pdf $f(\cdot)$, but they use multivariate normal distribution in their code.

it has been found that $f(y|0)$ gives much better approximation in most cases.[2]

2.2 The data generating process

The state variables are generated by the following VAR(2) estimate of the logarithm of annual per-capita, real non-durable consumption growth ($\lambda_t = c_t/c_{t-1}$) and the logarithm of annual S&P 500 aggregate, real dividend growth ($\xi_t = d_t/d_{t-1}$) over the period 1888-1978.

$$\begin{aligned} \ln \lambda_t &= 0.021 + 0.017 \ln \xi_{t-1} - 0.161 \ln \lambda_{t-1} + \varepsilon_t^1 \\ \ln \xi_t &= 0.004 + 0.117 \ln \lambda_{t-1} + 0.414 \ln \xi_{t-1} + \varepsilon_t^2 \\ Var(\epsilon) &= \begin{bmatrix} 0.01400 & 0.00177 \\ 0.00177 & 0.00120 \end{bmatrix} \end{aligned} \tag{29}$$

The parameters from the VAR(2) model above are from Kocherlakota (1990b), who uses the data from Mehra and Prescott (1985). The error terms $\varepsilon = (\varepsilon_t^1, \varepsilon_t^2)$ are assumed to be jointly normally distributed with the covariance matrix given by $Var(\varepsilon)$. Given the state variables λ_t and ξ_t following the assumed processes above, along with values for the constant discount factor β, CRRA θ, and EIS ρ, we can simulate the returns for three assets. They are the claim on aggregate wealth R_w, the claim on the S&P 500 dividend stream R_{sp}, and the risk free rate R_f (which is defined to be the return on a claim paying a unit of consumption for sure one period forward). Tauchen and Hussey's (1991) method calibrates a vector of state values $\tilde{\lambda}_j$ and $\tilde{\xi}_j$, $j = 1, 2, ..., S$, a vector of stationary probabilities Π_i, $i = 1, 2, ..., S$, and a set of probability weights $\pi_{ij} = Pr(s_{t+1} = \tilde{\lambda}_j, \tilde{\xi}_j | s_t = \tilde{\lambda}_i, \tilde{\xi}_i)$ to mimic the continuous vector autoregression process of the endowment variable λ_t and ξ_t. Where the s_t represents the tth state drawn from the set S of states. Following Smith (1999), I assume R_w is the return on a claim whose dividend equals per-capita consumption. That is, the representative agent receives only dividends as income and consumes all dividends when

[2] Tauchen provides the code of the Markov chain approximation technique in both Gauss and Fortran in his personal website. I have translated the code in R, and the code is available upon request.

received, i.e., $d_{i,t}/d_{i,t-1} = c_{i,t}/c_{i,t-1} = \lambda_t$. Recall that the finite-states Markov representation of the Euler equation (23) takes the form

$$\begin{aligned} p_{w,i} &= \sum_{j=1}^{S} \pi_{ij}[(p_{w,j} + d_{w,j})m_{ij}(c_i, c_j)] \\ m_{ij} &= \left(\beta\left(\frac{c_j}{c_i}\right)^{-\rho}\right)^{\frac{1-\theta}{1-\rho}} (\mathcal{R}_{w,ij})^{\frac{1-\theta}{1-\rho}-1} \end{aligned} \tag{30}$$

Let $v_{w,t} = p_{w,t}/d_{w,t}$ be the price-dividend ratio from the claim on aggregate wealth, along with previous assumption that dividend growth is equal to the consumption growth. So equation (30) can be rewritten as

$$v_{w,i}^{(1-\theta)/(1-\rho)} = \sum_{j=1}^{S} \pi_{ij}[\beta^{(1-\theta)/(1-\rho)} \lambda_j^{1-\theta} (1 + v_{w,j})^{(1-\theta)/(1-\rho)}], \quad i = 1, 2, ..., S. \tag{31}$$

Equation (31) comprises a system of S linear equations in $v_{w,i}$ which can be solve directly. Their solutions provide the values of the approximate solutions to the integral equation (24) at each of the quadrature points. Let v_w^* be the S-vector of solutions to (31), for fixed values of β, θ, and ρ. The sets of state-returns can be represent as the function of v_w^*. The return on aggregate wealth over state i and j is given by

$$R_{w,ij} = \frac{1 - v_{w,j}^*}{v_{w,i}^*} \lambda_j \tag{32}$$

The return on risk free asset, which is defined as the price in state i of a bond paying sure unity next period is

$$\frac{1}{R_{fi}} = \beta^{(1-\theta)(1-\rho)} \sum_{j=1}^{S} \pi_{ij} \lambda_j^{-\theta} \left(\frac{1 - v_{w,j}^*}{v_{w,i}^*}\right)^{(\rho-\theta)(1-\rho)} \tag{33}$$

For the return on risky asset, everything is the same with return on aggregate wealth except that consumption growth is replaced by the simulated S&P 500 dividend stream. The state returns are given by

$$R_{sp,ij} = \frac{1 - v^*_{w,j}}{v^*_{w,i}} \xi_j \tag{34}$$

Once the three state returns are obtained based on the process above, the realized time series return are drawn using the following algorithm:

(1) Draw a random variable u_0 from a uniform distribution $U(0, 1)$. Let the initial state, n_0, be the smallest number such that $\Pi(1) + \Pi(2) + \cdots + \Pi(n_0) \geq u_0$, where $\Pi(i)$ is the stationary probability of being in state i (which is from Tauchen and Hussey's (1991) method), $1 \leq n_0 \leq S$.

(2) Let n' be the previous state ($n' = n_0$ for the initial state) and n'' be the next state to be drawn. Draw u'' from $U(0, 1)$ and let the n'' be the smallest number such that $\sum_i^{n'} \pi(n', i) \geq u''$, where $\pi(i, j)$ is the transition probability from state i to state j.

(3) Set $n' = n''$ and then return to step 2 until $t = T$. The time series of three returns are obtained by choosing $R_w(n', n'')$, $R_f(n', n'')$, and $R_{sp}(n', n'')$.

The initial values of parameters are based on the range of CRRA θ and EIS ρ estimated by Epstein Zin (1991). To verify Smith's (1999) method, I use exactly the same values of parameters. The true value of θ is either 0.80 or 1.35 and the range of true value of ρ is over 0.80, 1.35 and 5.20. As this range of parameter values is usually considered to be economically reasonable, (see Hall (1988), for example).[3]

In Table 1, Panel A summarizes the mean, across 500 draws of 90 observations, of the VAR(2) coefficient estimates, standard errors, and the residual covariance matrix across 500 simulations. The mean VAR(2) coefficient and residual covariances are close to the actual value. With 90 observations, the VAR(2) regressions can adequately summarize the underlying distribution of state variables. Panel B contains the average mean and standard

[3]I wrote the code of Monte Carlo simulation in R from scratch, and the code is available upon request.

Table 1: Summary information on endowment variables and security returns from Monte Carlo simulations using 500 draws of 90 observations from annual DGP.

