THREE ESSAYS IN MACROECONOMICS AND EMPIRICAL FINANCE

A Dissertation

by

HYOSUNG YEO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ke-Li Xu
Co-Chair of Committee,	Hwagyun Kim
Committee Members,	Tatevik Sekhposyan
	Sarah Zubairy
Head of Department,	Timothy Gronberg

May 2016

Major Subject: Economics

ABSTRACT

We measure macroeconomic uncertainty and study its link to asset returns via a consumption-based model employing recursive preferences. We introduce a stochastic volatility model with two asymptotic regimes and smooth transition. Smooth transition in regimes produces sizable equity premiums for even a small amount of consumption volatility if uncertainty unravels slowly. The relative risk aversion is estimated around two, the estimated elasticity of intertemporal substitution is greater than one, and the simulation suggests that our volatility channel matters in explaining asset returns.

Next, we propose to use the Hodrick-Prescott filter to nonparametrically extract the conditional mean and volatility process of a time series. We find an optimal smoothing parameter for HP filter by minimizing the first order sample correlation of the residuals. The process extracted from our HP filter is therefore defined as the predictable component of the given time series, while the conventional HP filter decomposes a time series into trend and cyclical components. By simulations, we show that our HP filter performs better than the local linear estimator in terms of average mean squared error for both discrete and continuous time models.

Finally, we develop a novel methodology to test for stock return predictability using multiple predictors. It has been reported that the conventional least squares approach has an unacceptable level of size distortions and over-reject the null hypothesis of no predictability. Previous literatures which tried to resolve the Endogeneity problem with a persistent predictor have failed to allow multiple covariates in the predictive regression. We propose to apply Heteroskedasticity and Endogeneity correction sequentially to tackle the issue. Our approach not only makes it possible to

correctly test for the predictability of stock returns by multiple predictors but also reveals the marginal predictive power of each predictor. Using our new test, we find strong evidence for joint return predictability by dividend-price ratio, earnings-price ratio, short-term interest rates and term spread.

DEDICATION

To My Beloved Parents.

ACKNOWLEDGEMENTS

I am indebted to my advisor Prof. Hwagyun Kim for his insightful guidance and patience. I also would like to thank my committee chair, Prof. Ke-Li Xu, and committee members, Prof. Tatevik Sekhposyan and Prof. Sarah Zubairy, for their willingness to help and providing invaluable comments.

Prof. Joon Y. Park at the Indiana University mentored me wholeheartedly throughout my pursuit of a Ph.D. degree. I thank him for his tireless encouragement. Also, I cannot thank enough Prof. Yoosoon Chang at the Indiana University who kindly arranged a trip to College Station to attended my defense. I met Prof. Park and Chang when they were our faculty at Texas A&M University, and I'm very fortunate to have them as my mentor.

Without a doubt, I also owe much to my colleagues Yongok Choi, Minsoo Jeong, Eunhee Lee and Hyoung Il Lee for their valuable discussion and friendship. Especially, I want to thank Yongok Choi and Hyoung-Il Lee generously provided me with their program and shared their invaluable experience. I also would like to thank Dou-young Lee for his help and a warm heart.

Finally, I must express my profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Time-varying Macroeconomic Uncertainty and Implication for Asset Pricing

In Section 2, we measure macroeconomic uncertainty and study its link to asset returns via a consumption-based model employing recursive preferences. Time-varying macroeconomic uncertainty can be an important ingredient for asset valuation. The episode of great recession in 2008 shows that this is indeed a key link between macroeconomic variables and asset markets. When there exists a great deal of uncertainty about how an economy will evolve over time, financial markets demand premiums for bearing this uncertainty. Empirical studies often use volatility to gauge the degree of uncertainty. However, the previous studies using macroeconomic models to explain asset prices did not pay much attention to the macroeconomic volatility channel of generating equity premiums.[1]

The standard macroeconomic asset pricing models identify consumption growth as the link between macroeconomic variables and asset returns via a simple Euler equation. But, the level of aggregate consumption behaves like a random walk, and the sizes of unconditional and conditional volatilities of the consumption growth are fairly modest, which makes measuring macroeconomic uncertainty using the consumption growth volatility difficult. Moreover, the connections between consumption and dividend growths are rather weak unlike the assumption typically made in the literature. Thus, the standard consumption-based models fail to justify a high average equity premium with low and stable interest rates because of the small con-

[1] Recently, there is a burgeoning literature on ambiguity aversion which is closely related to this issue. Epstein and Schneider (2010) is a nice survey on this topic. However, there are very few empirical studies on this and related, the measure of ambiguity is still not well defined. Please see Jeong et al. (2015) for more details.

sumption volatility.[2] Overall, the role of the macroeconomic volatility related to the aggregate consumption and dividend growths appears to be minimal in explaining asset prices under this setting.

In Section 2, we tackle this issue by introducing a novel stochastic volatility model within the consumption-based asset pricing framework. We set the volatility processes for consumption and dividend growths to be driven by bounded, non-linear functions of latent factors consisting of a common and idiosyncratic components. Specifically, the volatility-generating common factor produces macroeconomic volatility through a parametric logistic function. This setup yields volatility processes having two asymptotic levels (the high and low volatility regimes) with smooth transitions between them. We show that the existence of smooth transition across the high and low regimes of macroeconomic volatility produces an additional source of uncertainty about which regimes an economy will end up with in the next period. In this light, we call the common stochastic component "macroeconomic uncertainty".

Epstein and Zin (1989) show that economic agents prefer early resolution of uncertainty if the risk aversion parameter is greater than the reciprocal of the elasticity of intertemporal substitution for the recursive utility function of Kreps and Porteus (1978). Incorporating this utility function, our non-linear volatility setup can create high expected returns even with a modest level of consumption volatility, because economic agents dislike stochastic changes in regimes of volatility, especially if the perceived macroeconomic uncertainty unravels slowly.

For the statistical analysis, we formulate our shock processes as a state space model with stochastic volatility driven by persistent latent factors. Then, the resulting asset pricing model is expressed as the Euler equation with both observable macro

[2] These puzzles related to the consumption based asset pricing models have been one of the main research questions in finance and macroeconomics since Hansen and Singleton (1983) and Mehra and Prescott (1985).

variables and latent factors. To estimate the model, we adopt a Bayesian approach by developing an algorithm to filter the stochastic volatilities and the key preference parameters. We use a Markov-Chain-Monte-Carlo (MCMC) method relying on the Metropolis-Hasting algorithm within Gibbs sampler.

Regarding the stochastic volatility component, there exist several other non-linear filtering techniques in the conventional approach that we may apply to estimate the state space models with non-linear measurement equations, such as the extended non-linear Kalman filter and the density-based filter as explained in Tanizaki (1996). However, the extended Kalman filter is not working well in the presence of persistent state variables, since the Kalman gain may vanish in case the latent factors diverge without bounds. Kim et al. (2009) to develop a density-based filter for the estimation of highly persistent stochastic volatility models. But, it is extremely difficult and costly, if not impossible, to implement the density-based filter for large dimensional models like ours. The MCMC method we use to estimate our Bayesian model is significantly less prone to the problem of multi-dimensionality, because it utilizes a uni-variate conditional density function in every step. Furthermore, it appears that the Bayesian approach in general yields more stable estimates than the conventional approach, especially in the presence of persistent latent factors.

With a sample of stochastic volatility and observables, we can draw a sample of the preference parameters. Chernozhukov and Hong (2003) show that drawing samples of parameters from a conditional posterior distribution made out of the moment condition is asymptotically equivalent to obtaining a sample from the asymptotic distribution of the GMM estimator under mild conditions. Gallant et al. (2014) show that the above procedure is valid with latent factors included in the moment conditions. We exploit these theoretical results to develop our Bayesian econometric method to produce a Markov chain. This approach is useful to econometrically evaluate complex asset pricing models including highly non-linear latent factors.

With the goal of evaluating the empirical performance of the proposed asset pricing model, we extract a common factor and idiosyncratic volatility factors from the consumption and dividend growth rates. We find that the common factor delineating the macroeconomic uncertainty predicts post-war business cycle recessions quite well. In addition, our common volatility factor captures the great moderation periods beginning around 1984. Fama and French (1989) and many other studies report that the expected returns on stocks and long-term bonds have counter-cyclical variations. Thus, we believe that the close link between the estimated macro uncertainty and business cycle can shed light on the dynamic behaviors of asset prices. According to our results, the coefficient of relative risk aversion is estimated around two, and the estimated coefficient of the intertemporal elasticity of substitution is greater than one. This implies that the representative investor prefers early resolution of uncertainty and requests a premium for bearing macroeconomic uncertainty that moves slowly over time. Simulation results show that the model matches the first and the second moments of the stock returns and the risk-free rate hence, it explains the equity premium puzzle, the risk-free rate puzzle, and the volatility puzzle addressed in the literature.

Section 2 is related to at least three strands of literature. First and most directly related is the long-run risks asset price model. Bansal and Yaron (2004), and Hansen et al. (2008) set consumption and dividend growth processes to contain a small, but persistent process in their means, and show that they can explain many stylized facts in asset market. In their papers they emphasize the long-run risks channel, which is based on the common portion of the conditional expectations that vary slowly over time. Due to the long-run risks channel together with the preferences on the early resolution of uncertainty, these models can generate a sufficiently high risk premium. Bansal and Yaron (2004) also added a stochastic volatility term to their model, but its role is mostly confined to explaining time-variability of risk premium. In the recent empirical paper by Bansal et al. (2007), the estimates of risk aversion and the elasticity of intertemporal substitution are around 10-15, and 0.5-1.5 respectively. Their results including the simulations vary little regardless of the existence of stochastic volatility. Thus, the role of stochastic volatility is restricted in their models. Moreover, correlations between the consumption and dividend growth rates are known to be small, which we confirm in our empirical study. On the contrary, we focus on a realistic and flexible volatility setup to see how the preferences on the timing of uncertainty resolution can produce a channel of volatility premium. Our empirical results show that the time-varying macroeconomic uncertainty matters to explain asset prices.

Another important class of asset pricing models uses the habit formation preferences. For instance, Campbell and Cochrane (1999) can generate a high equity premium with low and stable interest rates via time-varying risk aversion. A major difference from our approach is that the habit formation emphasizes the near unit root behavior of consumption and assumes the constant conditional volatility for consumption growth. To produce time-varying expected returns, instead, relative risk aversion is changing over business conditions.[3]

In addition, many papers consider relaxing other assumptions of the standard consumption-based model to address various asset market behaviors. For example, Constantinides et al. (2002) report that the existence of limited participation in the asset market increases the equity premium. Our model can be extended to incorporate this feature without a major modification, but the availability of consumption

[3] Given that our empirical result shows existence of non-trivial time-varying stochastic volatilities from consumption, it would be interesting to study the implications of our stochastic volatilities under the habit formation. We leave this for future research.

data for asset market participants is somewhat limited.

Another related literature would be recent works on rare events. Barro (2006) and Rietz (1988) report that the possibility of a rare disaster or event can make equity riskier. This approach yields a fat-tail of asset returns by assigning small but positive probabilities to extremely rare events. Our model can also generate fat-tails due to the presence of persistent stochastic volatilities.[4] In comparison, Section 2 differs from this literature in that our model focuses on small but persistent volatility changes, while still allowing shifts between two extreme regimes. Related, Eraker and Shaliastovich (2008) theoretically study asset pricing implications of Epstein-Zin preferences with an affine jump diffusion. They emphasize the role of jumps to generate a volatility channel, similar to the rare events. Analyzing the variance premium using jumps or rare disasters is another related area, studied in Bollerslev and Todorov (2011) and Eraker (2008).

1.2 Estimating the Conditional Mean Process Using the Hodrick-Prescott Filter

In Section 3, we consider to use the Hodrick-Prescott (HP) filter to estimate the conditional mean process of a time series. We propose the use of a modified version of the Hodrick-Prescott filter as an alternative way of estimating conditional mean process nonparametrically. The HP filter in Economics was initially introduced by Hodrick and Prescott (1980) and Hodrick and Prescott (1997). It conveniently decomposes a time-series variable into trends and cycles and has been widely used as a practical tool to study business cycles. Here we are suggesting to use the HP approach for a new objective to filter out the conditional mean process, which is the predictable part of the observed time series. For a given filtration, a time series

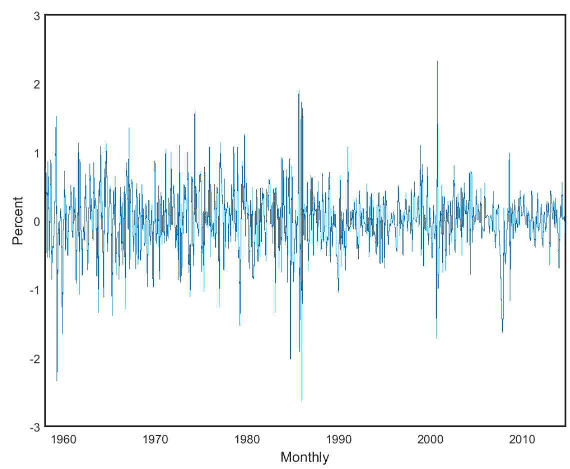
[4] The reader is referred to Park and Phillips (2001) and Park (2002) for the related asymptotic results and more detailed explanations on how the presence of persistent stochastic volatilities can generate fat-tails.

can be decomposed into its conditional mean that includes both trend and cyclical component, and a sequence of martingale difference unpredictable error component. The key identifying characteristics of a martingale difference sequence is that it has no serial correlation. Thus, if we apply HP procedure with an optimal smoothing parameter that produces the fitted residuals with the first-order sample auto-correlation zero, then the fitted process will capture all of the predictable components of the time series given the past information, and therefore may successfully filter out the conditional mean process nonparametrically, leaving only the martingale difference residuals.

Our HP estimator of the conditional mean process is therefore defined simply as the predictable component of (y_t) extracted using the HP filter with the penalty parameter minimizing the first order sample correlation of the residuals. We also show how we may use our approach of using the HP filter with such an optimal smoothing parameter to estimate conditional variance process of the given time series. Our HP approach can also be used to estimate the conditional mean and variance processes in continuous time model. This approach is the first attempt to applying HP filter to estimating conditional mean and variance processes.

For nonparametric estimation of the conditional mean of a time series, the Local Linear estimator with a cross-validation bandwidth smoothing has been commonly used as a standard method. It is well known that such data-dependent cross-validation method yields an asymptotically optimal bandwidth and the resulting Local Linear estimator performs well in finite samples for a wide class of models. See Li and Racine (2004) among others for more discussions. However, it has also been reported that in some important cases the Local Linear estimator produces a less than desirable fit for the conditional mean process. For example, Jeong et al. (2015) estimate the time varying volatility process of the monthly US consumption over the period 1960-2006 using the Local Linear estimator with the bandwidth chosen by the data-dependent cross-validation method, and report overly smoothed consumption volatility process. It is shown clearly in Figure 1.2 which shows the time series of the monthly consumption growth rates on the top panel, the time series of the realized consumption volatility and the Local Linear estimator for this in the middle panel.

Figure 1.1: Monthly Consumption Growth Rates

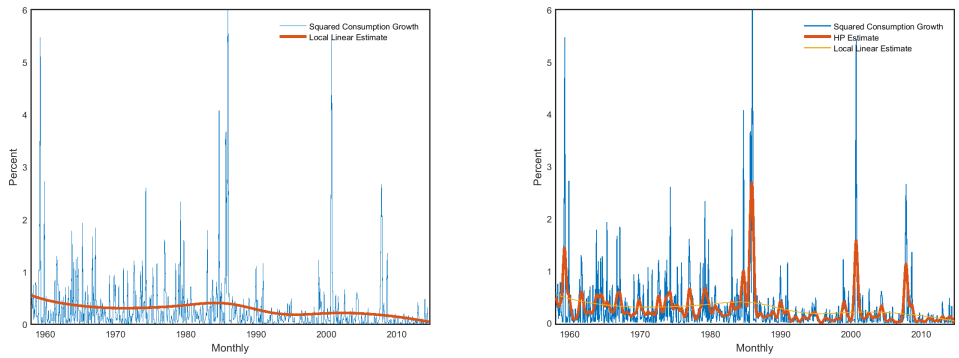


Notes: Figure 1.1 displays a time series plot of monthly consumption growth rate between February 1959 and December 2013. Aggregate consumption is defined as the sum of personal consumption expenditure for the non-durable goods(PCEND) and services(PCES). Our data is obtained from the Federal Reserve Banks of St. Louis (FRED) (http://research.stlouisfed.org).

Precisely estimating the time-varying consumption volatility is crucial for explaining risk premium for equity investment. As is well known, a simple consumption-based Capital Asset Pricing Model(CAPM) yields

$$\mathbb{E}_t(R^m_{t+1} - R^f_{t+1}) = \gamma \, Cov\left(R^m_{t+1}, \frac{C_{t+1}}{C_t}\right),$$

Figure 1.2: Fitted Monthly Consumption Volatility



(a) Local Linear Estimator (b) HP Estimator

Notes: Figure 1.2 compares the fitted consumption volatilities using the Local Linear estimator and our proposed HP estimator. In Panel (a), the thin solid line represents squared consumption growth rates after it has been de-trended. The solid thick line represents the fitted trends using the Local Linear method. We estimate the consumption volatility nonparametrically as a function of time. In Panel (b), the solid thick line represents the estimated consumption volatility using the HP filter.

where (R^m_{t+1}) is aggregate market return and (R^f_{t+1}) is risk-free rate, and the parameter γ measures investor's constant relative risk aversion. The observed value of the standard deviation of annual consumption growth is 1.2 % during the period 1960-2006, while that of excess market return is 14%. To match with the observed 6–7% of excess market return, investors' risk-aversion parameter γ should be over 200, which is hard to rationalize, and this produces the well known risk-premium puzzle.

Much effort has been devoted to better capture the time-varying nature of the consumption volatility, but estimating consumption volatility remains a challenging task since the annualized percentage change in aggregate consumption growth is very small. This is indeed the main motivation for developing our new HP estimator to nonparametrically fit the conditional mean process. It turns out that, as clearly shown in the bottom panel of Figure 1.2, the fitted consumption volatility by our HP estimator shows substantially more variability, and therefore captures the dynamics of the original consumption volatility process significantly better than the standard Local Linear estimator. The resulting larger variation in the consumption volatility may provide an important channel in explaining the risk-premium puzzle.

By simulations, we show that our estimator performs better in general in both discrete and continuous time models than the commonly used Local Linear estimator. For the discrete time model, we consider a simple regression model with a regressor generated as an AR(1) process for the discrete time model, and for the continuous time model, we consider the commonly used Ornstein-Uhlenbeck process as a continuous time analogue of our discrete time model. In both discrete and continuous time models, our estimator performs better than the Local Linear estimator except when the conditional mean component carries very little signal, i.e., when the regression coefficient is extremely small, or excessively persistent, i.e., when the auto-regressive coefficient for the regressor process in the discrete time model is close to 1 or when the mean-reversion parameter of the Ornstein-Uhlenbeck process is too close to zero.

Finally, we demonstrate our new HP approach by applying it to fit conditional mean process of the four key US macroeconomic variables - GDP growth rates, inflation rates, unemployment rates and short term interest rates at both monthly and quarterly frequencies. We find that the values of the optimal smoothing parameter for the quarterly series are substantially smaller than the value 1,600 recommended for the quarterly US macroeconomic variables, and consequently the fitted mean processes exhibit higher time varying volatilities for all macro time series considered.

1.3 Stock Return Predictability with Multiple Predictors

In Section 4, we develop a new methodology for testing stock return predictability with multiple predictors. The predictability of stock returns seems to be widely accepted by both academics and practitioners. For instance, it was stated as new facts in finance by Cochrane (1999, 2005) and reiterated also in Lettau and Ludvigson (2001), among many others, with dividend-price ratios and earnings-price ratios suggested typically as common predictors. Moreover, there is clear and unambiguous consensus in the financial industry that prediction based on such common predictors is highly useful, see, e.g., Wilcox (2007) and Ferson et al. (2003). In fact, predictive regressions seem to be routinely used in making financial decisions including tactical asset allocation, active portfolio management, conditional performance evaluation, and market timing, and others. From the theoretical point of view that stock prices should reflect the discounted current value of future dividends, it of course natural to expect that stock returns are predictable. Naturally, the return predictability is supported by many economic models explaining various aspects of financial decisions, including consumption smoothing in Balvers et al. (1990), habit formation in Campbell and Cochrane (1999), heterogeneous preferences in Chan and Kogan (2002), time-varying risk preferences in Menzly et al. (2004), and prior beliefs in Avramov (2004).

Nevertheless, the empirical evidence of stock return predictability is mixed and still far from being conclusive. Several authors including Bossaerts and Hillion (1999) and Welch and Goyal (2008) show that the standard predictive regressions do not provide any meaningful improvements in terms of predictive power compared to simple statistical models. In contrast, however, many others also provide positive evidence especially by adding a new set of predictors or employing a novel methodology. For instance, Avramov and Chordia (2006) find evidence of out-of-sample predictability of stock returns by the dividend yield, the term spread, the default spread, and the Treasury bill yield, and Bollerslev et al. (2009), Bollerslev et al. (2011), Drechsler and Yaron (2011) and Bollerslev et al. (2014) show that the return predictability becomes stronger if the variance of risk premium is used as an additional predictor. Moreover, Avramov (2002) and Cremers (2002) each find in-sample and out-of-sample stock return predictability using Bayesian model averaging. See also Campbell and Thompson (2008) and Guo (2006, 2009). Ang and Bekaert (2007) find that long horizon predictability to be statistically insignificant and not robust across countries and sample periods. Though, at the same time, they still find that stock return predictability is real, albeit at shorter horizons. See also Hjalmarsson (2011), Moon et al. (2004), and Valkanov (2003).

In Section 4, we provide an in-depth empirical analysis of stock return predictability. From an econometric point of view, testing for return predictability is extremely challenging, and we should employ a robust as well as powerful methodology. It has been known, widely and for long, that the usual least squares estimator yields a test that is not controllable and acceptable level of rejection probabilities. There are two important econometric issues involved in testing for return predictability: persistent and endogenous predictors, and nonstationary stochastic volatility in returns. The problem of having persistent and endogenous covariates has long been well recognized and has been fully analyzed especially for the return predictability regression by Stambaugh (1999), Campbell and Yogo (2006), Chen and Deo (2009) and Phillips and Lee (2013). The presence of nonstationary stochastic volatilities in stock returns have also been well demonstrated in Schaller and Norden (1997), Jacquier et al. (2004) and Cavaliere (2004), among others. However, it has not been considered specifically for the predictive regression of stock returns until recently

Choi et al. (2016) show that it also poses a serious challenge in testing for stock return predictability.

Now we are all very well aware of econometric challenges in return predictive regression, and available are several methodologies we may use to effectively deal with these challenges. In Section 4, we employ a novel methodology developed and investigated in Chang et al. (2016) to test for return predictability using multiple predictors. Very few existing methodologies, which are robust and valid against the aforementioned two main challenges of return predictive regression, most allow only for single predictor.[5] It is therefore necessary to test the predictability of return sequentially on each predictor. Needless to say, this is a serious drawback, which does not permit many interesting joint hypotheses regarding on return predictability. Our approach not only makes it possible to test for the predictability of stock returns jointly by multiple predictors, but also reveals the marginal predictive power of each predictor. Using a new methodology, we find some strong evidence for stock return predictability jointly by dividend-price ratio, earnings-price ration, short-term interest rates and term spread of interest rates.

[5] Amihud et al. (2009) demonstrated that the stock returns are predictable by two variables, income-to-consumption and dividend yield. Recent development by Phillips and Lee (2013) and Lee (2016) propose to use quantile regression in combination with IVX-filtering for a multivariate predictability test.

2. MACROECONOMIC UNCERTAINTY AND ASSET PRICES: A STOCHASTIC VOLATILITY MODEL

Section 2 is organized as follows. In Section 2.1, we develop our asset pricing model. A new stochastic volatility model is proposed and integrated into a consumption-based asset pricing model. Section 2.2 develops our Bayesian econometric methodology. Section 2.3 describes the data set, and Section 2.4 explains our main quantitative result. Then, we conclude in Section 2.5. Appendix A contains the derivations of theoretical results.

2.1 Asset Pricing with Time-Varying Uncertainty

2.1.1 Basic Model

We consider a simple closed economy in which the representative agent has a recursive preference (Epstein and Zin (1989) and Weil (1989)) given by

$$U_t = \left[(1-\delta)C_t^{\frac{1-\gamma}{\chi}} + \delta(\mathbb{E}_t U_{t+1}^{1-\gamma})^{\frac{1}{\chi}}\right]^{\frac{\chi}{1-\gamma}},$$

where $\chi = (1-\gamma)/(1-1/\psi)$, $\gamma \geq 0$ is the coefficient of relative risk aversion (RRA), $\psi \geq 0$ measures the elasticity of intertemporal substitution (EIS), and $0 < \delta < 1$ is the time discount factor. Compared with a conventional power utility function, Epstein-Zin-Weil utility function permits more flexibility in that it breaks the tight link between the parameters of risk aversion γ and intertemporal substitution ψ. In case of the power utility function, EIS is the reciprocal of the risk aversion parameter so that $\chi = 1$ should hold. Another useful property of this preference function is that the decision maker cares about timing for the resolution of uncertainty. It is well known that if $\gamma > (<)\psi^{-1}$ holds, then economic agents prefer early (late) resolution of uncertainty.[1] Thus, if both RRA and EIS parameters are greater than one, this implies that the representative agent prefers early resolution, and $\chi < 0$ holds.

The intertemporal budget constraint for the representative agent can be written as $A_{t+1} = R_{a,t+1}(A_t - C_t)$, where A_t is the wealth at time t, and $R_{a,t+1}$ is the gross return on the portfolio of all invested wealth on consumption claims between t and $t+1$. Epstein and Zin (1989) derive an Euler equation

$$1 = \mathbb{E}_t \left[\delta^{\chi} \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{\chi}{\psi}} R_{a,t+1}^{\chi-1} R_{i,t+1} \right] \tag{2.1}$$

for the gross rate of return $R_{i,t+1}$ on asset i between t and $t+1$. From the Euler equation (2.1), we have the logarithm of the intertemporal marginal rate of substitution (IMRS)

$$m_{t+1} = \chi \log \delta - \frac{\chi}{\psi} g_{c,t+1} + (\chi - 1) r_{a,t+1}, \tag{2.2}$$

where $r_{a,t+1} = \log R_{a,t+1}$ is the log return on the portfolio of all invested wealth, and $g_{c,t+1}$ is the log consumption growth between t and $t+1$. By defining $R_{a,t+1}$ as the returns on holding the aggregate wealth that pays the consumption good (C_t) as dividend, we can write it down as

$$R_{a,t+1} = \frac{C_t}{P_{a,t}} \left(1 + \frac{P_{a,t+1}}{C_{t+1}} \right) \frac{C_{t+1}}{C_t},$$

where $P_{a,t+1}$ is the price of a consumption claim at $t+1$. Applying the log-linear approximation as in Campbell and Shiller (1988), we find a theoretical relation among

[1] This result carries over to the case of general recursive preferences with minor modifications. See Brown and Kim (2013) for more results.

$r_{a,t+1}$, the price-consumption ratio, and the consumption growth as

$$r_{a,t+1} = \phi_{0,c} + \phi_{1,c} z_{c,t+1} - z_{c,t} + g_{c,t+1}, \tag{2.3}$$

$$\phi_{0,c} = \log(1 + \exp(\bar{z})) - \phi_{1,c}\bar{z}, \tag{2.4}$$

$$\phi_{1,c} = \exp(\bar{z})/(1 + \exp(\bar{z})), \tag{2.5}$$

where $z_{c,t} = \log(P_{a,t}/C_t)$ is the log price-consumption ratio, and $\bar{z}$ is a constant. They show that the approximation is highly accurate. We assume that the consumption growth ($g_{c,t+1}$) and the dividend growth ($g_{d,t+1}$) have a conditional mean component, denoted as $\mu_j + \nu_{j,t}$, $j = c, d$, and a logistic volatility function f driven by scalar processes $x_{c,t}$ and $x_{d,t}$ respectively.[2] We have the following form.

$$g_{j,t+1} = \mu_j + \nu_{j,t} + \sqrt{f_j(x_{j,t})}\, \varepsilon_{j,t+1}, \tag{2.6}$$

$$\nu_{j,t+1} = \rho_j \nu_{j,t} + \varphi_j \sqrt{f_j(x_{j,t})}\, \eta_{j,t+1}, \tag{2.7}$$

$$x_{j,t} = \lambda_j w_t + e_{j,t}, \tag{2.8}$$

$$w_t = \rho_w w_{t-1} + u_t, \tag{2.9}$$

$$f_j(x_{j,t}) = \alpha_j + \frac{\beta_j}{1 + \exp\left[-(x_{j,t} - \kappa_j)\right]}, \tag{2.10}$$

with $\alpha_j > 0, \beta_j > 0$ for $j = c, d$. We let the error terms be characterized as

$$\begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{d,t} \end{pmatrix} \sim iid\ \mathbb{N}\left(0, \Sigma\right), \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{2.11}$$

and $e_{j,t}$ is $i.i.d.$ $\mathbb{N}(0, \sigma_j^2)$ for $j = c, d$, and u_t and $\eta_{j,t}$ are $i.i.d.$ $\mathbb{N}(0, 1)$. Moreover, we assume that $\varepsilon_{j,t}, \eta_{j,t}, u_t$ and $e_{j,t}$ are mutually independent of each other. The

[2] From now on, subscripts c and d refer to consumption growth and dividend growth.

variances of $\varepsilon_{j,t}$, $\eta_{j,t}$ and u_t are all set to be unity to identify the parameters α_j, β_j, φ_j and λ_j. w_0 is assumed to be independent of u_t, $e_{j,t}$ and $\varepsilon_{j,t}$. In what follows, we denote by $\theta_j = (\alpha_i, \beta_j, \kappa_j)$ for $j = c, d$, mainly for our econometric analysis.

