

Maximum Entropy Analysis of Consumption-based Capital Asset Pricing Model and Volatility*

Tae-Hwy Lee[†] Millie Yi Mao[‡] and Aman Ullah[§]

First version: September 2019

Revised: March 2020

This version: July 2020

Abstract

Based on the maximum entropy (ME) method, we introduce an information theoretic approach to estimating conditional moment functions with incorporating a theoretical constraint implied from the consumption-based capital asset pricing model (CCAPM). Using the ME conditional mean/variance functions obtained from the ME density, we analyze the relationship between asset returns and consumption growth under the theoretical constraint of the CCAPM. We evaluate the predictability of asset return using consumption growth through in-sample estimation and out-of-sample prediction in the ME mean regression function. We also examine the ME variance regression function for the asset return volatility as a function of the consumption growth. Our findings suggest that incorporating the CCAPM constraint can capture the nonlinear predictability of asset returns in mean especially in tails, and that the consumption growth has an effect on reducing stock return volatility, indicating the counter-cyclical variation of stock market volatility.

Key Words: Information theory, Stock return and consumption growth, CCAPM theoretical constraint, ME mean regression function, ME variance regression function.

JEL Classification: C1, C5, G1.

*The authors are thankful to Raffaella Giacomini (the editor) and two anonymous referees for many valuable comments, to Amos Golan, Claudio Morana, and Liangjun Su for discussions on the subject matter of this paper, and to the seminar participants at the World Finance & Banking Symposium, New Delhi.

[†]Department of Economics, University of California, Riverside, CA 92521. E-mail: taelee@ucr.edu

[‡]Department of Mathematics, Physics and Statistics, Azusa Pacific University, Azusa, CA 91702. E-mail: ymao@apu.edu

[§]Department of Economics, University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu

1 Introduction

When the bivariate normal distribution assumption is made on the dependent and independent variables, it implies the linear mean regression function between them. However, many economic theories suggest nonlinear relationships between dependent and independent variables. For example, in the consumption-based capital asset pricing model (CCAPM), a representative household is solving a utility maximization problem subject to budget constraints. The optimization constraint obtained from the consumption Euler equation explains the trade-off between consumption in two consecutive periods, and the household's consumption and investment decision at each time period. This theoretical constraint from the consumption Euler equation suggests that the consumption growth rate affects the pricing of assets in a nonlinear way.

In this paper, we examine how important this theoretical constraint is in estimating and predicting asset returns in the conditional mean and conditional variance when using the consumption growth as a predictor. We start from a bivariate normal density, and then incorporate the nonlinearity implied by the consumption Euler equation into the estimation of the joint distribution of the consumption growth and asset return. That is to tilt the bivariate normal density towards a target density satisfying the consumption-based asset pricing model, which will produce the maximum entropy (ME) density. In using the exponentially tilted ME density, our paper is closely related to Robertson et al (2005) and especially to Giacomini and Ragusa (2014) in economics literature. However we derive the conditional mean and variance functions analytically from the ME density, instead of drawing them numerically from the ME density. We implement the “recursive integration” method proposed in Mao and Ullah (2020) to estimate the conditional mean and the conditional variance regression functions of asset returns conditional on the consumption growth.

Since we wish to examine the importance of a theoretical constraint in asset pricing, we determine the conditional moment functions of asset returns according to the estimated ME density subject to the CCAPM constraint. Our ME conditional mean and variance estimators are purely data-driven subject to the theoretical CCAPM constraint. We then carry out both in-sample estimation and out-of-sample prediction to investigate the importance of the theoretical constraint in asset pricing. Our findings suggest that incorporating the theoretical constraint does improve the

predictability of asset returns in mean especially in tails, and that the consumption growth has an effect on reducing stock return volatility, explaining the counter-cyclical variation of stock market volatility with a consumption-based asset pricing model as presented in Campbell and Cochrane (1999).

We organize this paper in the following order. In Section 2, we present the conditional mean regression function obtained from using the bivariate maximum entropy distribution with incorporating the theoretical constraint from the CCAPM. We use the recursive integration algorithm to estimate the conditional mean regression function. In Section 3, we carry out a simulation to illustrate our method. In Section 4, we estimate the ME mean regression function of the US stock returns conditional on the consumption growth and analyze the role of the CCAPM constraint in the nonlinearity of the ME mean function. In Section 5, we examine the out-of-sample predictability of US stock returns in the conditional mean. In Section 6, we introduce the ME conditional variance function and examine the effect of consumption growth on the stock return volatility. Section 7 concludes. The mathematical details of the CCAPM constraint and the recursive integration algorithm described in Section 2 are explained in Appendix (Section 8).

2 Maximum Entropy Regression under a Theoretical Constraint

We consider $\{y_t, x_t\}$, $t = 1, \dots, T$, to be independent and identically distributed observations from an absolutely continuous bivariate distribution $f(y, x)$. Suppose the conditional mean of y given x exists and it provides a formulation for the regression model as

$$\begin{aligned} y &= E(y|x) + u \\ &\equiv m(x) + u, \end{aligned} \tag{1}$$

where the error term u is such that $E(u|x) = 0$, and the mean regression function (conditional mean) is

$$m(x) = \int_y y \frac{f(y, x)}{f(x)} dy = \frac{1}{f(x)} \int_y y f(y, x) dy, \tag{2}$$

where $f(x)$ is the marginal density of x .

When the joint density of y and x is not known, which is often the case, we propose the maximum entropy method based on the information theory to estimate $f(y, x)$ and the recursive integration method to solve the conditional mean/variance of y given x . We explain the method in the following three subsections.

2.1 Maximum Entropy Distribution

We start with a “proposed” joint density $f(y, x)$. Our goal is to get the optimal “target” density $\tilde{f}(y, x)$ which incorporates the importance of theoretical constraints. The optimal target density $\tilde{f}(y, x)$ is achieved by maximizing the Kullback-Leibler type entropy (information), which has the form

$$\text{Max}_{\tilde{f}} H(\tilde{f}) \equiv - \int_x \int_y \tilde{f}(y, x) \log \frac{\tilde{f}(y, x)}{f(y, x)} dy dx \quad (3)$$

subject to

$$\int_x \int_y y^i x^j \tilde{f}(y, x) dy dx = \mu_{ij}, \quad 0 \leq i + j \leq k, \quad (4)$$

$$\int_x \int_y g_m(y, x) \tilde{f}(y, x) dy dx = 0, \quad m = 1, \dots, M, \quad (5)$$

where k is the highest order of moments conditions. $g_m(y, x)$ is the m th constraint based on economic theory such as the CCAPM. M represents the total number of theoretical constraints. In general, the maximum entropy approach can be applied to various examples in economics and finance to incorporate the importance of theoretical constraints, such as the production-based asset pricing model in Cochrane (1991), portfolio allocation and option pricing problems in Golan (2017).

In this paper, we are interested in analyzing the relationship between the asset return y and the consumption growth x according to the consumption-based asset pricing model. Hence, we use the consumption Euler equation as the theoretical constraint in this paper. With the single constraint from the CCAPM, i.e. $M = 1$, and $g_m(y, x)$ will be denoted as $g(y, x)$ for simplicity.

According to Hansen and Singleton (1983), consumers solve the time-additive utility maximiza-

tion problem, which takes the form

$$E_t \sum_{\tau=0}^{\infty} \beta^{t+\tau} \frac{C_{t+\tau}^{1-\alpha}}{1-\alpha}, \quad (6)$$

subject to the sequence of budget constraints

$$C_{t+\tau} + B_{t+\tau+1} \leq (1 + r_{t+\tau}) B_{t+\tau} + W_{t+\tau}. \quad (7)$$

Consumers have a constant relative risk aversion (CRRA) type utility function with risk aversion parameter α . β is the discount factor. $C_{t+\tau}$, $B_{t+\tau}$ and $W_{t+\tau}$ are the real consumption, saving and wealth at time $t + \tau$. $1 + r_{t+\tau}$ represents the gross asset return at time $t + \tau$. Forming the Lagrangian,

$$\mathcal{L} = E_t \sum_{\tau=0}^{\infty} \left(\beta^{t+\tau} \frac{C_{t+\tau}^{1-\alpha}}{1-\alpha} + \delta_{t+\tau} ((1 + r_{t+\tau}) B_{t+\tau} + W_{t+\tau} - C_{t+\tau} - B_{t+\tau+1}) \right), \quad (8)$$

where $\delta_{t+\tau}$ is the Lagrangian multiplier of budget constraint at time $t + \tau$. The first-order necessary conditions for the above maximization problem are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial C_{t+\tau}} &= E_t (\beta^{t+\tau} C_{t+\tau}^{-\alpha} - \delta_{t+\tau}) = 0, \\ \frac{\partial \mathcal{L}}{\partial C_{t+\tau+1}} &= E_t (\beta^{t+\tau+1} C_{t+\tau+1}^{-\alpha} - \delta_{t+\tau+1}) = 0, \\ \frac{\partial \mathcal{L}}{\partial B_{t+\tau+1}} &= E_t (\delta_{t+\tau} + \delta_{t+\tau+1} (1 + r_{t+\tau+1})) = 0. \end{aligned} \quad (9)$$

Substituting out $\delta_{t+\tau}$ and $\delta_{t+\tau+1}$ from the first two equations in (9) and rearranging the third equation, the consumption Euler equation is

$$E_t (\beta X_{t+\tau}^{-\alpha} Y_{t+\tau} - 1) = 0, \quad (10)$$

where $X_{t+\tau} \equiv \frac{C_{t+\tau}}{C_t}$, $Y_{t+\tau} \equiv 1 + r_{t+\tau}$. (10) shows the optimal trade-off between consumption at time t and time $t + \tau$. The marginal utility of consumption at time t is equal to the product of

marginal utility of consumption at time $t + \tau$ and total investment return at time t .