Panel A: Mean slope and the standard error of estimates from annual VAR(2) regressions of consumption and dividend growth.						
		Estimates				
Variable		Intercept	ξ_{t-1}	λ_{t-1}	$Var(\varepsilon)$	$Cov(\varepsilon)$
λ_t	Estimator	1.172	0.011	-0.162	0.001	0.002
	Std.error	0.116	0.034	0.109		
ξ_t	Estimator	0.460	0.097	0.454	0.015	0.002
	Std.error	0.396	0.115	0.373		

Panel B: Average sample mean, standard error and equity premium of three simulated returns using the parameter values from Epstein and Zin (1991).					
θ, ρ		R_w	R_{sp}	R_f	Equity premium
0.80, 0.80	Estimator	0.036	0.038	0.035	0.003
	Std.error	0.035	0.125	0.004	
0.80, 5.20	Estimator	0.124	0.126	0.122	0.004
	Std.error	0.063	0.137	0.030	
1.35, 1.35	Estimator	0.047	0.048	0.045	0.003
	Std.error	0.038	0.126	0.007	
1.35, 5.20	Estimator	0.123	0.123	0.120	0.003
	Std.error	0.063	0.136	0.030	

error of three simulated returns of assets by using different values of parameters θ and ρ. β is fixed at 0.98. The results are very close to Smith (1999), which verifies my programming is consistent with his.

Ferson and Foerster (1994) claim that the large number of instrument variables may lead to biased GMM estimation. On the other hand, too few instruments reduces the asymptotic efficiency. To study the effects of instruments on the GMM estimators of Epstein-Zin model, I follow Smith(1999), who uses three different instrument sets containing two, three and four lagged variables. Instrument series 1 includes a constant and the lagged one period return on the S&P 500, (1, $R_{sp,t-1}$). Series 2 also contains the lagged consumption growth, (1, $R_{sp,t-1}$, λ_{t-1}). Series 3 adds the return on risk-free asset on period t to series 2 variables, (1, $R_{sp,t-1}$, λ_{t-1}, $R_{f,t}$). The behavior of the instruments are summarized in table A1 and A2 in the appendix.

Table A1 shows the average slope estimates, standard errors and t-values from instrument autoregressions using five lags, 500 draws of 90 observations from annual data generation process (DGP), and four combinations of θ and ρ from $(\theta, \rho) = (0.80, 0.80)$, $(0.80, 5.20)$, $(1.35, 1.35)$, $(1.35, 5.20)$. The results exhibit a negative serial correlation among lags and the correlation dies out quickly after the first lag. However, the average t-value does not reject the null hypothesis that the autocorrelation is zero. There is no significant change of the behavior of λ, R_{sp}, and R_f, when I use different combination of θ and ρ. This implies the instruments are independent of the magnitude of parameters. However, Smith (1999) found that when parameters are fixed at some "extreme" values (e.g. $\theta = 29$, $\rho = 2$), then R_{sp} and R_{R_f} would exhibit a significant negative serial correlation. Table A2 reports the estimates from the regression of λ, R_{sp}, and R_f on instruments. Based on both t-values and adjusted R^2, the instruments explain little of the linear variation in all three dependent variables. Nelson and Startz (1990a,b) claim that considerable biasness could be generated if the instruments are not highly correlated with the explanatory variables. In the next part, I will report the performance of GMM estimator under different size of "poor" instruments.

2.3 Simulation results

Table 2 reports the simulation results from 500 draws of 90 annual observations using the values of θ and ρ in Epstein and Zin (1991). I also estimate the ratio $\gamma \equiv (1-\theta)/(1-\rho)$. The estimation method to be investigated is the Hansen (1982) two-step procedure described in the previous section. In most cases, the program gets "stuck" and the valid results are not available if I use heteroskedasticity and autocorrelation consistent (HAC) estimators. One reason for this problem could be that the instruments contain a high degree of collinearity, in that case, the optimal weighting matrix would be singular and so not invertible. Even if the weighting matrix is invertible, in practice, the infinite GMM sandwich form condition may be produced if the weighting matrix is too small. This issue is more likely to occur in case of nonlinear models. When the optimal weighting matrix is singular, the parameters can still be estimated consistent by using identity matrix as the weighting matrix, but the inference is no longer valid. Tauchen (1986) reports the similar issue and he also attributes it to the weak instruments. He claims that the GMM method uses lagged endogenous variables as instruments, and there are strong nonlinearities inherent in the structure and the estimation method. In addition, the restriction that the estimator must lie in a compact set, though common in theoretical work, is rarely operational in practice and is not imposed in the simulation. Also, Hansen two-step estimator does not possess moments. What's more, the natural moment-based statistics are also reported and, as it turns out, infinite moment problems generally are not apparent, which is consistent with Sargan (1982). Smith (1999) attempts to address the issue of instrument quality by using the "optimal" instruments in the sense of minimizing, over the choice of all feasible instruments, the asymptotic covariance matrix of the GMM estimator. This kind of instruments are a set of indicator variables that span the information set at time t and thus yield on average the conditional expectation $E_t[u_j(x_{t+1}, b_0)]$ for large T. However, his simulation results are strongly biased towards the true values. Even though the efficiency has been improved, it does not provide us more information.

In addition to the issue of weak instruments, how does the computer find out the optimal estimators in the case of the highly nonlinear model is another problem that we need to address. Most software uses the Newton-Raphson algorithm to pin down the optimal points, either maximization or minimization. The basic idea behind the algorithm is as following. First, construct a quadratic approximation to the function of interest around some initial values (The closer to the true value the better). Next, adjust the parameter value to which maximizes the quadratic approximation. This procedure is iterated until the parameter values converge. To be more specific, suppose we want to find the value of x that maximizes some twice continuously differentiable function $f(x)$. Then the second order Taylor approximation of $f(x)$ can be written as

$$f(x+h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$$

The first order condition for the value h (denote h^*) maximizes $f(x+h)$ is $h^* = -f'(x)/f''(x)$. (Simply by solving the first order condition $f'(x) + f''(x)h = 0$.) In other words, the value that maximize the second order Taylor approximation to f at x is

$$x + h^* = x - \frac{f'(x)}{f''(x)}$$

That is, h^* points out the fastest direction towards the optimal points. Note that the Newton-Raphson algorithm doesn't check the second order conditions necessary for $\hat{x}$ to be a maximizer. This means a bad initial value may lead us anywhere, either minimum or maximum or some saddle points. Recall the surface of GMM object function in Figure 1. The surface does not look like a standard concave one with a unique optimal point. So it's reasonable to doubt that the Newton based algorithm can find us the "optimal" estimates.