Equations (2.6) and (2.7) dictate that the conditional expectation of the consumption growth contains a time-varying component $\nu_{c,t}$, and shocks to this component affect the future consumption profile. Therefore, if this shock is highly persistent, it can have a long-run effect. For this shock to be transmitted to financial markets, Bansal and Yaron (2004) and Hansen et al. (2008) additionally assume that the corresponding component of the dividend growth $\nu_{d,t}$ equals a constant multiple of $\nu_{c,t}$, and this common component turns out to play an essential part in generating equity premium. This is so called the long-run risks channel following Bansal and Yaron (2004). Unlike their case, however, we do not assume a common component between $\nu_{c,t}$ and $\nu_{d,t}$ for the reasons we explain below.

First, it is one of our main objectives to examine whether the existence of the common volatility factor in consumption and dividend growth rates, which we interpret as the macroeconomic uncertainty, alone may explain the equity premium. Imposing the presence of commonality in the conditional means of consumption and dividend growth rates has a potential to have a fatal impact on the validity of our results. If either misspecified or imprecisely estimated, the common variation restriction on conditional mean would introduce a spurious common factor in volatility. Such a hazard is likely to occur in our analysis since it is well known to be hard to precisely estimate the conditional mean components of consumption and dividend growth rates. For this reason, we do not impose any restriction on the conditional means of the consumption and dividend growths and leave their specifications as flexible as possible.

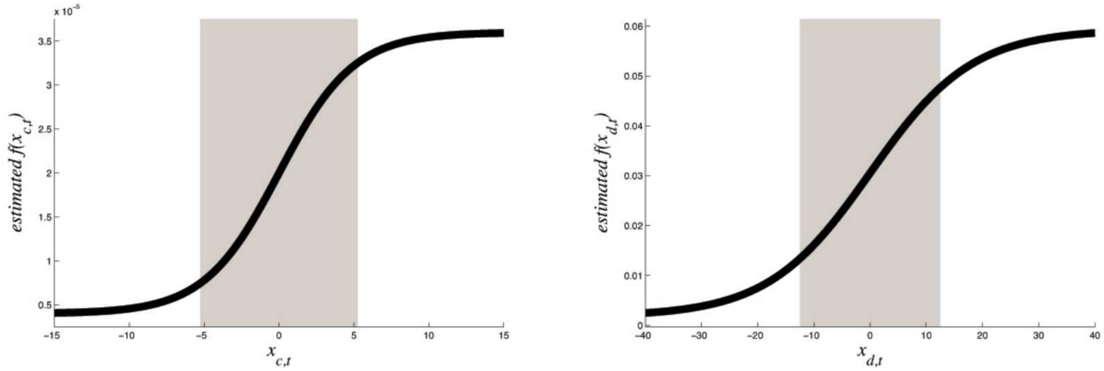
Second, as mentioned above, the long-run risk models use a parameter, say Γ, connecting $\nu_{c,t}$ and $\nu_{d,t}$ ($\nu_{d,t} = \Gamma\nu_{c,t}$) to make a common component in the conditional mean of the consumption and dividend growths.[3] These models need a counterfactually large value, often greater than 2-3 to generate a sizable equity premium. We later show that relaxing the assumption may produce even a negative relation between risk and return. Therefore, without a priori restriction on the commonality of the conditional expectation processes, our model allows us to analyze how important this macroeconomic volatility is as another source of long-run risk in explaining the asset pricing anomalies, even if the conventional long-run risk channel is weak or not working properly.

We now turn our attention to stochastic volatility. Equations (2.6) to (2.11) state that the stochastic volatilities in our model are largely generated by the latent common factor w_t. The common factor is set to be an autoregressive model of order 1 (AR(1)), whereas the idiosyncratic factor is assumed to be *i.i.d.* Our volatility model features two asymptotic levels of volatility with smooth transitions. It is regarded as a generalization of a two-state Markov switching model, and our model identifies smooth transitions as well as the two asymptotic levels of volatility generated by a highly persistent common factor. This specification is motivated by the facts that are well known in the literature on the stochastic volatilities of consumption growth, dividend growth, and asset returns. The reader is referred to Park (2002), Jacquier et al. (2004), Kim et al. (2009) and Jeong et al. (2015) for more details.

Equation (2.10) states that the logistic volatility function f has two asymptotes α_j and $\alpha_j + \beta_j$, which represent the two extreme regimes, i.e., the low and high volatility regimes for $j = c$, d. We impose the identifying restriction $\beta_j > 0$, so that a larger realized value of the latent volatility factor implies higher volatility.

[3] This is often called, consumption leverage parameter following Abel (1990) and Bansal and Yaron (2004). More generic form includes an additional factor as well as consumption and dividend growths. Menzly et al. (2004) find that this parameter may be close to 1 for some industry portfolios.

Figure 2.1: Estimated Logistic Volatility Functions for Consumption Growth and Dividend Growth



Notes: The figure displays the estimated volatility function $f(x_j)$ for $j = c, d$. The estimated parameters are available in Table 2.2. The left panel is the estimated logistic volatility for consumption growth (c), and the right panel is that of dividend growth (d), respectively. The horizontal axis stands for the latent factor $x_{j,t}$ for $j = c, d$. The shaded area refers to the transition period. The end points of the transition period are defined by the points at which $f'''(x) = 0$ holds.

The parameters κ_j and λ_j in (2.8) characterize the transition between two regimes, i.e., the location and speed of the transition. As λ_j gets larger, the transition speed becomes faster and, to a larger extent, the actual volatilities are generated by one of the two asymptotic regimes. Indeed, we may view the regime switching model as the limiting case of our model with $\lambda = \infty$. Figure 2.1 plots the estimated volatility function of consumption and dividend growth rates, showing the features explained above. Finally, κ designates the location of the transition. Because w_t is an autoregressive model of order 1 (AR(1)) with the starting value w_0 independent of all other stochastic components in the model, κ and w_0 are not individually identified. We must therefore set either $w_0 = 0$ or $\kappa_j = 0$.

The smooth transition combined with a common factor in our stochastic volatility setup provides new insight into the relation between asset prices and macroeconomic uncertainty. There exists little doubt that the amount of risk and uncertainty is larger in the high volatility regime compared with the low volatility regime, and abrupt regime changes can create a sizable risk. However, low and stable consumption volatility over time suggests that models relying on high volatility and the possibility of sudden, big changes are not consistent with data. Instead, our model emphasizes the transitions between regimes in macroeconomic volatility. If the level of volatility is low and in between the two extreme regimes, it can create a source of risk and uncertainty, because economic agents are unsure about future volatility regimes they are heading toward. Intuitively, this effect becomes smaller when the economy gets closer to either of the two asymptotic regimes. In particular, if the speed of volatility changes is time-varying, this effect can be significant. Conventional regime shifting models allow only sudden changes and ignore the role of transition states, and other stochastic volatility models have no boundaries, and hence are subject to the possibility of explosive dynamics.

Our model overcomes the drawbacks and predicts that an economy with relatively low-to-medium levels of the macroeconomic volatility can request a higher premium than otherwise, due to the additional uncertainty different from the conventional sources of risk, provided that economic agents prefer early resolution of uncertainty. In sum, our stochastic volatility setup effectively prevents the volatility from exploding over time, yet paves the way to generate a sufficient amount of risk and uncertainty using the Epstein-Zin-Weil preferences.

We believe that this is a novel way of understanding the time-varying uncertainty in macroeconomic variables. Similar to the conventional method of stochastic volatility, the level of volatility matters to measure the degree of uncertainty plagued in an economy. However, our model says that the distance to either the maximum or minimum uncertainty as well as the speed of changes in uncertainty are critical in quantifying the effect of risk and uncertainty. To offer theoretical explanations on this channel, we follow Campbell and Shiller (1988) and Bansal and Yaron (2004) to solve for the log price-consumption ratio $z_{c,t}$.

$$z_{c,t} = A_{0,c} + A_{1,c}\nu_{c,t} + A_{2,c} f_c(x_{c,t}) \tag{2.12}$$

with

$$\begin{aligned}
A_{0,c} &= \frac{\log\delta + \phi_{0,c} + \left(1 - \frac{1}{\psi}\right)\mu_c + \phi_{1,c}A_{2,c}\zeta_0 + \frac{\chi}{2}\left(\phi_{1,c}A_{2,c}\right)^2 \lambda_c^2\beta_c^2\zeta_1}{(1-\phi_{1,c})}, \\
A_{1,c} &= \frac{(1-1/\psi)}{(1-\phi_{1,c}\rho_c)}, \\
A_{2,c} &= \frac{(1-\phi_{1,c}\rho_\omega) + \sqrt{(1-\phi_{1,c}\rho_\omega)^2 - \lambda_c^2\beta_c^2\zeta_2\left(\phi_{1,c}\right)^2\left[(1-\gamma)^2 + \left(\frac{(1-\gamma)\phi_{1,c}\varphi_c}{(1-\phi_{1,c}\rho_c)}\right)^2\right]}}{\lambda_c^2\beta_c^2\zeta_2\chi\left(\phi_{1,c}\right)^2},
\end{aligned}$$

where ζ_0, ζ_1, ζ_2 are positive constants describing the conditional moments of $f_{c,t}$ process defined in Appendix A.1.[4] If $\gamma > 1$ and $\psi > 1$, then $\chi < 0$, $A_{1,c} > 0$ and $A_{2,c} < 0$ hold. This means that the representative investor prefers higher expected future consumption growth, but does not like a rise in macroeconomic volatility. In addition to the conventional channel of volatility ($f_c(x_{c,t})$) in which a higher value of $x_{c,t}$ means higher uncertainty, we can observe from the definition of $A_{2,c}$ that if λ_c is small, i.e., when there exists a slow and smooth transition, the absolute value of $A_{2,c}$ increases via decreases in λ_c. That is, given the traditional risk-return trade-off relation, our model has an extra layer of the volatility channel: a low level of macroeconomic volatility does not necessarily imply that economic agents perceive the state of the economy as being safe. When uncertainty unfolds in a sluggish fashion and is in transit, the representative investor dislikes the obscure nature of

[4]The derivations of $A_{0,c}$, $A_{1,c}$, and $A_{2,c}$, and the accuracy of the approximations are discussed in Appendix A.1.

macro volatility regime, and this amplifies the premium for bearing risks associated with macroeconomic volatility.

Once all the coefficients are verified, we can derive the innovations in the stochastic discount factor, or IMRS as

$$m_{t+1} - \mathbb{E}_t(m_{t+1}) = -\gamma\sqrt{f_c(x_{c,t})}\varepsilon_{c,t+1} - \Lambda_{m,\eta}\sqrt{f_c(x_{c,t})}\eta_{c,t+1} - \Lambda_{m,u}(t)u_{t+1} \quad (2.13)$$

with

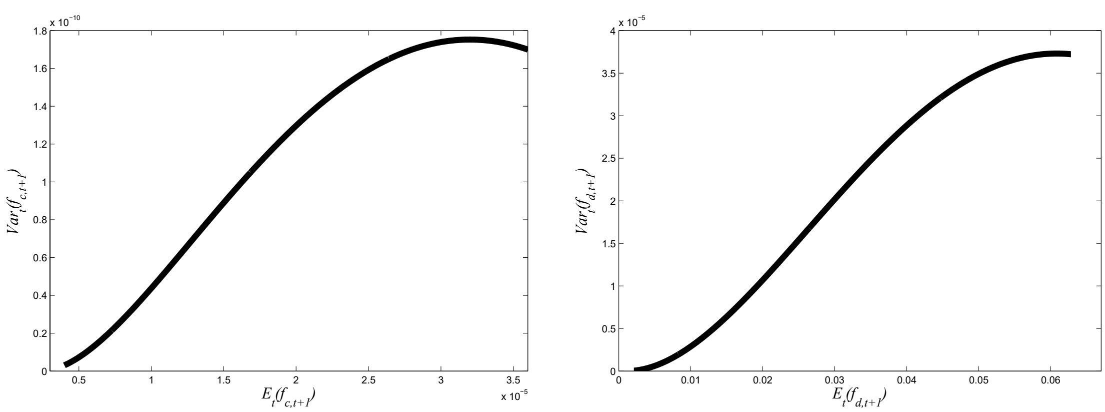
$$\Lambda_{m,\eta} \equiv (1-\chi)\phi_{1,c}A_{1,c}\varphi_c$$
$$\Lambda_{m,u}(t) \equiv (1-\chi)\phi_{1,c}A_{2,c}\sqrt{Var_t\left(f_{t+1}\right)},$$

where the coefficient terms are derived in Appendix A.2. Here the risk sources are represented by three terms, labeled as the short-run risk ε_{t+1}, the long-run risk η_{t+1}, and the macroeconomic volatility risk u_{t+1}. Similar to (2.12), positive innovations in both short-term and long-term consumption growth lead to lower discount rates for futures, while higher volatility innovations refer to higher discount for futures again if $A_{2,c} < 0$. Note that the conditional variance of macroeconomic volatility f_t, denoted as $Var_t\left(f_{t+1}\right)$ prevails in equation (2.13). Appendix A.1 derives the conditional variance in terms of the conditional expectation of macroeconomic volatility as follows:

$$Var_t(f_{t+1}) = \left[\sqrt{\left(\frac{\beta}{2\lambda}\right)^2 - \alpha\left(\alpha+\beta\right) + 2\left(1+\alpha\right)\beta\mathbb{E}_t\left(f_{t+1}\right) - \left[\mathbb{E}_t\left(f_{t+1}\right)\right]^2} - \frac{\beta}{2\lambda}\right]^2. \quad (2.14)$$

$Var_t\left(f_{t+1}\right)$ is time-varying and stochastic, despite the fact that the common macroeconomic uncertainty process w_t is a simple AR(1) process with *i.i.d.* shocks. As explained early, this additional risk comes from the fact that macroeconomic volatility is bounded and varies non-linearly over time between the two regimes. This risk is priced via the uncertainty resolution channel ($\chi < 0$), according to equation (2.13). To describe equation (2.14), Figure 2.2 plots the relations between $Var_t(f_{t+1})$ and $\mathbb{E}_t\left(f_{t+1}\right)$ for consumption and dividend using the parameters estimated in the next section. This shows that within the interval of $[\alpha, \alpha+\beta]$, $Var_t(f_{t+1})$ increases as $\mathbb{E}_t\left(f_{t+1}\right)$ increases, yet the marginal effect becomes smaller in an accelerating fashion.

Figure 2.2: Time-Varying Uncertainty of Volatility Regimes



Notes: The figure displays the relation between the conditional expectation and the conditional variance of the macroeconomic volatility for consumption (c) and dividend growths (d), or $f_{j,t+1}, j = c, d$. For parameters values, the estimation results in Table 2.2 are used. The range of the horizontal axis is defined by the estimated minimum and the estimated maximum of the macroeconomic volatilities for consumption and dividend growth rates, respectively.

With the IMRS, we can express the risk-free rate as

$$\begin{aligned} r_{f,t} &= -\log\left[\mathbb{E}_t \exp(m_{t+1})\right] \\ &= -\log\delta + \frac{1}{\psi}\mathbb{E}_t g_{c,t+1} + \frac{1-\chi}{\chi}\mathbb{E}_t\left[r_{a,t+1} - r_{f,t}\right] - \frac{1}{2\chi}\mathrm{Var}_t(m_{t+1}). \end{aligned} \tag{2.15}$$

The negative relation between the risk-free rate and the intertemporal elasticity of substitution is clear from the first term in (2.15). But as the risk aversion coefficient becomes larger (i.e., χ being more negative), hence both the conditional mean of risk premium on wealth and the conditional variance of IMRS increase, the effect on the risk-free rate is not clear from equation (2.15) immediately because the last two terms have opposite signs. If the third term dominates the fourth term, which is usually the case, the risk-free rate decreases. We will verify this via simulations later in Section 2.3.

2.1.2 Asset Returns Dynamics

In Section 2.1.2, we theoretically explain how the model produces equity premiums. We begin with a consumption claim, earning $r_{a,t}$ in each period. Solving the model with the consumption claim, we can show the following:

$$\mathbb{E}_t\left[r_{a,t+1} - r_{f,t}\right] = \gamma f_{c,t} + \Phi_1 f_{c,t} + \Phi_2 Var_t\left(f_{c,t+1}\right) - \frac{1}{2}Var_t(r_{a,t+1}) \tag{2.16}$$

with

$$\begin{aligned} \Phi_1 &= \frac{\left(\gamma - \frac{1}{\psi}\right)\phi_{1,c}\varphi_c}{(1-\phi_{1,c}\rho_c)} \\ \Phi_2 &= (1-\chi)^2\,\phi_{1,c}\left(A_{2,c}\right)^2. \end{aligned}$$

The first term in the right hand side of equation (2.16) refers to the conventional risk premium and is known to be very small. Φ_1 is related to the long-run risk channel of Bansal and Yaron (2004), and the coefficient depends on the speed of uncertainty resolution. That is, if the investor prefers early resolution of uncertainty ($\gamma > \frac{1}{\psi}$), $\Phi_1 > 0$ holds, and the risk premium is big, if ρ_c is close to 1. The third term $\Phi_2 Var_t\left(f_{c,t+1}\right)$ is the premium arising from the macroeconomic uncertainty channel we explained early.

Now, we derive an expression similar to (2.16) for the market equity premium. Firstly, we assume that $\nu_{c,t}$ and $\nu_{d,t}$ are uncorrelated to each other. As with equation (2.3), we follow Campbell and Shiller (1988) to define the market rate of return as

$$r_{m,t+1} = \phi_{0,m} + \phi_{1,m} z_{m,t+1} - z_{m,t} + g_{d,t+1} \tag{2.17}$$

where $z_{m,t}$ is the log price-dividend ratio. Then, we have

$$\mathbb{E}_t\left[r_{m,t+1} - r_{f,t}\right] = \gamma\rho\sqrt{f_{c,t}}\sqrt{f_{d,t}} + \Phi_1 f_{c,t} + \Phi_2 Var_t\left(f_{c,t+1}\right) - \frac{1}{2}Var_t(r_{m,t+1}) \tag{2.18}$$

with

$$\Phi_1 = -\frac{\left(\gamma - \frac{1}{\psi}\right)\phi_{1,c}\phi_{1,m}\left(\varphi_c\right)^2}{\left(1-\phi_{1,c}\rho_c\right)\left(1-\phi_{1,m}\rho_c\right)\psi}$$

$$\Phi_2 = (1-\chi)\phi_{1,c}\phi_{1,m}A_{2,c}A_{2,d},$$

where π_d and $A_{2,d}$ are defined in Appendix A.3. The first term in the right hand side is the conventional risk premium term found in consumption-based models. The second term refers to the contribution by the long-run risks channel. However, note that $\Phi_1 < 0$ with early resolution of uncertainty. This implies that this channel gives a negative risk-return trade-off with $\gamma > \psi$. This mainly comes from the zero correlation of the conditional mean components between consumption growth and dividend growth. If we instead assume a long-run relation as $\nu_{d,t} = \Gamma \nu_{c,t}$ as in Bansal and Yaron (2004), we obtain

$$\mathbb{E}_t \left[r_{m,t+1} - r_{f,t} \right] = \gamma \rho \sqrt{f_{c,t}} \sqrt{f_{d,t}} + \Phi_1 f_{c,t} + \Phi_2 Var_t \left(f_{c,t+1} \right) - \frac{1}{2} Var_{tt}(r_{m,t+1}) \quad (2.19)$$

with

$$\Phi_1 = \frac{\left(\gamma - \frac{1}{\psi}\right)\left(\Gamma - \frac{1}{\psi}\right)\phi_{1,c}\phi_{1,m}(\varphi_c)^2}{(1-\phi_{1,c}\rho_c)(1-\phi_{1,m}\rho_c)},$$

and Φ_2 is defined the same as for (2.18). In this case, $\Phi_1 > 0$ can hold if $\Gamma > \psi^{-1}$ is satisfied. Existing studies use Γ greater than 2-3 to match the data. Recall that we did not restrict the long-run common component between two conditional expectations and solve our cases numerically. Further, we do not resort to either of the above approximate solutions when we evaluate the empirical performance of the model. Therefore our model allows cases in which the conventional long-run risks channel is weak.

The main thrust of our model lies in the third term in equations (2.16), (2.18), and (2.19). This term describes how the time-varying macroeconomic uncertainty is related to the expected excess returns for holding equity. One can see from the similarity of Φ_2 that this channel is robust regardless of the conditional mean specifications. An interplay between $\chi < 0$ (preference for early resolution of uncertainty) and $A_{2,c}, A_{2,d} < 0$ produces this uncertainty premium ($\Phi_2 > 0$). Thus, even a fairly modest amount of the macroeconomic volatility does not imply a small equity premium requested by investors, because there exists time-varying uncertainty about the volatility regimes.

2.2 Econometric Methodology

This section proposes an econometric procedure to estimate an asset pricing model with both observables and latent factors. Our asset pricing model consists of the moment condition (2.1) and the specification of economic variables, (2.3) to (2.11). For the purpose of identifying models cast in moment conditions, the generalized method of moment (GMM) by Hansen (1982) is the standard method. However, we need a new econometric methodology, because our setup contains latent factors of a non-linear volatility function inside the Euler equation. The latent variables need identifying, and the non-linearity makes worsen the curse of dimensionality. Further, the estimation of the preference parameters depends upon the latent variables that determine stochastic volatility. To overcome the difficulties, we extend a Bayesian approach called Markov Chain Monte Carlo (MCMC) methods to incorporate the GMM criterion. Our procedure is outlined as follows:

First, we begin by drawing samples for the nonlinear, stochastic volatility factors. To this end, we select initial points for both latent factors and parameters. Then, we obtain samples of the latent factors given the initial set of parameters. With the latent factors in hand, samples of volatility parameters are drawn. We use a combination of Gibbs sampling and Metropolis-Hastings (MH) algorithm to implement this step.[5] Although this is a common Bayesian approach, our algorithm is new because our non-linear volatility function is different from the conventional stochastic volatility models.

The second step of this method is about sampling the preference parameters (γ and ψ) given the sample of stochastic volatility and observables.[6] We use the GMM

[5]For more details on the methods, see Chib and Greenberg (1995).

[6]In evaluating the moment condition, we compute the coefficients $\phi_{0,c}$ and $\phi_{1,c}$ using the definitions, $\phi_{0,c} = \log(1 + \exp(\bar{z})) - \phi_{1,c}\bar{z}$ and $\phi_{1,c} = \exp(\bar{z})/(1 + \exp(\bar{z}))$ with the choice of $\bar{z} = \mathbb{E}(z_t)$. Specifically, equation (2.12) implies that $\bar{z} = A_{0,c} + A_{1,c}\mathbb{E}(v_{c,t}) + A_{2,c}\mathbb{E}(f_{c,t})$ holds. By plugging

criterion to draw samples. That is, we obtain a sample of preference parameters from a quasi-posterior function based on the Euler equation (2.1) and previous draws of latent factors and parameters to implement our MCMC method. Clearly, this step is reminiscent of frequentist approach and pragmatically motivated. However, this procedure turns out to be statistically sensible as well. Papers by Chernozhukov and Hong (2003), and Gallant et al. (2014) show that drawing samples of parameters from a conditional posterior distribution made out of the moment condition is asymptotically equivalent to obtaining a sample from the asymptotic distribution of the GMM estimator under mild conditions. Thus, this step can be viewed as a GMM within Gibbs sampling given observable data and previously drawn latent factors and parameters.

Third, we repeat the first and the second steps to create a MCMC chain. Once we have enough samples, then we use Monte Carlo integration to compute the moments of interest. In the below, we explain our algorithm to identify stochastic volatility function, followed by a Bayesian GMM procedure to draw a sample of the preference parameters.

2.2.1 Stochastic Volatility

According to the asset pricing model developed in the previous section, asset returns evolve over time because of the time-varying uncertainty inherent in the consumption and dividend growth rates. For this purpose, we rewrite (2.6) as

$$y_{j,t+1} = \sqrt{f_j(x_{j,t})}\,\varepsilon_{j,t+1} \tag{2.20}$$

with $y_{j,t+1} = g_{j,t+1} - \mu_j - \nu_{j,t}$ for $j = c, d$. Note that $y_{j,t+1}$ is a martingale difference sequence defined from $g_{j,t+1}$, net of its slow moving conditional mean component

the definitions of $\phi_{0,c}$ and $\phi_{1,c}$ into $A_{0,c}$, $A_{1,c}$, and $A_{2,c}$, we solve for $\bar{z}$.

$\mu_j + \nu_{j,t}$. We use the Hodrick-Prescott (HP) filter to estimate the moving conditional mean component of $g_{j,t+1}$, which is subtracted from $g_{j,t+1}$ to obtain the estimated $y_{j,t+1}$. In particular, to exploit the fact that $y_{j,t+1}$ has no trend, we optimally select a weight for the Hodrick-Prescott filter that makes $y_{j,t+1}$ as close as possible to a martingale difference sequence. We also experimented with the method suggested by Bansal and Yaron (2004) and Bansal et al. (2007), which regress the consumption growth onto the interest rate and the price-dividend ratio, exploiting the theoretical structures. Although both results are compatible with each other, we find that the statistical filtering seems to identify the long-run components better, especially with monthly frequency data.[7]

Now with the estimates of $y_{j,t}$ in hand, one can easily see that (2.20) as well as (2.7) to (2.10) form a usual state space model with latent factors. The resulting state space model, however, is nonstandard in two aspects: It is non-linear, and includes non-stationary latent factors. There are several methodologies that can be used to statistically analyze non-linear state space models with latent factors. The most conventional approach is to apply the extended Kalman filter for the linearized version of the measurement equation (2.20) after taking squares and logs. However, we find that the conventional extended Kalman filter performs rather poorly in the presence of highly persistent latent factors.[8] To avoid this difficulty, Kim et al. (2009) developed a density-based filter, relying on the updates, predictions and smoothings based on the estimated densities of latent factors, for the model similar to, but much simpler than, what we consider in Section 2. At least in theory, it is straightforward to extend their filter to make it applicable for our model. However, computationally it

[7] Therefore, we estimate ρ_c and ρ_d from the trend component of the consumption and dividend growth rates. Our estimates are 0.993 and 0.992 respectively.

[8] This is well expected, since some of the partial derivatives of our volatility function are integrable functions of highly persistent latent factors. Therefore, the Kalman gains get very close to zero, when the extracted factors take large values, providing no further updates for the latent factors.

seems extremely difficult and time demanding to use this approach for the estimation of our model.

In Section 2, we take the Bayesian approach. To introduce our approach, it will be convenient to define some additional notations. We let n be the sample size, and define $Y = (y_1, \cdots, y_n)$ with $y_t = (y_{c,t}, y_{d,t})'$, and $L = (X, W)$ with $X = (x_1, \cdots, x_n)$, $x_t = (x_{c,t}, x_{d,t})$ and $W = (w_1, \cdots, w_n)$. Moreover, we define $\Psi = (\theta, \rho, \lambda, \sigma^2, \rho_w)$ with $\theta = (\theta_c, \theta_d)$, $\lambda = (\lambda_c, \lambda_d)$ and $\sigma^2 = (\sigma_c^2, \sigma_d^2)$.[9] Note that Y, L and Ψ denote respectively the observed samples, latent factors and unknown parameters of our model. Moreover, we let $D_t = \operatorname{diag}\left(\sqrt{f_c(x_{c,t})}, \sqrt{f_d(x_{d,t})}\right)$, and signify the marginal and conditional densities by $p(\cdot)$ and $p(\cdot|\cdot)$ in what follows. Now we may easily deduce that the joint posterior density of the latent factors and unknown parameters is given by

$$\begin{aligned} p(L, \Psi|Y) &\propto p(L, Y|\Psi)p(\Psi) \\ &\propto \left(\prod_{t=1}^{n} p(y_t|x_t, \Psi)p(x_{c,t}|w_t, \Psi)p(x_{d,t}|w_t, \Psi)p(w_t|w_{t-1}) \right) p(\Psi). \end{aligned} \tag{2.21}$$

Following the usual Bayesian procedure, we will implement a Markov-Chain-Monte-Carlo (MCMC) method to sample (L, Ψ) from the joint posterior density $p(L, \Psi|Y)$ in (2.21). Once the observations of (L, Ψ) are drawn, we may readily obtain any posterior sample moments of L and Ψ using the standard Monte-Carlo numerical integration.[10]

For our MCMC procedure, we use the Gibbs sampler and the Metropolis-Hastings (MH) algorithm within the Gibbs sampler.[11] The procedure allows us to effectively

[9] For the purpose of identification, φ_c and φ_d are assumed to be 1.