According to Cochrane (2000), the value of an asset depends on the delay and risk of its payments. In consumption-based asset pricing models, risk corrections to asset prices should be driven by the relationship between asset payoffs and consumption. For a given expected payoff of an asset, the asset price is affected by the state of the economy. When the economy is going through a recession, investors feel poorer and contract their consumption. Risky assets cannot guarantee investors with certain amount of consumption. Thus, they are sold at lower prices comparing to an asset in a state where the economy is experiencing a boom when investors feel wealthy and expand their consumption. The decrease in asset prices during economic recession reflects a discount of their riskiness, which is determined by the relationship between asset payoff and consumption. Based on the above economic intuition, we are interested in predicting asset returns using the information included in the consumption Euler equation in (10), which indicates a certain relationship between asset price and consumption growth.

Since we are interested in studying the relationship between consumption growth and asset return at the same time, we drop the time subscripts of both variables. Define $x \equiv \log(X)$ and $y \equiv \log(Y)$. After a Taylor series expansion up to the 4th order¹, the consumption Euler equation in (10) becomes

$$g(y, x) = \log \beta - \alpha x + y + \frac{1}{2} \alpha (\alpha + 1) x^2 - \alpha xy - \frac{1}{6} \alpha (\alpha + 1) (\alpha + 2) x^3 + \frac{1}{2} \alpha (\alpha + 1) x^2 y + \frac{1}{24} \alpha (\alpha + 1) (\alpha + 2) (\alpha + 3) x^4 - \frac{1}{6} \alpha (\alpha + 1) (\alpha + 2) x^3 y. \quad (11)$$

The derivation of (11) is shown in subsection 8.1. When the risk aversion parameter α is recognized as 3, which has been commonly used in the literature, the consumption Euler equation in (11) becomes

$$g(y, x) = \log \beta - 3x + y + 6x^2 - 3xy - 10x^3 + 6x^2y + 15x^4 - 10x^3y. \quad (12)$$

In general, the risk aversion parameter α can be estimated using generalized method of moments

¹Generally, the function of theoretical restrictions can be expanded to any higher orders. For our consumption-based asset pricing model, we expand the Euler equation up to the 4th order. For other applications, expansions with higher than the 4th order may be useful.

(GMM) and generalized empirical likelihood (GEL) methods. See Kitamura and Stutzer (1997), Imbens et al (1998), Bera and Biliias (2002), and Kitamura et al (2004). Since it is not our primary interest to estimate the risk aversion parameter α in our empirical analysis in Sections 4-6, where we predict the stock returns using consumption growth rates, we set the α value to be 3 as well as a range of values from 0.1 to 4.6 as used in Giacomini and Ragusa (2014). We report the empirical results based on all these α choices in Sections 4-6.

When the true joint density of x and y is not known, parametric functional forms of their relationship are usually assumed. In particular, a regression function is often considered to be linear. Under the linearity assumption, the highest order of moment conditions k in (4) is 2, as implied by a bivariate normal density for y and x . In this paper, we assume that the initially proposed joint density $f(y, x)$ follows the bivariate normal distribution, and thus $m(x)$ is proposed to be linear.

In our consumption-based asset pricing application, all moment conditions in (4) are finite. With a reasonable choice of the risk aversion parameter α , the theoretical constraint in (11) is also finite. Therefore, according to Komunjer and Ragusa (2016), the target maximum entropy density exists and is unique.

By solving the maximization problem in (3), we obtain the target ME density $\tilde{f}(y, x)$ of the form

$$\tilde{f}(y, x) = \exp \left\{ - \left[\lambda_{00} + \lambda_{10}y + \lambda_{01}x + \lambda_{20}y^2 + \lambda_{02}x^2 + \lambda_{11}yx + \lambda_g g(y, x) \right] \right\}, \quad (13)$$

where λ_{ij} is the Lagrangian multiplier that corresponds to the moment condition $\int_x \int_y y^i x^j \tilde{f}(y, x) dy dx = \mu_{ij}$ in (4), and λ_g is the Lagrangian multiplier that corresponds to the theoretical constraint in (5). Given the joint density form in (13), the marginal density of x is achieved by integrating y out, and the marginal density of y by integrating x out, viz.,

$$\begin{aligned} \tilde{f}(x) &= \int_y \exp \left\{ - \left[\lambda_{00} + \lambda_{10}y + \lambda_{01}x + \lambda_{20}y^2 + \lambda_{02}x^2 + \lambda_{11}yx + \lambda_g g(y, x) \right] \right\} dy, \\ \tilde{f}(y) &= \int_x \exp \left\{ - \left[\lambda_{00} + \lambda_{10}y + \lambda_{01}x + \lambda_{20}y^2 + \lambda_{02}x^2 + \lambda_{11}yx + \lambda_g g(y, x) \right] \right\} dx. \end{aligned} \quad (14)$$

The Lagrangian multipliers not only characterize the above density functions, but also represent the relative strength of each moment condition and the theoretical constraint. See Golan et al (1996), Ullah (1996), and Golan (2017). To estimate the Lagrangian multipliers, we use the Newton method considered by Wu (2003, 2010). According to the functional form of the theoretical constraint of the consumption Euler equation in (11), the third order and fourth order terms bring non-normality into the maximum entropy joint density $\tilde{f}(y, x)$, which tilts the initially proposed bivariate normal density $f(y, x)$. It will make a nonlinear mean regression model. We then examine the gains from incorporating the theoretical constraint $g(y, x)$ in estimation and prediction of conditional mean.

2.2 ME Mean Regression Function

First, let us define

$$\lambda_{10}(x, \alpha) \equiv -\frac{1}{6}\lambda_g\alpha(\alpha+1)(\alpha+2)x^3 + \frac{1}{2}\lambda_g\alpha(\alpha+1)x^2 + (\lambda_{11} - \lambda_g\alpha)x + (\lambda_{10} + \lambda_g). \quad (15)$$

$\lambda_{10}(x, \alpha)$ is a function of the log consumption growth rate x , the risk aversion parameter α , and the Lagrange multipliers λ_g , λ_{11} and λ_{10} . Based on the ME joint density in (13) and ME marginal density in (14), the conditional mean of y given x is represented as²

$$m(x, \boldsymbol{\lambda}, \alpha) = E(y|x) = \frac{1}{\tilde{f}(x)} \int_y y \tilde{f}(y, x) dy \quad (16)$$

$$\begin{aligned} &= \frac{\int_y y \exp\{-[\lambda_{00} + \lambda_{10}y + \lambda_{01}x + \lambda_{20}y^2 + \lambda_{02}x^2 + \lambda_{11}yx + \lambda_g g(y, x)]\} dy}{\int_y \exp\{-[\lambda_{00} + \lambda_{10}y + \lambda_{01}x + \lambda_{20}y^2 + \lambda_{02}x^2 + \lambda_{11}yx + \lambda_g g(y, x)]\} dy} \\ &= \frac{\int_y y \exp\{-[\lambda_{20}y^2 + \lambda_{10}(x, \alpha)y]\} dy}{\int_y \exp\{-[\lambda_{20}y^2 + \lambda_{10}(x, \alpha)y]\} dy}, \end{aligned} \quad (17)$$

where $\boldsymbol{\lambda} = (\lambda_{20}, \lambda_{11}, \lambda_{10}, \lambda_g)'$. When $\alpha = 3$,

$$\lambda_{10}(x, 3) = -10\lambda_g x^3 + 6\lambda_g x^2 + (\lambda_{11} - 3\lambda_g)x + (\lambda_{10} + \lambda_g). \quad (18)$$

²Since the discount factor β in (10) does not play a role in Equation (16), our prediction results are robust to the choice of β . When β is unknown, one can refer to Hahn et al. (2019) on estimation of β .