To address this concern, I use two different optimization algorithms to find the estimates of parameters β, θ, ρ, and γ using the same GMM moment condition. The first method is call *nlminb* in R, which is a Newton based algorithm described above. The details are somewhat difficult to understand from the manual, but it depends heavily on the gradient. Another method I use is called *Nelder-Mead simplex* (also known as the *nonlinear simplex*) method. This method is gradient-free, that is, it requires no derivatives to be computed and it does not require the objective function to be smooth. The weakness of this approach is that it is slow particularly for problems with more than 10 variables. The basic idea of *Nelder-Mead* algorithm is to use a simplex, which is a structure in n-dimensional space formed by $n+1$ point that are not in the same plane (e.g., a line segment is a 1-dimensional simplex, a triangle is a 2-dimensional simplex and a tetrahedron forms a 3-dimensional simplex.), to search the extremum directly. First, start with a simplex given an initial guess x_0. Then

Table 2: Summary of estimates of EZ parameters from Monte Carlo simulations using parameter values from Epstein and Zin (1991) and 500 draws of 90 observations DGP.

		nlminb				Nelder-Mead			
		$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\gamma}$
(0.98, 0.80, 0.80, 1.00)									
Series1	Mean	0.970	0.763	0.818	2.626	0.972	0.872	0.776	0.696
	Median	0.976	0.765	0.865	1.765	0.975	0.865	0.794	0.633
	std.dev	0.019	0.128	0.244	2.821	0.017	0.047	0.144	0.393
Series2	Mean	0.970	0.418	0.732	2.639	0.971	0.870	0.776	0.701
	Median	0.975	0.752	0.834	1.911	0.975	0.865	0.796	0.632
	std.dev	0.019	1.206	0.395	2.643	0.017	0.045	0.145	0.382
Series3	Mean	0.968	-1.656	0.388	3.804	0.971	0.842	0.767	1.017
	Median	0.968	-0.735	0.415	2.753	0.975	0.865	0.808	0.753
	std.dev	0.017	3.712	0.743	3.478	0.018	0.209	0.300	0.918
(0.98, 0.80, 5.20, -0.05)									
Series1	Mean	0.970	-2.227	5.136	-0.784	0.936	0.752	5.613	-0.050
	Median	0.979	-1.590	5.137	-0.623	0.931	0.755	5.685	-0.052
	std.dev	0.026	1.935	0.047	0.478	0.027	0.376	0.326	0.077
Series2	Mean	0.968	0.265	1.469	2.010	0.936	0.769	5.578	-0.046
	Median	0.974	0.736	0.878	1.389	0.933	0.770	5.705	-0.049
	std.dev	0.021	1.226	1.742	2.699	0.026	0.367	0.416	0.074
Series3	Mean	0.963	-1.510	0.913	2.895	0.935	0.863	4.798	0.039
	Median	0.967	-0.654	0.634	2.189	0.929	0.789	5.370	-0.045
	std.dev	0.020	3.383	1.591	3.730	0.026	0.643	1.731	0.497
(0.98, 1.35, 1.35, 1.00)									
Series1	Mean	0.967	1.415	1.284	2.461	0.973	1.316	1.490	0.700
	Median	0.970	1.427	1.276	1.720	0.977	1.318	1.480	0.656
	std.dev	0.015	0.229	0.142	2.584	0.016	0.122	0.235	0.296
Series2	Mean	0.967	0.962	1.452	2.130	0.973	1.314	1.487	0.694
	Median	0.971	1.012	1.115	1.418	0.977	1.320	1.481	0.654
	std.dev	0.021	1.225	1.300	2.647	0.016	0.090	0.169	0.273
Series3	Mean	0.962	-0.907	1.140	2.883	0.969	1.266	1.421	0.846
	Median	0.966	0.248	0.745	2.064	0.972	1.306	1.478	0.643
	std.dev	0.019	3.317	6.633	3.261	0.019	0.379	0.520	0.734
(0.98, 1.35, 5.20, 0.08)									
Series1	Mean	0.960	3.745	5.078	0.713	0.949	1.443	5.405	0.108
	Median	0.960	2.965	5.165	0.471	0.949	1.467	5.537	0.104
	std.dev	0.023	2.215	0.310	0.684	0.028	0.246	0.752	0.100
Series2	Mean	0.958	1.495	2.199	1.727	0.948	1.442	5.385	0.110
	Median	0.966	1.168	1.350	1.000	0.943	1.447	5.537	0.102
	std.dev	0.027	1.601	1.884	2.283	0.026	0.230	0.812	0.087
Series3	Mean	0.950	0.028	1.619	2.593	0.938	1.663	4.762	0.308
	Median	0.958	1.000	1.000	1.728	0.932	1.542	5.479	0.122
	std.dev	0.027	3.451	6.374	2.954	0.025	0.580	1.982	0.445

Table 3: Summary of estimates of EZ parameters from Monte Carlo simulations using the same true parameter values (0.98,0.8,0.8,1.0) but different starting value for the estimation.

		nlminb				Nelder-Mead			
		$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\gamma}$
(0.98, 0.80, 0.80, 1.00)									
Series1	Mean	0.970	0.763	0.818	2.626	0.972	0.872	0.776	0.696
	Median	0.976	0.765	0.865	1.765	0.975	0.865	0.794	0.633
	std.dev	0.019	0.128	0.244	2.821	0.017	0.047	0.144	0.393
Series2	Mean	0.970	0.418	0.732	2.639	0.971	0.870	0.776	0.701
	Median	0.975	0.752	0.834	1.911	0.975	0.865	0.796	0.632
	std.dev	0.019	1.206	0.395	2.643	0.017	0.045	0.145	0.382
Series3	Mean	0.968	-1.656	0.388	3.804	0.971	0.842	0.767	1.017
	Median	0.968	-0.735	0.415	2.753	0.975	0.865	0.808	0.753
	std.dev	0.017	3.712	0.743	3.478	0.018	0.209	0.300	0.918
(0.98, 0.80, 5.20, -0.05)									
Series1	Mean	0.958	0.318	5.176	-0.164	0.954	1.076	5.264	0.013
	Median	0.961	0.364	5.185	-0.152	0.960	1.236	5.238	0.056
	std.dev	0.021	0.455	0.031	0.110	0.029	0.24	0.651	0.088
Series2	Mean	0.968	0.418	1.093	2.266	0.953	1.087	5.227	0.015
	Median	0.974	0.750	0.843	1.651	0.954	1.245	5.240	0.058
	std.dev	0.020	1.181	1.283	2.447	0.029	0.241	0.769	0.086
Series3	Mean	0.969	-1.256	0.711	2.912	0.953	1.146	4.845	-0.013
	Median	0.969	-0.056	0.599	2.202	0.957	1.267	5.234	0.063
	std.dev	0.018	3.491	1.242	3.785	0.029	0.235	1.367	0.416
(0.98, 1.35, 1.35, 1.00)									
Series1	Mean	0.975	1.377	1.279	2.403	0.979	1.310	1.534	0.674
	Median	0.979	1.424	1.271	1.633	0.981	1.317	1.482	0.667
	std.dev	0.018	0.321	0.177	2.551	0.013	0.103	0.480	0.288
Series2	Mean	0.971	1.044	1.211	2.379	0.979	1.312	1.570	0.682
	Median	0.977	1.054	1.076	1.550	0.980	1.317	1.479	0.669
	std.dev	0.020	0.953	0.882	2.648	0.013	0.093	0.940	0.280
Series3	Mean	0.969	-1.044	0.679	3.078	0.977	1.287	24.697	0.869
	Median	0.971	0.218	0.617	2.220	0.980	1.292	1.467	0.681
	std.dev	0.017	3.517	1.348	3.467	0.015	0.221	399.502	0.750
(0.98, 1.35, 5.20, 0.08)									
Series1	Mean	0.960	2.175	4.949	0.551	0.958	1.355	5.445	0.083
	Median	0.959	1.361	5.200	0.086	0.967	1.418	5.413	0.094
	std.dev	0.028	1.847	0.816	1.342	0.026	0.298	0.517	0.068
Series2	Mean	0.971	1.142	1.607	2.169	0.960	1.355	5.421	0.087
	Median	0.977	1.091	1.140	1.394	0.967	1.414	5.413	0.093
	std.dev	0.021	0.954	1.462	2.606	0.025	0.298	0.859	0.087
Series3	Mean	0.971	-0.344	1.016	2.695	0.960	1.371	5.188	0.129
	Median	0.973	0.784	0.929	1.957	0.967	1.435	5.367	0.102
	std.dev	0.018	3.452	1.600	3.134	0.024	0.353	1.185	0.249