[10] For related Bayesian procedures, the reader is referred to Jacquier et al. (1994) Jacquier et al. (2004), So et al. (1998) and Geweke and Tanizaki (2001). Our procedure follows Jacquier et al. (2004) most closely.

[11] As explained in Chib and Greenberg (1995), we may regard our entire procedure as the MH

deal with the multi-dimensionality of our latent factors and parameters and difficulties in drawing samples from the complicated target densities. The Gibbs sampler simplifies the required sampling procedure by reducing our multi-dimensional transition to the composition of a sequence of uni-dimensional transitions. This is very helpful for the analysis of our model, which includes a large number of latent factors and unknown parameters. Subsequently to the application of Gibbs sampler, we rely on the MH algorithm in case it is difficult to proceed by sampling directly from the required transition. The MH algorithm allows us to implement the required transition simply by drawing samples from a candidate-generating density, together with a random selection rule. See Chib and Greenberg (1995) for a nice introduction of the algorithm. For the choice of candidate-generating distributions, we follow the suggestion by Geweke and Tanizaki (2001). For the candidate-generating distributions of latent factors, we use the distributions given by the transition equation of our model. On the other hand, the prior distributions are used for the candidate-generating distributions for the unknown parameters. As is well known, our MCMC procedure is valid, since the underlying Markov chain is irreducible and aperiodic.

Now we derive the conditional posteriors and the required sampling methods to implement the Gibbs sampler and MH algorithm, which are needed to execute our MCMC procedure. For the common latent factor w_t, we have for $t = 1, \ldots, n-1$

$$\begin{aligned} p(w_t|X, W_{\setminus t}, Y, \Psi) &= p(w_t|x_{c,t}, x_{d,t}, w_{t+1}, w_{t-1}, \Psi) \\ &\propto p(x_{c,t}, x_{d,t}|w_t, \Psi)p(w_{t+1}|w_t, \Psi)p(w_t|w_{t-1}, \Psi) \\ &= p(x_{c,t}|w_t, \lambda_c, \sigma_c^2)p(x_{d,t}|w_t, \lambda_d, \sigma_d^2)p(w_{t+1}|w_t, \rho_w)p(w_t|w_{t-1}, \rho_w), \end{aligned} \tag{2.22}$$

algorithm applied in turn to one variable at a time.

where $W_{\setminus t}$ denotes W with w_t deleted. We may readily deduce from (2.22) that the conditional distribution of w_t given $X, W_{\setminus t}, Y$, and Ψ is indeed $\mathbb{N}(RS^{-1}, S^{-1})$, where

$$R = \sum_j \frac{\lambda_j x_{j,t}}{\sigma_j^2} + \rho_w \left(w_{t+1} + w_{t-1}\right) \qquad \text{and} \qquad S = \sum_j \frac{\lambda_j^2}{\sigma_j^2} + 1 + \rho_w^2.$$

Therefore, it is quite straightforward to draw samples for the common factor given all values of other latent factors and unknown parameters.[12]

In case of the latent factor $x_{j,t}$, the conditional posterior density is given by

$$\begin{aligned}
&p(x_{j,t}|X_{\setminus j,t}, W, Y, \Psi) = p(x_{j,t}|x_{\setminus j,t}, w_t, y_t, \Psi) \\
&\quad \propto p(y_t|x_t, w_t, \Psi)p(x_t|w_t, \Psi) = p(y_t|x_t, \theta, \rho)p(x_{j,t}|w_t, \lambda_j, \sigma_j^2) \\
&\quad = \frac{1}{2\pi} \det(D_t \Sigma D_t)^{-1/2} \exp\left(-\frac{y_t'(D_t\Sigma D_t)^{-1} y_t}{2}\right) \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_{j,t} - \lambda_j w_t)^2}{2\sigma_j^2}\right) \\
&\quad \propto \det(D_t\Sigma D_t)^{-1/2} \exp\left(-\frac{y_t'(D_t\Sigma D_t)^{-1} y_t}{2} - \frac{(x_{j,t} - \lambda_j w_t)^2}{2\sigma_j^2}\right).
\end{aligned} \tag{2.23}$$

Note that the conditional distribution of y_t given x_t, θ, ρ and that of $x_{j,t}$ given $w_t, \lambda_j, \sigma_j^2$ are given by $\mathbb{N}(0, D_t\Sigma D_t)$ and $\mathbb{N}(\lambda_j w_t, \sigma_j^2)$ respectively. It is difficult to directly draw samples from the conditional density in (2.23), and therefore, we use the MH algorithm. In particular, we apply the algorithm with $\mathbb{N}(\lambda_j w_t, \sigma_j^2)$ as the candidate-generating density.

Finally, we derive the conditional posterior distributions of $(\theta, \rho, \lambda, \sigma^2, \rho_w)$. The

[12]For $t = n$, we can similarly show that the conditional distribution of the common latent factor is also $\mathbb{N}(RS^{-1}, S^{-1})$ with $R = \sum_j (\lambda_j x_{j,n}/\sigma_j^2) + \rho_w w_{n-1}$ and $S = \sum_j (\lambda_j^2/\sigma_j^2) + 1 + \rho_w^2$.

conditional posterior density of θ_j, $j = c, d$, is given by

$$
\begin{aligned}
p(\theta_j | X, W, Y, \Psi \backslash \theta_j) &\propto p(Y|X, \theta, \rho) p(\theta_j) \\
&= \prod_{t=1}^{n} \frac{1}{2\pi} \det(D_t \Sigma D_t)^{-1/2} \exp\left(-\frac{y_t'(D_t \Sigma D_t)^{-1} y_t}{2} \right) p(\theta_j).
\end{aligned}
$$

Similarly, the conditional posterior density of ρ is

$$
\begin{aligned}
p(\rho | X, W, Y, \Psi \backslash \rho) &\propto p(Y|X, \theta, \rho) p(\rho) \\
&= \prod_{t=1}^{n} \frac{1}{2\pi} \det(D_t \Sigma D_t)^{-1/2} \exp\left(-\frac{y_t'(D_t \Sigma D_t)^{-1} y_t}{2} \right) p(\rho).
\end{aligned}
$$

Moreover, the conditional posterior densities of λ_j and σ_j^2 are given by

$$
\begin{aligned}
p(\lambda_j | X, W, Y, \Psi \backslash \lambda_j) &\propto p(X_j | W, \lambda_j, \sigma_j^2) p(\lambda_j) \\
&= \prod_{t=1}^{n} \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{(x_{j,t} - \lambda_j w_t)^2}{2\sigma_j^2} \right) p(\lambda_j),
\end{aligned}
$$

$$
\begin{aligned}
p(\sigma_j^2 | X, W, Y, \Psi \backslash \sigma_j^2) &\propto p(X_j | W, \lambda_j, \sigma_j^2) p(\sigma_j^2) \\
&= \prod_{t=1}^{n} \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{(x_{j,t} - \lambda_j w_t)^2}{2\sigma_j^2} \right) p(\sigma_j^2)
\end{aligned}
$$

and

$$
\begin{aligned}
p(\rho_w | W, \Psi \backslash \rho_w) &\propto p(w_t | w_{t-1}, \Psi) p(\rho_w) \\
&= exp[-\frac{1}{2}\{(\rho_w)^2 (\sum_{t=1}^{n} (w_{t-1})^2) - 2\rho_w (\sum_{t=1}^{n} w_t w_{t-1})\}] p(\rho_w)
\end{aligned}
$$

for $j = c, d$. We apply the MH algorithm to sample from the conditional posterior distributions of $(\theta, \rho, \lambda, \sigma^2)$, using their prior distributions as the candidate-generating densities. We use Gamma priors for the parameters α, β, σ^2 and λ, and normal priors for the parameters κ and ρ.

2.2.2 Bayesian GMM Estimator with Latent Factors

An appealing characteristic of GMM estimator is that non-linear structural models are estimable without knowing the exact form of conditional distribution by relying only on moment conditions. The main issue of our moment condition (2.1) lies in the existence of latent factors associated with stochastic volatility. Bayesian approach of identification is basically to apply Monte Carlo integration, after generating a sequence of samples using posterior marginal densities of the model parameters. However, it is often difficult to have an exact form of posterior densities with more complex models, and Gibbs sampling and Metropolis-Hastings algorithms are helpful tools to overcome this hardship. In a nut shell, this approach draws samples based on the previous samples to produce a sequence of estimates. Chernozhukov and Hong (2003) develop a Bayesian estimator using general statistical criterion functions, instead of the conventional likelihood function. This allows the GMM criterion to be applied in an MCMC framework. Specifically, they define transformations of statistical functions as quasi-posterior distributions, then they form an MCMC chain to draw samples. They show that this not only reduces computational burden, but also this estimator is as efficient as the extremum estimators, consistent with the large sample theory. Gallant et al. (2014) suggest that this Bayesian GMM approach can be extended to incorporate latent factors under some mild conditions. The key point is that drawing samples from the conditional posterior implied by the GMM criterion function is equivalent to identifying a model using the conventional GMM estimator when observable time-series data are sufficiently long. Following these papers, we estimate the preference parameters using an MCMC approach, as briefly explained in the beginning of the section. To be concrete, we explain this procedure in the below. Using the MCMC procedure developed in Section 2.2.1, we have a sample of the volatility function $f(\dot{x})$. Then, together with the observable data Y, we form a criterion function as

$$\mathcal{L}_T = -\frac{1}{2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Omega_t\left(\varsigma\right)\right)' H_T\left(\varsigma\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Omega_t\left(\varsigma\right)\right), \tag{2.24}$$

where $\Omega_t\left(\varsigma\right)$ is a moment condition that satisfies equation (2.1), i.e., $\mathbb{E}\left[\Omega_t\left(\varsigma\right)\right] = 0$ holds, and $\varsigma = (\gamma, \psi)$ is the vector of preference parameters, and $H_T\left(\varsigma\right)$ is a weighting function. Equation (2.24) is the objective function to be maximized for a GMM estimation. Following Chernozhukov and Hong (2003), quasi-posterior $Q_T\left(\varsigma\right)$ is defined as

$$Q_T\left(\varsigma\right) = \frac{e^{\mathcal{L}_T(\varsigma)}}{\int_\varsigma e^{\mathcal{L}_T(\varsigma)}d\varsigma.} \tag{2.25}$$

Equation (2.25) states that $Q_T\left(\varsigma\right)$ is a normalized GMM criterion function. Note also that equation (2.25) is not a true posterior, because it does not come from the conditional likelihood. Chernozhukov and Hong (2003) and Gallant et al. (2014) show that we can draw a Markov chain $\left(\varsigma^{(1)}, \cdots, \varsigma^{(n)}\right)$ from a quasi-posterior Q based on the GMM criterion function. They also show that this method is valid regardless of the likelihood formulation, and it is basically identical to drawing from the GMM estimation asymptotically.

2.3 Data and Summary Statistics

We use monthly consumption and dividend series from January 1959 to January 2013 to apply the method we developed. The consumption data for non-durable goods and service expenditure series are obtained from the Federal Reserve Banks of St. Louis web site (http://research.stlouisfed.org) and the dividend, the stock market index, and the risk-free rate data come from the Center for Research in Security Prices (CRSP). Dividend growth is generated from the seasonally-adjusted real dividend series, which is created using value-weighted returns with and without dividend.[13] Additionally, both series were adjusted as per capita values. There are total 650 observations. Table 2.1 provides summary statistics for the consumption and dividend growth rates, the market return, and the risk-free rate using monthly data.

The annualized average consumption growth (g_c) is around 1.8%, with the annualized standard deviation of 1.42%, and the annualized equity premium is over 5.5% with the annualized standard deviation of over 15%. Correlation between the consumption growth and the excess market return is about 0.17. Thus, the covariance between the two variables is much smaller compared to the size of equity premium, and this is the essence of the prominent asset pricing puzzle. In addition, correlation between consumption growth and dividend growth is around 0.033, which suggests that the link between the two variables is rather tenuous. Instead, we focuses on the common part of the stochastic volatility of consumption and dividend growth rates.

2.4 Quantitative Results

2.4.1 Macroeconomic Uncertainty

Top panels of Figure 2.3 display the monthly consumption and dividend growth rates. Bottom panels show the estimated conditional mean components ($\mu_j + \nu_{j,t}$, $j = c, d$). The estimated conditional mean components do not appear to move

[13] Real dividend is computed using (returns with dividend - returns without dividend) $\times$ market price index / CPI, where CPI is the consumer price index from the St. Louis Fed. We use cash dividends to measure dividend series and this may be insufficient to capture the total payouts distributed by corporations. Since we want to compare our results with the existing studies, we use the conventional definition of the dividend series. It would be interesting to analyze how the volatilities of the total payouts differ from those of the cash dividends.

Table 2.1: Summary Statistics

	(a) Basic Stat.					
	mean	std.	max.	min.	skewness	kurtosis
g_c	0.0015	0.0041	0.0178	-0.0178	-0.0628	4.9741
g_d	0.0012	0.1425	0.4868	-0.5373	-0.0061	4.3409
r_m	0.0055	0.0449	0.1582	-0.2290	-0.4938	4.6704
r_f	0.0009	0.0029	0.0182	-0.0109	0.0265	6.1380
$r_m - r_f$	0.0047	0.0446	0.1610	-0.5088	0.0557	4.8256

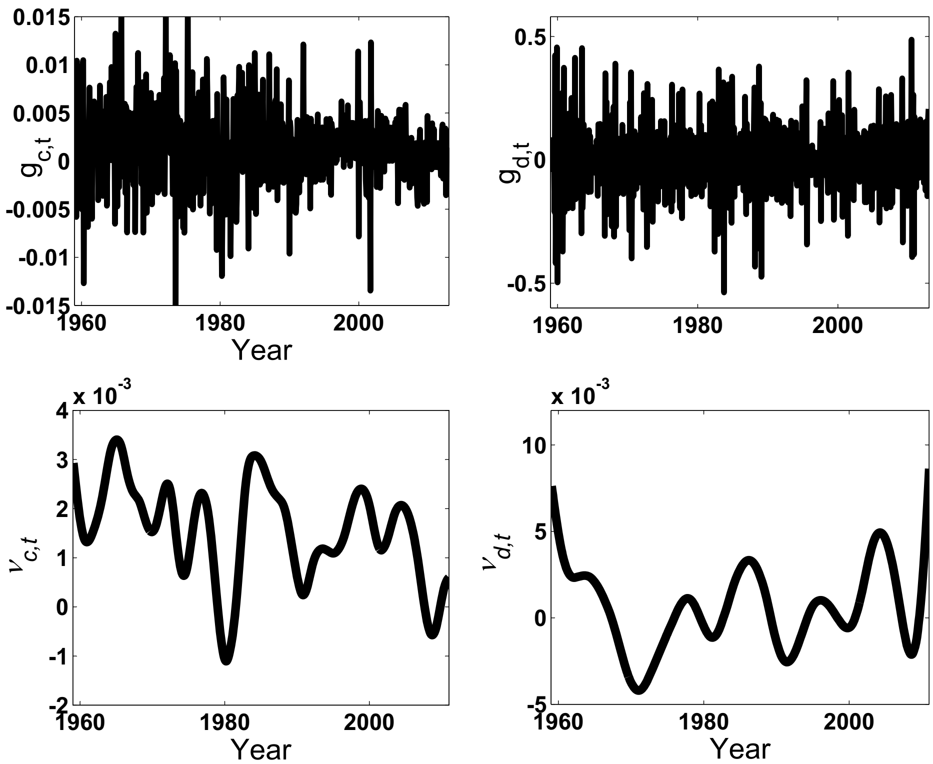
	(b) Correlations			
	g_c	g_d	$r_m - r_f$	r_f
g_c	1	0.033	0.166	0.327
g_d		1	0.252	0.068
$r_m - r_f$			1	0.067
r_f				1

Notes: Table 2.1 displays summary statistics of data used in Section 2. In panel (a), presented are the mean, the standard deviation (std.), the maximum (max), the minimum (min), the skewness, and the kurtosis. Panel (b) displays correlations of these variables. In the table, g_c, g_d, r_m, and r_f refer to the consumption growth, dividend growth, stock market return, and risk-free rate, respectively. All the variables are at monthly frequency and adjusted in real terms. Both consumption and dividend growth rates are per capita. For the market returns, dividends and risk-free rate, the database from the center for research in security prices (CRSP) is used, and the consumption data is from the St. Louis Federal Reserve database (FRED). The period covered is from January 1959 to January 2013.

closely together, and the correlation between the components is around 0.103. When $\nu_{d,t}$ is projected onto $\nu_{c,t}$ with a constant, the estimated coefficient is 0.45 with a Newey-West corrected t-statistic of 1.73. Thus, the long-run risk channel from the conditional mean growth component may exist, yet it is quantitatively weak to generate a sizable equity premium, as explained in Section 2.2.

For our Bayesian estimation, we draw 120000 samples for each parameters and latent variables by the Gibbs sampler and the MH algorithm, and discard the first 40000 samples, which are considered as samples in the burn-in period. Panel (a) of Table 2.2 shows that the estimated parameters are well converged and mostly significant. Standard errors are computed using the method of conventional Monte

Figure 2.3: Consumption and Dividend Growth Rates



Notes: The figures in the top panel present the consumption and dividend growth rates ($g_{c,t}$ and $g_{d,t}$) during the period of January 1959 to January 2013. The figures in the bottom panel show the estimated conditional mean components of consumption and dividend growth rates ($\nu_{c,t}$ and $\nu_{d,t}$).

Carlo standard error to adjust for possible autocorrelations in the MCMC samples. Though λ_c has a relatively high value of convergence diagnostic, their small standard errors show that sample series are quite concentrated after the burn-in period.[14]

We can see the α_c and β_c are smaller than α_d and β_d due to the volatile dividend process. The lower bounds of volatility for both consumption and dividend growths are $\sqrt{\hat{\alpha}_c} = 0.2\%$ and $\sqrt{\hat{\alpha}_d} = 4.24\%$, and the upper bounds of volatility are $\sqrt{\hat{\alpha}_c + \hat{\beta}_c} = 0.58\%$ and $\sqrt{\hat{\alpha}_d + \hat{\beta}_d} = 24.35\%$ in a month, respectively. The values of $\hat{\lambda}_c$ and $\hat{\lambda}_d$ indicate that the volatility generating processes $x_{c,t}$ and $x_{d,t}$ are scaled down compared to the magnitude of the macro uncertainty w_t. As mentioned early, Figure 2.1 displays the estimated volatility functions of the consumption and dividend growths, and it shows that the volatilities in the macro level evolve over time via neither an explosive process nor an abrupt regime-shifting process. Thus, when there exists a shock to the macroeconomic uncertainty w_t, this causes the volatilities of the consumption growth and the dividend growth to migrate into a transition period, and the uncertainty about the regimes of consumption and dividend growth is likely to persist over time. This increases the equity premium. Regarding the volatility of idiosyncratic component, $\hat{\sigma}_c^2$ is larger than $\hat{\sigma}_d^2$, which implies that the signal of the dividend process is stronger than that of consumption process such that the dividend volatility generating process $x_{d,t}$ is closer to the common macro volatil-

[14]The measure of convergence diagnostic (CD) is given by

$$CD = \frac{\overline{\theta}_A - \overline{\theta}_B}{\sqrt{\frac{\hat{\omega}_A^2}{n_A} + \frac{\hat{\omega}_B^2}{n_B}}},$$

where A is the set of Gibbs samples with n_A iterations after burn-in period and B is the set of Gibbs samples with last n_B observations, and $\hat{\omega}_A^2$ and $\hat{\omega}_B^2$ are their HAC estimates obtained using Parzen window. By the convention we set $n_A/n = 0.1$ and $n_B/n = 0.5$, where n denotes the number of Gibbs samples after burn-in periods. As shown in Geweke (1992), CD converges to the standard normal distribution as the number of samples goes to infinity, if the sequence of Gibbs samples for a parameter is stationary. Jacquier et al. (2004) warn that one must perform careful sampling experiments to establish convergence across a wide range of empirically relevant parameter values. Our samples show that they wander a lot at the beginning, but converge after the burn-in period.

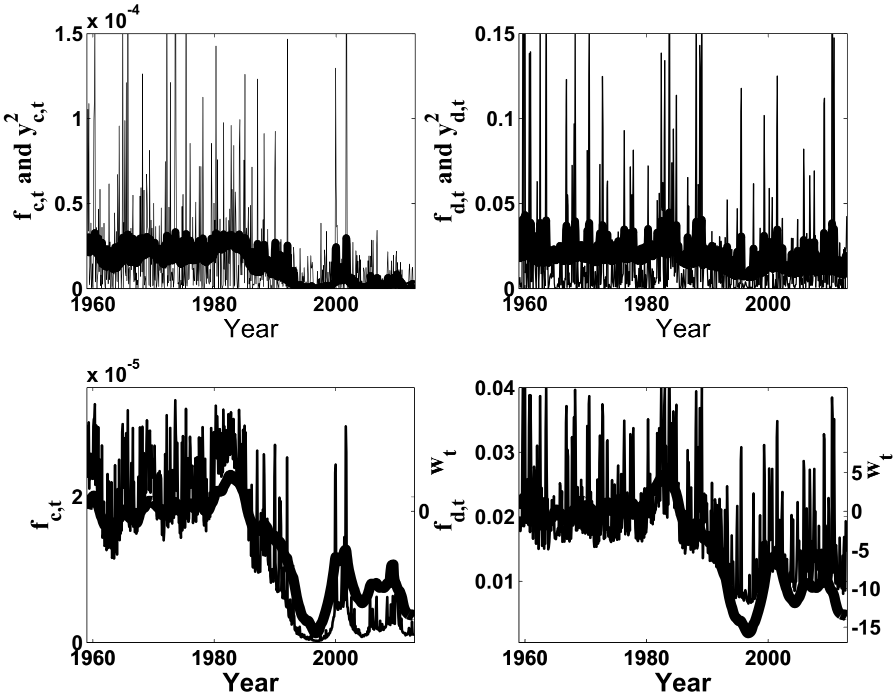
Table 2.2: Estimation Results

	Posterior				
Parameters	Mean	Std. Err.	5%	95%	Convergence
			(a) Stochastic Volatility		
α_c	0.000004	0.0000006	0.000003	0.000005	-0.1742
β_c	0.000030	0.000005	0.000022	0.000041	-0.3439
κ_c	-1.1473	0.4529	-2.0349	-0.2597	-0.6198
λ_c	0.6181	0.0910	0.4397	0.7966	-1.8360
ρ	0.0394	0.0428	-0.0436	0.1232	0.8676
σ_c	2.6602	0.8205	1.0521	4.2683	-0.3369
α_d	0.0018	0.0016	-0.0013	0.0050	0.2167
β_d	0.0575	0.0191	0.0201	0.0949	0.2283
κ_d	0.5369	0.6406	-0.7186	1.7925	0.1306
λ_d	0.1087	0.0612	-0.0112	0.2286	-0.7868
σ_d	1.7705	0.9702	-0.1312	3.6721	0.0816
ρ_w	0.9940	0.0056	0.9831	1.0050	0.7509
			(b) Preferences		
γ	2.2957	1.0371	-0.7663	5.4948	-0.5720
ψ	7.6839	5.6899	-16.7343	36.9302	0.4702

Notes: This table presents the parameter estimates of the model using the Bayesian estimation method developed in Section 2. Panel (a) displays the estimates of the stochastic volatility model for consumption growth and dividend growth, and Panel (b) reports those of the preference function. Mean, Std. Err., 5%, 95%, and Convergence refer to the sample means, the standard errors, the 5th quantiles, the 95th quantiles, and Geweke (1992)'s convergence diagnostics, respectively. Conventional conjugate priors are used to draw samples in the beginning of the estimation, and the first 40,000 samples out of 120,000 samples are considered as the burn-in period and discarded.

ity generating process w_t. The estimated value of $\hat{\rho}$ indicates that the correlation between the short-run shocks is virtually zero.

Figure 2.4: Volatility Factors of Consumption and Dividend Growth Rates.



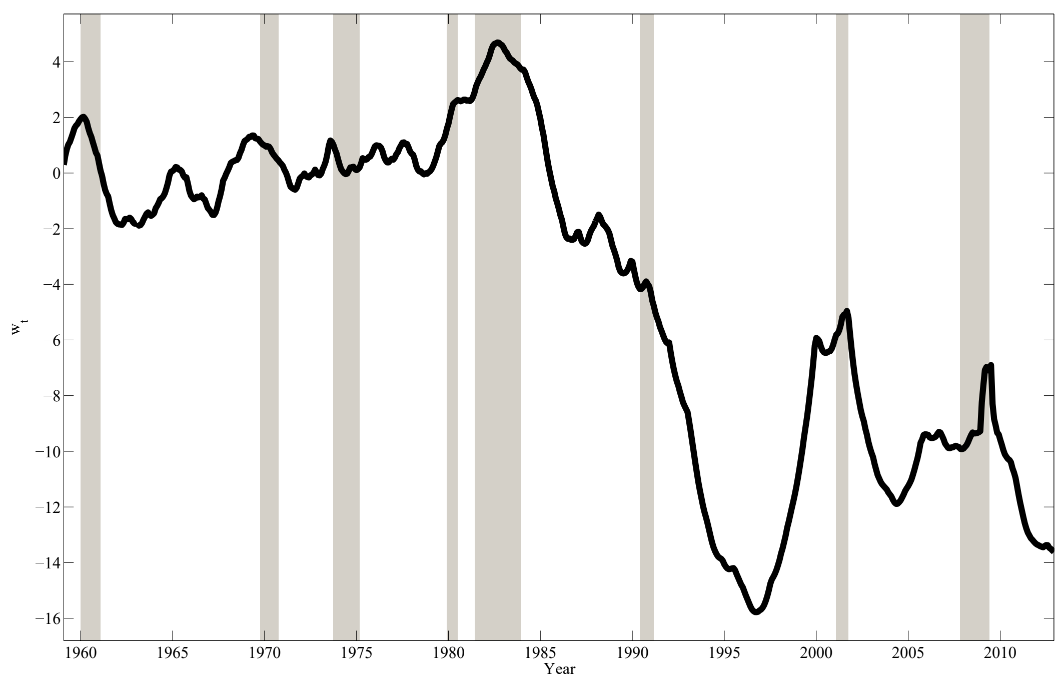
Notes: The figures in the top panel show the estimated conditional variances ($f_{c,t}$ and $f_{d,t}$) in conjunction with $y^2_{c,t}$ and $y^2_{d,t}$, where $y_{j,t}$, $j = c, d$ is defined as $y_{j,t} = \sqrt{f_j(x_{j,t})}\,\varepsilon_{j,t+1}$ in equation (2.20). The bottom panel displays the extracted common latent factor w_t in thick line and the estimated conditional variances of consumption and dividend growth rates.

Figure 2.4 displays the estimated latent common macroeconomic uncertainty process w_t and the estimated volatility generating processes for the consumption and dividend growth rates $x_{c,t}$, $x_{d,t}$. Volatility-generating processes of both consumption growth and dividend growth are highly persistent, due to the common macroeconomic uncertainty process w_t. In fact, the value of the AR(1) coefficient, ρ_w is around 0.99. We believe that this is another long-run risk channel that affects both the stochastic discount factor and the cash flow volatility.

Figures 2.1 and 2.4 allow us to think about thresholds for a regime change. To be more specific, we define the transition period by selecting the endpoints of this interval at $f'''(x) = 0$. Then the area above the upper boundary can be regarded as the high volatility regime and the region below the lower boundary as the low volatility regime. We can interpret that the fundamental factors that generate volatility stayed in the transition regimes in the 1960s and 1970s, then entered into the high volatility regime in the early 1980s. From the mid-1980s until the early 1990s the common factor remained again in the transition period with sharply decreasing trends. Except for the 2001-2002 period and the 2008 financial crisis period, the US economy was in the low volatility period, often referred to as the great moderation period. Notably, during most of the post-war period, the macroeconomic volatility was in transition periods. Thus, models without incorporating the intermediate regimes are likely to be misspecified in capturing business cycles. At the same time, the existence of the upper and lower boundaries is necessary to keep this volatility process from exploding, given its strong persistence. Knowing the maximum distance of consumption volatility from a current level of uncertainty, what matters to consumers would be how fast news on uncertainty evolves over time and how closely the current level of uncertainty approaches the maximum or the minimum level. In this regard, economic agents' preferences on different persistence of a fundamental shock are crucial to account for the empirical evidence of high equity premium, despite the small level of the consumption volatility.