Observing the numerator and denominator in (17), we define the following integrals as functions of x

$$F_r(x, \boldsymbol{\lambda}, \alpha) \equiv \int_y y^r \exp \{ -[\lambda_{20}y^2 + \lambda_{10}(x, \alpha)y] \} dy, \quad (19)$$

where $r = 0, 1, 2, \dots$. Then the conditional mean function in (16) is expressed as

$$m(x, \boldsymbol{\lambda}, \alpha) = \frac{F_1(x, \boldsymbol{\lambda}, \alpha)}{F_0(x, \boldsymbol{\lambda}, \alpha)}. \quad (20)$$

When the Lagrangian multipliers $\boldsymbol{\lambda}$ are estimated by the Newton method,

$$m(x, \hat{\boldsymbol{\lambda}}, \alpha) = \frac{F_1(x, \hat{\boldsymbol{\lambda}}, \alpha)}{F_0(x, \hat{\boldsymbol{\lambda}}, \alpha)}. \quad (21)$$

This is our maximum entropy conditional mean estimator.

Moreover, the response function $\beta(x, \boldsymbol{\lambda}, \alpha) \equiv \frac{dm(x, \boldsymbol{\lambda}, \alpha)}{dx}$ (the derivative of the conditional mean) can be written as

$$\beta(x, \boldsymbol{\lambda}, \alpha) = \frac{F_1'(x, \boldsymbol{\lambda}, \alpha) F_0(x, \boldsymbol{\lambda}, \alpha) - F_1(x, \boldsymbol{\lambda}, \alpha) F_0'(x, \boldsymbol{\lambda}, \alpha)}{F_0^2(x, \boldsymbol{\lambda}, \alpha)}, \quad (22)$$

and its estimator is given by $\beta(x, \hat{\boldsymbol{\lambda}}, \alpha)$. When the Lagrangian multipliers are given and the risk aversion parameter α is fixed, $F_r(x, \boldsymbol{\lambda}, \alpha)$ is only a function in terms of x , and can be written simply as $F_r(x, \boldsymbol{\lambda}, \alpha) = F_r(x)$, $m(x, \boldsymbol{\lambda}, \alpha) = m(x)$, and $\beta(x, \boldsymbol{\lambda}, \alpha) = \beta(x)$.

2.3 Recursive Integration Algorithm

Since the theoretical constraint $g(y, x)$ in (11) contains the third and fourth order terms, it is not likely to solve out the exponential polynomial integrals in both the numerator and denominator in (17). One can use numerical methods to compute the integrals at every value of x . However, it can be very computationally expensive in large samples and thus is unsatisfactory. As an alternative, Mao and Ullah (2020) introduce a recursive integration method which not only solves the conditional

mean function $m(x)$ but also reduces the computational cost.

We present the recursive integration method as the following algorithm. Given the definition of $F_r(x)$ in (19), we use $F'_r(x)$ to represent the first derivative of $F_r(x)$.

Recursive Integration Algorithm

1. Write the first derivatives of $F_0(x)$, $F_1(x)$ and $F_2(x)$ in a linear system,

$$\begin{aligned} F'_0(x) &= \Lambda_{01}(x)F_1(x), \\ F'_1(x) &= \Lambda_{12}(x)F_2(x), \\ F'_2(x) &= \Lambda_{21}(x)F_1(x) + \Lambda_{22}(x)F_2(x), \end{aligned} \tag{23}$$

where $\Lambda(x)$'s denote the corresponding coefficients as functions of x .

2. Starting from a given initial value x_0 , trace out functions $F_0(x)$, $F_1(x)$ and $F_2(x)$ over the entire range of x with a small increment h ,

$$\begin{aligned} F_0(x_0 + h) &\approx F_0(x_0) + F'_0(x)h, \\ F_1(x_0 + h) &\approx F_1(x_0) + F'_1(x)h, \\ F_2(x_0 + h) &\approx F_2(x_0) + F'_2(x)h, \end{aligned} \tag{24}$$

3. Given that $m(x) = \frac{F_1(x)}{F_0(x)}$ and $\beta(x) = \frac{dm(x)}{dx}$, trace out the conditional mean and response functions.
-

The mathematical derivation of (23) is provided in subsection 8.2. Note that Equation (23) is about integrating (19) over indefinite domain. The results for the definite domain integration are similar to the above. For the step size h in Equation (24), we have used several fixed small values and found that our estimation results are robust to the choice of h . Computing time of our algorithm does not vary much when h value changes.

3 Simulation

In this section, we design a simulation to evaluate the performance of our maximum entropy conditional mean estimator and illustrate the usefulness of the recursive integration method shown in subsection 2.3.

We first generate data $\{y_t, x_t\}$, $t = 1, \dots, T$, from a bivariate normal distribution

$$N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_x \\ \rho\sigma_y\sigma_x & \sigma_x^2 \end{pmatrix} \right), \quad (25)$$

where ρ captures the linear correlation between y and x . (25) has a bivariate normal joint density $f(y, x)$. Then, with $k = 2$ in (4), the target density $\tilde{f}(y, x)$ tilts the proposed bivariate normal density by incorporating the theoretical constraint $g(y, x)$ in (11). Thus, $\tilde{f}(y, x)$ has the form

$$\tilde{f}(y, x) = f(y, x) \times \exp[-\lambda_g g(y, x)]. \quad (26)$$

We then use the Monte-Carlo Markov-Chain (MCMC) method to draw data from the target density $\tilde{f}(y, x)$, under which the relationship between y and x is no longer linear. See Chib and Greenberg (1995) for the usefulness of the MCMC method in drawing samples from unnamed densities. We first estimate $\tilde{f}(y, x)$ by solving the entropy maximization problem in (3). After we obtain the estimated maximum entropy joint density, we use the recursive integration method introduced in subsection 2.3 to compute $m(x)$. As a comparison, we run a linear regression of y on x , which corresponds to the initially proposed bivariate normal density. Fixing the parameter values at $\lambda_g = 0.1$, $\alpha = 3$, $\sigma_y = 1$, we evaluate the estimation results based on both proposed and target densities when $\sigma_x \in \{0.1, 0.2\}$, and the correlation coefficient $\rho \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. In this case, the theoretical constraint we use is Equation (12). Define the root mean squared error (RMSE) under the proposed and target densities as

$$\text{RMSE}_\xi = \left(\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t^\xi)^2 \right)^{1/2}, \quad (27)$$

where $\xi = \text{LN}$ or ME . Here, \hat{y}_t^{LN} is estimated from the linear (denoted as LN) regression based on the proposed density. \hat{y}_t^{ME} is estimated by the maximum entropy conditional mean $\hat{m}(x_t)$ shown in (21). We compute the Monte Carlo average of the RMSEs over 100 replications with sample size $T = 500$. Table 1 reports results under $\sigma_x = 0.1$ and 0.2 while the correlation coefficient ρ is

changing from 0 to 0.9.

The results in Table 1 show that incorporating the theoretical constraint improves the estimation of the conditional mean of y given x regardless of the variance of x and coefficient of correlation ρ . As the coefficient of correlation ρ increases, RMSE under both proposed and target density decreases because stronger linearity prevails. Even when ρ gets very large, which indicates a strong linear relationship between y and x , the maximum entropy conditional mean estimator still outperforms the simple linear regression. Since the true data generating process is never known in practice, the result indicates that our ME mean regression estimator would provide a better estimation method. In the next section we apply the ME method for estimating the mean regression function of stock returns conditional on consumption growth.

4 Estimating Stock Returns: ME Mean Regression Function

Based on the economic theory in subsection 2.1, we are interested in analyzing the relationship between stock return and consumption growth. To examine the role of the theoretical constraint in (11) to predict stock return, we conduct in-sample estimation in this section. We investigate how consumption growth rate serves as a stock return predictor by using quarterly and monthly data. Quarterly aggregate real consumption data are available on Martin Lettau's website (<https://sites.google.com/view/martinlettau>), containing observations from 1952Q1 to 2016Q3. Monthly aggregate real consumption data from 1959JAN to 2016DEC are downloaded from the webpage of FRED of the Federal Reserve Bank at St. Louis. Stock index data at different frequencies are obtained from Amit Goyal's website (<http://www.hec.unil.ch/agoyal/>). According to the introduction of economic theory in subsection 2.1, x represents the real consumption growth rate and y represents the asset return. We take logarithms for both the aggregate real consumption and stock index data, and then compute the first-differences to obtain the consumption growth and stock return. Each variable is then scaled by dividing its sample standard deviation.

Having the data of y and x available, we first obtain $\tilde{f}(y, x)$, the maximum entropy joint density between y and x . Then, we plug in $\tilde{f}(y, x)$ into (16) and (22) to estimate our ME conditional mean

function and its response function. We also report the estimation results under the simple linear regression. Because the ME conditional mean estimator exhibits nonlinearity as the maximum entropy density incorporates the theoretical constraint, the improvement in the ME nonlinear model over the linear model sheds light on the contribution of the theoretical constraint in the regression function of the stock return when the consumption growth is a regressor.