modify the simplex at each iteration using four simple operations: reflection, expansion, outside contraction, inside contraction and shirking. We can consider these operations are directions of searching. Each of these operations generates a new point and the the sequence of operations performed in one iteration depends on the value of the objective at the new point relative to the other key points. For example, let x_w, x_l, and x_b denote the worst, the second worst and best points, respectively, among all n+1 points of the simplex. We first calculate the average of all the points that exclude the worst, i.e., $x_a = 1/n \sum_{i \neq w} x_i$. After computing x_a, it's fair to conjecture that the line from x_w to x_a is a descent direction. A new point found on this line by reflection is given by

$$x_r = x_a + \alpha(x_a - x_w)$$

Similarly, the other new points generated by expansion, inside contraction, and outside contraction, are given by

$$\begin{aligned} x_e &= x_r + \gamma(x_r - x_w) \\ x_{in} &= x_a - \beta(x_a - x_w) \\ x_{out} &= x_a + \beta(x_a - x_w) \end{aligned}$$

respectively. By comparing the objective value at each new point, we can figure out the best direction to the extremum.

In table 2, the numbers in the parenthesis are the true values of the parameters β, θ, ρ, and $\gamma \equiv (1-\theta)/(1-\rho)$. Series 1, series 2 and series 3 have everything in common except different instruments. Following Tauchen's (1985) suggestion, in addition to the mean, I also report the median since the Epstein-Zin model inherent high nonlinearities. I use the actual values of the parameters as the starting values as is common in the literature. As we can see in table 2, the estimates are surprisingly accurate. The Nelder-Mead algorithm, which is the gradient-free method works much better than the gradient-based nlminb method in most cases. Another finding is that, as we add more "weak" instruments, the difference between the real values and the estimates are further and further away. Can we simply draw a conclusion that the Epstein-Zin recursive preference model is well identified because the estimates are close to the true values? The simple answer is no. First, in most cases, the HAC estimates are not available. One has to use identity weighting matrix to get some estimates. Though the regular estimates are still unbiased, inference is not valid any more. Another concern is that most computer algorithm can only find the local optimal point given some starting values. In other words, the estimates are very sensitive to initial values, particularly in case of nonlinear model. To check what would happen if I use different initial values for the estimation, I fix the true values of parameters but use different initial values to estimate the parameters. The results are summarized in table 3. In the table 3, the true values of the parameters are fixed at $\beta = 0.98$, $\theta = 0.80$, $\rho = 0.80$, and $\gamma = 1.00$. The numbers in the parenthesis are initial values. From the results in table 3 we find that the estimates are actually close to the initial values instead of the values of population parameters. We can also notice that the estimates are always near the initial values if we use Nelder-Mead approach, but the estimates have a tendency to get closer to the truth if the gradient-based method is used. Though the estimates get further away from the starting points as we add more instruments. We cannot draw a conclusion as to how the instruments would affect the estimation results. Figure 2 plot the surface of GMM objective function using different instruments. The data is generated using the value of ($\beta = 0.98$, $\theta = 0.80$, $\rho = 0.80$). (Surface for other combinations of parameters are provided in the appendix). To see more details of the surface, I use log form of GMM objective function. Otherwise some large point would drive the rest of surface into a plain. In figure 2, adding more instruments does not change the surface too much. Also, the surface looks flat and does not have an unique minimum even when I use simulated data. In a Monte Carlo simulation, it's not a problem to choose a good initial point since we know exactly the data generation process. In practice, however, it is impossible to know where to start particularly when estimating nonlinear model. Researchers usually use the results in the literature as the starting values and estimate, and then report the similar results with the literature and say they provide some new evidence for the literature. However, it's possible that literature gives some local optimal point, and then just keep converging to this point simply because they all use the similar local optimal point as the starting point.

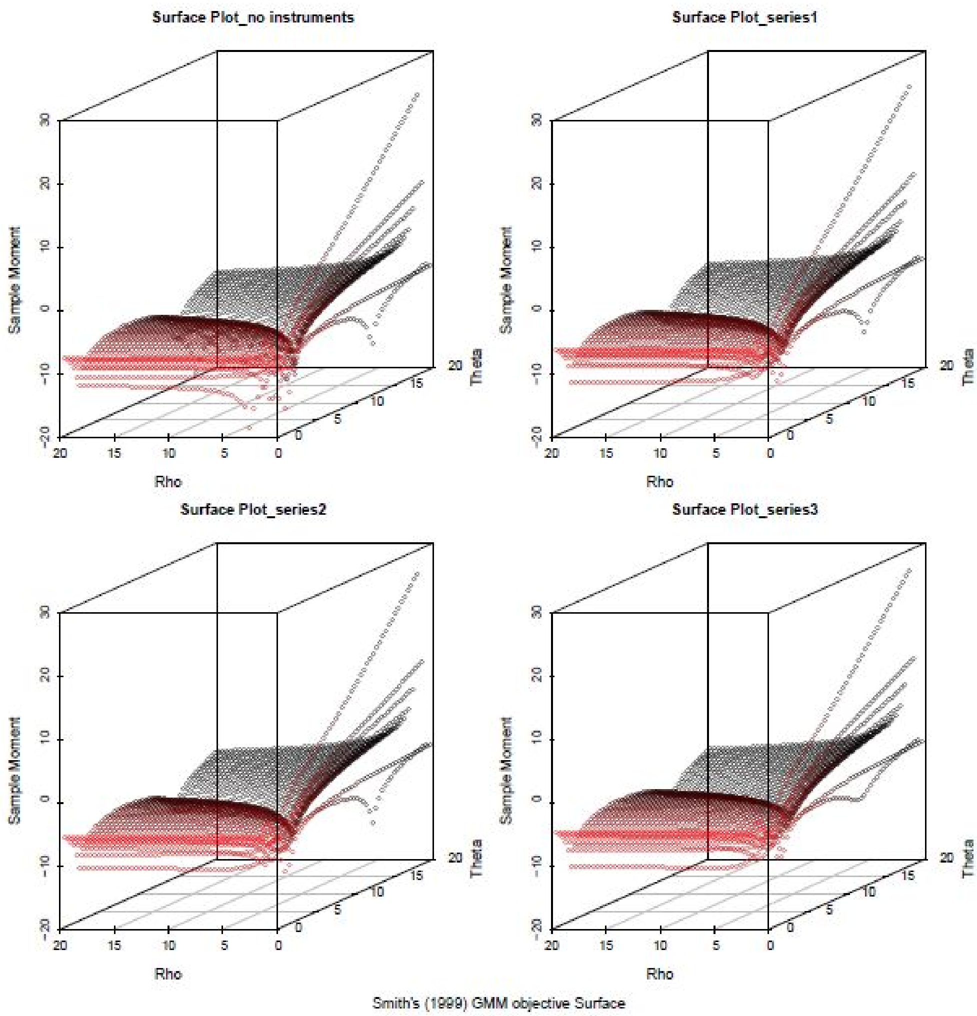

Figure 2: Surface of GMM object function ($\theta = 0.8$, $\rho = 0.8$)