Figure 2.5 shows the relation between the common macro uncertainty factor w_t and business cycle fluctuations recorded by the National Bureau of Economic

Figure 2.5: Business Cycles and the Macroeconomic Uncertainty



Notes: The solid line represents the common volatility generating process w_t and the shadow area stands for the recession period from the NBER.

Research (NBER). We find that the recession periods are well-matched with the local peaks of the common volatility generating process. In fact, the common volatility factor reached its local peak just before the entry into the 1970s, 1980s, the 2001 recession, and the 2007-2008 financial crisis. Even the 1991 mini credit crunch was predicted by a small, but a conspicuous increase in the macro uncertainty factor w_t. According to the previous literature, one of the most foundational links between asset returns and macroeconomic variables is that the expected asset returns are higher at business cycle troughs, reflecting the risk-return trade-offs with counter-cyclical variations. Our macroeconomic uncertainty factor rises and reaches its local maximum when a recession begins, i.e., an economic downturn starts around the time of rising aggregate uncertainty, and the degree of the uncertainty becomes lower, as the recession period is coming to an end. Thus, we expect that this is going to be well-connected to a higher risk premium demanded by asset market participants. In the next section, we discuss this link.

2.4.2 Asset Pricing Implications

In this subsection, we study the implications of our model on asset prices, using numerical simulations. For this purpose, we focus on the roles played by two core preference parameters (γ and ψ) as well as the parameters of the volatility function. Panel (b) of Table 2.2 shows that the estimate of the relative risk aversion parameter γ is around 2.3 and accurately measured. Mehra and Prescott (1985) consider that a maximal level for relative risk aversion is around 10 and Barro (2006) argues that the usual view in the related literature is between 2 and 5. In this light, our result for the risk aversion parameter is reasonable and it is, in fact, a significant improvement over the existing studies which reported the estimated values of 15 to over 100.[15]

[15] Although the external habit formation allows a low value of γ, the relative risk aversion in this setup is not γ, but associated with γ/X_t, where $X_t$is a time-varying consumption surplus factor

For the elasticity of intertemporal substitution ψ, we have the estimates of 7.7, and the measurement is somewhat noisy. Several papers consider that a plausible value of ψ exceeds 1. For example, the benchmark value in case of Bansal and Yaron (2004) is 1.5, and Hansen (1982), Attanasio and Weber (1989), Vissing-Jorgensen (2002), Jeong et al. (2015), *etc* estimate the parameter in excess of 1.5. In most studies, however, their estimates have large standard errors, and our case is not an exception. At least, we confirm that our finding for the EIS is similar to the results from the existing studies. One important point related to our estimates is that $\hat{\gamma} > 1/\hat{\psi}$ holds. This implies that early resolution is preferred in this economy.[16] Thus, our volatility setup can generate a channel for risk premium as discussed in the previous section.

To evaluate the performance of our asset pricing model, we simulate the 650 data points series of consumption and dividend for 1000 times using the parameters in Table 2.2 as our baseline case. From these simulated data, we generate asset returns, which are reported as the baseline result in Table 2.3. Our simulated equity premium is 5.35%, and the risk free rate is around 1.5% per year. This is consistent with most of the empirical results. We match the second moments of the equity premium and the risk free rate as well. When compared with the existing studies, our model performs well. We believe that the results come from a realistic model of the time-varying macroeconomic uncertainty, combined with the recursive preferences.

We run the sensitivity check to understand the roles of the risk aversion coefficient and the EIS parameter. First, we increase the risk aversion coefficient from 1 to 10. Not so surprisingly, larger values of the risk aversion parameter generate higher equity premiums. In fact, the equity premium becomes too large compared with the actual

and the relative risk aversion is still very high.

[16] In a recent experimental study by Brown and Kim (2013), they show that most subjects prefer early resolution and the risk aversion is greater than the reciprocal of the estimated EIS.

Table 2.3: Model Implied Annual Asset Returns

	RRA(γ)	EIS(ψ)	$E(r_f)$	$\sigma(r_f)$	$E(r_m)$	$\sigma(r_m)$	$E(r_m - r_f)$
Data			0.0108	0.0099	0.0660	0.1887	0.0552
Base Line	2.30	7.70	0.0147	0.0015	0.0682	0.2073	0.0535
Change in γ	1.1	7.70	0.0232	0.0017	0.0509	0.1682	0.0276
	3		0.0131	0.0020	0.0790	0.3758	0.0659
	5		0.0081	0.0025	0.1225	0.5554	0.1143
	10		0.0003	0.0036	0.2805	0.9010	0.2802
Change in ψ	2.30	0.67	0.0579	0.0124	0.0757	0.0916	0.0178
		1.5	0.0365	0.0055	0.0688	0.2066	0.0323
		5	0.0185	0.0020	0.0688	0.2925	0.0503
		10	0.0143	0.0014	0.0689	0.3132	0.0556

Notes: This table shows simulation results of the model, using the estimated parameters. For simulation, 650 data points of the aggregate consumption and dividend processes are generated for 1000 times, plugging the estimated parameters into the asset pricing model. Then, asset prices are numerically computed. The first row (Data) refers to the sample moments from the data in annual terms, the second row (Base Line) reports the simulation results with the estimated parameters. The third row (Change in γ) displays the simulation results when different relative risk aversion (RRA)coefficients are used, holding constant the coefficient of the elasticity of intertemporal substitution (EIS), and the fourth row (Change in ψ) varies the EIS, fixing the RRA.

data when the risk aversion is over 5. Next, we change the EIS parameter. We vary the EIS parameter between 0.67 and 10, holding the risk aversion parameter fixed at the benchmark value. We borrow the range of the EIS parameter from the existing studies, such as Bansal and Yaron (2004), Jeong et al. (2015) and Hansen et al. (2007). According to the result, a low value of EIS decreases the equity premium by increasing the risk-free rate. Once the EIS is greater than one, the equity premium starts increasing significantly. However, as the EIS parameters reaches 5, the equity premium barely increases. Therefore, one can infer that a weak identification problem can exist in case of identifying the EIS parameter.[17]

Finally, we study in detail the volatility channel of our model in producing the key

[17]For a more discussion on the nature of the weak identification for the EIS in the context of Epstein-Zin-Weil preferences, see Jeong et al. (2015).

Table 2.4: Volatility Channel and Asset Returns: Robustness Check

Parameters			$\mathbb{E}(r_f)$	$\sigma(r_f)$	$\mathbb{E}(r_m)$	$\sigma(r_m)$	$\mathbb{E}(r_m - r_f)$
Changes in ρ_ω	0.9		0.0164	0.0012	0.0610	0.2007	0.0446
	0.7		0.0165	0.0013	0.0582	0.1203	0.0417
	0.5		0.0131	0.0006	0.0529	0.1189	0.0398
Changes in λ_c, λ_d	0.1	0.1	0.0165	0.0012	0.0661	0.3044	0.0496
	0.5	0.5	0.0159	0.0015	0.0682	0.3073	0.0523
	100	100	0.0147	0.0018	0.0680	0.3130	0.0533

Notes: Table 2.4 reports simulation results when volatility parameters vary. The top panel shows the moments of asset returns when the persistence of the volatility factor changes, and the bottom panel refers to the case with different degrees of smooth transitions. That is, the parameter ρ_ω refers to the $AR(1)$ coefficient for the latent volatility factor ω_t (i.e., $\omega_t = \rho_\omega \omega_{t-1} + u_t$) and λ_j determines the slope of the transition as w changes. The conditional volatilities for consumption and dividend ($j = c, d$) are defined as

$$f_j(w_{j,t}) = \alpha_j + \frac{\beta_j}{1 + \exp\left[-(\lambda_j \omega_t + e_{j,t} - \kappa_j)\right]}.$$

moments of asset returns by simulating the model with alternative parameter values of the volatility setup. Table 2.4 tabulates the result. First, we simulate the model with different values of the persistence of the latent volatility factor ω_t denoted as ρ_ω. When the latent factor becomes less persistent, the model has difficulty in matching data, especially in terms of the average equity premium and the average market volatility. Since this parameter determines how slowly macroeconomic uncertainty unfolds over time, a lower value of this parameter weakens the volatility channel. For instance, with a value of ρ_ω being 0.5, the model produces 3.98% of the equity premium per year, and the volatility of the market return is low around 12%.

The second set of simulations try different values of the smooth transition parameter λ. As discussed early, this parameter has two effects on the conditional volatility of consumption and dividend growth rates. From the definition of volatility function $f(\lambda\omega)$, a lower λ decreases the level of volatility which reduces equity premium. At the same time, however, from equations (2.16) to (2.19), a lower λ creates a slower and more dampened transition that leads to a higher premium from volatility risk. Thus, if the second channel is quantitatively important, simulated equity premiums is high, despite a small amount of volatility. Indeed, this is the case, according to the bottom panel of Table 2.4. With the low value of both λ_c, λ_d at 0.1, the model can generate an equity premium of 5%. Note that higher values of these parameters (λ_c and λ_d) refer to stochastic volatility models close to a two-regime switching setup. With counterfactual values of 100 for both λ_c and λ_d, the simulated equity premium is around 5.33%, but the average variance of excess returns is much higher than the empirical evidence. This again shows the importance of the smooth transition that amplifies the uncertainty premium. Overall, our asset pricing model accounts for the key characteristics of the asset return data with plausible values of the key preference parameters.

2.5 Conclusion

Macroeconomic models have strong theoretical implications on asset prices. However, empirical links between macroeconomic and financial variables appear to be weak according to the existing studies which report a small level of macroeconomic volatility as the main suspect. Macroeconomic uncertainty is latent to econometricians, and therefore, a measurement problem exists. Furthermore, a valid transmission mechanism is required to amplify its small and smooth dynamics into large fluctuations in financial variables. In Section 2 we addressed the issue by developing and estimating a consumption-based asset pricing model with an emphasis on new stochastic volatility function. To quantify the macroeconomic uncertainty, we introduce a non-linear stochastic volatility model with consumption and dividend processes. In particular, we assume that the volatility function takes a logistic form with a common factor as well as a stationary idiosyncratic factor. That is, this setup has two asymptotic regimes with a continuum of smooth transitions and the fluctuations of the common latent factor. The main advantage of this setup is that volatility can be persistent without explosive dynamics. This is not only realistic to describe macroeconomic volatility but also relevant for producing uncertainty premiums because it allows slow and potentially long transition periods, perceived to be risky by consumers, even if the level of volatility is not high. We develop a Bayesian algorithm and identify this model numerically to avoid the difficulties with multidimensional integrations. We find that the extracted volatility factor explains well the realized volatility series of both the consumption and the dividend data. In addition, we see a counter-cyclical relation of the extracted macroeconomic uncertainty. To make full use of our volatility setup, we then combine this volatility model with a recursive utility function which has a preference ordering on the persistence of shocks. Therefore, the key insight of our model is that news about how rapidly the macro uncertainty reaches its highest or lowest point can also be a source of risk because the future outlook on an economy is unknown, and this becomes riskier if the current level of uncertainty is in transit.

We show that the estimated risk-aversion coefficient is around two, and the intertemporal elasticity of substitution is greater than one. Recent works adopting Epstein-Zin-Weil preferences report the relatively high estimates of the risk aversion coefficient to generate high risk premium through the persistent long-run risk channel. However, our model produces a high equity premium even without resorting to the common long-run risk component in the conditional means of consumption and dividend growth rates. Based on our theoretical and empirical results, we argue that the common, non-linear stochastic volatility of macroeconomic variables are crucial in understanding business cycle fluctuations and the related asset return dynamics.

3. USING HODRICK-PRESCOTT FILTER TO NONPARAMETRICALLY ESTIMATE CONDITIONAL MEAN PROCESS

Section 3 is organized as follows. In Section 3.1, we introduce our model and methodology where we compare it with the conventional HP filter and the Local Linear estimator commonly used for fitting conditional mean nonparametrically. We present our new approach in both discrete and continuous time models. Section 3.2 reports our simulation studies, which evaluate the performance of our new HP estimator relative to the conventional Local Linear estimator in both discrete and continuous time models considered in Section 3.1. The empirical illustrations presented in 3.3 consist of the analysis of the US GDP growth rates, inflation rates, unemployment rates and short term interest rates at both quarterly and monthly frequencies. Section 3.4 concludes this Section, and Appendix B collects additional figures.

3.1 Model and Methodology

3.1.1 Discrete Time Model

We consider a time series (y_t) defined as

$$y_t = m_t + u_t, \tag{3.1}$$

where (u_t) is a martingale different sequence with respect to the filtration $(\mathcal{F}_t)$ signifying the information available at time t. It follows immediately that we have

$$m_t = \mathbb{E}(y_t|\mathcal{F}_{t-1}),$$

since $\mathbb{E}(u_t|\mathcal{F}_{t-1}) = 0$, and (m_t) is the conditional mean process of (y_t) with respect to the filtration $(\mathcal{F}_t)$. Clearly, (m_t) and (u_t) in (3.1) represent respectively the predictable and unpredictable component of (y_t). Note that the conditional mean process (m_t) includes the cyclical component as well as the trend component of (y_t), i.e.,

$$m_t = \tau_t + c_t,$$

where τ_t and c_t signify the trend and cyclical components of y_t.

In what follows, we assume that (y_t) is observed for $t = 1, \ldots, n$, where n denotes the sample size.

3.1.1.1 Hodrick-Prescott Filter In Section 3, we propose to use the Hodrick-Prescott (HP) filter to estimate the conditional mean process m_t of y_t. We propose the use of a modified version of the Hodrick-Prescott filter as an alternative way of estimating conditional mean process m_t nonparametrically. The conventional HP filter decomposes a variable into trend and cyclical component, and has been widely used as an effective tool to study business cycles.

Here we are suggesting to use the HP approach for a different objective and use to filter out the conditional mean process m_t, which is the predictable part of the observed time series y_t. For a given filtration $\mathcal{F}_t$, y_t can be decomposed of its conditional mean $m_t = E(y_t|\mathcal{F}_{t-1})$ that includes both trend τ_t and cyclical component c_t of y_t, and a sequence of martingale difference unpredictable error component, u_t. One of the key characteristics of a martingale difference sequence is that it has no serial correlation. Thus, if we apply HP procedure with an optimal smoothing parameter λ^* such that the filtered-out residuals' sample auto-correlation to be zero, then the fitted process m_t^* will capture all of the predictable components of the time series given the past information and therefore may successfully filter out the conditional mean process nonparametrically.

To introduce our estimator, we let

$$m(\lambda) = \min_{m} \left(\sum_{t=1}^{n} (y_t - m_t)^2 + \lambda \sum_{t=2}^{n-1} \Big[(m_{t+1} - m_t) - (m_t - m_{t-1}) \Big]^2 \right) \tag{3.2}$$

with $m(\lambda) = \Big(m_t(\lambda)\Big)$ for each $\lambda > 0$. The first term in the objective function is the sum of squared residuals $\sum_{t=1}^{n} (u_t)^2$ that penalizes poor fit. The second term puts a prescribed penalty to the lack of smoothness in m_t. Note that we can allow for any deterministic linear trend without penalty in HP filter. We define

$$\lambda^* = \underset{\lambda>0}{\operatorname{argmin}} \Big|\rho(\lambda)\Big|$$

where

$$\rho(\lambda) = \left(\sum_{t=1}^{n} u_t^2(\lambda) \right)^{-1} \sum_{t=2}^{n} u_t(\lambda) u_{t-1}(\lambda)$$

with $u_t(\lambda) = y_t - m_t(\lambda)$ is the first order sample auto-correlation of the residuals from the trend extraction of a time series (y_t) using the HP filter with the penalty parameter λ. The estimator $(\hat{m}_t)$ of (m_t) we propose is then given by

$$\hat{m}_t = m_t(\lambda^*). \tag{3.3}$$

Therefore, our estimator $(\hat{m}_t)$ of (m_t) is defined simply as the predictable component of (y_t) extracted using the HP filter with the penalty parameter minimizing the first order sample correlation of the residuals.

3.1.1.2 Choice of Smoothing Parameter λ Let $u_t(\lambda) = y_t - m_t(\lambda)$ be the residuals from y_t after extracting $m_t(\lambda)$ using the HP smoothing parameter λ. In estimating conditional mean process, we use the optimal smoothing parameter λ^* which makes the fitted residuals as close as possible to martingale difference sequence process by minimizing the absolute value of the first order sample correlation of the fitted residual process as described above.

Here we make two comments on the smoothing parameter λ used for the HP filter and the bandwidth parameter h used for the Local Linear estimator. First, we note that the HP filter smoothing parameter λ. As $\lambda \to \infty$, $\hat{m}_t$ approaches to the linear trend. i.e.,

$$\hat{m}_t = \hat{\alpha} + \hat{\beta}(t/n),$$

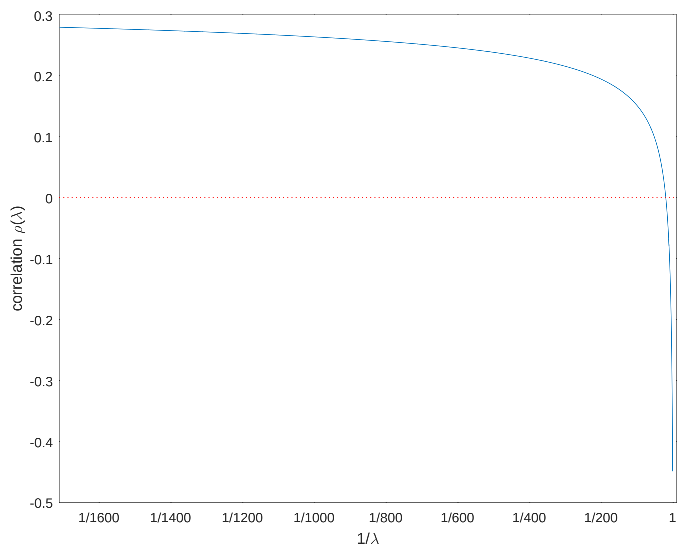
where $\hat{\alpha}$ and $\hat{\beta}$ are the least squares estimators for the intercept and slope parameters. On the other hand, $\hat{m}_t$ becomes equivalent to the process itself y_t as $\lambda \to 0$. We also note that as the bandwidth parameter h for the Local Linear estimator increases to infinity, $\tilde{m}_t$ approaches to the sample mean. i.e.,

$$\tilde{m}_t = \frac{1}{n}\sum_{t=1}^{n} y_t.$$

However, for $h = 0$, $\tilde{m}_t$ becomes equivalent to y_t.

As an illustration, Figure 3.1 present $\rho(\lambda)$ for the quarterly US real GDP growth rates as a function of $1/\lambda$. Note that $\rho(\lambda)$ is positive for $1/\lambda$ small, decreases monotonically as $1/\lambda$ increases and crosses the horizontal axis to become negative eventually for $1/\lambda$ large. We have λ^* given by $\lambda^* = 20.03$. For the US quarterly macroeconomic data including GDP growth rates, it is recommended to use $\lambda_Q = 1,600$. It is natural and well expected that $\lambda^* \ll \lambda_Q$, since the HP filter is used originally to extract only the persistent trend component, while we adopt it here to filter out the transitory cyclical component as well, of a given time series. We have $\rho(\lambda^*) = 0$, though $\rho(\lambda_Q) = 0.278$. For comparisons, we provide in Figure 3.2 the extracted

Figure 3.1: Optimal Smoothing Parameter λ^* for Quarterly GDP Growth Rates



Notes: In this figure, we present $\rho(\lambda)$ for the quarterly US real GDP growth rates as a function of $1/\lambda$. The $x-$axis is for inverse of λ so that the residual autocorrelation starts from some positive value at the origin. The $y-$axis is for the residual sample autocorrelation $rho(\lambda)$ after the estimated $\hat{m}_t(\lambda)$ has been subtracted from the quarterly GDP growth rates.

components with $\lambda = \lambda_Q$ and λ^*. Also presented there are the original time series and the fitted residual process obtained with $\lambda = \lambda_Q$ and λ^*.

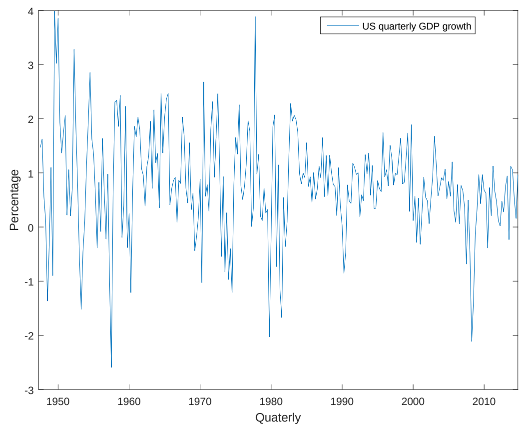
The pattern we observe in Figure 3.1 for the quarterly US real GDP growth rates is typical for many economic time series. It is indeed well expected. Note that we have

$$u_t(\infty) = y_t - m_t(\infty) = y_t - \bar{y}_n$$

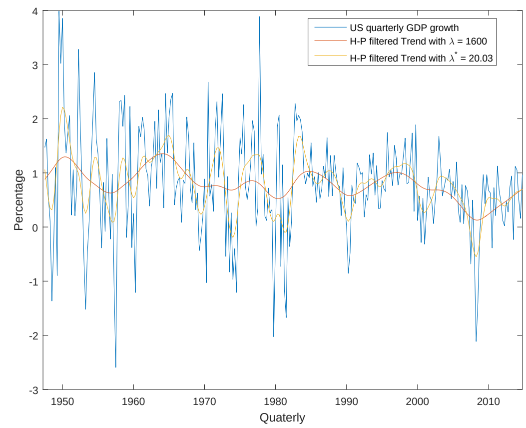
for $t = 1, \ldots, n$, where $\bar{y}_n = (1/n)\sum_{t=1}^n y_t$. It follows that

$$\rho(\infty) = \left(\sum_{t=1}^n (y_t - \bar{y}_n)^2\right)^{-1} \sum_{t=2}^n (y_t - \bar{y}_n)(y_{t-1} - \bar{y}_n),$$

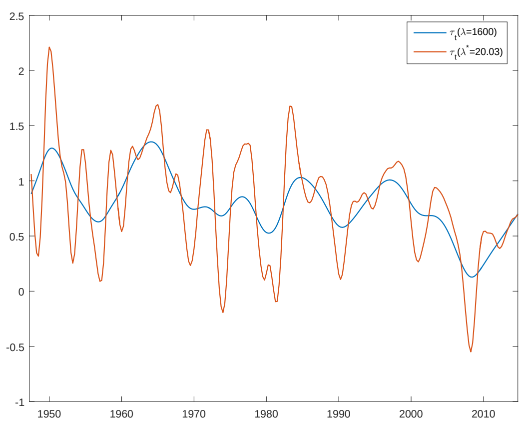
Figure 3.2: Real GDP Growth Rates (Quarterly)



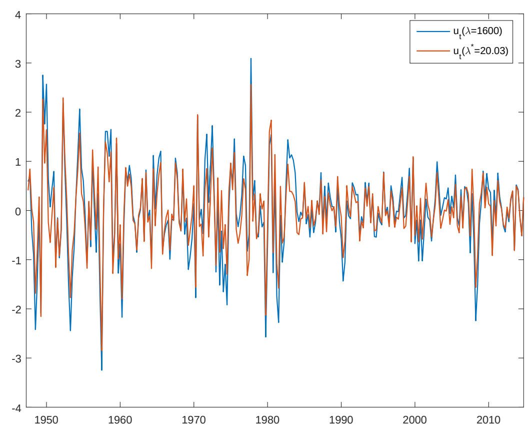
(a) Quarterly Real GDP Growth Rates



(b) Fitted GDP Growth Rates



(c) Fitted Means with λ_Q and λ^*



(d) Fitted Residuals: $u_t(\lambda_Q)$ and $u_t(\lambda^*)$

Notes: In this figure, we illustrate our empirical result using Quaternary real GDP growth rates. In panel (a), time series of U.S. GDP growth rate is plotted. Panel (b) and (c) compares the HP filtered trend with λ_Q and λ^*. Our optimal smoothing parameter λ^* is found at 20.03, which is well below the conventionally used 1600. Panel (d) compares the fitted residuals $u_t(\lambda_Q)$ and $u_t(\lambda^*)$ after subtracting $\tau_t(\lambda_Q)$ and $\tau_t(\lambda^*)$.

i.e., the usual sample autocorrelation of (y_t), and therefore, we expect $\rho(\lambda) > 0$ at $1/\lambda = 0$ for many economic time series as we have for the US GDP growth rates. Moreover, we naturally expect that the positive first order sample autocorrelation $\rho(\lambda)$ of the residuals $\big(u_t(\lambda)\big)$ decreases as $1/\lambda$ increases from zero, since $\big(m_t(\lambda)\big)$

becomes time varying and starts to explain some positive correlation $\rho(\lambda)$ of (y_t). However, if $1/\lambda$ becomes too large, $\big(m_t(\lambda)\big)$ becomes too wiggly and generates spurious negative sample autocorrelation $\rho(\lambda)$ of the residuals $\big(u_t(\lambda)\big)$.

3.1.1.3 Nonparametric Estimation of Conditional Variance We may use the HP filter to estimate the conditional variance process as well. To show this, we let

$$z_t = \sigma_t w_t$$

where it is assumed that (σ_t) is adapted to a filtration $(\mathcal{F}_{t-1})$ such that $\mathbb{E}(w_t|\mathcal{F}_{t-1}) = 0$ and $\mathbb{E}(w_t^2|\mathcal{F}_{t-1}) = 1$. Under the assumption, (σ_t^2) is identified as the conditional variance process of (z_t), and we have

$$z_t^2 = \sigma_t^2 + \sigma_t^2(w_t^2 - 1),$$

which we may write as in (3.1) with

$$y_t = z_t^2, \quad m_t = \sigma_t^2, \quad u_t = \sigma_t^2(w_t^2 - 1).$$

Note in particular that (u_t) with $u_t = \sigma_t^2(w_t^2 - 1)$ is a martingale difference sequence with respect to the filtration $(\mathcal{F}_t)$. Of course, it is also possible to estimate both the conditional mean and conditional variance processes of a given time series by applying the HP filter successively. We estimate the conditional mean process of a given time series in the first step with our choice of the smoothing parameter λ^*, and successively in the second step we get the estimated conditional variance of the given time series by estimating the conditional mean of the squared residual process obtained from the first step.

3.1.1.4 Conventional HP filter The conventional HP filter decompose y_t into trend and cyclical components:

$$y_t = \tau_t + c_t$$

By taking derivatives of the objective function w.r.t $(\tau_t)_{t=1}^n$ and rearranging them, the solution can be written as following matrix form

$$y = (\lambda F + I_n)\tau.$$

The trend and cyclical component τ and c can be written as

$$\tau = (\lambda F + I_n)^{-1} y$$
$$c = y - \tau$$

If we assume that c_t and $\triangle^2\tau_t$ are i.i.d $\mathbb{N}(0, \sigma_1^2)$ and $\mathbb{N}(0, \sigma_2^2)$, then the conditional expectation of τ_t given the observations would be the solution to the program when $\sqrt{\lambda} = \sigma_1/\sigma_2$. Hodrick and Prescott (1980) and Hodrick and Prescott (1997) assume 5% variability in cyclical component and (1/8)% change in growth rate of trend component in a quarter, and suggest to use $\lambda = 1600$, i.e., $\sqrt{\lambda} = 5/(1/8) = 40$. King and Rebelo (1993) discuss the properties of the filter in both time and frequency domain. Ravn and Uhlig (2002) study how to adjust HP smoothing parameter when changing the frequency of observations and suggest to use $\lambda = 6.25$ for annual data. de Jong and Sakarya (2015) proposed a new representation for HP filter without using ARMA based approximation.

3.1.1.5 Local Linear Kernel Estimator The conditional mean process (m_t) of (y_t) can also be estimated nonparametrically using the Local Linear kernel estimator

$(\tilde{m}_t)$ with cross-validation bandwidth smoothing from the nonparametric regression of (y_t) on time t for $t = 1, \ldots, n$, which is given by

$$\tilde{m}_t = \underset{-\infty<a<\infty}{\operatorname{argmin}} \sum_{s=1}^{n} \Big(y_s - a - b(t-s) \Big)^2 K\left(\frac{t-s}{nh}\right) \tag{3.4}$$

for $t = 1, \ldots, n$, where K is the kernel function and h is the bandwidth parameter. The selection of bandwidth parameter h is critical for the finite sample performance of $(\tilde{m}_t)$ as we discuss further below.