As a comparison, we also consider the nonparametric (NP) kernel methods in this empirical study, for which \hat{y}_t is computed by the local constant least squares (LCLS) and local linear least squares (LLLS) estimators. The local constant (Nadaraya-Watson) nonparametric kernel estimator is $\tilde{m}(x) = \sum y_t w_t(x)$, where $w_t(x) = \frac{K((x_t-x)/b)}{\sum K((x_t-x)/b)}$ in which $K(\cdot)$ is a kernel function and b is the bandwidth. For more detail, see Pagan and Ullah (1999). Minimizing the local constant weighted squared errors $\sum (y_t - m(x))^2 K((x_t - x)/b)$ with respect to $m(x)$ provides the local constant nonparametric kernel estimator $\tilde{m}(x)$. The local linear nonparametric kernel estimators $m^*(x)$ and $\beta^*(x)$ are obtained by minimizing the local linear weighted squared errors $\sum (y_t - m(x) - (x_t - x)\beta(x))^2 K((x_t - x)/b)$ with respect to $m(x)$ and $\beta(x)$.

4.1 Quarterly Data

We use quarterly data for both variables from 1952Q1 to 2016Q3. After taking differences of the raw data set, there are $T = 258$ observations totally. When the risk aversion parameter α takes the value 3, we show the conditional mean of stock return plot in Figure 1. The black curve represents estimated conditional mean of stock return under the ME conditional mean estimator, which shows clear nonlinearity at both left and right tails. We define the average response coefficient $\bar{\beta} = \int_x \beta(x) \tilde{f}(x) dx$ with $\tilde{f}(x)$ in (14) is approximated by $\bar{\beta} \approx \frac{1}{T} \sum_{t=1}^T \beta(x_t)$. Over the range of x , the ME conditional mean estimator has an average response coefficient $\bar{\beta}_{ME} = 0.1554$. In this case where $\alpha = 3$, if consumption growth increases by 1%, on average stock return goes up by 0.1554% under the ME model. The estimated Lagrangian multiplier of the theoretical constraint $\hat{\lambda}_g$ is -0.0013 . The slope of maximum entropy conditional mean curve changes according to x , the consumption growth rate. It shows that the marginal effect of consumption growth on stock return is varying rather than being fixed as in the linear regression. In Figure 1, the red line

represents the estimated stock return based on linear model (LN), which has the regression result $\hat{y} = 0.1142 + 0.1414x$, where the slope is $\hat{\beta}_{LN} = 0.1414$. Comparing the ME regression with the linear regression, the ME mean function has a larger average marginal effect of consumption growth on stock return than the linear model.

RMSE from conditional stock return estimations under maximum entropy (black line) and initial (red line) densities are $\text{RMSE}_{ME} = 0.9837$ and $\text{RMSE}_{LN} = 0.9880$, respectively. As a comparison, the NP kernel methods have RMSE under LCLS and LLLS are 0.9857 (with the average response coefficient $\bar{\beta}_{LCLS} = 0.0868$) and 0.9848 (with the average response coefficient $\bar{\beta}_{LLLS} = 0.1829$), respectively.

To examine the significance of $\hat{\lambda}_g$, i.e. the importance of the theoretical constraint, we expand choice of the risk aversion parameter α according to a range of 10 values from 0.1 to 4.6 considered in Giacomini and Ragusa (2014). In Table 2, we report the RMSE of the conditional mean stock return estimation under the ME density with various values of the risk aversion parameters α , the average response coefficient $\bar{\beta}$, as well as the corresponding $\hat{\lambda}_g$. The RMSE of conditional mean stock return estimation under the ME density does not change a lot under different risk aversion levels α since the Lagrangian multiplier λ_g also changes with α when implementing the Newton method to estimate the Lagrangian multiplier. As shown in Table 2, the estimation efficiency is improved over the linear model ($\text{RMSE}_{LN} = 0.9880$) regardless of the choice of α , which indicates that nonlinearity arising from imposing the theoretical constraint provides extra information for the stock return mean function for all values of α .

4.2 Monthly Data

Following the analysis based on quarterly data, we want to check whether increasing data frequency may strengthen the importance of incorporating the theoretical constraint. We use monthly data from 1959JAN to 2016DEC. After differencing the raw data set, there are 695 observations totally. We conduct the in-sample estimation based on all the $T = 695$ observations. When the risk aversion parameter α is 3, we display the conditional mean regression functions of stock return in Figure 2. The black curve represents the estimated ME conditional mean of stock return. Over the range

of x , it has an average response coefficient $\bar{\beta}_{ME} = 0.1137$. When $\alpha = 3$, if consumption growth increases by 1%, stock return goes up by 0.1137% on average under maximum entropy estimation. The estimated Lagrangian multiplier of the theoretical constraint $\hat{\lambda}_g$ is -0.0013 .

In Figure 2, the red line represents the estimated stock return based on the linear model, which is characterized by $\hat{y} = 0.0897 + 0.1311x$ with slope $\hat{\beta}_{LN} = 0.1311$. The slope of the ME conditional mean curve changes with the consumption growth while the slope is constant in the linear regression. On average, consumption growth has a similar marginal effect on stock return under the ME estimation compared to that in the linear regression when using monthly data. The RMSE from conditional mean stock return estimations under target (black line) and proposed (red line) densities are $RMSE_{ME} = 0.9904$ and $RMSE_{LN} = 0.9930$, respectively. RMSE from conditional stock return estimation under LCLS and LLLS are 0.9913 (with $\bar{\beta}_{LCLS} = 0.0652$) and 0.9907 (with $\bar{\beta}_{LLLS} = 0.1311$) for a comparison.

Similarly to what we have examined with the quarterly data, we check the contribution of the theoretical constraint at different risk aversion levels α . In Table 3, we have reported RMSE, $\bar{\beta}$, and $\hat{\lambda}_g$ when α is changing from 0.1 to 4.6. Our ME model outperforms the linear model at all risk aversion levels in terms of estimation efficiency in RMSE.

5 Predicting Stock Returns: ME Mean Regression Function

In addition to the in-sample analysis of the relationship between stock return and consumption growth rate, we study the out-of-sample predictability of stock return given consumption growth rate as a predictor. We report the prediction results under the ME conditional mean estimator, simple linear regression, and nonparametric kernel methods. Improvement in the prediction by the ME model over the linear model shows the relevance of theoretical constraint in prediction of stock return using consumption growth rate.

Suppose T is the total sample size. We divide the entire sample into two periods. We use R to denote the number of regression periods and P to denote the number of prediction periods. The forecast horizon is denoted as s . The number of prediction periods is thus $T - R - s$. We use the

root mean squared forecast error (RMSFE) as the evaluating criterion

$$\text{RMSFE}_\xi = \left(\frac{1}{T - R - s} \sum_{t=R+s+1}^T (y_t - \hat{y}_t^\xi)^2 \right)^{1/2}, \quad (28)$$

where $\xi = \text{LN, ME, or NP}$. LN refers to the linear prediction model. Under the ME distributions, $y_{t+s} = m(x_t) + u_{t+s}$, where u_{t+s} is the error term at time $t + s$. \hat{y}_t^{ME} is computed as $\hat{m}(x_t) = m(x_t, \hat{\lambda}, \alpha)$ according to (21). NP refers to the nonparametric kernel method. In the out-of-sample prediction, we report only LCLS. Examination of the predictability of stock return is conducted by using the same quarterly and monthly data.

5.1 Quarterly Data

The quarterly data has the sample size $T = 258$. We consider three choices of regression periods with $R \in \{80, 120, 160\}$, which correspond to 20, 30, and 40 years. We consider two forecast horizons $s \in \{1, 4\}$, for one quarter and one year ahead forecasts. The results are presented in Table 4. Prediction results under the choice of risk aversion parameter $\alpha = 3$, which is often used in literature, are reported on the top row. In the following rows, we report the prediction results when risk aversion parameter α is changing between 0.1 to 4.6. As a comparison, out-of-sample RMSFE based on simple linear models (LN) and nonparametric kernel methods (NP) are shown in the last two rows of Table 4. All the numbers shown in Table 4 are RMSFE values. According to Table 4, forecast accuracy improves under the ME model by incorporating the theoretical constraint compared to the linear model for different values of α , even though the improvement is not large. This result is robust for all values of R and s .

We report p -values of the Giacomini-White (GW 2006) test for the conditional predictive ability of the ME and NP models in comparison to the linear model (LN) using quarterly data in Table 5. The GW statistics are constructed with the test function using a constant term plus one lagged loss-difference. The results show that most of the GW p -values appears to be systematic yet not very significant, which indicates a weak gain in predictive ability of the ME model from incorporating the CCAPM theoretical constraint.