2.4 Revisiting Chen, Favilukis, and Ludvigson (2013)

To double check the issue that the estimation results are very sensitive to the given initial values, I apply the CFL's (2013) method described in the previous section. With the same data as CFL's (2013), I estimate the parameters β, θ, and ρ using different starting values. The results are reported in table 4.

Table 4: Value function statistics and preference parameter estimates using CFL's (2013) method with different initial values.

(0.98, 50.00, 0.70)			
Value function statistics	Mean	Std	Autocorrelation
W = I	2.41	0.12	0.39
Second step estimation	β	θ	ρ
W = I	0.73	60.55	0.51
(0.98, 20.00, 0.90)			
Value function statistics	Mean	Std	Autocorrelation
W = I	1.32	0.14	0.14
Second step estimation	β	θ	ρ
W = I	0.99	1.14	1.69
(0.98, 0.80, 0.80)			
Value function statistics	Mean	Std	Autocorrelation
W = I	1.16	0.08	0.33
Second step estimation	β	θ	ρ
W = I	0.84	-0.55	0.48
(0.98, 5.20, 1.20)			
Value function statistics	Mean	Std	Autocorrelation
W = I	0.02	0.00	0.26
Second step estimation	β	θ	ρ
W = I	1.00	5.88	1.13

CFL's (2013) approach requires a two-step estimation procedure. First, given any parameter values, estimate the coefficients of known basis function, $\{a_j\}_{j=0}^{K}$, that minimizing the sandwich form GMM objective condition. In the second step, given the previously estimated value function, one can estimate the parameters based on the regular GMM procedure. This two-step process are iterated back and forth until the estimated value function and parameters converge. However, during the process I find that the estimates depend also on the initial values. For example, I can get similar results with CFL(2013) only when I choose starting values around their results, but the estimates change as I use different initials. This implies that the surface of GMM objective function using real life data are also flat. So computer only find the minimizer around where we start. The global minimizer is difficult to be reached unless we know exactly how the data is generated. Another issue happens in the estimation procedure using real life data is that, θ converges to ρ. This implies that the EZW model has a tendency to collapse to the time-separable preference model. In other words, the EZW recursive preference model enable us to release CRRA form EIS. In practice, however, the EZW model does provide more information than the time-separable model. In addition, CFL's (2013) method has the advantage that we can estimate the latent value function directly without finding proxy for the observable return on aggregate wealth. In practice, however, GMM objective function created by applying their approach are more fragile. Recall equation (17), the moment condition including a risk factor part

$$\text{Risk factor} = \left(\frac{\frac{V_{t+1}}{C_{t+1}} \frac{C_{t+1}}{C_t}}{\left(\frac{1}{\beta} \left[\left(\frac{V_t}{C_t} \right)^{1-\rho} - (1-\beta) \right] \right)^{\frac{1}{1-\rho}}} \right)^{\rho-\theta} \tag{35}$$

This risk factor is also the main difference between the recursive preference model and the time separable one, since the latter is a special case of the former when $\rho = \theta$. During my experiment process, one problem is that the risk factor is not available for some combination of parameters. For example, once the term

$$\left(\frac{V_t}{C_t} \right)^{1-\rho} - (1-\beta) \tag{36}$$

in the denominator becomes a very small positive number, the infinite risk factor would be generated and then terminate the estimation process. However, there is no theoretical reason to rule out these combinations. One way to solve this issue is to try another initial value, we may have some results if we are lucky. However, it's not convincing that the estimates that work are what we want. Another concerns I have when with CFL's (2013) method is that the two-step estimation procedure needs to go back and forth between value function and the parameters until they converge, then we can pin down each of them. However, in practice, it's very likely that θ and ρ get very close to each other during the process. As a result, the risk factor will be near zero and one can doubt such a small number can still provide enough information to pin down the value function, which including multiple parameters. In another word, we are actually trying to estimate value function based on nothing if the only source of information, risk factor, is too small. So, once the ρ and θ get close enough in any of the iteration of the process, the estimation hereafter would be meaningless.

3 Discussion and conclusion

In this paper I investigate some of the problems that may be encountered, and the cause behind the problem when using GMM to estimate parameters in EZW model. First, I find that some widely used GMM moment condition, particularly in the case of the nonlinear model, may be well defined theoretically but may not have a numerical result. For example, for some combinations of parameters, there could be infinite objective function or moment conditions may be undefined. In some cases, this issue could be solved by imposing some restriction on parameters. However, this restriction can be small in scope and is highly subjective. In other words, one estimates the value of a parameter in a subjectively predetermined, possibly small, interval and then get results. This is fine if the restricted domain can contain all the theoretically allowed points, and there is only one extreme point within the interval chosen. In the context of Epstein-Zin model, however, I find that when the model inherent the high degree of nonlinearity, neither the real-life data nor the simulated one can provide large enough variation to identify each parameter. Graphically speaking, the surface of GMM objective function is too flat to pin down the parameters. Second, since almost every computer algorithm requires a given starting point, this may lead to a result that the estimates are close to the starting point when the surface of the objective function is flat. This would not be a problem if we can have a nice concave surface with unique extremum. But, in practice, this may not be the case so one may worry that it is not convincing to say there is new proof when we use the results in the literature as the initials and finally get the similar results. From my experiment, I find the estimates vary significantly once we use much different initial values. This problem is considerable when the theoretically allowed range is broad, and the required accuracy of the estimate is high. Third, GMM method using lagged endogenous variables as instruments may suffer from the collinearity problem. The direct result is that the "optimal" weighting matrix is not invertible. Actually, the main reason is that the component $g'g$ is not invertible (where g is the GMM moment condition). This component is not only in HAC, but also used in some software as identity matrix. Once $g'g$ is not invertible, the estimation process will terminates and the error will be reported. Still, this issue is more likely to happen in the nonlinear model case. In most of the cases, the estimates are always available when we use identity matrix $w = I$. However, the inference is no longer valid, and all tests would have zero explanatory power.