To more formally discuss the properties of estimators, we often assume that the conditional mean process (m_t) is generated from a function $\omega : [0,1] \to \mathbb{R}$ as

$$m_t = \omega(t/n) \tag{3.5}$$

for $t = 1, \ldots, n$. Strictly speaking, the conditional mean process in this case should be defined more formally as a triangle array of random variables and denoted by $(m_{n,t})$ instead of (m_t). However, for notational simplicity, we continue to denote it by (m_t). This should cause no confusion.

If the conditional mean process (m_t) is generated as in (3.5), we may define $\tilde{\omega} : [0,1] \to \mathbb{R}$ correspondingly as

$$\tilde{\omega}(r) = \underset{-\infty<a<\infty}{\operatorname{argmin}} \sum_{s=1}^{n} \Big(y_s - a - b(r - s/n) \Big)^2 K\left(\frac{r - s/n}{h}\right),$$

for $r \in [0,1]$, so that $\tilde{m}_t = \tilde{\omega}(t/n)$. In this case, $\tilde{\omega}$ reduces to a usual nonparametric estimator of ω, for which all well known asymptotic properties of kernel estimator hold under appropriate regularity conditions for the kernel function K, the bandwidth parameter h and the regression function ω, as well as for the observed time series (y_t). Selection of bandwidth is critical for the performance of $\tilde{m}_t$. We may use the data-dependent, leave-one-out cross-validation method, and search for the bandwidth $h = cn^{-1/5}$ over the range $0 \leq c \leq 10$. It is well known that such data driven cross-validation method yields an asymptotically optimal bandwidth, which also performs well in finite samples, for a wide class of models. See Li and Racine (2004) for more details.

3.1.2 Continuous Time Model

Our approach can also be used to estimate the conditional mean process in a continuous time model

$$dY(r) = \mu(r)dr + dU(r), \tag{3.6}$$

where U is a martingale with respect to a filtration $(\mathcal{F}(r))$ so that $\mathbb{E}(dU(r)|\mathcal{F}(r)) = 0$, and $\mu(r)dr$ represents the instantaneous conditional mean of $dY(r)$. To estimate μ using the HP filter, we note

$$\begin{aligned} Y(t\delta) - Y\Big((t-1)\delta\Big) &= \int_{(t-1)\delta}^{t\delta} \mu(r)dr + \Big(U(t\delta) - U\Big((t-1)\delta\Big)\Big) \\ &\approx \delta\mu((t-1)\delta) + \Big(U(t\delta) - U\Big((t-1)\delta\Big)\Big), \end{aligned}$$

and define $y_t = Y(t\delta) - Y((t-1)\delta)$, $m_t = \delta\mu((t-1)\delta)$ and $u_t = U(t\delta) - U((t-1)\delta)$ to have

$$y_t \approx m_t + u_t \tag{3.7}$$

for small $\delta > 0$. If $Y = \Big(Y(r)\Big)$ is observed at sampling interval $\delta > 0$ over $r \in [0, T]$ with $T = n\delta$, then we have (y_t) for $t = 1, \ldots, n$, and therefore, we may estimate (m_t) using the HP filter for $t = 1, \ldots, n$, from which we may obtain an estimate for $\mu = \Big(\mu(r)\Big)$ at $r = 0, \ldots, (n-1)\delta$.

If we have

$$dZ(r) = \sigma(r)dW(r) \tag{3.8}$$

in continuous time, where W is the standard Brownian motion with respect to some filtration $\Big(\mathcal{F}(r)\Big)$, to which σ is adapted, we have

$$\begin{aligned}\Big(Z(t\delta) - Z\big((t-1)\delta\big)\Big)^2 &= \int_{(t-1)\delta}^{t\delta} \sigma^2(r)dr + 2\int_{(t-1)\delta}^{t\delta} \Big(Z(r) - Z\big((t-1)\delta\big)\Big)dZ(r) \\ &\approx \delta\sigma^2\big((t-1)\delta\big) + 2\int_{(t-1)\delta}^{t\delta} \Big(Z(r) - Z\big((t-1)\delta\big)\Big)dZ(r)\end{aligned}$$

for small $\delta > 0$. Therefore, we may let $y_t = \Big(Z(t\delta) - Z\big((t-1)\delta\big)\Big)^2$, $m_t = \delta\sigma^2\big((t-1)\delta\big)$ and

$$u_t = 2\int_{(t-1)\delta}^{t\delta} \Big(Z(r) - Z\big((t-1)\delta\big)\Big)dZ(r)$$

to have (3.7). Therefore, we may estimate the conditional variance process σ^2 from the continuous time model (3.8) using the HP filter.

It is also obvious that we may estimate the conditional mean process μ and variance process σ^2 from the continuous time model

$$dY(r) = \mu(r)dr + \sigma(r)dW(r) \tag{3.9}$$

by successively applying the HP filter. The process Y specified as in (3.9) is often referred to as an Ito process and used widely in economics and finance.

3.2 Relative Performance of HP Filter

Our simulations show the performance of the HP filter relative to the Local Linear kernel estimator for two baseline models in discrete time and continuous time considered in the previous section.

3.2.1 Discrete Time Model

3.2.1.1 Model DT0 For the discrete time model (3.1), we consider (y_t) generated as

$$y_t = \beta x_{t-1} + u_t \quad \text{with} \quad x_t = \alpha x_{t-1} + v_t, \tag{3.10}$$

where (u_t) and (v_t) are both i.i.d. standard normals with $\mathrm{cor}(u_t, v_t) = \rho$. Clearly, we may write (y_t) as in (3.1) with $m_t = \beta x_{t-1}$. The regressor x_t is generated as an auto-regressive process with auto-regressive parameter α and the innovation v_t generating the regressor is allowed to correlated with the error process u_t. In this simple regression setup, the conditional mean of y_t is βx_{t-1}. The regression coefficient β represents the magnitude of the conditional mean for a given noise level, and the signal becomes clearer as β increases. On the other hand, the AR coefficient α controls the persistency of the regressor x_t, and therefore both α and β magnifies m_t.

3.2.1.2 Simulation Setup and Results In our simulations, we consider stationary x_t with the values of the AR coefficient α that controls the persistency of the regressor are taken from the values starting from 0.1 to 0.9 with the increment of 0.1, excluding the unit-root case with $\alpha = 1$. One period lagged x_{t-1} is then multiplied by β and the observed y_t is a mixture of the conditional mean βx_{t-1} and an iid normal error process u_t. The parameter β determining the magnitude of the conditional mean is chosen from the set of values $(0.1, 0.5, 1, 5, 10, 20)$. We jointly generate the error process u_t and the innovation for the regressor process x_t as

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} =_d \mathbf{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

with the endogeneity parameter ρ selected from $(0, 0.5)$.

We simulated 10,000 sample paths of y_t with two sample sizes $n = 100$ and $n = 500$.[1] For all our simulations, we use the leave-one-out cross-validation method, we search for the bandwidth $h = cn^{-1/5}$ over the range $0 \leq c \leq 10$ to find the optimal bandwidth for the Local Linear estimator.

3.2.1.3 Model DT1 To disentangle the joint effects of the two parameters β and α on the magnitude of the conditional mean, we slightly modify the data generation process given in 3.10 as

$$y_t = \beta(x_{t-1}\sqrt{1-\alpha^2}) + u_t \tag{3.11}$$

$$x_t = \alpha x_{t-1} + v_t \tag{3.12}$$

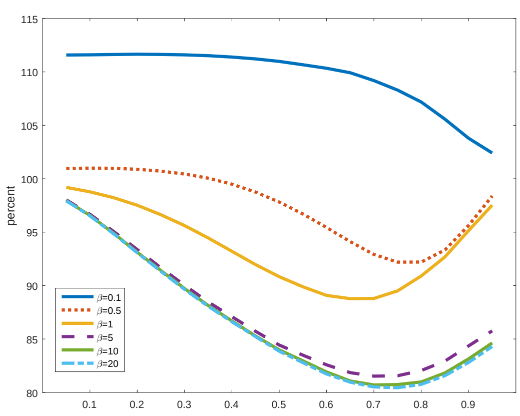
where we normalize the regressor x_t by the square root of its long-run variance $1/(1-\alpha^2)$. Now only β controls the magnitude of the conditional mean, and increasing β will increase the signal-to-noise ratio, and α controls only the persistency of the regressor. We plot the relative efficiency of HP procedure against the standard nonparametric Local Linear estimation procedure in Table 3.1 and Figure 3.1. In Table 3.1, the relative performance of the HP filter when there is no endogeneity with $\rho = 0$ is presented in the upper panel, and those from the endogenous model with $\rho = 0.5$ in the lower panels presented. For easier comparison, we present the same results in the $\alpha - \beta$ plane in Figure 3.3 where the results from the exogenous and endogenous models are show in the upper and lower panels respectively. It is shown clearly that the relative performance of the HP filter improves as the signal from the regressor measured by β gets stronger. This tendency is monotone for all the values

[1] We may generate y_t and x_{t-1} either by staring from the stationary distribution of x_t or discarding burn-in periods after drawing more samples than the required sample size n initially.

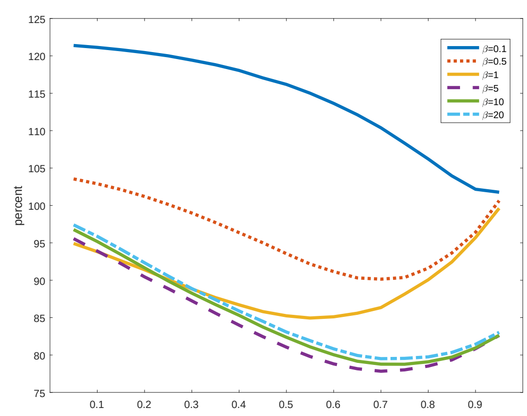
of β we considered for the exogenous cases. For the endogenous cases, however, the monotonously improving pattern is reversed, though not by much, when the signal is very large with $\beta = 5, 10, 20$.

The relative performance of the HP filter against the Local Linear estimator improves as the persistence of the regressor α gets large until the value of α is around 0.7 and becomes worse as it becomes more persistent. This is the case for all values of the magnitude parameter β except when the signal is extremely low with $\beta = 0.1$. This reversal in the relative performance across the values of the persistency parameter α happens for both exogenous and endogenous model as can be seen in the upper and lower panels of Table 3.1 and Figure 3.3.

Figure 3.3: Relative Performance of HP Estimator with $\rho = 0.0$ and 0.5



(a) $\rho = 0$



(b) $\rho = 0.5$

Notes: In this figure, we report the relative performance of the HP estimator compared to Local Linear estimator. Panel (a) shows the simulation results for $\rho = 0$ and panel (b) is for $\rho = 0.5$ case. In both panel (a) and (b), the persistency parameter α is on the x-axis and the relative performance on the $y-$axis where 100 represent equivalent average Means Squared Error(MSE) with our bench-mark Local Linear estimator. Here we consider the magnitude parameter $\beta = [0.1, 0.5, 1.0, 5.0, 10.0, 20.0]$ and plot the relative performance for each β. Sample size is $n = 100$.

Table 3.1: Relative Efficiency of HP Filter against Local Linear Estimator

(a) $\rho = 0.0$

	α								
β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	111.6	111.7	111.6	111.4	111.0	110.3	109.2	107.2	103.8
0.5	101.0	100.9	100.4	99.5	97.8	95.4	92.9	92.2	95.6
1.0	98.8	97.5	95.6	93.2	90.8	89.1	88.8	90.9	95.1
5.0	96.7	93.4	90.0	87.1	84.4	82.6	81.5	82.1	84.4
10.0	96.6	93.1	89.7	86.7	84.0	81.9	80.7	81.0	83.1
20.0	96.5	93.1	89.7	86.6	83.9	81.7	80.5	80.8	82.8

(b) $\rho = 0.5$

	α								
β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	121.1	120.4	119.4	118.1	116.2	113.7	110.4	106.2	102.2
0.5	102.9	101.2	99.0	96.4	93.6	91.2	90.2	91.6	96.4
1.0	93.8	91.4	88.8	86.7	85.2	85.1	86.3	90.1	95.7
5.0	93.9	90.5	87.2	84.0	81.1	78.8	77.8	78.5	80.8
10.0	95.2	91.7	88.3	85.3	82.4	80.0	78.8	79.1	81.0
20.0	95.9	92.3	88.9	85.9	83.1	80.8	79.5	79.8	81.4

Notes: We simulated y_t of sample size $n = 100$ with two values of the endogeneity parameter $\rho = (0.0, 0.5)$. The persistency parameter α is selected from a set of values starting from 0.1 to 0.9 with the increment of 0.1, and the magnitude parameter β from the set of values $(0.1, 0.5, 1.0, 5.0, 10.0, 20.0)$ in the range $(0, 20)$.

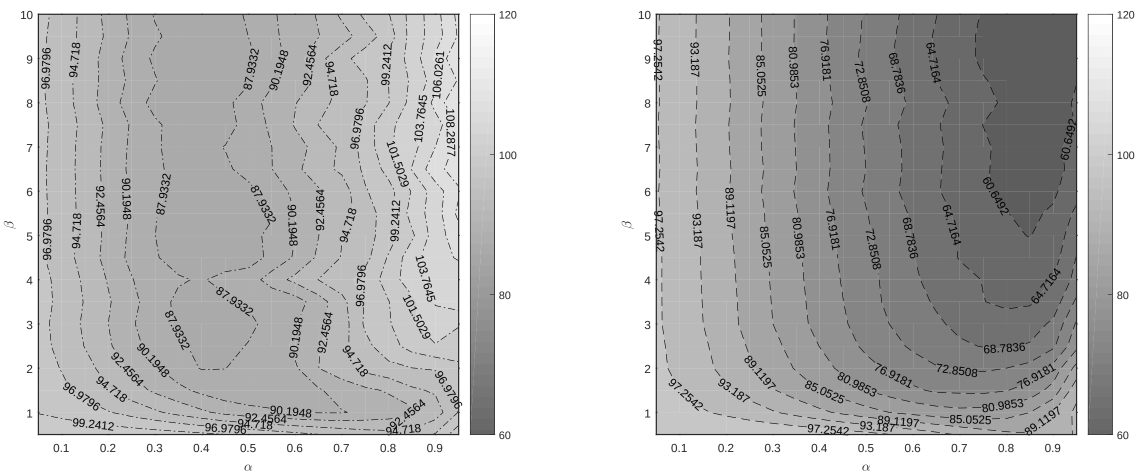
Figure 3.4 shows in the α-β plane the performance of the HP filter relative to the Local Linear estimator as we increase the sample size from $n = 100$ to $n = 500$. In this case, we focus on the exogenous cases with $\rho = 0$, and continue to consider the same set of values for the magnitude parameter β and the persistency parameter α. As expected, the relative performance of the HP filter improves substantially. When the sample size is small, the HP performs best when the regressor is not persistent with the values of α around $0.3 - 0.5$ as long as the signal is not too small with the values of β greater than 2. However, when the sample size is large at $n = 500$, the HP filter performs better when the regressors are more persistent with $\alpha > 0.7$ and with the signal relatively high with $\beta > 5$.

3.2.1.4 Model DT2 We also considered the models with linear time trend $a + bt$, where a and b are constants, to see how the performance of the HP filter against the standard nonlinear Local Linear estimation procedure may change in the presence of the linear trend.

$$y_t = a + bt + \beta(x_{t-1}\sqrt{1-\alpha^2}) + u_t$$
$$x_t = \alpha x_{t-1} + v_t$$

For this model, we also focus on the exogenous model with $\rho = 0$ and sample sizes $n = 100$ and 500. We set $a = 0$ and $b = 1$ and consider the same set of values for the magnitude β and the persistency α parameters as for the models without linear trend. As can be seen clearly the results presented in Figure 3.5, adding a linear trend did not change the relative performance of the HP filter. The patterns of the contours are almost identical to those presented in Figure 3.4 for the models without linear trend.

Figure 3.4: Relative Performance of HP Estimator with Sample Size $n = 100$ and 500



(a) Sample Size $n = 100$ (b) Sample Size $n = 500$

Notes: In this figure, we report the relative performance of the HP estimator compared to Local Linear estimator on α-β plane. We choose α between 0.05 and 0.95. For β range, we consider $[0.5, 10]$. Panel (a) shows the simulation result with sample size $n = 100$ and panel (b) reports the results for $n = 500$ case. In both panel (a) and (b), the persistency parameter α is on the x-axis and the magnitude parameter β is on the $y-$axis. The relative performance of 100 means that our HP estimator performs equally well with Local linear estimator in terms of average Means Squared Error(MSE). Contour regions under 100 means that the HP estimator performs better than our benchmark.

3.2.2 Continuous Time Model

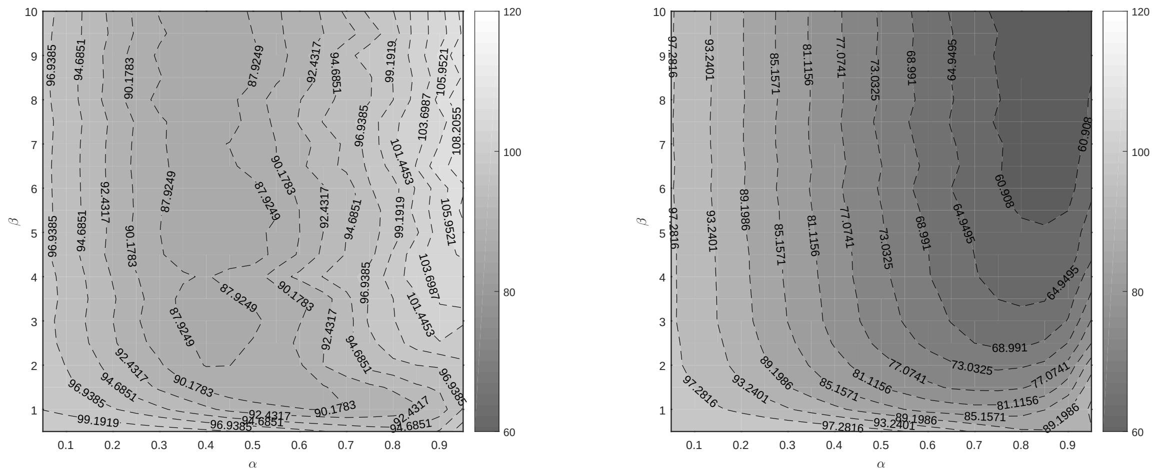
3.2.2.1 Model CT For (3.6), we consider the continuous time model

$$dY(r) = \beta X(r)dr + dU(r) \quad \text{with} \quad dX(r) = -\alpha X(r)dr + dV(r),$$

where α and β are parameters, and U and V are standard Brownian motions with $\text{cor}(U, V) = \rho$.

As a continuous time analog of the discrete time model, we consider the Ornstein-

Figure 3.5: Relative Performance of HP Estimator with Linear Trend



(a) $n = 100$ with Linear Time Trend

(b) $n = 500$ with Linear Time Trend

Notes: In this figure, we report the relative performance of the HP estimator compared to Local Linear estimator on α-β plane when the simulated series include a linear time trend. We choose α between 0.05 and 0.95. For β range, we consider $[0.5, 10]$. Panel (a) shows the simulation result with sample size $n = 100$ and panel (b) reports the results for $n = 500$ case. In both panel (a) and (b), the persistency parameter α is on the x-axis and the magnitude parameter β is on the $y-$axis. The relative performance of 100 means that our HP estimator performs equally well with Local linear estimator in terms of average Means Squared Error(MSE). Contour regions under 100 means that the HP estimator performs better than our benchmark.

Uhlenbeck process. Our continuous time data generating process is given as

$$dY_t = \beta(X_t\sqrt{2\kappa})dt + dU_t$$

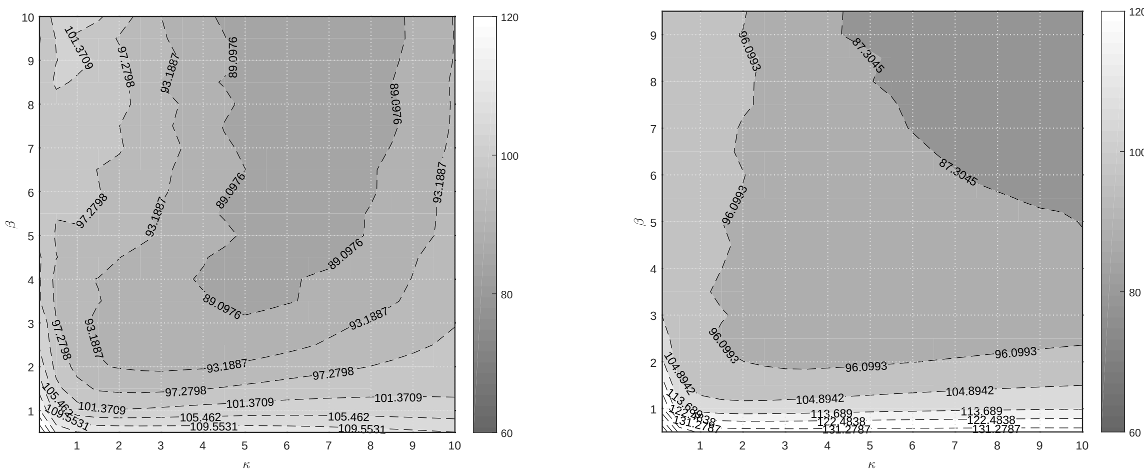
$$dX_t = -\kappa X_t dt + dV_t,$$

where U and V are standard Brownian motions with $corr(U, V) = \rho$. The mean-reversion parameter κ plays a same role of α in discrete time model. Note that for $\kappa = 0$, X_t becomes a Brownian motion. X_t has time invariant distribution $\mathbb{N}(0, 1/2\kappa)$ and we normalize X_t by multiply $\sqrt{2\kappa}$. In continuous time models, we could have more observations either by increasing time span T or sampling frequency Δ. For our simulations, we consider $T = 10$ and two sampling frequencies at monthly with $\Delta = 1/12$ and at weekly with $\Delta = 1/52$.

The values of the coefficient β determining the magnitude of the conditional mean and therefore the signal-to-noise ratio of the model are selected from the range $(0, 10)$ of the values starting from 0.5 to 10 in the increment of 0.5. The values of the mean-reversion parameter κ controlling the persistency of the regressor process x_t considered are also selected from the same set of values, but we additionally consider $\kappa = 0.1$ to see how the HP filter performs when the regressor process is very close to being non-stationary. For the simulations, focus on the exogenous model where the error process u_t and the regressor innovation v_t are uncorrelated with the endogeneity parameter $\rho = 0$.

We generate 10,000 simulation samples for y_t, and the relative performance of our HP estimator to the standard Local Linear estimator is summarized in the the κ-β plane in Figure 3.6 below. The results obtained from the monthly observations are presented in the upper while those from the weekly observations in the lower panel of Figure 3.6. At both monthly and weekly frequencies, the HP filter performs better than the Local Linear estimator in most of the cases except when the parameter β determining the signal-to-noise ratio is very small. The HP filter performs better than the Local Linear estimator for most of the values we considered for the mean-reversion parameter κ. For the observations collected at the lower monthly frequency as can be seen in the upper panel of Figure 3.6, the HP filter performs better for a wide range of κ in the middle of the range (0.10). For the higher frequency at weekly, on the other hand, the HP filter also performs well for a wide range of κ, but in the right side the range (1,10). This is more apparent when the signal-to-noise ratio is smaller as can be seen in the lower panel.

Figure 3.6: Relative Performance of HP Estimator in Continuous Time



(a) $T = 10$, Monthly Frequency

(b) $T = 10$, Weekly Frequency

Notes: In this figure, we report the relative performance of the HP estimator compared to Local Linear estimator on κ-β plane. We choose the mean reverting parameter κ between 0.0 and 10. For β range, we consider $[0.5, 10]$. Panel (a) shows the simulation result with monthly frequency and panel (b) reports the results for weekly frequency. In both panel (a) and (b), the persistency parameter α is on the x-axis and the magnitude parameter β is on the $y-$axis. Considered sample span is $T = 10$. The relative performance of 100 means that our HP estimator performs equally well with Local linear estimator in terms of average Means Squared Error(MSE). Contour regions under 100 means that the HP estimator performs better than our benchmark.

3.3 Empirical Illustrations

To illustrate our new HP estimator, we consider four key macroeconomic variables - GDP growth rates, inflation rate, unemployment rate and 3 month Treasury Bill (3MTB) rate - over the 67 years of time span covering the sample period 1948-01 to 2015-08 where data for all four variables exist at both quarterly and monthly frequencies.

The data for GDP growth rate is obtained from the FRED Economic Data of the Federal Reserve Bank of St. Louis - Series GDPC96 on Quarterly Real GDP

Growth. The series is provided at quarterly frequency only, and seasonally adjusted annual rate in percentage change units. Inflation measured by percentage change of monthly Consumer Price Index for All Urban Consumers (CPIAUC). For quarterly Inflation, monthly CPI was averaged in each quarter and percentage change was calculated. CPI : FRED Series ID (CPIAUCSL), Seasonally adjusted, Monthly, Units: Percent change, Whole Sample span: 1947-01 to 2015-08. For unemployment rate, we use Monthly Civilian Unemployment Rate (FRED Series UNRATE) which is also seasonally adjusted rate in percentage units, and provided at monthly frequency. To get the quarterly unemployment rate, we take the simple average of the monthly measures in each quarter. We obtain the 3MTB also from the FRED data base (Series TB3MS). The series is not seasonally adjusted, and available at monthly frequency. For the quarterly data, we again take the simple average of the monthly measures in each quarter.

For the four key macro variables we consider - GDP growth rates, inflation rate, unemployment rate and 3 month T-Bill rate, we first report the results the quarterly data in Figures 3.1-3.2 and Figures B.3-B.5. We present the smoothing parameter λ, the extracted components with $\lambda = \lambda_Q$ and λ^* along with the original time series and the fitted residual process obtained with $\lambda = \lambda_Q$ and λ^* for all four time series. We note that for all four quarterly series, $\rho(\lambda)$ is positive for $1/\lambda$ small, decreases monotonically as $1/\lambda$ increases and crosses the horizontal axis to become negative eventually for $1/\lambda$ large.

The estimated values of the optimal smoothing parameter λ^* are reported below in Table 3.2. For the US quarterly macroeconomic data, it is recommended to use $\lambda_Q = 1,600$ for the HP filter. We have zero first-order autocorrelation, i.e., $\rho(\lambda^*) = 0$ for the fitted processes from our HP filter with the optimal smoothing parameter λ^* for all four quarterly macro series; however, those fitted series from the conventional

HP with λ_Q exhibit significant amount of serial correlations with the estimate values of the first-order autocorrelations varying from $\rho(\lambda_Q) = 0.278 - 0.826$.

Table 3.2: Optimal Smoothing Parameter λ^* (Quarterly Frequency)

Variable	Sample auto-corr	$u_t(\lambda_Q)$ auto-corr	Optimal λ^*	$u_t(\lambda^*)$ auto-corr
GDP growth	0.372	0.278	20.03	0.00
Inflation	0.726	0.430	18.99	0.00
Unemployment	0.971	0.899	0.56	0.00
Interest rate	0.975	0.826	1.73	0.00

Notes: Table 3.2 report the values of optimal smoothing parameter λ^* for quarterly GDP growth, inflation, unemployment and interest rate. For quarterly frequency data, Hodrick-Prescott filter's penalty parameter $\lambda_Q = 1600$ is commonly suggested. Here we compared the auto-correlation of residuals when λ_Q and λ^* are applied. We search the optimal λ^* over $[0, 10^7]$ range, thus we only impose non-negativity restriction on λ.

We do the same analysis for the data at monthly frequency for the three macro time series for which we obtain monthly frequency data - inflation, unemployment and 3-month Treasury Bill rate.[2] The results are report in Table 3.3 and Figures B.6 - B.8.

3.3.1 Extensions to More General Models

We may consider various models including threshold and smooth transition AR models such as TAR and ESTAR. We may also study the effect of the endogeneity parameter ρ on the performance of the HP filter. We could also try a different objective function. Instead of minimizing the first order auto-correlation, we may use a Portmanteau-type test that minimizes a weighted average of the first few

[2]For GDP growth, quarterly is the most high frequency data and omitted from this table.

Table 3.3: Optimal Smoothing Parameter λ^* - Monthly Frequency

Variable	Sample auto-corr	$u_t(\lambda_M)$ auto-corr	Optimal λ^*	$u_t(\lambda^*)$ auto-corr
Inflation	0.593	0.392	17.00	0.00
Unemployment	0.992	0.969	9.68	0.00
Interest rate	0.991	0.940	0.92	0.00

Notes: Table 3.3 reports the values of optimal smoothing parameter λ^* and the associated residual auto-correlation $u_t(\lambda^*)$ for monthly inflation, unemployment and interest rate. Ravn and Uhlig (2002) suggested $\lambda_M = 129600$ for monthly frequency. Here we compared the auto-correlation of residuals when λ_M and λ^* are applied. We search the optimal λ^* over $[0, 10^7]$ range, thus we only impose non-negativity restriction on λ.

sample auto-correlations to find the penalty parameter λ. One of the widely used Portmanteau tests is the Ljung-Box q-test. In continuous time models, the sample size n depends on both time span T and sampling frequency δ, i.e. $n = T\delta$. We may therefore further study the effects of the sampling frequency δ as well as time span T on the performance of the HP filter.