5.2 Monthly Data

The monthly data has the sample size $T = 695$. We consider $R \in \{240, 360, 480\}$, corresponding to 20, 30, and 40 years. We consider four forecast horizons $s \in \{1, 3, 6, 12\}$. We report the results in Table 6. For a comparison, RMSFE are also reported for LN and NP in the last two rows. All the numbers shown in Table 6 are RMSFE values, which show the forecast accuracy improvement based on the ME density. However, the gain from incorporating the theoretical constraint is relatively small. The gain becomes somewhat weaker when the data frequency increases from quarterly to monthly. In Table 7, we report the GW p -values using monthly data to compare predictive ability of the ME models with the linear model. The p -values become less significant when the data frequency increases to monthly from quarterly. Incorporating the theoretical constraint becomes less helpful with a higher data frequency in predicting stock returns.

6 Estimating Stock Volatility: ME Variance Regression Function

Similarly to Equation (16), we compute the conditional second moment of stock return by

$$\begin{aligned}
 h(x, \boldsymbol{\lambda}, \alpha) &\equiv E(y^2|x) = \frac{1}{\tilde{f}(x)} \int_y y^2 \tilde{f}(y, x) dy \\
 &= \frac{\int_y y^2 \exp\{-[\lambda_{20}y^2 + \lambda_{10}(x, \alpha)y]\} dy}{\int_y \exp\{-[\lambda_{20}y^2 + \lambda_{10}(x, \alpha)y]\} dy} \\
 &= \frac{F_2(x, \boldsymbol{\lambda}, \alpha)}{F_0(x, \boldsymbol{\lambda}, \alpha)} = \frac{F_2(x)}{F_0(x)}.
 \end{aligned} \tag{29}$$

The conditional volatility of stock return y given consumption growth x can be obtain by

$$V(x, \boldsymbol{\lambda}, \alpha) \equiv h(x, \boldsymbol{\lambda}, \alpha) - (m(x, \boldsymbol{\lambda}, \alpha))^2. \tag{30}$$

According to Equation (20), the conditional volatility function is represented by

$$V(x, \boldsymbol{\lambda}, \alpha) = \frac{F_2(x)}{F_0(x)} - \left(\frac{F_1(x)}{F_0(x)}\right)^2. \tag{31}$$

To examine the effect of the consumption growth on stock return volatility, we use quarterly data for both variables from 1952Q1 to 2016Q3 with 258 observations in total. When the risk aversion parameter α takes the value 3, we report the ME variance regression function of stock return conditional on the consumption growth in Figure 3, over the range of x from -2.5 to 2.5 . According to Figure 3, the conditional stock return volatility is a decreasing function in the consumption growth. It indicates that the consumption growth has an effect on reducing stock return volatility. We approximate the average slope coefficient $\bar{\gamma} \approx \frac{1}{T} \sum_{t=1}^T \frac{dV(x_t)}{dx_t}$, where $\frac{dV(x_t)}{dx_t}$ is computed based on Equation (31) as follows

$$\frac{dV(x)}{dx} = \frac{F_2'(x) F_0(x) - F_2(x) F_0'(x)}{F_0^2(x)} - 2 \frac{F_1(x) F_1'(x) F_0(x) - F_1(x) F_0'(x)}{F_0^2(x)}. \quad (32)$$

Over the range of x , the ME conditional volatility estimator has an average slope coefficient $\bar{\gamma} = -0.0227$. Our empirical result helps explaining the counter-cyclical variation of stock market volatility with a consumption-based asset pricing model as presented in Campbell and Cochrane (1999).

We have also examined the effect of the consumption growth rate on stock return volatility using monthly data, which we do not report for space. We find the monthly effect to be smaller as the frequency of data gets higher because the monthly consumption growth is smaller and noisier than the quarterly consumption growth.

7 Conclusions

In this paper, we estimate the conditional moment functions of asset returns through an information-theoretic maximum entropy method, which incorporates the consumption Euler equation from the consumption-based capital asset pricing model into the maximum entropy density estimation. We have demonstrated the simplicity and efficiency of the ME estimator of the mean and variance function of the asset returns. We find that incorporating the CCAPM theoretical constraint improves the mean prediction, although this improvement is small or mostly at tails. We also find the

ME volatility regression suggests that there exists a negative relationship between the consumption growth and the stock return volatility.

In order to further examine the predictability of asset returns using the same ME method, one may consider more than just one theoretical constraint from the CCAPM as done in this paper. Alternative theoretical models to the consumption-based asset pricing model, such the dividend-based or the production-based asset pricing models may be investigated, which we leave for other work.

8 Appendix

8.1 4th-Order Taylor Expansion of Theoretical Constraint

In this subsection, we show the mathematical derivation of the 4th-order Taylor expansion of consumption Euler equation in (10) to obtain (11). After we drop the subscripts of X and Y , we define $G(Y, X) \equiv \beta X^{-\alpha} Y - 1$. According to the Taylor Theorem, $G(Y, X)$ is approximated as

$$\begin{aligned}
G(Y, X) \approx & G(Y_0, X_0) + \frac{\partial G}{\partial Y} \cdot (Y - Y_0) + \frac{\partial G}{\partial X} \cdot (X - X_0) + \frac{\partial^2 G}{\partial Y^2} \cdot \frac{(Y - Y_0)^2}{2} \\
& + \frac{\partial^2 G}{\partial X^2} \cdot \frac{(X - X_0)^2}{2} + \frac{\partial^2 G}{\partial Y \partial X} \cdot (Y - Y_0) \cdot (X - X_0) \\
& + \frac{\partial^3 G}{\partial Y^3} \cdot \frac{(Y - Y_0)^3}{3!} + \frac{\partial^3 G}{\partial Y^2 \partial X} \cdot \frac{(Y - Y_0)^2}{2} \cdot (X - X_0) \\
& + \frac{\partial^3 G}{\partial Y \partial X^2} \cdot (Y - Y_0) \cdot \frac{(X - X_0)^2}{2} + \frac{\partial^3 G}{\partial X^3} \cdot \frac{(X - X_0)^3}{3!} \\
& + \frac{\partial^4 G}{\partial Y^4} \cdot \frac{(Y - Y_0)^4}{4!} + \frac{\partial^4 G}{\partial Y^3 \partial X} \cdot \frac{(Y - Y_0)^3}{3!} \cdot (X - X_0) \\
& + \frac{\partial^4 G}{\partial Y^2 \partial X^2} \cdot \frac{(Y - Y_0)^2}{2} \cdot \frac{(X - X_0)^2}{2} \\
& + \frac{\partial^4 G}{\partial Y \partial X^3} \cdot (Y - Y_0) \cdot \frac{(X - X_0)^3}{3!} + \frac{\partial^4 G}{\partial X^4} \cdot \frac{(X - X_0)^4}{4!},
\end{aligned}$$

under the 4th-order expansion. The partial derivatives are computed as

$$\begin{aligned}
\frac{\partial G}{\partial Y} &= \beta X^{-\alpha}, \\
\frac{\partial G}{\partial X} &= -\beta \alpha X^{-\alpha-1} Y, \\
\frac{\partial^2 G}{\partial Y^2} &= 0, \\
\frac{\partial^2 G}{\partial X^2} &= \beta \alpha (\alpha + 1) X^{-\alpha-2} Y, \\
\frac{\partial^2 G}{\partial Y \partial X} &= -\beta \alpha X^{-\alpha-1}, \\
\frac{\partial^3 G}{\partial Y^3} &= \frac{\partial^3 G}{\partial Y^2 \partial X} = 0, \\
\frac{\partial^3 G}{\partial Y \partial X^2} &= \beta \alpha (\alpha + 1) X^{-\alpha-2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 G}{\partial X^3} &= -\beta\alpha(\alpha+1)(\alpha+2)X^{-\alpha-3}Y, \\
\frac{\partial^4 G}{\partial Y^4} &= \frac{\partial^4 G}{\partial Y^3 \partial X} = \frac{\partial^4 G}{\partial Y^2 \partial X^2} = 0, \\
\frac{\partial^4 G}{\partial Y \partial X^3} &= -\beta\alpha(\alpha+1)(\alpha+2)X^{-\alpha-3}, \\
\frac{\partial^4 G}{\partial X^4} &= \beta\alpha(\alpha+1)(\alpha+2)(\alpha+3)X^{-\alpha-4}Y.
\end{aligned}$$

All the above partial derivatives are evaluated at (X_0, Y_0) . After rearranging all the terms, (11) is obtained.

8.2 Recursive Integration

In this subsection, we explain the mathematical details of the recursive integration method in subsection 2.3 with indefinite range. When the integral range of y is definite from a to b , the procedure to compute $m(x)$ and $\beta(x)$ is similar.