I apply both Smith's (1999) and CFL's (2013) method to estimate the parameters in the Epstein-Zin asset pricing model and check the quality of moments generated through these methods. I find that Smith's (1999) method, which is finding a proxy for return to aggregate wealth, is more stable and compatible with more combination of parameters. In other words, it is less likely to produce infinite or undefined moments. CFL's (2013) approach has the advantage that it allows us to estimate the latent value function directly. However, this method is more fragile in the sense that it frequently produces "bad" moments and thus is unable to provide any results. We have to impose more restrictions on the parameters in order to get a numerical solution and continue the estimation process. However, excessive restrictions on the parameters may lead to the decline of the objectivity of the estimation. Finally, from my experiments, no matter which method we use, whether we are using real or simulated data, it is difficult to identify a unique and robust minimizer of GMM objective function. We can only find a local minimizer around the starting point.

Reference

Chauss, P. (2010). Computing Generalized Method of Moments and Generalized Empirical Likelihood with R. Journal of Statistical Software.

Chen, X., Favilukis, J., & Ludvigson, S. C. (2013). An estimation of economic models with recursive preferences. Quantitative Economics.

Epstein, L. G., & Zin, S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. Econometrica.

Epstein, L. G., & Zin, S. E. (1991). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis. Journal of Political Economy.

Ferson, W. E., & Foerster, S. R. (1994). Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. Journal of Financial Economics.

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. Journal of Business & Economic Statistics. Journal of Political Economy.

Hansen, L. P., Heaton, J., & Li, N. (2008). Consumption Strikes Back?: Measuring Long-Run Risk. Journal of Political Economy.

Nelson, C. R., & Startz, R. (1990). The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One. The Journal of Business.

Nelson, C., & Startz, R. (1990). Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator. Econometrica.

Smith, D. C. (1999). Finite Sample Properties of Tests of the Epstein-Zin Asset Pricing Model. Journal of Econometrics.

Stock, J. H., & Wright, J. H. (2000). GMM with Weak Identification. Econometrica.

Tauchen, G. (1986). Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data. Journal of Business Economic Statistics.

Tauchen, G., & Hussey, R. (1991). Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models. Econometrica.

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. Journal of Statistical Software.

Appendix

Table A1: Summary of estimates of instrument autoregressions of consumption growth and returns from Monte Carlo simulations using 500 draws of 90 observations from annual DGP.

θ, ρ	Variable		Intercept	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
0.80, 0.80	λ	Estimator	1.019	-0.152	-0.022	-0.011	-0.019	-0.020
		Std.error	0.003	0.105	0.107	0.108	0.109	0.108
		T-value	348.682	-1.439	-0.196	-0.097	-0.172	-0.190
	R_{sp}	Estimator	1.038	-0.011	-0.028	-0.014	-0.016	-0.006
		Std.error	0.013	0.105	0.106	0.107	0.107	0.108
		T-value	88.061	-0.103	-0.259	-0.136	-0.152	-0.055
	R_f	Estimator	1.035	-0.012	-0.030	-0.008	-0.023	-0.015
		Std.error	0.000	0.105	0.106	0.107	0.107	0.108
		T-value	2496.328	-0.111	-0.280	-0.076	-0.218	-0.140
0.80, 5.20	λ	Estimator	1.019	-0.144	-0.007	-0.011	-0.026	-0.017
		Std.error	0.003	0.105	0.107	0.108	0.109	0.108
		T-value	342.005	-1.364	-0.056	-0.095	-0.238	-0.158
	R_{sp}	Estimator	1.126	-0.005	-0.026	-0.018	-0.027	-0.014
		Std.error	0.014	0.105	0.106	0.107	0.107	0.108
		T-value	87.949	-0.049	-0.239	-0.172	-0.252	-0.138
	R_f	Estimator	1.122	-0.009	-0.026	-0.015	-0.018	-0.019
		Std.error	0.003	0.106	0.106	0.107	0.108	0.108
		T-value	399.454	-0.088	-0.238	-0.141	-0.168	-0.182
1.35, 1.35	λ	Estimator	1.019	-0.142	-0.013	-0.024	-0.029	-0.018
		Std.error	0.003	0.105	0.107	0.108	0.108	0.108
		T-value	345.289	-1.348	-0.115	-0.224	-0.264	-0.164
	R_{sp}	Estimator	1.048	-0.022	-0.013	-0.013	-0.014	-0.014
		Std.error	0.013	0.105	0.106	0.107	0.107	0.108
		T-value	88.077	-0.206	-0.118	-0.119	-0.135	-0.131
	R_f	Estimator	1.045	-0.013	-0.021	-0.008	-0.021	-0.011
		Std.error	0.001	0.105	0.106	0.107	0.107	0.108
		T-value	1502.682	-0.12	-0.203	-0.077	-0.201	-0.108
1.35, 5.20	λ	Estimator	1.019	-0.147	-0.014	-0.018	-0.033	-0.014
		Std.error	0.003	0.105	0.107	0.108	0.108	0.108
		T-value	347.353	-1.394	-0.130	-0.172	-0.305	-0.132
	R_{sp}	Estimator	1.123	-0.010	-0.015	-0.010	-0.032	-0.005
		Std.error	0.014	0.105	0.106	0.107	0.107	0.108
		T-value	86.86	-0.092	-0.137	-0.092	-0.304	-0.044
	R_f	Estimator	1.120	-0.010	-0.010	-0.012	-0.013	-0.010
		Std.error	0.003	0.105	0.106	0.107	0.107	0.108
		T-value	386.187	-0.097	-0.095	-0.114	-0.120	-0.097

Table A2: Summary of estimates of instrument regressions of consumption growth and returns from Monte Carlo simulations using 500 draws of 90 observations from annual DGP.

θ, ρ	Variable		Intercept	$R_{sp,t-1}$	λ_{t-1}	$R_{f,t}$	$Adj - R^2$
0.80, 0.80	λ	Estimator	1.150	0.001	-0.147	0.017	0.023
		Std.error	0.932	0.031	0.107	0.892	
		T-value	1.251	0.044	-1.401	0.017	
	R_{sp}	Estimator	1.375	-0.008	-0.024	-0.294	0.001
		Std.error	3.307	0.108	0.380	3.167	
		T-value	0.409	-0.076	-0.070	-0.082	
	R_w	Estimator	1.028	0.001	0.006	0.001	0.002
		Std.error	0.939	0.031	0.108	0.900	
		T-value	1.106	0.035	0.055	0.002	
0.80, 5.20	λ	Estimator	1.155	-0.002	-0.141	0.008	0.019
		Std.error	0.183	0.028	0.107	0.128	
		T-value	6.357	-0.064	-1.331	0.066	
	R_{sp}	Estimator	1.162	-0.003	-0.002	-0.027	-0.003
		Std.error	0.714	0.108	0.418	0.496	
		T-value	1.650	-0.031	0.002	-0.064	
	R_w	Estimator	1.123	-0.002	0.015	-0.010	-0.003
		Std.error	0.330	0.050	0.193	0.229	
		T-value	3.442	-0.043	0.075	-0.045	
1.35, 1.35	λ	Estimator	1.184	0.001	-0.139	-0.023	0.018
		Std.error	0.568	0.031	0.107	0.531	
		T-value	2.104	0.044	-1.317	-0.041	
	R_{sp}	Estimator	1.122	-0.018	-0.005	-0.047	0.000
		Std.error	2.017	0.108	0.382	1.887	
		T-value	0.563	-0.172	-0.010	-0.025	
	R_w	Estimator	1.066	0.001	0.000	-0.019	-0.001
		Std.error	0.610	0.033	0.116	0.571	
		T-value	1.764	0.042	-0.007	-0.030	
1.35, 5.20	λ	Estimator	1.166	-0.003	-0.144	0.003	0.018
		Std.error	0.183	0.028	0.107	0.127	
		T-value	6.427	-0.103	-1.361	0.025	
	R_{sp}	Estimator	1.142	-0.008	-0.011	0.002	0.001
		Std.error	0.705	0.108	0.414	0.490	
		T-value	1.641	-0.072	-0.030	0.001	
	R_w	Estimator	1.117	0.000	0.003	0.003	0.001
		Std.error	0.327	0.050	0.192	0.228	
		T-value	3.451	-0.004	0.017	0.019	