We may try other measures to evaluate the performance of the HP filter. We may estimate the spectral density of the fitted residual and compute the distance between the estimated density and the constant function. This will reflect better how close as a process the fitted residual is to the white noise process. The HP filter is not concerned about obtaining the conditional mean that minimize the distance to the true trend at each point, rather it tries to fit the mean as a process that leaves a white noise residual. Hence the root mean square error measure favors such methods as the Local Linear method that minimizes the aforementioned distance point-wise. To more fairly evaluate the HP filter, we may consider an alternative measure which reflects how closely the fitted trend preserves the dynamics of the original trend. If the trend was generated as an AR(1) process as in our simulations, we may estimate the AR coefficient of the filtered trend and compare it with the original AR coefficient to see how well the HP filter preserves the original dynamics.

We may view our simulation model as a state space model with only measurement equation without specifying the transition equation. Hence we are estimating the measurement equation without information about the dynamics of the state variable. Hence our estimation of the conditional mean will be less efficient than the ideal, well-specified state space model with both measurement and transition equation well specified to be estimated by the Kalman Filter. However, the HP may serve as an excellent alternative since it doesn't require a transition equation to specify the dynamic structure, if HP filter can estimate the conditional mean part similar to the estimate by the Kalman Filter, albeit less efficient.

3.4 Conclusion

In Section 3, we explore the possibility of using the Hodrick-Prescott filter for a new objective of extracting the conditional mean process of a time series in both discrete and continuous time framework. While the HP filter is conventionally used to decompose a time series into trend and cyclical components, our new approach aims to filter out the conditional mean process of a given time series, which includes both trend and cyclical components. To this end, our HP filter uses the penalty parameter minimizing the first order sample correlation of the residuals. The process extracted from our HP filter is therefore defined as the predictable component of the given time series.

Our HP filter may also be used to estimate the conditional volatility of a time series. We also show that our HP approach can also be used to estimate the conditional mean and variance processes in continuous time model. This approach is the first attempt to applying the HP filter to extracting the conditional mean and variance processes. By simulations, we show that our HP filter generally performs better for both discrete and continuous time models than the commonly used Local Linear estimator. Hence our new HP estimator may serve as a useful and effective alternative to the conventional Local Linear estimator in nonparametric estimation of both conditional mean and variance processes of a time series from a wide class of models.

4. TESTING FOR RETURN PREDICTABILITY WITH MULTIPLE PREDICTORS

Section 4 is organized as follows. Section 4.1 introduces the model and main econometric challenges with relevant backgrounds and statistical theory. Our novel econometric methodologies based on endogeneity correction and Cauchy inference are presented in Section 4.2, and simulations are reported in Section 4.3. Section 4.4 summarizes and interprets our empirical results on stock return predictability and Section 4.5 concludes.

4.1 Model and Econometric Challenges

4.1.1 Model and Background

We consider a continuous time predictive regression of stock return, which is given by

$$dY_t = \Big(\alpha + \beta' X_t\Big) dt + \sigma_t dW_t, \tag{4.1}$$

where α and β are respectively 1- and K-dimensional unknown parameters, $X = (X_t)$ is an K-dimensional covariate continuous time process, $\sigma = (\sigma_t)$ is a stochastic volatility process and $W = (W_t)$ is a standard Brownian motion. In regression (4.1), we set $dY = (dY_t)$ to be the excess log returns of stock, i.e., the log return of stock net of risk free rate. Clearly, the hypothesis

$$\mathrm{H}_0 : \ \beta = 0 \tag{4.2}$$

implies no predictability of stock return.

Throughout the Section 4, we assume that

$$(Y_{i\Delta}, X_{i\Delta})$$

are available for $i = 1, \ldots, N$ in time intervals of length Δ over the time span $[0, T]$ with $T = N\Delta$. [1] This provides a discrete time approximation of our model as

$$Y_{i\Delta} - Y_{(i-1)\Delta} \approx \Big(\alpha + \beta' X_{(i-1)\Delta}\Big)\Delta + \sigma_{(i-1)\Delta}\Big(W_{i\Delta} - W_{(i-1)\Delta}\Big)$$

for $i = 1, \ldots, N$, which corresponds to the usual predictive regression in discrete time used to test the null hypothesis of the unpredictability of stock returns. If stock returns are not predictable, clearly we will have that $\beta = 0$. The most commonly used covariates are those which make the most economic sense, such as the dividend-price ratio and the earnings-price ratio.

The two econometric challenges for the test of hypothesis (4.2) in regression (4.1) are

(a) persistence and endogeneity of X, and

(b) nonstationarity in σ,

which will be discussed subsequently in detail.

The covariate process X, commonly used in the predictive regressions all show strong persistency. This is easily observable and has been well noted by many authors in discrete time framework. For instance, see Goyal and Welch (2003) or Torous et al. (2004). Indeed, it is routinely modeled as an autoregressive process with the autoregressive parameter that is close to one, or more rigorously as a local-to-unity process with autoregressive parameter approaching unity as the sample size increases.

[1] The assumption that our discrete time observations $(Y_{i\Delta})$ are given in an equi-spaced time intervals is made simply for expositional convenience and entirely inconsequential.

As long as the predictors are exogenous and independent of the error process, the persistency of predictors does not matter and the limit distribution of standard t-statistics is given by standard normal distribution. However, in case it has nonzero correlation with the error process, the limit distribution of standard t-statistics is no longer normal and the test relying on the standard t-test yield a considerable size distortion.

This problem has been clearly demonstrated by Campbell and Yogo (2006). In particular, they note that the size distortion is quite severe when the predictor is generated as a near-unit root process from innovations that are heavily correlated with regression errors. For instance, if there is an exact unit root in the covariate and perfect long-run correlation between the innovations of the covariate and regression errors, they find that the asymptotic size of the one-sided t-test at a 5% significance level test is as large as 46%. The reality of the data does not seem to be far from this worst case scenario. Upon examination, one can easily see that the predictors are highly persistent. More formally, Campbell and Yogo (2006) report that for the commonly used predictors such as dividend-price ratio and earnings-price ratio, unity lies outside the 95% confidence interval for only ten out of twenty eight data combinations. Even when a unit root may be rejected, the predictors are highly persistent. Of the ten for which unity lies outside the 95% confidence bounds, seven include an autoregressive parameter above 0.95. Moreover, innovations of the predictors seem to be highly correlated with stock returns in the long-run. For instance, the empirical long-run correlation between differences in the dividend-price ratio and stock returns is -0.98.

The volatility in stock returns, which is clearly stochastic in nature, shows a strong tendency to be time varying and nonstationary. Though it fluctuates within a bounded range, its time series behavior is rather nonstationary showing some clear evidence of clusterings. Naturally, various nonstationary stochastic volatility models have been considered for stock returns in the literature. Indeed, many authors have found that the AR parameter for the volatility process is close to unity under some appropriate functional transformations. See, for instance, Jacquier et al. (2004), who provide convincing evidence that the log of volatility process follows a near-unit root process for a very wide range of equity and exchange rate time series. We may then conclude that the true volatility process is highly nonstationary, since it will be the exponential of a near-unit root process. See also Cavaliere (2004) or Cavaliere and Taylor (2007), who studied the unit root test in the presence of stochastic volatility in the innovations.

The limit null distribution of the standard t-statistic is non-normal, even when the innovations of the covariate and those of the volatility factor are completely independent of the innovations of errors. How far it is away from the standard normal depends on many factors including the volatility function, asymptotic covariances of the innovations, and local-to-unity parameters. Given the previous simulation studies by Chung and Park (2007) and Cavaliere and Taylor (2007), we may expect substantial size distortion from using the standard normal critical values in the predictive regression setting.

In the presence of time varying stochastic volatility σ, the ordinary least squares (OLS) estimator is generally inefficient. Furthermore, we have

$$d[Y]_t = \sigma_t^2 dt,$$

and therefore, the volatility process σ in our model is consistently estimable by the realized volatility of stock returns, which implies that the generalized least squares (GLS) is always feasible using high frequency observations of stock returns. Consequently, we will not consider the OLS estimation except in our simulation study, where we use it as a benchmark to compare with the GLS estimation and our new methodologies. To implement the GLS procedure, we use the realized volatility as an estimate for the volatility process σ, i.e.,

$$\hat{\sigma}^2_{i\Delta} = \sum_{(i-1)\Delta \leq j\delta < i\Delta} \Big(Y_{j\delta} - Y_{(j-1)\delta} \Big)^2, \tag{4.3}$$

where $\delta = o(\Delta)$. In what follows, we call δ and Δ the sampling frequency and the regression frequency respectively. Chang et al. (2016) show that the use of the realized volatility, in place of the true volatility process, does not affect the asymptotics as long as $\delta/\Delta \to 0$.

4.1.2 Statistical Theory

For the expositional convenience, we redefine X and β so that we write (4.1) more compactly as $dY_t = \beta' X_t dt + \sigma_t dW_t$, and its continuous-time GLS transform as

$$\frac{1}{\sigma_t} dY_t = \beta' \frac{X_t}{\sigma_t} dt + dW_t. \tag{4.4}$$

Clearly, the discrete time analogue of the continuous-time GLS in (4.4) is given by

$$\frac{1}{\sigma_{(i-1)\Delta}} \Delta_i Y = \beta' \left(\Delta \frac{X_{(i-1)\Delta}}{\sigma_{(i-1)\Delta}} \right) + \Delta_i W, \tag{4.5}$$

where and hereafter we write $\Delta_i Z = Z_{i\Delta} - Z_{(i-1)\Delta}$, $i = 1, \ldots, N$, for any continuous-time process $Z = (Z_t)$. In our subsequent discussions, we mainly consider regression (4.5) with $\sigma_{(i-1)\Delta}$ replaced by $\hat{\sigma}_{(i-1)\Delta}$ defined in (4.3), and denote by $\hat{\beta}_T$ the GLS estimator of β.

To discuss the distributional aspects of our testing procedures in predictive regressions, we introduce some high level assumptions. The reader is referred to Chang et al. (2016) and the references cited there for the discussions on more precise technical conditions required for our subsequent assumptions to hold. In particular, we let

$$c_T^{-1}\left(\Delta\sum_{i=1}^{N}\frac{X_{i\Delta}X_{i\Delta}'}{\sigma_{i\Delta}^2}\right)c_T^{-1\prime}\sim_d c_T^{-1}\left(\int_0^T\frac{X_tX_t'}{\sigma_t^2}dt\right)c_T^{-1\prime} \tag{4.6}$$

and

$$c_T^{-1}\sum_{i=1}^{N}\frac{X_{i\Delta}}{\sigma_{i\Delta}}\Big(W_{i\Delta}-W_{(i-1)\Delta}\Big)\sim_d c_T^{-1}\int_0^T\frac{X_t}{\sigma_t}dW_t \tag{4.7}$$

as $T\to\infty$ for some non-stochastic and invertible sequence (c_T) of matrices such that $c_T\to\infty$ as $T\to\infty$, where and elsewhere in Section 4 $A_T\sim_d B_T$ implies that A_T and B_T have the same limit distributions as $T\to\infty$ with Δ fixed or $\Delta\to 0$.

Our assumptions in (4.6) and (4.7) are very mild and expected to hold widely. If Δ is given as a function of T such that $\Delta\to 0$ sufficiently fast as $T\to\infty$, they are generally satisfied as long as the Riemann integral $\int_0^T(X_tX_t'/\sigma_t^2)\,dt$ is well defined a.s. However, it is not necessary to require $\Delta\to 0$ as $T\to\infty$. As we discuss in more detail later, if X/σ is either nonstationary or stationary with infinite second moment, they hold also with fixed $\Delta>0$ for a large class of diffusion and jump-diffusion type models.

Under assumptions in (4.6) and (4.7), for the GLS estimator $\hat\beta_T$ of β, we have

$$c_T'\Big(\hat\beta_T-\beta\Big)\sim_d\left[c_T^{-1}\left(\int_0^T\frac{X_tX_t'}{\sigma_t^2}dt\right)c_T^{-1\prime}\right]^{-1}c_T^{-1}\int_0^T\frac{X_t}{\sigma_t}dW_t, \tag{4.8}$$

and therefore, the limit distribution of $\hat\beta_T$ can be easily obtained from the asymptotics of

$$\int_0^T\frac{X_tX_t'}{\sigma_t^2}dt\quad\text{and}\quad\int_0^T\frac{X_t}{\sigma_t}dW_t. \tag{4.9}$$

The asymptotics for the additive functional and martingale transform in (4.9) are provided by Kim and Park (2016a) and Jeong and Park (2016) respectively for the diffusion and jump-diffusion type processes.

If X/σ is stationary and has finite second moment, we have

$$\frac{1}{T}\int_0^T \frac{X_t X_t'}{\sigma_t^2}dt \to_{a.s.} \mathbb{E}\left(\frac{X_t X_t'}{\sigma_t^2}\right)$$
$$\frac{1}{\sqrt{T}}\int_0^T \frac{X_t}{\sigma_t}dW_t \to_d \mathbb{N}\left(0, \mathbb{E}\left(\frac{X_t X_t'}{\sigma_t^2}\right)\right)$$

by the standard LLN and CLT for continuous time processes, and therefore, we may easily deduce from (4.8) that

$$\sqrt{T}\left(\hat{\beta}_T - \beta\right) \to_d \mathbb{N}\left(0, \left[\mathbb{E}\left(\frac{X_t X_t'}{\sigma_t^2}\right)\right]^{-1}\right)$$

as $T \to \infty$, assuming that $\mathbb{E}\left(X_t X_t'/\sigma_t^2\right)$ is nonsingular. In this case, therefore, the standard tests are valid.

However, in case that X/σ is either nonstationary or stationary with infinite second moment, the asymptotics for the additive functional and martingale transform in (4.9) are generally nonstandard and non-normal. Consequently, the limit distribution of the GLS estimator $\hat{\beta}_T$ is generally nonstandard and non-normal, and the standard tests based on the t-ratio and Wald statistics are invalid. Since it is widely agreed in the literature on testing for return predictability that the aforementioned two econometric challenges make the standard inferences in predictive regressions are invalid, we focus on the case that $\hat{\beta}_T$ has non-standard and non-normal distribution, so that the t-test and Wald test are invalid asymptotically. Therefore, for the rest of the Section 4, we let X/σ be either stationary without finite second moment or nonstationary.

There is a very important exception. If X/σ is asymptotically independent of

W, the limit distribution of the GLS estimator $\hat{\beta}_T$ is mixed-normal and the standard tests become valid. This is well expected from (4.8) since, conditional on X/σ, we have

$$c_T^{-1} \int_0^T \frac{X_t}{\sigma_t} dW_t =_d \mathbb{N}\left(0, \left[c_T^{-1}\left(\int_0^T \frac{X_t X_t'}{\sigma_t^2} dt\right) c_T^{-1\prime}\right]\right)$$

in this case. Of course, it is highly unrealistic to assume that X/σ is asymptotically independent of W in predictive regressions, and also we may hardly expect that the condition holds for actual predictive regressions. However, it provides an obvious and strong motivation for the endogeneity correction methodology developed by Chang et al. (2016), on which our empirical study heavily relies. This will be explained in detail later.

4.2 New Econometric Approaches

To test for the predictability of stock returns using multiple predictors, we use two different novel approaches: the direct endogeneity correction developed in Chang et al. (2016) and the multivariate modification of Cauchy estimation studied in Choi et al. (2016). Both of them are highly robust against the two econometric challenges we have in testing for return predictability.

4.2.1 Endogeneity Correction

As in Chang et al. (2016), we assume that individual components of X/σ be given as general Ito processes, so that we may specify the k-th component $(X/\sigma)^k$ of X/σ as

$$d(X/\sigma)_t^k = \mu_t^k dt + \nu_t^k dV_t^k \tag{4.10}$$

with two processes $\mu^k = (\mu_t^k)$ and $\nu^k = (\nu_t^k)$ representing the drift and volatility of $(X/\sigma)^k$ for each $k = 1, \ldots, K$. If V, $V = (V_1, \ldots, V_K)'$, is observable, we may consider

$$dY_t^* = \Big(\alpha + \beta' X_t\Big)dt + \sigma_t dW_t^* \tag{4.11}$$

with

$$dY_t^* = dY_t - \sigma_t\Big(\rho' dV\Big),$$

in place of (4.1), where $\rho = \mathbb{E}\Big(dV_t dW_t\Big)$. In the continuous-time regression (4.11), note that $dW_t^* = dW_t - \rho' dV_t$ and W^* is independent of V.

Following our earlier convention, we redefine X and β so that we may write (4.11) simply as $dY_t^* = \beta' X_t dt + \sigma_t dW_t^*$, which yields the continuous-time GLS transform

$$\frac{1}{\sigma_t} dY_t^* = \beta' \frac{X_t}{\sigma_t} dt + dW_t^* \tag{4.12}$$

as in (4.4). The discrete time analogue of the continuous-time GLS in (4.12) is given by

$$\frac{1}{\sigma_{(i-1)\Delta}} \Delta_i Y - \rho' \Delta_i V = \beta' \left(\Delta \frac{X_{(i-1)\Delta}}{\sigma_{(i-1)\Delta}} \right) + \Delta_i W^*, \tag{4.13}$$

which is in contrast with (4.5). In regressions (4.5) and (4.13), we have error terms $\Delta_i W$ and $\Delta_i W^* = \Delta_i W - \rho' \Delta_i V$, respectively.

Now we may well expect that the GLS estimator $\hat{\beta}_T^*$ from the modified regression in (4.13) for the continuous-time regression in (4.12) has the limit distribution

$$c_T'\Big(\hat{\beta}_T - \beta\Big) \sim_d \left[c_T^{-1} \left(\int_0^T \frac{X_t X_t'}{\sigma_t^2} dt \right) c_T^{-1\prime} \right]^{-1} c_T^{-1} \int_0^T \frac{X_t}{\sigma_t} dW_t^*, \tag{4.14}$$

which is always mixed normal since, conditional on X/σ,

$$c_T^{-1} \int_0^T \frac{X_t}{\sigma_t} dW_t^* =_d \mathbb{N} \left(0, \left[c_T^{-1} \left(\int_0^T \frac{X_t X_t'}{\sigma_t^2} dt \right) c_T^{-1\prime} \right] \right).$$

Note that the limit distribution of the modified GLS estimator $\hat{\beta}_T^*$ of β is mixed normal, though we allow X/σ and W to be dependent each other. The modification we introduce here is completely analogous to Phillips and Hansen (1990) and Park (1992).

Clearly, it is not feasible to run regression (4.13), since ρ and $\left(\Delta_i V\right)$, as well as $\left(\sigma_{i\Delta}\right)$, are not directly observable. As discussed earlier, we may simply replace $\left(\sigma_{i\Delta}\right)$ with $\left(\hat{\sigma}_{i\Delta}\right)$ defined in (4.3). Though the required procedure is more involved and should be followed with much caution, it is also possible to replace ρ and $\left(\Delta_i V\right)$ with their estimates without affecting the limit distribution of the infeasible GLS estimator $\hat{\beta}_T^*$ introduced above. The actual procedure we use to estimate ρ and $\left(\Delta_i V\right)$ will be given subsequently.

By estimating each component of $\left(\Delta_i V\right)$ separately, we may assume without loss of generality that $K = 1$ and $\left(\Delta_i V\right)$ is univariate. Our estimate of $\Delta_i V$ is given by

$$\widehat{\Delta_i V} = \frac{(X/\hat{\sigma})_{i\Delta} - (X/\hat{\sigma})_{(i-1)\Delta}}{\hat{\nu}_{(i-1)\Delta}}, \tag{4.15}$$

where

$$\hat{\nu}_{i\Delta}^2 = \sum_{(i-1)\Delta \leq j\delta < i\Delta} \Big((X/\hat{\sigma})_{j\delta} - (X/\hat{\sigma})_{(j-1)\delta} \Big)^2$$

with $\delta = o(\Delta)$, analogously as in (4.3). Moreover, we define an estimate for $\Delta_i W$ as

$$\widehat{\Delta_i W} = \frac{Y_{i\Delta} - Y_{(i-1)\Delta}}{\hat{\sigma}_{(i-1)\Delta}},$$

similarly as $\widehat{\Delta_i V}$, and use

$$\hat{\rho} = \frac{1}{T} \sum_{i=1}^{N} \widehat{\Delta_i V} \widehat{\Delta_i W} \tag{4.16}$$

as an estimate for ρ. The feasible GLS estimator $\hat{\beta}_T^*$ we use in this Section is defined from regression (4.13) with $\left(1/\sigma_{(i-1)\Delta}\right)\Delta_i Y - \rho'\Delta_i V$ replaced by

$$\frac{1}{\hat{\sigma}_{(i-1)\Delta}}\Delta_i Y - \hat{\rho}'\widehat{\Delta_i V},$$

and $\Delta X_{(i-1)\Delta}/\sigma_{(i-1)\Delta}$ by $\Delta X_{(i-1)\Delta}/\hat{\sigma}_{(i-1)\Delta}$, where $\hat{\sigma}_{i\Delta}$, $\widehat{\Delta_i V}$ and $\hat{\rho}$ are defined respectively in (4.3), (4.15) and (4.16).

As shown in Chang et al. (2016), the approximation errors in replacing ρ and $\left(\Delta_i V\right)$ by their estimates are asymptotically negligible and do not affect the limit distribution of the GLS estimator $\hat{\beta}_T^*$ of β under very general set of technical conditions. There is only one critical condition. It is well expected from (4.10) and (4.15) that

$$\widehat{\Delta_i V} \approx \frac{\mu_{(i-1)\Delta}}{\nu_{(i-1)\Delta}} + \Delta_i V,$$

and therefore, we should require that μ/ν be negligible asymptotically for our feasible GLS estimator $\hat{\beta}_T^*$ to perform properly. We believe that the required condition is met widely in the predictive regressions of stock returns, since all commonly used predictors appear to have the volatilities that are of much bigger stochastic magnitudes than those of their drifts.

4.2.2 Cauchy Inference

Cauchy inference suggested by Choi et al. (2016) is extremely simple and just uses as instruments the sign functions of the predictors themselves. More precisely, if we let $\tilde{X}$ be the recursively demeaned X defined as

$$\tilde{X}_t = X_t - \frac{1}{t}\int_0^t X_s ds,$$

which has the discrete time analog

$$\tilde{X}_{i\Delta} = X_{i\Delta} - \frac{1}{i}\sum_{j=1}^{i} X_{j\Delta}$$

for $i = 1, \ldots, N$, then the Cauchy estimator $\tilde{\beta}_T$ of β in regression (4.5) is defined as the simple IV estimator using $\Big(\operatorname{sgn}(\tilde{X}_{(i-1)\Delta})\Big)$ as instruments.

If there is a single regressor without a constant term, the Cauchy estimator $\tilde{\beta}_T$ has standard normal distribution as shown in Choi et al. (2016) for a very broad class of predictors including both stationary and nonstationary processes possibly with jumps, structural breaks, regime switching and many other possible aberrant characteristics. This is because

$$\frac{1}{\sqrt{T}}\sum_{i=1}^{N}\operatorname{sgn}(\tilde{X}_{(i-1)\Delta})\Delta_i W \sim_d \frac{1}{\sqrt{T}}\int_0^T \operatorname{sgn}(\tilde{X}_t)dW_t =_d \mathbb{N}(0,1). \tag{4.17}$$

Note that, if we define a process U by $dU = \operatorname{sgn}(\tilde{X}_t)dW_t$, then U is a continuous martingale such that $d[U]_t = dt$, and therefore, U becomes Brownian motion. Therefore, in particular, normality in (4.17) holds even for finite T as long as Δ is small. It is possible to allow for the presence of a constant term in regression while preserving the nice property of the Cauchy estimator, if we recursively demean both the regressor $\Big(X_{(i-1)\Delta}/\sigma_{(i-1)\Delta}\Big)$ and the regressand $\Big(\Delta_i Y/\sigma_{(i-1)\Delta}\Big)$ in regression (4.5) and fit the regression without the constant term.

If we use the Cauchy estimator in regressions with multiple regressors, it no longer has the robustness and asymptotic normality discussed above. To see why, we suppose that there are two regressors X^1 and X^2, and define U^1 and U^2 by

$$dU_t^1 = \operatorname{sgn}(\tilde{X}_t^1)dW_t \quad \text{and} \quad dU_t^2 = \operatorname{sgn}(\tilde{X}_t^2)dW_t.$$

Clearly, U^1 and U^2 are individually Brownian motions. However, they are not joint Brownian motions. Indeed, we have

$$d[U^1, U^2]_t = \operatorname{sgn}(\tilde{X}_t^1) \operatorname{sgn}(\tilde{X}_t^2) dt,$$

and therefore, $U = (U^1, U^2)'$ is not a bivariate Brownian motion. Consequently,

$$\begin{aligned}\left(\frac{1}{\sqrt{T}} \sum_{i=1}^{N} \operatorname{sgn}(\tilde{X}^1_{(i-1)\Delta}) \Delta_i W, \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \operatorname{sgn}(\tilde{X}^2_{(i-1)\Delta}) \Delta_i W \right) \\ \sim_d \left(\frac{1}{\sqrt{T}} \int_0^T \operatorname{sgn}(\tilde{X}_t^1) dW_t, \frac{1}{\sqrt{T}} \int_0^T \operatorname{sgn}(\tilde{X}_t^2) dW_t \right)\end{aligned}$$

is generally not bivariate normal for any fixed T. Of course, we have asymptotic bivariate normality, if X^1 and X^2 are stationary and

$$\frac{1}{T}[U^1, U^2]_T = \frac{1}{T} \int_0^T \operatorname{sgn}(\tilde{X}_t^1) \operatorname{sgn}(\tilde{X}_t^2) dt$$

converges to a non-random constant as $T \to \infty$. It is now clear that the Cauchy estimator $\tilde{\beta}_T$ permits normal inference only asymptotically and only when predictors are stationary. Of course, they allow X/σ to have infinite second moments as long as they are stationary.

It is straightforward to modify the Cauchy estimator $\tilde{\beta}_T$ to deal with nonstationarity of predictors. We let

$$\Delta_i \tilde{X} = \Delta_i X - \frac{X_i - X_0}{i}, \tag{4.18}$$

and define $\tilde{\beta}_T^*$ to be the IV estimator using $\left(\operatorname{sgn}(\Delta_i \tilde{X})\right)$ as instruments, which is called the modified Cauchy estimator. Note that the heteroskedasticity correction is unnecessary in (4.18), and note in particular that $\operatorname{sgn}(\Delta_i \tilde{X}) = \operatorname{sgn}(\Delta_i \tilde{V})$, where

$\Delta_i \tilde{V} = \Big(\Delta_i X - (X_i - X_0)/i\Big)/\nu_{(i-1)\Delta}$. It is well expected that the modified Cauchy estimator is asymptotically normal for a large class of nonstationary predictors, since $\Big(\Delta_i \tilde{V}\Big)$ is asymptotically stationary for a variety of nonstationary processes.

4.3 Monte Carlo Simulations

In this section, we examine by simulation the finite sample performances of the endogeneity correction and the Cauchy inference, which are introduced to deal with the two econometric challenges discussed earlier.

4.3.1 Simulation Model

The performance of our testing procedure depends critically on the specification of our model. For our simulation, we consider X defined as the solution of the stochastic differential equation

$$dX_t = \frac{acX_t}{(1 + cX_t^2)^{1-b}} dt + (1 + cX_t^2)^{b/2} dV_t \tag{4.19}$$

with appropriately chosen parameters a, b and $c \geq 0$, where V is Brownian motion. The process X defined in (4.19) is a slightly modified version of the generalized Höpfner-Kutoyants (GHK) diffusion introduced earlier in Kim and Park (2016b). The GHK diffusion is quite flexible and generates diffusions with very diverse characteristics, including stationary and nonstationary, and mean-reverting and non-mean-reverting processes, depending upon the values of parameters a, b and $c \geq 0$.

The scale density s' and speed density m of the GHK diffusion X in (4.19) are given explicitly as $s'(x) = \Big(1 + cx^2\Big)^{-a}$ and $m(x) = \Big(1 + cx^2\Big)^{a-b}$. Therefore, we may easily see that X is recurrent if $a \leq 1/2$, and stationary if $a - b < -1/2$. The endogeneity correction we employ is expected to work properly only when

$$\frac{\mu_t}{\nu_t} = \frac{acX_t}{(1 + cX_t^2)^{1-3b/2}}$$

is small, where μ and ν denote respectively the drift and volatility components of X. If $a - b > -1/2$ and X is nonstationary, the normalized process X^T, defined by $X_t^T = c_T^{-1} X_{Tt}$ with $c_T = T^{1/2(1-b)}$, converges weakly to a skew Bessel process. Finally, X becomes mean-reverting if and only if either $a \leq -1/2$ or $a - b \leq -1/2$. In our simulation, we set various combinations of the values of a and b for $a = -1/4, 0, 1/4$ and $b = -1/4, 0, 1/4$, and $c = c_T^{-2}$ using the normalizing sequence $c_T = T^{1/2(1-b)}$ so that the stochastic volatility term cX^2 is stochastically bounded.