When the range for y is from $-\infty$ to $+\infty$, define the following integrals as functions of x .

$$F_r(x) \equiv \int_{-\infty}^{+\infty} y^r \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y]dy,$$

where $r = 0, 1, 2, \dots$

$$\begin{aligned}
F_0(x) &\equiv \int_{-\infty}^{+\infty} \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y]dy, \\
F_1(x) &\equiv \int_{-\infty}^{+\infty} y \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y]dy, \\
F_2(x) &\equiv \int_{-\infty}^{+\infty} y^2 \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y]dy.
\end{aligned}$$

Firstly, assuming $\lambda_{20} > 0$,

$$\begin{aligned}
0 &= \int_{-\infty}^{+\infty} d \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] \\
&= \int_{-\infty}^{+\infty} (-2\lambda_{20}y - \lambda_{10}(x))e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} dy \\
&= -2\lambda_{20}F_1(x) - \lambda_{10}(x)F_0(x).
\end{aligned}$$

Secondly,

$$\begin{aligned}
F_0(x) &= \int_{-\infty}^{+\infty} \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\
&= ye^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} y de^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} \\
&= \int_{-\infty}^{+\infty} (2\lambda_{20}y^2 + \lambda_{10}(x)y)e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} dy \\
&= 2\lambda_{20}F_2(x) + \lambda_{10}(x)F_1(x).
\end{aligned}$$

Thirdly,

$$\begin{aligned}
F_1(x) &= \int_{-\infty}^{+\infty} y \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\
&= \int_{-\infty}^{+\infty} \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] d\left(\frac{1}{2}y^2\right) \\
&= \frac{1}{2}y^2 e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{1}{2}y^2 de^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} \\
&= \int_{-\infty}^{+\infty} \frac{1}{2}y^2 (2\lambda_{20}y + \lambda_{10}(x))e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} dy \\
&= \lambda_{20}F_3(x) + \frac{1}{2}\lambda_{10}(x)F_2(x).
\end{aligned}$$

Thus,

$$F_3(x) = \frac{1}{\lambda_{20}}F_1(x) - \frac{\lambda_{10}(x)}{2\lambda_{20}}F_2(x). \tag{33}$$

Define

$$F'_0(x) \equiv \frac{dF_0(x)}{dx}, \quad F'_1(x) \equiv \frac{dF_1(x)}{dx}, \quad F'_2(x) \equiv \frac{dF_2(x)}{dx}, \quad \text{and} \quad \lambda'_{10}(x) \equiv \frac{d\lambda_{10}(x)}{dx}.$$

Firstly, solve for $F'_0(x)$

$$\begin{aligned} F'_0(x) &\equiv \frac{d}{dx} \int_{-\infty}^{+\infty} \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\ &= \int_{-\infty}^{+\infty} \frac{d}{dx} \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\ &= - \int_{-\infty}^{+\infty} \lambda'_{10}(x) y e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} dy \\ &= -\lambda'_{10}(x) F_1(x). \end{aligned}$$

Secondly, solve for $F'_1(x)$

$$\begin{aligned} F'_1(x) &\equiv \frac{d}{dx} \int_{-\infty}^{+\infty} y \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\ &= \int_{-\infty}^{+\infty} \frac{d}{dx} y \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\ &= - \int_{-\infty}^{+\infty} \lambda'_{10}(x) y^2 e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} dy \\ &= -\lambda'_{10}(x) F_2(x). \end{aligned}$$

Thirdly, solve for $F'_2(x)$

$$\begin{aligned} F'_2(x) &\equiv \frac{d}{dx} \int_{-\infty}^{+\infty} y^2 \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\ &= \int_{-\infty}^{+\infty} \frac{d}{dx} y^2 \exp -[\lambda_{20}y^2 + \lambda_{10}(x)y] dy \\ &= - \int_{-\infty}^{+\infty} \lambda'_{10}(x) y^3 e^{-[\lambda_{20}y^2 + \lambda_{10}(x)y]} dy \\ &= -\lambda'_{10}(x) F_3(x). \end{aligned}$$

Substitute $F_3(x)$ for (33) to obtain

$$F_2'(x) = \frac{\lambda_{10}'(x)\lambda_{10}(x)}{2\lambda_{20}}F_2(x) - \frac{\lambda_{10}'(x)}{\lambda_{20}}F_1(x).$$

The expressions for $F_0'(x)$, $F_1'(x)$, and $F_2'(x)$ can be written in a linear system,

$$\begin{aligned} F_0'(x) &= \Lambda_{01}(x)F_1(x), \\ F_1'(x) &= \Lambda_{12}(x)F_2(x), \\ F_2'(x) &= \Lambda_{21}(x)F_1(x) + \Lambda_{22}(x)F_2(x), \end{aligned}$$

where Λ 's denote the corresponding coefficients.

Since

$$\begin{aligned} F_0(x_0 + h) &\approx F_0(x_0) + F_0'(x)h, \\ F_1(x_0 + h) &\approx F_1(x_0) + F_1'(x)h, \\ F_2(x_0 + h) &\approx F_2(x_0) + F_2'(x)h, \end{aligned}$$

for a given initial value x_0 and a small increment h , the functions of x , $F_0(x)$, $F_1(x)$ and $F_2(x)$ can be traced out. Given that

$$\begin{aligned} m(x) &= \frac{F_1(x)}{F_0(x)}, \\ \beta(x) &= \frac{F_1'(x)F_0(x) - F_1(x)F_0'(x)}{F_0^2(x)}, \end{aligned}$$

the ME conditional mean function and its response function can be traced out as well.

References

- Bera, A.K. and Biliyas, Y. (2002). “The MM, ME, ML, EL, EF and GMM Approaches to Estimation: A Synthesis.” *Journal of Econometrics* 107: 51-86.
- Campbell, J.Y. and Cochrane, J.H. (1999). “By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior.” *Journal of Political Economy* 107: 205-251.
- Chib, S. and Greenberg, E. (1995). “Understanding the Metropolis-Hastings Algorithm.” *The American Statistician* 49: 327-335.
- Cochrane, J. (1991). “Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations.” *The Journal of Finance* 46: 209-237.
- Cochrane, J. (2000). *Asset Pricing*. Princeton University Press.
- Giacomini, R. and Ragusa, G. (2014). “Theory-Coherent Forecasting.” *Journal of Econometrics* 182: 145-155.
- Giacomini, R. and White, H. (2006). “Tests of Conditional Predictive Ability.” *Econometrica* 74: 1545-1578.
- Golan, A., Judge, G. and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley & Sons.
- Golan, A. (2017). *Foundations of Info-Metrics*. Oxford University Press, New York.
- Hahn, J., Kuersteiner, G. and Mazzocco, M. (2019). “Estimation with Aggregate Shocks.” *The Review of Economic Studies* 87(3): 1365-1398.
- Hansen, L.P. and Singleton, K.J. (1983). “Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns.” *Journal of Political Economy* 91: 249-265.
- Imbens, G.W., Spady, R.H. and Johnson, P. (1998). “Information Theoretic Approaches to Inference in Moment Condition Models.” *Econometrica* 66: 333-357.
- Kitamura, Y. and Stutzer, M. (1997). “An Information-Theoretic Alternative to Generalized Method of Moments Estimation.” *Econometrica* 65: 861-874.
- Kitamura, Y., Tripathi, G. and Ahn, H. (2004). “Empirical Likelihood-Based Inference in Conditional Moment Restriction Models.” *Econometrica* 72: 1667-1714.
- Komunjer, I. and Ragusa, G. (2016). “Existence and Characterization of Conditional Density Projections.” *Econometric Theory* 32: 947-987.

- Kullback, S. and Leibler, R.A. (1951). "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22: 79-86.
- Mao, Y. and Ullah, A. (2020). "Information Theoretic Estimation of Econometric Functions." Forthcoming in *Advances in Info-Metrics: Information and Information Processing across Disciplines* (eds. Chen, M., Dunn, J.M., Golan, A., and Ullah, A.). Oxford University Press.
- Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.
- Robertson, J.C., Tallman, E.W. and Whiteman, C.H. (2005). "Forecasting Using Relative Entropy." *Journal of Money, Credit and Banking* 37: 383-401.
- Ullah, A. (1996). "Entropy, Divergence and Distance Measures with Econometric Applications." *Journal of Statistical Planning and Inference* 49: 137-162.
- Wu, X. (2003). "Calculation of Maximum Entropy Densities with Application to Income Distribution." *Journal of Econometrics* 115: 347-354.
- Wu, X. (2010). "Exponential Series Estimator of Multivariate Densities." *Journal of Econometrics* 156: 354-366.