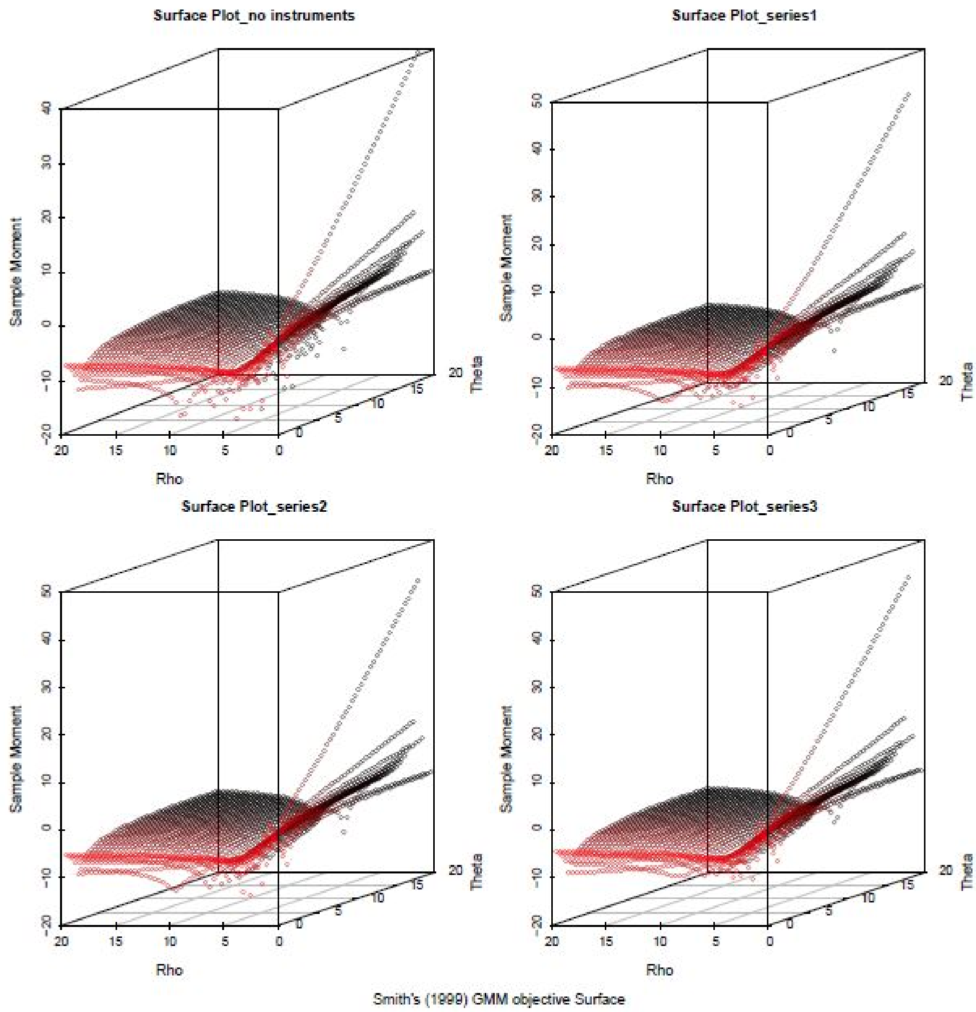

Figure 3: Surface of GMM object function ($\theta = 0.8$, $\rho = 5.2$)

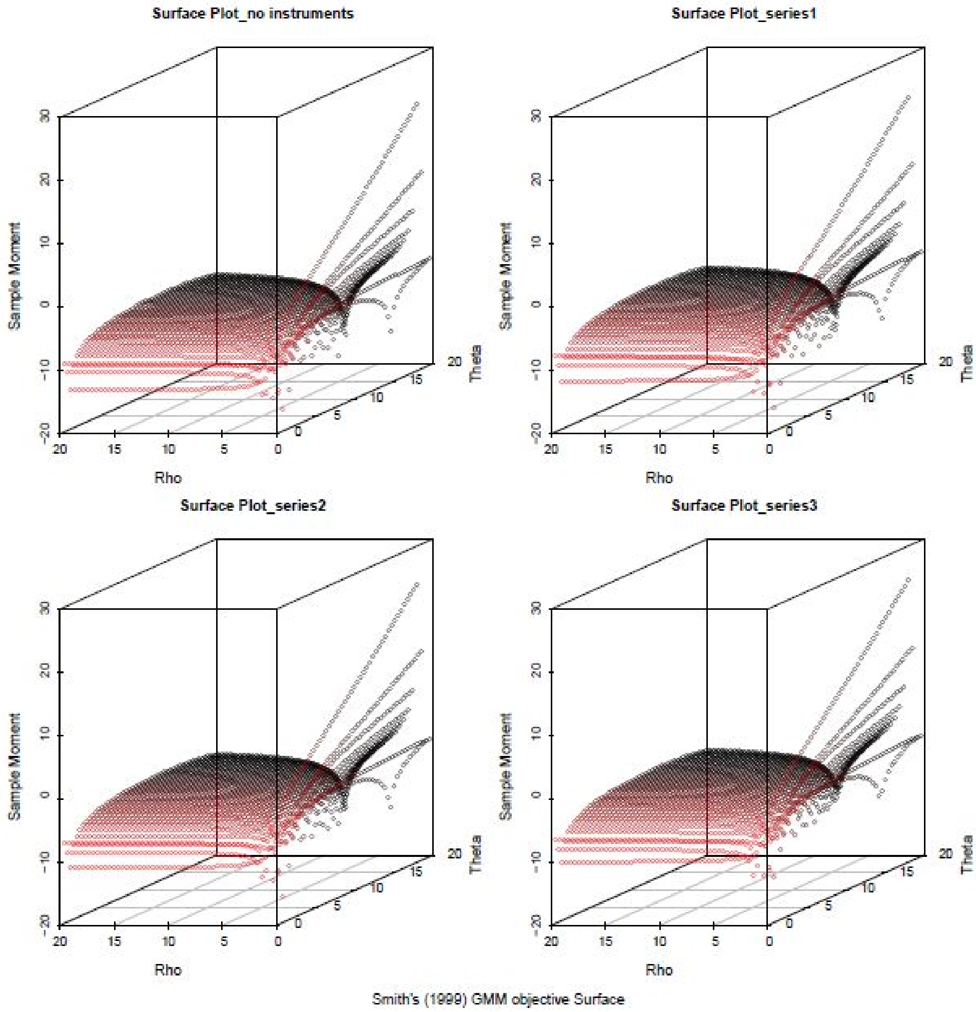

Figure 4: Surface of GMM object function ($\theta = 1.35$, $\rho = 1.35$)