Furthermore, we set

$$\sigma_t = (1 + (1/T)Z_t^2)^{1/2}$$

with

$$dZ_t = -0.5dW_t + \sqrt{1 - 0.5^2}dU_t,$$

where U is a standard Brownian motion independent of W. Our specification of σ implies that the Brownian motions generating returns and return volatilities are correlated to yield the leverage effect observed in actual stock returns.

In our simulations, we consider X to be both one dimensional and two dimensional. For one dimensional X given by (4.19), we set the correlation coefficient ρ between V and W, denoting respectively the Brownian motions driving the regressor X and the error process U, to be -0.95, approximately as the sample correlation between the excess returns and dividend-price ratio after heterogeneity correction. For two dimensional X, $X = (X^1, X^2)$, with X^1 and X^2 generated individually as (4.19) from two Brownian motions V^1 and V^2, we set the correlation coefficients ρ_1

and ρ_2 between $V = (V^1, V^2)$ and W to be -0.95 and -0.10. We choose the correlation coefficient -0.10 for the additional regressor to make it roughly the same as the actual sample correlation coefficient between the Brownian motions driving the T-bill rate. Note that the endogeneity of the first regressor is much stronger than that of the second regressor.

Throughout our simulations, we look at the 5% t and Wald tests respectively for the predictability of a single and multiple predictors. We consider the tests based on the OLS, GLS, GLS with endogeneity correction (GLS-EC), Cauchy and modified Cauchy estimators. To obtain the actual finite sample sizes of the tests, we let $\beta = 0$. On the other hand, for the finite sample powers of the tests, we specify

$$\beta = \frac{\bar{\beta}}{T}$$

for $\bar{\beta}$ fixed at $0 \leq \bar{\beta} \leq 10$, and use the critical values that are size-adjusted. Therefore, all tests have exact 5% power in case $\bar{\beta} = 0$. Of course, the powers of the tests are expected to increase as $\bar{\beta}$ deviates from 0. The time span T in our simulations are set to be $T = 10$ and 50. Finally, the number of iterations is 10,000 in all our simulations.

4.3.2 Simulation Results

Our simulation results are summarized in Tables 4.1-4.2 and Figures 4.1-4.4. The actual finite sample sizes of the 5% t- and Wald tests are reported in the tables and their size corrected powers are presented in the figures.

In terms of finite sample size, the tests based on the Cauchy and modified Cauchy estimators perform best. Under the null hypothesis of no predictability, both of them produce the actual rejection probabilities that are quite close to the nominal 5% size in all cases we consider here. This is true for both $T = 10$ and $T = 50$. However, the finite sample size of the tests based on the modified Cauchy estimator becomes virtually identical to the 5% nominal size in every case when $T = 50$. It appears that the size performance of the tests based on the modified Cauchy estimator is truly good if the same size is moderately large. In contrast, the performances of the tests based on the OLS and GLS estimators are unacceptable. The actual sizes of the 5% tests are over 30% and can be well above 40%. Therefore, they are expected to severely over-reject the null hypothesis. Clearly, the evidence for return predictability from these tests should not be taken seriously.

Our new tests based on the GLS-EC perform reasonably well. The actual finite sample sizes of the tests are not as accurate as the tests based on the Cauchy and modified Cauchy estimators. However, they are acceptably close to the 5% nominal size in all cases considered in our simulations. In some cases their actual rejection probabilities are around 4% and below 5%, but in other cases they become 7% or slightly larger and above 5%. There is no systematic tendency to under-reject or over-reject the null hypothesis of no return predictability.

The comparison based on the size-adjusted finite sample powers is clear and unambiguous. The new tests based on the GLS-EC outperform, unanimously and significantly, all other tests in all model parameters and time spans used in our simulations. It is obvious that the tests based on the GLS-EC are generally more effective than all other tests in terms of discriminatory powers.

The discriminatory powers of the tests based on the OLS, GLS and Cauchy estimators are roughly comparable. In regressions with a single regressor, the tests based on the GLS are indeed not clearly better than the tests based on the Cauchy estimator. Indeed, the latter even outperform the former in some range of $\bar{\beta}$ and other model parameter values. This is contradicting to what we may expect from the asymptotic theory, since the Cauchy estimator, being an IV estimator, is less efficient than the GLS estimator. However, these anomalies generally disappear, though not entirely, for the tests in regressions with two regressors. The tests based on the GLS and OLS generally outperform the tests based on the Cauchy estimator.

Unfortunately, the tests based on the modified Cauchy estimator perform very poorly in terms of power. Recall that they perform best in terms of size. The discriminatory power increases as $\bar{\beta}$, but only very slowly. The tests do not seem to be useful at all for practical applications. We may say that the evidence of return predictability is very strong, if it is supported by the tests based on the modified Cauchy estimator.

4.4 Empirical Results

4.4.1 Description of Data

We borrow commonly used predictors from the Welch and Goyal (2008)'s data library. Our monthly frequency data spans from January 1954 to December 2014. January 1954 is the earliest month that daily 3 Month T-bill rate is available. We use daily 3 Month T-bill rate to measure interest volatility. Dividend are a twelve month moving sum of dividends paid on the S&P 500 Index. Prices from S&P 500 index prices from CRSP's month end values. The log of dividend-price ratio $\log(D/P)$ is the difference between log of dividend and log of price. Likewise, earnings are twelve month moving sum of earnings paid on the S&P 500 Index. Dividend Payout ratio ($\log(D/E)$) is the difference between log of dividend and log of earnings. Book to Market ratio(BM) is the ratio of book value to market value for the Dow Jones Industrial Average. Default Yield Spread (DFY) is the difference between BAA- and AAA- rated corporate bond yields, and yields on AAA- and BAA-rated bonds are from the economic research database at Federal Reserve Bank at St. Louis (FRED). T-bill rates are the 3-Month Treasury Bill: Secondary Market Rate from

Table 4.1: Actual Sizes of 5% of t-Test

(a) $T = 10$

b	a	OLS	GLS	GLS-EC	Cauchy	Modified Cauchy
	-0.25	40.6	41.0	3.6	4.2	6.9
-0.25	0	42.2	41.8	4.8	4.3	6.8
	0.25	42.5	43.5	6.3	4.7	6.7
	-0.25	40.6	41.1	4.5	4.5	6.5
0	0	42.1	41.5	5.3	4.4	6.9
	0.25	42.5	43.0	7.3	4.3	6.6
	-0.25	40.8	41.3	5.0	4.4	6.8
0.25	0	41.9	41.5	5.5	4.5	6.8
	0.25	42.1	42.0	6.9	4.7	6.9

(b) $T = 50$

b	a	OLS	GLS	GLS-EC	Cauchy	Modified Cauchy
	-0.25	41.4	41.8	3.6	4.0	5.2
-0.25	0	43.4	43.4	5.2	3.9	4.8
	0.25	44.4	44.9	7.3	4.2	5.1
	-0.25	41.6	41.9	4.3	4.3	5.1
0	0	43.5	43.1	5.2	4.2	5.3
	0.25	44.4	43.5	7.2	3.9	4.8
	-0.25	42.6	42.3	5.0	4.0	5.1
0.25	0	43.6	43.0	5.3	4.5	5.3
	0.25	44.0	43.6	5.9	4.0	5.4

Notes: Table 4.1 reports the simulated sizes of 5% t-Test for the OLS, GLS, GLS-EC, Cauchy and Modified Cauchy tests. Simulation sample scan we consider here is a $T = 10$ and 50. The value of endogeneity parameter ρ is -0.95. Table 4.1 considers different combination of the GHK parameter value $b = [-1/4, 0, 1/4]$ and $a = [-1/4, 0, 1/4]$.

Table 4.2: Actual Sizes of 5% Wald Test

(a) $T = 10$

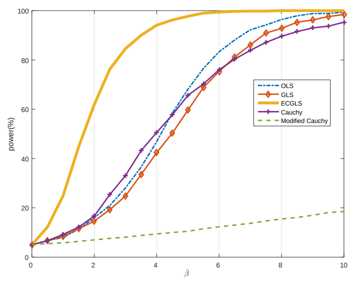
X_1 b	X_1 a	X_2 b	X_2 a	OLS	GLS	GLS-EC	Cauchy	Modified Cauchy
-0.25	-0.25	0.25	0.25	29.4	32.5	7.4	6.9	5.8
		0	0	30.3	33.4	5.6	6.5	5.4
		-0.25	-0.25	30.5	33.2	5.7	6.6	5.4
0	0	0.25	0.25	31.3	34.0	5.4	6.5	5.5
		0	0	31.3	33.9	5.4	6.4	5.5
		-0.25	-0.25	31.2	33.7	5.3	6.6	5.4
0.25	0.25	0.25	0.25	31.7	34.3	5.7	6.4	5.5
		0	0	31.7	34.3	5.6	6.4	5.5
		-0.25	-0.25	31.5	34.0	5.5	6.5	5.3

(b) $T = 50$

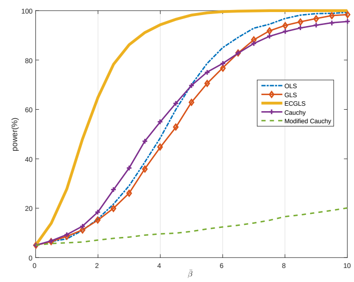
X_1 b	X_1 a	X_2 b	X_2 a	OLS	GLS	GLS-EC	Cauchy	Modified Cauchy
-0.25	-0.25	0.25	0.25	28.7	31.4	5.7	5.4	5.7
		0	0	30.4	32.5	6.2	6.0	4.6
		-0.25	-0.25	30.2	32.7	6.3	5.7	4.9
0	0	0.25	0.25	31.3	33.6	6.0	5.7	4.8
		0	0	31.2	33.7	6.0	5.7	4.9
		-0.25	-0.25	30.9	33.4	6.1	5.8	4.9
0.25	0.25	0.25	0.25	31.7	33.8	6.0	5.7	4.9
		0	0	31.7	33.9	6.0	5.7	4.9
		-0.25	-0.25	31.0	33.8	6.0	5.8	5.0

Notes: Table 4.2 reports simulated sizes of 5% Wald Test for the OLS, GLS, GLS-EC, Cauchy and Modified Cauchy tests. We consider the sample span of $T = 10$ and $T = 50$. The values of endogeneity parameter ρ_1 and ρ_2, which is respectively for the first and second covariate X_1 and X_2, are considered at -0.95 and -0.1. Table 4.2 report the simulated sizes of tests with different combinations of GHK parameter $b = [-1/4, 0, 1/4]$ and $a = [-1/4, 0, 1/4]$ for first and second covariates.

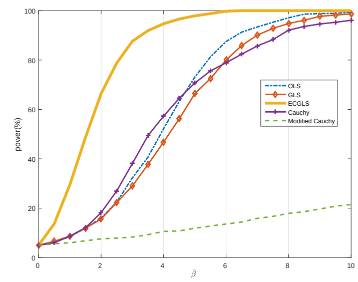
Figure 4.1: Simulated Power Function for $t-$Tests: $T = 10$



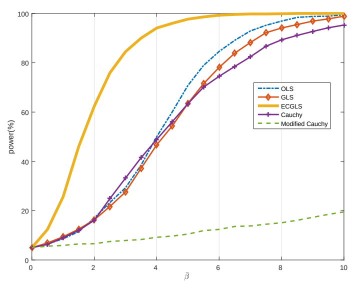
$(b, a) = (-1/4, -1/4)$



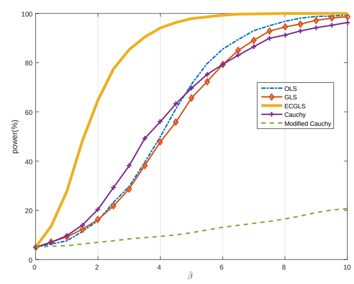
$(b, a) = (-1/4, 0)$



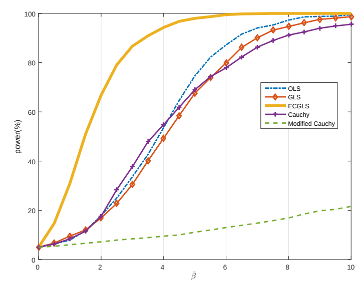
$(b, a) = (-1/4, 1/4)$



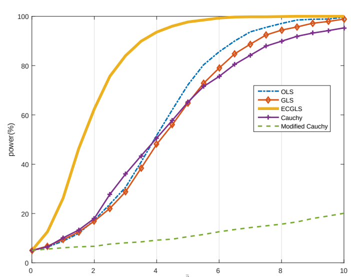
$(b, a) = (0, -1/4)$



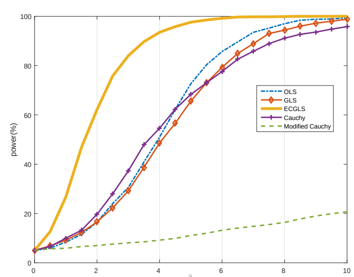
$(b, a) = (0, 0)$



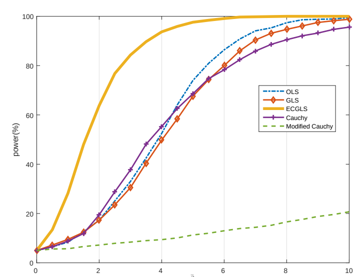
$(b, a) = (0, 1/4)$



$(b, a) = (1/4, -1/4)$



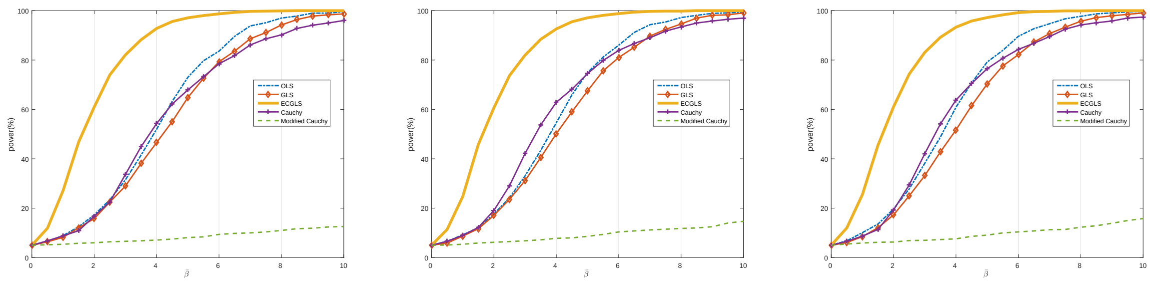
$(b, a) = (1/4, 0)$



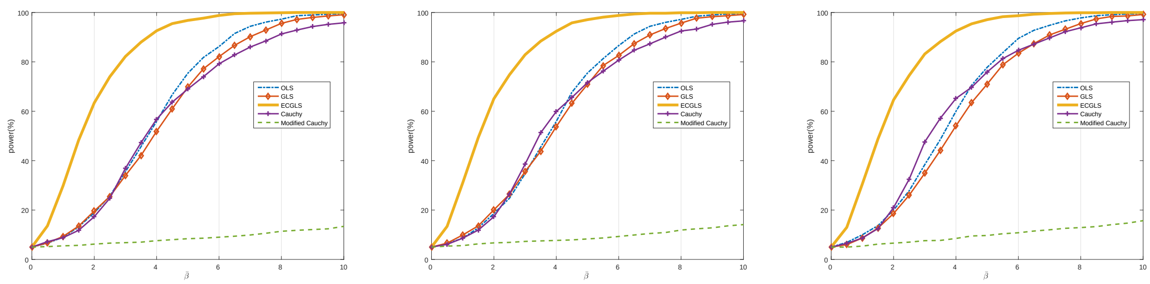
$(b, a) = (1/4, 1/4)$

Notes: Figure 4.1 reports the simulate power functions of 5% $t-$Tests with a sample span of $T = 10$. The considered value of endogeneity parameter ρ is 0.95. For values of GHK parameters, we consider $b, a = [-1/4, 0, 1/4]$ and compare the size adjusted power function of the OLS, GLS, GSL-EC, Cauchy and Modified Cauchy tests. The dash-dot lines represent the power of OLS test and the solid line with square marks represent the GLS test. The thick solid lines represent the power function of our GLS-EC test. The solid line with + marks and dashed lines represent the Cauchy and Modified Cauchy tests, respectively.

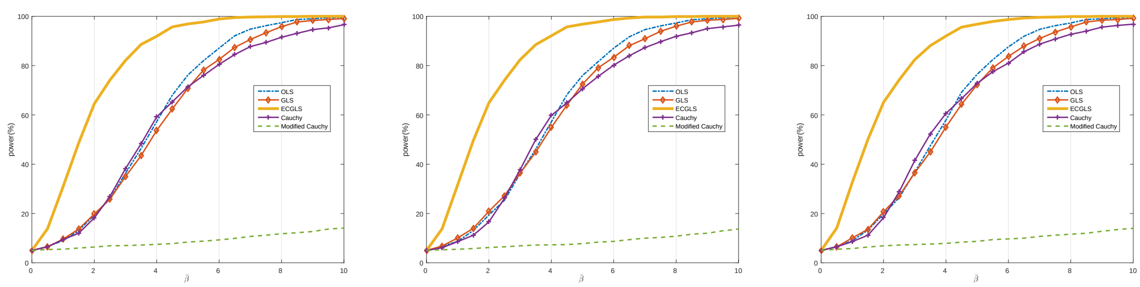
Figure 4.2: Simulated Power Function for $t-$Tests: $T = 50$



$(b, a) = (-1/4, -1/4)$ $(b, a) = (-1/4, 0)$ $(b, a) = (-1/4, 1/4)$



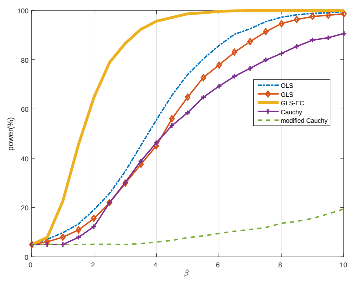
$(b, a) = (0, -1/4)$ $(b, a) = (0, 0)$ $(b, a) = (0, 1/4)$



$(b, a) = (1/4, -1/4)$ $(b, a) = (1/4, 0)$ $(b, a) = (1/4, 1/4)$

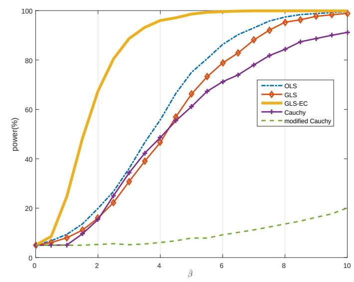
Notes: Figure 4.2 reports the simulate power functions of 5% $t-$Tests with a sample span of $T = 50$. The considered value of endogeneity parameter ρ is 0.95. For values of GHK parameters, we consider $b, a = [-1/4, 0, 1/4]$ and compare the size adjusted power function of the OLS, GLS, GSL-EC, Cauchy and Modified Cauchy tests. The dash-dot lines represent the power of OLS test and the solid line with square marks represent the GLS test. The thick solid lines represent the power function of our GLS-EC test. The solid line with + marks and dashed lines represent the Cauchy and Modified Cauchy tests, respectively.

Figure 4.3: Simulated Power Function for Joint Wald Tests: $T = 10$



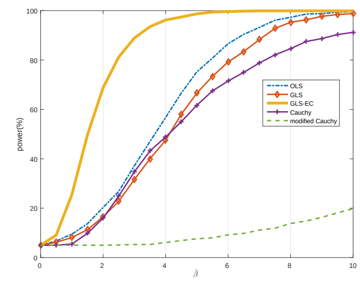
$(b1, a1) = (-1/4, -1/4)$

$(b2, a2) = (-1/4, -1/4)$



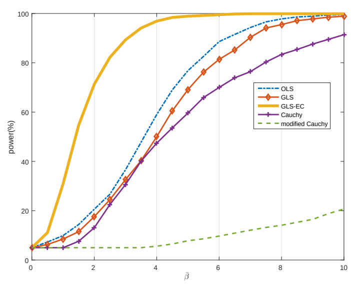
$(b1, a1) = (-1/4, -1/4)$

$(b2, a2) = (0, 0)$



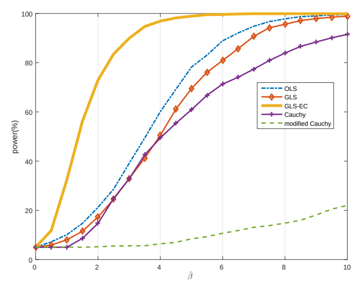
$(b1, a1) = (-1/4, -1/4)$

$(b2, a2) = (1/4, 1/4)$



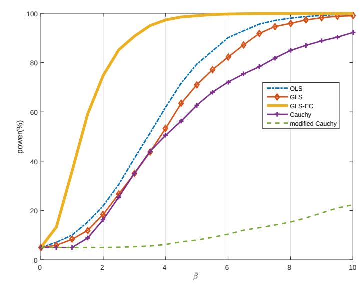
$(b1, a1) = (0, 0)$

$(b2, a2) = (-1/4, -1/4)$



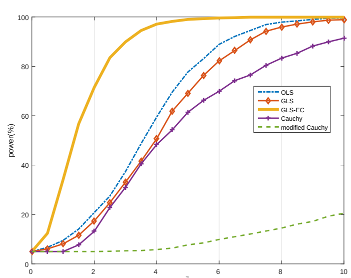
$(b1, a1) = (0, 0)$

$(b2, a2) = (0, 0)$



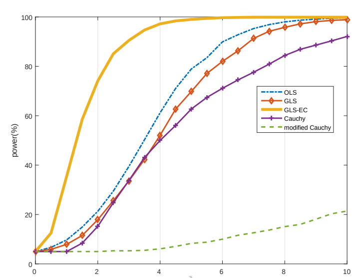
$(b1, a1) = (0, 0)$

$(b2, a2) = (-1/4, -1/4)$



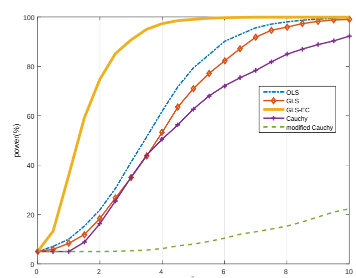
$(b1, a1) = (1/4, 1/4)$

$(b2, a2) = (-1/4, -1/4)$



$(b1, a1) = (1/4, 1/4)$

$(b2, a2) = (0, 0)$

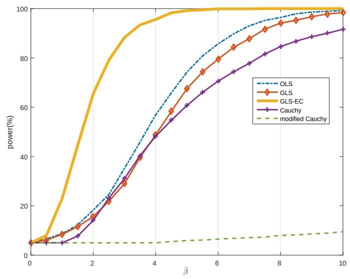


$(b1, a1) = (1/4, 1/4)$

$(b2, a2) = (1/4, 1/4)$

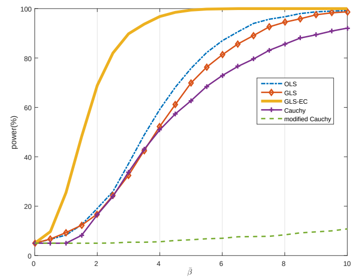
Notes: Figure 4.3 reports the simulated power functions of joint Wald Test with $T = 10$. The considered values of endogeneity parameter ρ_1 and ρ_2 are 0.95 and 0.10, respectively. For values of GHK parameters, we choose $b, a = [-1/4, 0, 1/4]$ for both X_1 and X_2 and compare the size adjusted power function of the OLS, GLS, GSL-EC, Cauchy and Modified Cauchy tests. The dash-dot lines represent the power function of OLS test and the solid line with square marks represent the power function of GLS test. The thick solid lines represent power function our GLS-EC test. The solid line with + marks and dashed lines represent the Cauchy and Modified Cauchy tests, respectively.

Figure 4.4: Simulated Power Function for Joint Wald Tests: $T = 50$



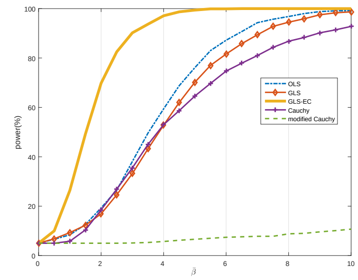
$(b1, a1) = (-1/4, -1/4)$

$(b2, a2) = (-1/4, -1/4)$



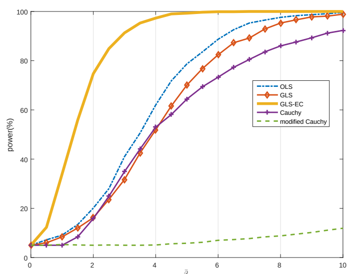
$(b1, a1) = (-1/4, -1/4)$

$(b2, a2) = (0, 0)$



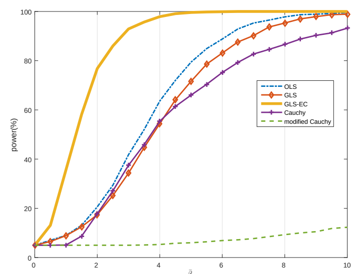
$(b1, a1) = (-1/4, -1/4)$

$(b2, a2) = (1/4, 1/4)$



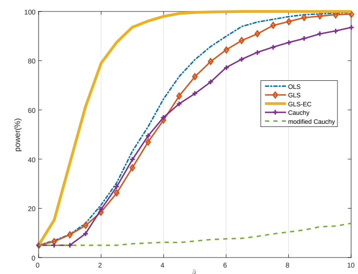
$(b1, a1) = (0, 0)$

$(b2, a2) = (-1/4, -1/4)$



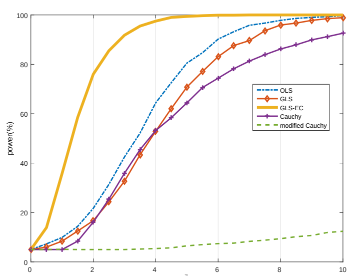
$(b1, a1) = (0, 0)$

$(b2, a2) = (0, 0)$



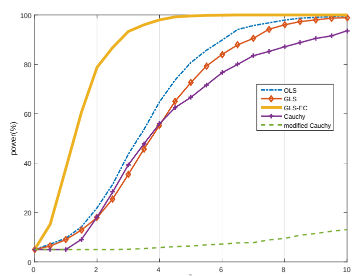
$(b1, a1) = (0, 0)$

$(b2, a2) = (-1/4, -1/4)$



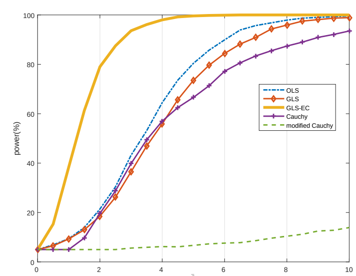
$(b1, a1) = (1/4, 1/4)$

$(b2, a2) = (-1/4, -1/4)$



$(b1, a1) = (1/4, 1/4)$

$(b2, a2) = (0, 0)$



$(b1, a1) = (1/4, 1/4)$

$(b2, a2) = (1/4, 1/4)$

Notes: Figure 4.4 reports the simulated power functions of joint Wald Test with $T = 50$. The considered values of endogeneity parameter ρ_1 and ρ_2 are 0.95 and 0.10, respectively. For values of GHK parameters, we choose $b, a = [-1/4, 0, 1/4]$ for both X_1 and X_2 and compare the size adjusted power function of the OLS, GLS, GSL-EC, Cauchy and Modified Cauchy tests. The dash-dot lines represent the power function of OLS test and the solid line with square marks represent the power function of GLS test. The thick solid lines represent power function our GLS-EC test. The solid line with + marks and dashed lines represent the Cauchy and Modified Cauchy tests, respectively.

the FRED. Long-term government bond returns for the period 1926 to 2013 are from IbbotsonâĂŹs Stocks, Bonds, Bills and Inflation Yearbook. The Term Spread (tms) is the difference between the long term yield on government bonds and the T-bill.

Table 4.3 describes our predictors for return predictability along with the estimated $\hat{\rho}$ for each predictor. The commonly used dividend-price(dp) and earnings-price(ep) ratios are highly correlated with the heteroskedasticity-corrected regression errors and their correlation coefficients are estimated as $\hat{\rho} = -0.95$ and -0.82, respectively. The book-market ratio(bm) has moderate endogeneity with -0.69. Other variables has low correlation with the heteroskedasticity-corrected regression errors. For example, the estimated $\hat{\rho}$ for Three month T-bill rate(tbl) is -0.13.