Table 1. Monte Carlo: Estimation of the LN and ME Mean Regression Functions

$\sigma_x = 0.1$										
ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
RMSE _{LN}	1.0026	1.0004	0.9818	0.9515	0.9191	0.8694	0.8013	0.7214	0.5881	0.4309
RMSE _{ME}	0.9933	0.9913	0.9744	0.9434	0.9120	0.8636	0.7997	0.7136	0.5839	0.4280
$\sigma_x = 0.2$										
ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
RMSE _{LN}	1.0097	1.0066	0.9897	0.9629	0.9253	0.8962	0.8178	0.7354	0.6322	0.4326
RMSE _{ME}	0.9958	0.9917	0.9761	0.9569	0.9244	0.8825	0.8148	0.7300	0.6295	0.4320

Notes: Reported are the Monte Carlo average of RMSEs for the linear (LN) and the maximum entropy (ME) mean regression functions over 100 replications, with the sample size $T = 500$, $\sigma_x \in \{0.1, 0.2\}$, and the correlation coefficient $\rho \in \{0, \dots, 0.9\}$.

Figure 1: Conditional Mean Regression Functions of Quarterly Stock Return

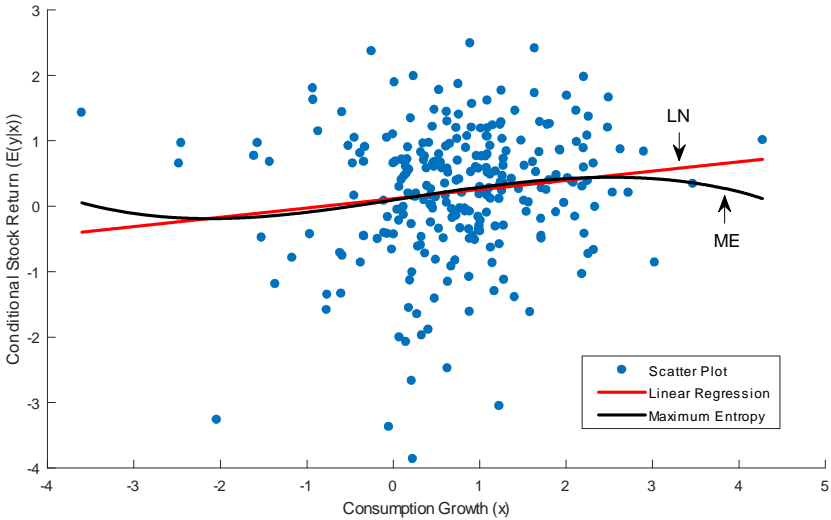


Table 2. Conditional Mean Regression Functions of Quarterly Stock Return

α	0.1	0.6	1.1	1.6	2.1	2.6	3.1	3.6	4.1	4.6
RMSE_{ME}	0.9788	0.9798	0.9809	0.9818	0.9826	0.9833	0.9838	0.9843	0.9847	0.9850
$\bar{\beta}_{ME}$	0.1832	0.1738	0.1676	0.1631	0.1597	0.1571	0.1550	0.1534	0.1520	0.1509
$\hat{\lambda}_g$	-.8885	-.0656	-.0188	-.0076	-.0037	-.0020	-.0012	-.0007	-.0005	-.0003

Notes: The RMSE_{ME} of the maximum entropy (ME) mean regression functions for quarterly data are reported with different values of α . For comparison, the linear (LN) mean regression gives $\text{RMSE}_{LN} = 0.9880$ and $\hat{\beta}_{LN} = 0.1414$.

Figure 2: Conditional Mean Regression Functions of Monthly Stock Return

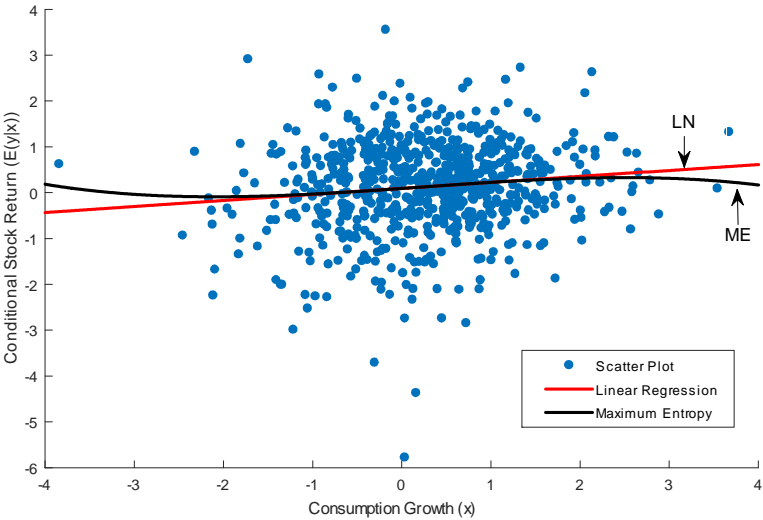


Table 3. Conditional Mean Regression Functions of Monthly Stock Return

α	0.1	0.6	1.1	1.6	2.1	2.6	3.1	3.6	4.1	4.6
RMSE_{ME}	0.9909	0.9907	0.9906	0.9905	0.9905	0.9904	0.9904	0.9904	0.9903	0.9903
$\bar{\beta}_{ME}$	0.1132	0.1144	0.1150	0.1150	0.1147	0.1142	0.1135	0.1128	0.1121	0.1113
$\hat{\lambda}_g$	-.4343	-.0388	-.0129	-.0059	-.0031	-.0018	-.0011	-.0008	-.0006	-.0004

Notes: The RMSE_{ME} of the maximum entropy (ME) mean regression functions for monthly data are reported with different values of α . For comparison, the linear (LN) mean regression gives $\text{RMSE}_{LN} = 0.9930$ and $\hat{\beta}_{LN} = 0.1311$.

Table 4. RMSFE under ME, LN and NP Models (Quarterly)

	$R = 80$		$R = 120$		$R = 160$	
	$s = 1$	$s = 4$	$s = 1$	$s = 4$	$s = 1$	$s = 4$
$\alpha = 3$	1.0528	1.0565	1.0270	1.0268	0.9980	1.0163
$\alpha = 0.1$	1.0498	1.0575	1.0243	1.0244	0.9996	1.0131
$\alpha = 0.6$	1.0507	1.0569	1.0234	1.0246	0.9978	1.0134
$\alpha = 1.1$	1.0513	1.0566	1.0236	1.0250	0.9971	1.0139
$\alpha = 1.6$	1.0519	1.0565	1.0242	1.0255	0.9970	1.0146
$\alpha = 2.1$	1.0522	1.0565	1.0253	1.0260	0.9972	1.0152
$\alpha = 2.6$	1.0526	1.0565	1.0263	1.0265	0.9976	1.0158
$\alpha = 3.1$	1.0528	1.0565	1.0272	1.0269	0.9981	1.0164
$\alpha = 3.6$	1.0530	1.0566	1.0280	1.0272	0.9986	1.0168
$\alpha = 4.1$	1.0532	1.0566	1.0286	1.0276	0.9990	1.0173
$\alpha = 4.6$	1.0533	1.0566	1.0292	1.0278	0.9995	1.0176
LN	1.0568	1.0577	1.0358	1.0318	1.0057	1.0216
NP	1.0645	1.0574	1.0312	1.0283	1.0011	1.0108

Notes: The quarterly data has $T = 258$. We consider the three sample sizes for the estimation period with $R \in \{80, 120, 160\}$, each corresponding to 20, 30, and 40 years. Two forecast horizons $s \in \{1, 4\}$, for one quarter and one year ahead forecasts, are reported.

Table 5. GW Test p -values in Comparison to LN (Quarterly)

	$R = 80$		$R = 120$		$R = 160$	
	$s = 1$	$s = 4$	$s = 1$	$s = 4$	$s = 1$	$s = 4$
$\alpha = 3$	0.1472	0.3397	0.0893	0.0560	0.2508	0.2557
$\alpha = 0.1$	0.1022	0.2993	0.3758	0.0431	0.7345	0.2853
$\alpha = 0.6$	0.0776	0.2963	0.2335	0.0370	0.5555	0.2835
$\alpha = 1.1$	0.0850	0.3062	0.1555	0.0377	0.4235	0.2747
$\alpha = 1.6$	0.1010	0.3174	0.1146	0.0412	0.3367	0.2668
$\alpha = 2.1$	0.1173	0.3252	0.0993	0.0460	0.2868	0.2614
$\alpha = 2.6$	0.1337	0.3335	0.0924	0.0514	0.2612	0.2578
$\alpha = 3.1$	0.1506	0.3412	0.0889	0.0571	0.2491	0.2552
$\alpha = 3.6$	0.1674	0.3480	0.0855	0.0630	0.2444	0.2533
$\alpha = 4.1$	0.1835	0.3541	0.0803	0.0688	0.2434	0.2517
$\alpha = 4.6$	0.1988	0.3595	0.0762	0.0746	0.2443	0.2503
NP	0.5917	0.9483	0.5607	0.8028	0.7083	0.5276

Notes: The quarterly data has $T = 258$. We consider the three sample sizes for the estimation period with $R \in \{80, 120, 160\}$, each corresponding to 20, 30, and 40 years. Two forecast horizons $s \in \{1, 4\}$, for one quarter and one year ahead forecasts, are reported.