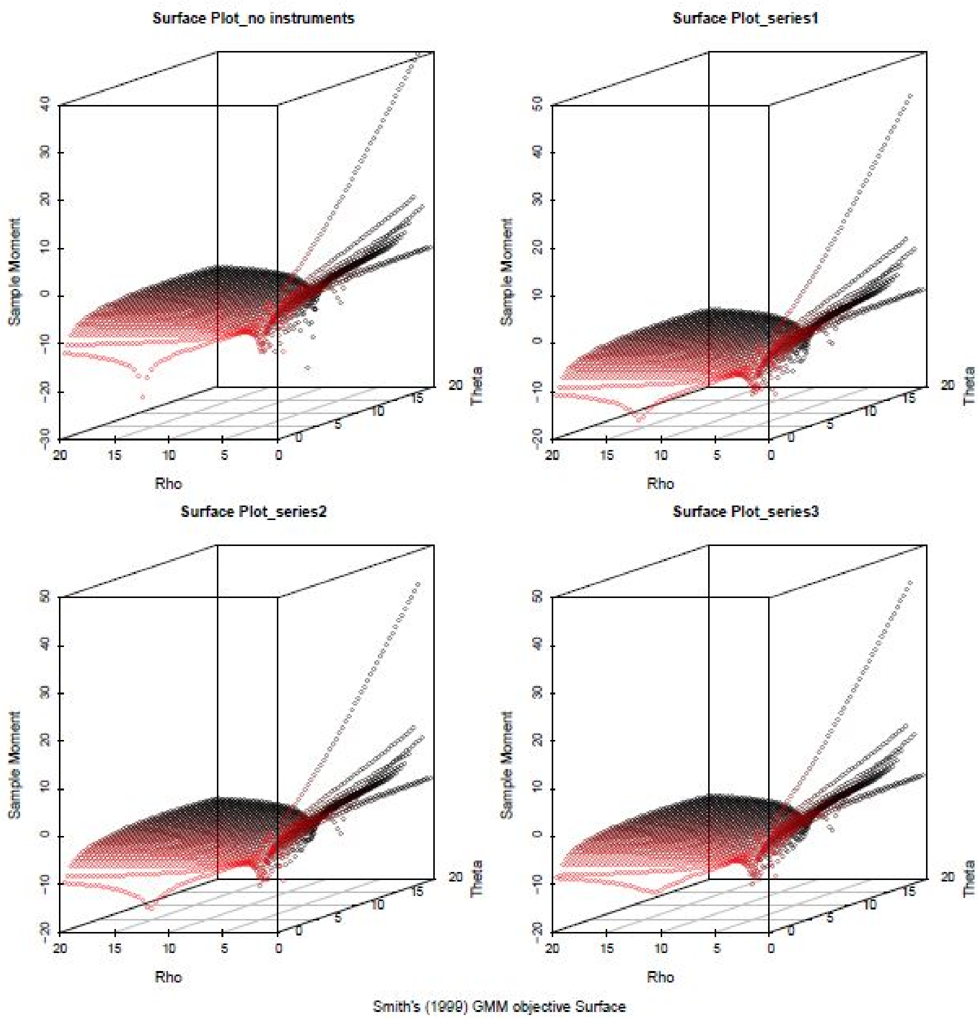

Figure 5: Surface of GMM object function ($\theta = 1.35$, $\rho = 5.2$)

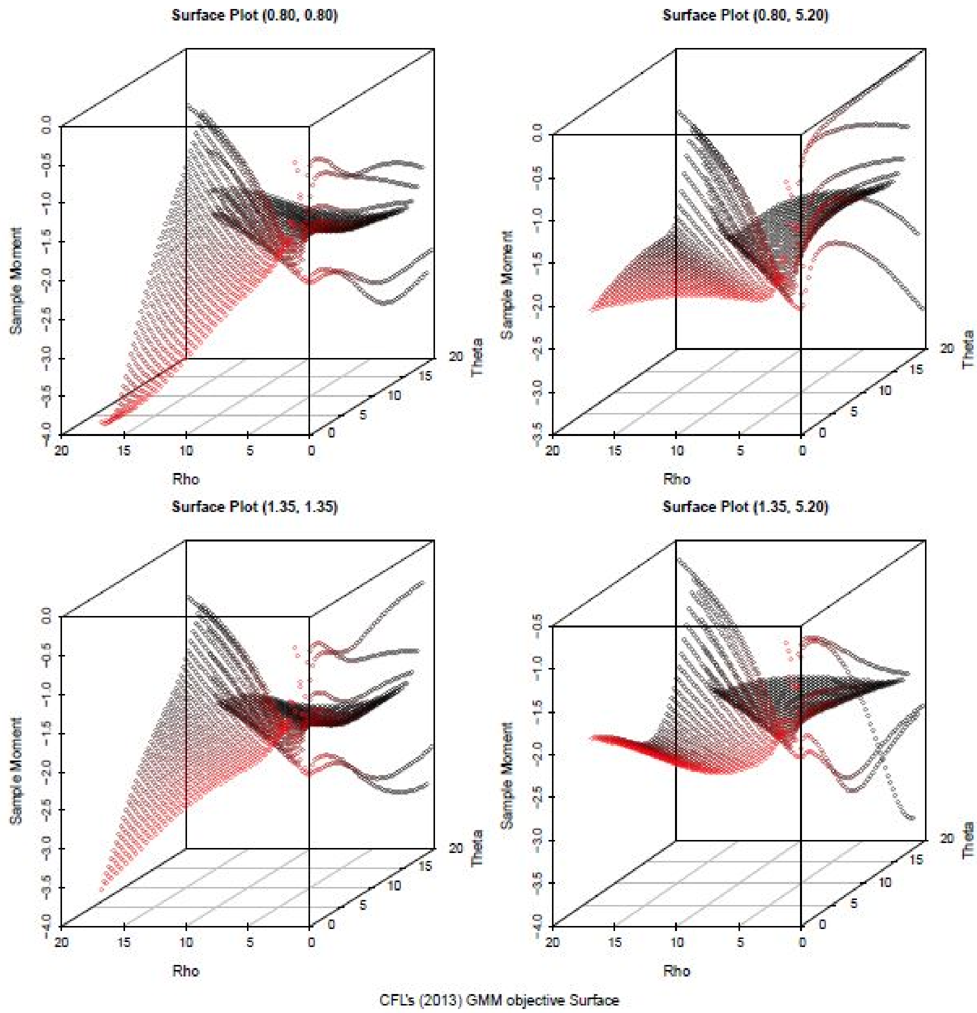

Figure 6: CFL surface of GMM objective function ($W = g'g$)