Table 6. RMSFE under ME, LN and NP Models (Monthly)

	$R = 240$		$R = 360$		$R = 480$	
	$s = 1$	$s = 3$	$s = 1$	$s = 3$	$s = 1$	$s = 3$
$\alpha = 3$	1.0322	1.0284	0.9837	0.9852	1.0207	1.0232
$\alpha = 0.1$	1.0336	1.0286	0.9840	0.9854	1.0210	1.0235
$\alpha = 0.6$	1.0333	1.0286	0.9839	0.9854	1.0209	1.0235
$\alpha = 1.1$	1.0329	1.0285	0.9838	0.9853	1.0208	1.0235
$\alpha = 1.6$	1.0327	1.0285	0.9838	0.9853	1.0208	1.0234
$\alpha = 2.1$	1.0325	1.0284	0.9837	0.9853	1.0208	1.0233
$\alpha = 2.6$	1.0323	1.0284	0.9837	0.9852	1.0207	1.0233
$\alpha = 3.1$	1.0322	1.0284	0.9837	0.9852	1.0207	1.0232
$\alpha = 3.6$	1.0321	1.0284	0.9837	0.9851	1.0207	1.0232
$\alpha = 4.1$	1.0320	1.0283	0.9836	0.9850	1.0206	1.0231
$\alpha = 4.6$	1.0320	1.0283	0.9836	0.9850	1.0206	1.0231
LN	1.0340	1.0301	0.9850	0.9864	1.0278	1.0220
NP	1.0405	1.0289	0.9947	0.9908	1.0387	1.0393

	$R = 240$		$R = 360$		$R = 480$	
	$s = 6$	$s = 12$	$s = 6$	$s = 12$	$s = 6$	$s = 12$
$\alpha = 3$	1.0345	1.0307	0.9832	0.9890	1.0244	1.0248
$\alpha = 0.1$	1.0355	1.0305	0.9835	0.9854	1.0246	1.0246
$\alpha = 0.6$	1.0354	1.0305	0.9834	0.9855	1.0246	1.0246
$\alpha = 1.1$	1.0351	1.0306	0.9834	0.9855	1.0245	1.0247
$\alpha = 1.6$	1.0349	1.0306	0.9833	0.9855	1.0245	1.0247
$\alpha = 2.1$	1.0347	1.0307	0.9833	0.9856	1.0245	1.0248
$\alpha = 2.6$	1.0346	1.0307	0.9832	0.9856	1.0244	1.0248
$\alpha = 3.1$	1.0345	1.0307	0.9832	0.9856	1.0244	1.0248
$\alpha = 3.6$	1.0344	1.0307	0.9832	0.9857	1.0244	1.0248
$\alpha = 4.1$	1.0344	1.0306	0.9831	0.9857	1.0243	1.0249
$\alpha = 4.6$	1.0343	1.0306	0.9831	0.9857	1.0243	1.0250
LN	1.0355	1.0322	0.9853	0.9901	1.0260	1.0281
NP	1.0347	1.0358	0.9824	1.0009	1.0274	1.0440

Notes: The monthly data has the sample size $T = 695$. We report for the estimation window sizes $R \in \{240, 360, 480\}$, each corresponding to 20, 30, and 40 years. The forecast horizons are $s \in \{1, 3, 6, 12\}$ months.

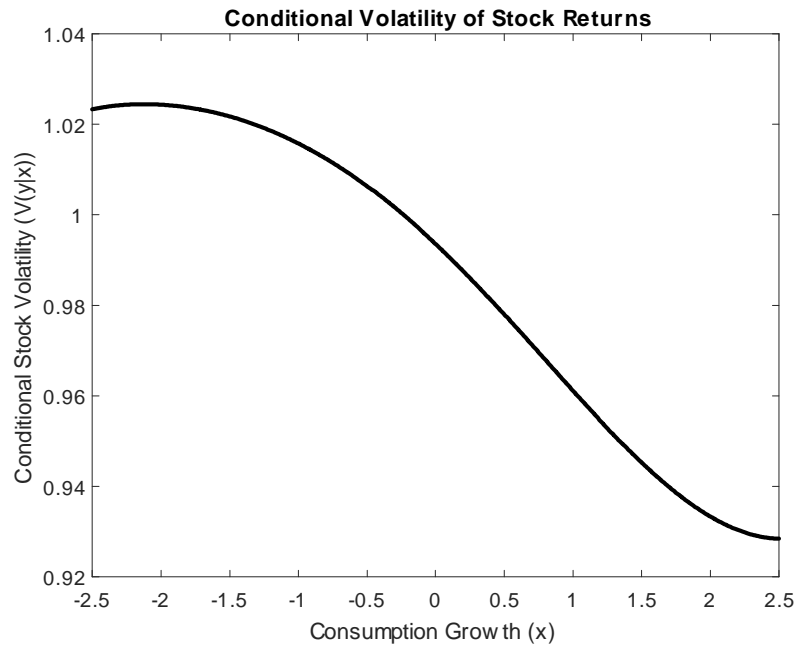
Table 7. GW Test p -values in Comparison to LN (Monthly)

	$R = 240$		$R = 360$		$R = 480$	
	$s = 1$	$s = 3$	$s = 1$	$s = 3$	$s = 1$	$s = 3$
$\alpha = 3$	0.3829	0.8862	0.7445	0.9121	0.3419	0.8397
$\alpha = 0.1$	0.3831	0.8234	0.8103	0.9135	0.3565	0.8374
$\alpha = 0.6$	0.3817	0.9192	0.8111	0.9116	0.3508	0.8397
$\alpha = 1.1$	0.3794	0.9045	0.8028	0.9110	0.3473	0.8411
$\alpha = 1.6$	0.3792	0.8940	0.7893	0.9111	0.3449	0.8417
$\alpha = 2.1$	0.3808	0.8881	0.7734	0.9115	0.3434	0.8415
$\alpha = 2.6$	0.3820	0.8860	0.7570	0.9118	0.3424	0.8407
$\alpha = 3.1$	0.3831	0.8865	0.7415	0.9121	0.3418	0.8394
$\alpha = 3.6$	0.3840	0.8885	0.7275	0.9124	0.3414	0.8379
$\alpha = 4.1$	0.3848	0.8912	0.7151	0.9125	0.3412	0.8361
$\alpha = 4.6$	0.3854	0.8941	0.7044	0.9126	0.3412	0.8342
NP	0.3957	0.7014	0.4985	0.3748	0.5730	0.5145

	$R = 240$		$R = 360$		$R = 480$	
	$s = 6$	$s = 12$	$s = 6$	$s = 12$	$s = 6$	$s = 12$
$\alpha = 3$	0.2555	0.2304	0.5082	0.6022	0.4412	0.4250
$\alpha = 0.1$	0.2561	0.2465	0.5606	0.6130	0.5027	0.4224
$\alpha = 0.6$	0.2529	0.2402	0.5531	0.6117	0.4864	0.4223
$\alpha = 1.1$	0.2520	0.2366	0.5445	0.6101	0.4734	0.4226
$\alpha = 1.6$	0.2523	0.2343	0.5352	0.6082	0.4627	0.4232
$\alpha = 2.1$	0.2532	0.2326	0.5254	0.6061	0.4537	0.4239
$\alpha = 2.6$	0.2544	0.2312	0.5157	0.6040	0.4463	0.4245
$\alpha = 3.1$	0.2558	0.2302	0.5063	0.6018	0.4400	0.4251
$\alpha = 3.6$	0.2571	0.2293	0.4975	0.5996	0.4349	0.4257
$\alpha = 4.1$	0.2584	0.2286	0.4894	0.5974	0.4305	0.4262
$\alpha = 4.6$	0.2596	0.2280	0.4819	0.5953	0.4269	0.4266
NP	0.3845	0.6830	0.2343	0.6142	0.7936	0.4292

Notes: The monthly data has the sample size $T = 695$. We report for the estimation window sizes $R \in \{240, 360, 480\}$, each corresponding to 20, 30, and 40 years. The forecast horizons are $s \in \{1, 3, 6, 12\}$ months.

Figure 3. ME Variance Regression Function



Notes: Reported is the estimated ME variance function of the stock return conditional on the consumption growth, using quarterly data from 1952Q1 to 2016Q3 with $T = 258$ and the risk aversion parameter $\alpha = 3$.