Table 4.3: Estimated $\hat{\rho}$ for Predictors

Predictors	$\hat{\rho}$	Description
dp	-0.95	log of the Dividend-Price ratio
ep	-0.82	log of the Earnings-Price ratio
bm	-0.69	log of the Book-Market ratio
de	0.02	log of the Payout ratio
tbl	-0.13	3 Month T-bill rate
tms	0.04	Term Spread between long-term Government Bond and 3 Month T-bill rate
dfy	0.11	Default Credit Spread between BAA- and AAA- rated corporate bond yield

Notes: Table 4.3 reports the estimated endogeneity parameter ρ for each predictor. We use data from the Welch and Goyal (2008)'s data library.

4.4.2 Testing for Predictability of Stock Returns

We summarize our test results for return predictability in Tables 4.4, 4.5 and 4.6. In general, our tests support return predictability, and suggest that stock returns are predictable if we choose relevant predictors. In the univariate predictive regression, our results show that three predictors have predictive powers: earnings-price ratio, interest rates and term spread. Seemingly more robust tests based on the Cauchy estimator only picks term spread only as a relevant predictor. As expected, the tests based on the modified Cauchy estimator find no evidence of return predictability. In the univariate predictive regression, though we see some evidence of return predictability, the evidence is not strong and robust.

The evidence for return predictability is much stronger in regressions with multiple predictors. In regressions with dividend-price ratio included as a key predictor, stock returns are mostly predictable with various choices of other predictors. More precisely, the regressions with interest rates, with default credit spread, with interest rates and term spread, and with interest rates, term spread and default credit spread, as additional predictors all yield strong evidence for the predictability of stock returns. Amongst them, the evidence from the regression with interest rates and term spread as additional predictors seem to yield the strongest evidence for return predictability.

The results from regressions including earnings-price ratio as a key predictor are also supportive of return predictability, though the evidence in these regressions is generally somewhat weaker than those including dividend-price ratio as a key predictor. All regressions including earnings-price ratio and other sets of predictors we consider show a positive evidence of the predictability of stock returns. Relatively, it seems that the evidence from the regression with interest rates and term spread as additional predictors is relatively stronger than that from other regressions. This is in parallel to the regressions with the dividend-price ratio as a key predictor.

In sum, it seems to be fair to say that there is a strong evidence of return predictability, with the dividend-price ratio, the earnings-price ratio, interest rates and term spread as most relevant predictors. Our finding is consistent with Ang and Bekaert (2007) who report that stock returns are predictable using the short rate and dividend yield.

Table 4.4: Univariate Predictive Regression

Predictor		OLS	GLS	GLS-EC	Cauchy	Modified Cauchy
dp	$\hat{\beta}$	0.08	0.07	-0.01	0.00	0.00
	s.e.	0.05	0.05	0.01	0.00	0.01
	t-value	1.69	1.48	-1.02	-0.02	0.11
ep	$\hat{\beta}$	0.05	0.08	-0.11	0.00	-0.01
	s.e.	0.05	0.05	0.03	0.00	0.01
	t-value	0.99	1.51	**-3.52**	-0.05	-0.62
bm	$\hat{\beta}$	0.03	0.00	-0.03	0.00	0.00
	s.e.	0.04	0.04	0.03	0.00	0.01
	t-value	0.71	0.05	-1.24	-1.31	-0.19
de	$\hat{\beta}$	0.05	0.00	-0.01	0.01	0.02
	s.e.	0.06	0.07	0.07	0.01	0.03
	t-value	0.81	-0.02	-0.19	1.03	0.71
tbl	$\hat{\beta}$	-1.43	-1.53	-1.69	-0.08	0.00
	s.e.	0.63	0.59	0.59	0.06	0.25
	t-value	**-2.26**	**-2.58**	**-2.86**	-1.46	0.00
tms	$\hat{\beta}$	2.95	2.72	2.84	0.29	1.32
	s.e.	1.33	1.16	1.16	0.12	1.75
	t-value	**2.21**	**2.35**	**2.45**	**2.38**	0.75
dfy	$\hat{\beta}$	4.80	4.47	4.97	0.42	5.40
	s.e.	4.31	4.18	4.18	0.43	13.76
	t-value	1.11	1.07	1.19	0.97	0.39

Notes: In this table, we report the univariate predictive regression results for OLS, GLS, GLS-EC, Cauchy and Modified Cauchy test. For each predictors, we report the estimated coefficient $\hat{\beta}$, standard error and $t-$ value and compare our test results. Bold faced $t-$values represent that the null of no predictability is rejected at a 5% significance level.

Table 4.5 and 4.6 report multivariate predictability regression result for dividend-price and earnings-price ratio using tbl, tms and dfy as the control variables. Our multivariate regression setup is similar to Ang and Bekaert (2007) who show convincing evidence of return predictability with bivariate regression using the short rate and dividend yield.

4.5 Conclusion

Though there has been much debate for over several decades, it remains to be still an open question whether stock returns are predictable. In Section 4, we apply a novel approach to empirically test the predictability of stock returns rigorously and effectively in a multivariate setting. It has long been recognized that there are some prominent features in the time series of stock returns and predictive ratios that can seriously distort the standard testing procedures, which can be summarized as the persistence and endogeneity of predictors and the nonstationarity of stochastic volatilities. There are some methodologies available in the existing literature that make it possible to do robust inferences on return predictability. Unfortunately, however, all existing methodologies we may use to efficiently deal with these two problems only allow for single predictor. Therefore, the test should be done successively for each single predictor. This is a serious limitation. Not only it is impossible to test for the predictability of stock returns jointly by a set predictors, but also we cannot measure the marginal predictive power of each predictor in the presence of other predictors.

In this section, we introduce a novel methodology to test for stock return predictability. The methodology is based on a simple endogeneity correction procedure and modified Cauchy inference, together with a basic heteroskedasticity correction. Through simulations, we show that our methodology is very robust against both per-

Table 4.5: Multivariate Predictive Regression with the Dividend-Price Ratio

	dp		tbl		tms		dfy		Wald
	$\hat{\beta}$	t-value	$\hat{\beta}$	t-value	$\hat{\beta}$	t-value	$\hat{\beta}$	t-value	Statistics
OLS	0.19	**3.31**	-2.64	**-3.64**					16.15
	0.11	**2.25**			3.64	**2.67**			10.00
	0.07	1.43					2.93	0.65	3.29
	0.18	**3.27**	-2.21	-2.73	1.78	1.17			17.56
	0.18	**3.07**	-2.56	-2.89	0.98	0.56	4.99	0.96	18.51
GLS	0.16	**2.90**	-2.35	-3.59					15.15
	0.10	1.92			3.10	2.65			9.23
	0.06	1.25					3.09	0.71	2.71
	0.16	**2.96**	-2.01	-2.95	2.04	1.67			18.01
	0.15	**2.76**	-2.86	-3.41	0.58	0.39	9.84	1.74	21.10
GLS-EC	0.04	**3.31**	-1.47	**-9.57**					92.54
	-0.01	-1.06			-0.07	-0.23			1.13
	0.01	0.58					-6.29	**-6.09**	38.21
	0.04	**3.23**	-1.63	**-10.21**	-0.89	**-3.14**			105.40
	0.04	**3.39**	-1.65	**-8.35**	-0.81	**-2.33**	-0.05	-0.04	106.40
Modified	0.00	0.11	0.02	0.06					0.01
Cauchy	0.01	0.73			0.66	0.67			0.57
	0.00	0.40					0.94	0.26	0.16
	0.01	0.82	0.46	0.85	0.97	0.77			0.74
	0.01	0.80	0.37	0.71	0.47	0.47	1.83	0.60	0.92

Notes: In this table, we summarize the multivariate predictive regression results for OLS, GLS, GLS-EC, Cauchy and Modified Cauchy test using dividend-price(dp) in combination with T-bill rate(tbl), term spread(tms) and default spread(dfy). Bold faced $t-$values represent that the null of no predictability is rejected at a 5% significance level. Wald statistics for each regression is reported in the last column.

Table 4.6: Multivariate Predictive Regression with the Earnings-Price Ratio

	ep		tbl		tms		dfy		Wald
	$\hat{\beta}$	t-value	$\hat{\beta}$	t-value	$\hat{\beta}$	t-value	$\hat{\beta}$	t-value	Statistics
OLS	0.16	**2.80**	-2.65	**-3.47**					13.04
	0.09	1.87			3.86	**2.72**			8.42
	0.04	0.90					4.47	1.03	2.06
	0.16	**2.93**	-2.17	**-2.63**	2.33	1.53			15.41
	0.17	**3.08**	-2.98	**-3.16**	0.88	0.51	9.06	1.77	18.60
GLS	0.20	**3.37**	-2.74	**-3.97**					18.09
	0.10	**1.96**			3.12	**2.66**			9.38
	0.07	1.21					2.54	0.57	2.60
	0.20	**3.37**	-2.41	**-3.34**	1.94	1.59			20.68
	0.18	**3.05**	-3.06	**-3.59**	0.72	0.49	8.22	1.44	22.80
GLS-EC	-0.01	-0.31	-2.17	**-5.05**					37.83
	-0.09	**-2.79**			2.91	**4.03**			28.82
	-0.09	**-2.80**					-3.24	-1.17	13.28
	-0.01	-0.32	-1.81	**-4.07**	2.04	**2.73**			45.73
	-0.01	-0.14	-1.57	**-2.98**	2.38	**2.61**	-2.94	-0.83	44.52
Modified	-0.01	-0.62	-0.11	-0.34					0.39
Cauchy	0.00	-0.27			1.49	0.91			0.90
	0.00	0.28					2.04	0.70	0.51
	0.00	-0.05	0.38	0.78	1.92	0.95			1.12
	0.01	0.66	0.26	0.60	0.40	0.42	3.18	1.08	1.29

Notes: In this table, we summarize the multivariate predictive regression results for OLS, GLS, GLS-EC, Cauchy and Modified Cauchy test using earnings-price(ep) in combination with T-bill rate(tbl), term spread(tms) and default spread(dfy). Bold faced $t-$values represent that the null of no predictability is rejected at a 5% significance level. Wald statistics for each regression is reported at the last column.

sistent and endogenous predictors and nonstationary stochastic volatilities. Moreover, it allows for multiple predictors, as well as it effectively deals with the two econometric challenges in testing for the predictability of stock returns. Our novel approach not only makes it possible to test for the predictability of stock returns jointly by multiple predictors, but also reveals the marginal predictive power of each predictor. Using a new methodology, we find some strong evidence for stock return predictability jointly by dividend-price ratio, earnings-price ration, short-term interest rates and term spread of interest rates.

5. SUMMARY

In Section 2, we develop and estimate a consumption-based asset pricing model with a stochastic volatility function. We assume that the volatility function takes a logistic form with a common factor as well as a stationary idiosyncratic factor. Our logistic volatility function has two asymptotic regimes with a continuum of smooth transitions and the fluctuations of the common latent factor. The main advantage of this setup is that volatility can be persistent without explosive dynamics. We find that the extracted volatility factor explains well the realized volatility series of both the consumption and the dividend data. In addition, we see a counter-cyclical relation of the extracted macroeconomic uncertainty.

We show that the estimated risk-aversion coefficient is around two, and the intertemporal elasticity of substitution is greater than one. Our model produces a high equity premium even without resorting to the common long-run risk component in the conditional means of consumption and dividend growth rates.

In Section 3, we explore the possibility of using the Hodrick-Prescott filter for a new objective of extracting the conditional mean process of a time series in both discrete and continuous time framework. Our HP filter uses the penalty parameter minimizing the first order sample correlation of the residuals. The process extracted from our HP filter is therefore defined as the predictable component of the given time series.

By simulations, we show that our HP filter generally performs better for both discrete and continuous time models than the commonly used Local Linear estimator. Hence our new HP estimator may serve as a useful and effective alternative to the conventional Local Linear estimator in nonparametric estimation of both conditional mean and variance processes of a time series for a wide class of models.

In Section 4, we apply a novel approach to empirically test the predictability of stock returns in a multivariate setting. Our methodology is based on a simple endogeneity correction procedure and modified Cauchy inference, combined with a simple heteroskedasticity correction. Through simulations, we show that our methodology is very robust against both persistent and endogenous predictors and nonstationary stochastic volatilities. Moreover, it allows for multiple predictors, as well as it effectively deals with the two econometric challenges in testing for the predictability of stock returns. Using a new methodology, we find some strong evidence for stock return predictability jointly by dividend-price ratio, earnings-price ration, short-term interest rates and term spread of interest rates.

REFERENCES

Abel, A. (1990). Asset Prices under Habit Formation and Catching up with the Joneses. *American Economic Review*, 80:38–42.

Amihud, Y., Hurvich, C. M., and Wang, Y. (2009). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies*, 22(1):413–434.

Ang, A. and Bekaert, G. (2007). Stock Return Predictability: Is It There? *Review of Financial Studies*, 20:651–707.

Attanasio, P. O. and Weber, G. (1989). Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption. *Economic Journal*, 99:59–73.

Avramov, D. (2002). Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics*, 64:423–458.

Avramov, D. (2004). Stock Return Predictability and Asset Pricing Models. *Review of Financial Studies*, 17:699–738.

Avramov, D. and Chordia, T. (2006). Predicting Stock Returns. *Journal of Financial Economics*, 82:387–415.

Balvers, R. J., Cosimano, T. F., and McDonald, B. (1990). Predicting Stock Returns in an Efficient Market. *Journal of Finance*, 45:1109–1128.

Bansal, R., Kiku, D., and Yaron, A. (2007). Risks for the Long Run: Estimation and Inference. Working Paper, Wharton School, University of Pennsylvania.

Bansal, R. and Yaron, A. (2004). Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance*, 59:1481–1509.

Barro, R. J. (2006). Rare Disasters and Asset Markets in the Twentieth Century. *Quarterly Journal of Economics*, 121:823–866.

Bernanke, B., Boivin, J., and Eliasz, P. S. (2005). Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data. *American Economic Review*, 99(1):350–84.

Bollerslev, T., Gibson, M., and Zhou, H. (2011). Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities. *Journal of Econometrics*, 160:235–245.

Bollerslev, T., Marrone, J., Xu, L., and Zhou, H. (2014). Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence. *Journal of Financial and Quantitative Analysis*, 49:633–661.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies*, 22:4463–4492.

Bollerslev, T. and Todorov, V. (2011). Tails, Fears, and Risk Premia. *Journal of Finance*, LXVI:2165–2211.

Bossaerts, P. and Hillion, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *Review of Financial Studies*, 12:405–428.

Brown, A. L. and Kim, H. (2013). Do Individuals Have Preferences Used in Macro-finance Models? An Experimental Investigation. *Management Science*, 60(4):939–958.

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.

Campbell, J. Y. and Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies*, 1:195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, 21:1509–1531.

Campbell, J. Y. and Yogo, M. (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics*, 81:27–60.

Cavaliere, G. (2004). Unit Root Tests Under Time-Varying Variances. *Econometric Reviews*, 23:259–292.

Cavaliere, G. and Taylor, A. M. R. (2007). Testing for Unit Roots in Time Series Models with Non-Stationary Volatility. *Journal of Econometrics*, 140:919–947.

Chan, Y. L. and Kogan, L. (2002). Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices. *Journal of Political Economy*, 110:1255–1285.

Chang, Y., Kim, H., and Park, J. Y. (2016). Econometric Analysis of Continuous Time Asset Pricing Models. Working Paper: Indiana University, Bloomington, IN.

Chen, W. W. and Deo, R. S. (2009). Bias Reduction and Likelihood-Based Almost Exactly Sized Hypothesis Testing in Predictive Regressions Using the Restricted Likelihood. *Econometric Theory*, 25:1143–1179.

Chernozhukov, V. and Hong, H. (2003). An MCMC Approach to Classical Estimation. *Journal of Econometrics*, 115 (2):293–346.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49:327–335.

Choi, Y., Jacewitz, S., and Park, J. Y. (2016). A Reexamination of Stock Return Predictability. *Journal of Econometrics*, 192:168–189.

Chung, H. and Park, J. Y. (2007). Nonstationary Nonlinear Heteroskedasticity in Regression. *Journal of Econometrics*, 137:230–259.

Cochrane, J. H. (1999). New Facts in Finance. *Economic Perspectives*, 23:36–58.

Cochrane, J. H. (2005). *Asset Pricing, (rev. ed.)*. Princeton University Press, Princeton: NJ.

Constantinides, G. M., Donalson, J. B., and Mehra, R. (2002). Junior Can't Borrow: A New Perspective on the Equity Premium. *Quarterly Journal of Economics*, 117:269–296.

Cremers, K. J. M. (2002). Stock Return Predictability: A Bayesian Model Selection Perspective. *Review of Financial Studies*, 15:1223–1249.

de Jong, R. M. and Sakarya, N. (2015). The Econometrics of the Hodrick-Prescott Filter. *Review of Economics and Statistics*, 98(2):310–317.

Drechsler, I. and Yaron, A. (2011). What's Vol Got to Do With It. *Review of Financial Studies*, 24:1–45.

Epstein, L. and Schneider, M. (2010). Ambiguity and Asset Markets. *NBER Working Paper*, 16181:1–51.

Epstein, L. and Zin, S. (1989). Substitution, Risk Aversion and the Temporal Behavior of Consumption and Aseet Return: A Theoretical Framework. *Econometrica*, 57:937–969.

Eraker, B. (2008). Volatility Premium. Working Paper,University of Wisconsin, Wisconsin School of Business.

Eraker, B. and Shaliastovich, I. (2008). An Equilibrium Guide to Designing Affine Pricing Models. *Mathematical Finance*, 18:519–543.

Fama, E. F. and French, K. R. (1989). Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics*, 25:23–49.

Ferson, W. E., Sarkissian, S., and Simin, T. (2003). Is Stock Return Predictability Spurious? *Journal of Investment Management*, 1:1–10.

Gallant, A. R., Giacomini, R., and Ragusa, G. (2014). Generalized Method of Moments with Latent Variables. Working Paper, Penn State University: State College, PA.

Geweke, J. and Tanizaki, H. (2001). Bayesian Estimation of State-Space Models Using the Metropolis-Hastings Algorithm within Gibbs Sampling. *Computational Statistics and Data Analysis*, 37:151–170.

Goyal, A. and Welch, I. (2003). Predicting the Equity Premium with Dividend Ratios. *Management Science*, 49:639–654.

Guo, H. (2006). On the Out-of-Sample Predictability of Stock Market Returns. *Journal of Business*, 79:645–670.

Guo, H. (2009). Data Revisions and Out-of-Sample Stock Return Predictability. *Economic Inquiry*, 47:81–97.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50:1029–1054.

Hansen, L. P., Heaton, J., Lee, J., and Roussanov, N. (2007). Intertemporal Substitution and Risk Aversion. *Handbook of Econometrics*, 6:3967–4056.

Hansen, L. P., Heaton, J., and Li, N. (2008). Consumption Strikes Back?: Measuring Long-Run Risk. *Journal of Political Economy*, 116:260–302.

Hansen, L. P. and Singleton, K. (1983). Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns. *Journal of Political Economy*, 91:249–268.

Hjalmarsson, E. (2011). New Methods for Inference in Long-Horizon Regressions. *Journal of Financial and Quantitative Analysis*, 46:815–839.

Hodrick, R. J. and Prescott, E. C. (1980). PostWar U.S. Business Cycles: An Empirical Investigation. Working Paper, Carnegie-Mellon University: Pittsburgh, PA.

Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1):pp. 1–16.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian Analysis of Stochastic Volatiltiy Model. *Journal of Business and Economic Statistics*, 12:371–389.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian Analysis of Stochastic Volatiltiy Models with Fat-Tails and Correlated Errors. *Journal of Econometrics*, 122:185–212.

Jeong, D., Kim, H., and Park, J. Y. (2015). Does Ambiguity Matter?: An Estimation of Asset Pricing Models with a Multiple-Priors Recursive Utility. *Journal of Financial Economics*, 115:361–382.

Jeong, M. and Park, J. Y. (2016). An Asymptotic Thoery of Jump Diffusion Model. Working Paper: Indiana University, Bloomington, IN.

Kim, H., Lee, H. I., and Park, J. Y. (2009). A General Approach to Extract Stochastic Volatilities with an Empirical Analysis of Volatility Premium. Working Paper: Texas A&M University, Mays Business School.

Kim, J. and Park, J. Y. (2016a). Asymptotics for Recurrent Diffusions with Application to High Frequency Regression. Forthcoming in *Journal of Econometrics*.

Kim, J. and Park, J. Y. (2016b). Unit Root and Mean Reversion in Economic Time Series Generated from Diffusion Model. Working Paper: Indiana University, Bloomington, IN.

King, R. G. and Rebelo, S. T. (1993). Low Frequency Filtering and Real Business Cycles. *Journal of Economic Dynamics and Control*, 17(1):207–231.

Kreps, D. M. and Porteus, E. L. (1978). Temporal Resolution of Uncertainty and Dynamic Choice Theory. *Econometrica*, 46:185–200.

Lee, J. H. (2016). Predictive Quantile Regression with Persistent Covariates: IVX-QR Approach. *Journal of Econometrics*, 192(1):105–118.

Lettau, M. and Ludvigson, S. (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance*, 56:815–849.

Li, Q. and Racine, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14(2):485–512.

Mehra, R. and Prescott, E. C. (1985). The Equity Premium: A Puzzle. *Journal of Monetary Economics*, 15:145–161.

Menzly, L., Santos, T., and Veronesi, P. (2004). Understanding Predictability. *Journal of Political Economy*, 112:1–47.

Moon, R., Rubia, A., and Valkanov, R. (2004). Long-Horizon Regressions when the Predictor is Slowly Varying. Working Paper, UCLA, Anderson School of Management.

Park, J. and Phillips, P. (2001). Nonlinear Regressions with Integrated Time Series. *Econometrica*, 69(1):117–161.

Park, J. Y. (1992). Canonical Cointegrating Regressions. *Econometrica*, 60:119–143.

Park, J. Y. (2002). Nonstationary Nonlinear Heteroskedasticity. *Journal of Econometrics*, 110:383–415.

Phillips, P. C. B. and Hansen, B. (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies*, 57:99–125.

Phillips, P. C. B. and Lee, J. H. (2013). Predictive Regression under Various Degrees of Persistence and Robust Long-Horizon Regression. *Journal of Econometrics*, 177:250–264.

Ravn, M. O. and Uhlig, H. (2002). On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *Review of Economics and Statistics*, 84:2:371–376.

Rietz, T. A. (1988). The Equity Risk Premium: A Solution. *Journal of Monetary Economics*, 22:117–131.

Schaller, H. and Norden, S. V. (1997). Regime Switching in Stock Market Returns. *Applied Financial Economics*, 7:177–191.

So, M. K. P., Lam, K., and Li, W. K. (1998). A Stochastic Volatility Model with Markov Switching. *Journal of Business and Economic Statistics*, 16:244–253.

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics*, 54:375–421.

Tanizaki, H. (1996). *Nonlinear Filters*. Springer, 2nd edition.

Torous, W., Valkanov, R., and Yan, S. (2004). On Predicting Stock Returns with Nearly Integrated Explanatory Variables. *Journal of Business*, 77:937–966.

Valkanov, R. (2003). Long-Horizon Regressions: Theoretical Results and Applications. *Journal of Financial Economics*, 68:201–232.

Vissing-Jorgensen, A. (2002). Limited Asset Market Participation and the Elasticity of Intertemporal Substitution. *Journal of Political Economy*, 110:825–853.

Weil, P. (1989). The Equity Premium Puzzle and the Risk-free Rate Puzzle. *Journal of Monetary Economics*, 24:401–421.

Welch, I. and Goyal, A. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 21:1455–1508.

Wilcox, S. E. (2007). The Adjusted Earnings Yield. *Financial Analysts Journal*, 63:54–68.

APPENDIX A

APPENDIX TO SECTION 2

A.1 Conditional Mean and Variance of the Logistic Volatility Function

In Appendix A.1, we compute the conditional mean and variance of the logistic volatility function. The key component of the stochastic volatility setup is given as

$$f_{t+1} = \alpha + \frac{\beta}{1 + \exp\left(-\lambda w_{t+1}\right)}, \tag{A.1}$$

$$w_{t+1} = \rho_w w_t + u_{t+1},$$

where u_t follows an *i.i.d.* standard normal distribution and $0 < \rho_w < 1$ holds. The conditional mean and the conditional variance of the stochastic volatility are denoted as $\mathbb{E}_t\left(f_{t+1}\right)$ and $Var_t\left(f_{t+1}\right)$, respectively. For future reference, we write down $\mathbb{E}_t\left(f_{t+1}\right)$ as

$$\mathbb{E}_t\left(f_{t+1}\right) = \alpha + \beta \mathbb{E}_t\left(G_{t+1}\right), \tag{A.2}$$

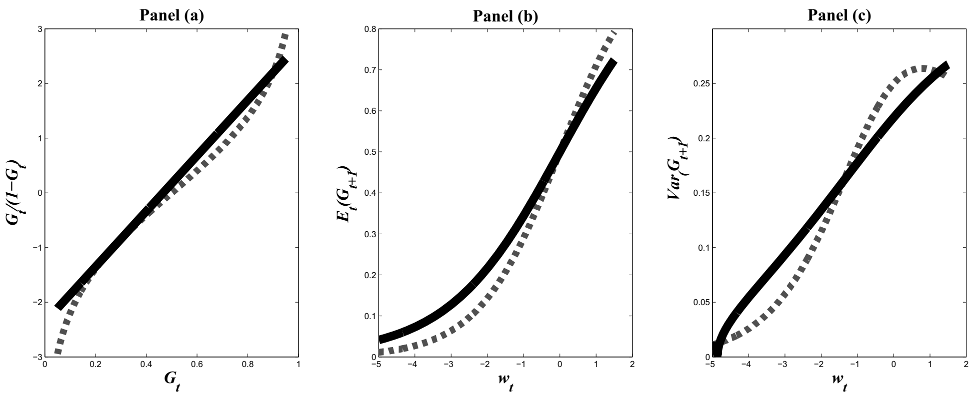
$$G_t = \frac{1}{1 + \exp\left(-\lambda w_t\right)}. \tag{A.3}$$

Although the numerical integration of this moment is easy and accurate, this type of integral does not have a closed-form solution. Instead, we compute an approximate closed-form solution as follows. First, we can show that

$$G_{t+1} \exp(-\lambda \omega_{t+1}) = \left(1 - G_{t+1}\right),$$

$$G_t{}^{\rho_\omega} \exp(-\lambda \rho_\omega \omega_t) = \left(1 - G_t\right)^{\rho_\omega}.$$

Figure A.1: Accuracy of Approximations



Notes: The figure displays comparisons between the true and approximated moments of the baseline logistic function G_t defined in equation (A.3) . Dotted lines refer to the true moments and the straight lines represent the approximated moments. All the moments are computed using the estimated parameters. Panel (a) reports the true and approximated functions of $G/(1-G)$, Panel (b) depicts the conditional expectation of G_t, and Panel (c) describes the conditional variance of G_t, respectively.

Then, by making a ratio of the two equations at t and $t+1$, we have

$$\left(\frac{G_{t+1}}{1-G_{t+1}}\right) = \left(\frac{G_t}{1-G_t}\right)^{\rho_\omega} \exp\left(\lambda u_{t+1}\right).$$

Now, we use a logarithmic linear approximation for $\left(\frac{G_{t+1}}{1-G_{t+1}}\right)$ as

$$\left(\frac{G_{t+1}}{1-G_{t+1}}\right) = \exp\left[\ln\left(\frac{\bar{G}}{1-\bar{G}}\right) + \left(\frac{1}{\bar{G}} + \frac{1}{1-\bar{G}}\right)\left(G_t - \bar{G}\right)\right],$$

for some $\bar{G}$. Then, the law of motion for G_t becomes

$$\begin{aligned}
&\exp\left[\ln\left(\frac{\bar{G}}{1-\bar{G}}\right) + \left(\frac{1}{\bar{G}} + \frac{1}{1-\bar{G}}\right)\left(G_{t+1} - \bar{G}\right)\right] \\
&= \exp\left[\rho_\omega \ln\left(\frac{\bar{G}}{1-\bar{G}}\right) + \rho_\omega\left(\frac{1}{\bar{G}} + \frac{1}{1-\bar{G}}\right)\left(G_t - \bar{G}\right) + \lambda u_{t+1}\right].
\end{aligned}$$

To check the accuracy of this approximation, we compare $\ln\left(\frac{G_{t+1}}{1-G_{t+1}}\right)$ with its approximation by plugging the estimated parameters. Panel (a) of Figure A.1 shows that the approximation is highly accurate for most of the range of G.

Using this result, we can compute the conditional expectation of G as follows:

$$\mathbb{E}_t\left(G_{t+1}\right) = \zeta_0^G + \rho_\omega G_t,$$

with

$$\zeta_0^G = (1-\rho_\omega)\left(\bar{G} - \left(\frac{1}{\bar{G}} + \frac{1}{1-\bar{G}}\right)^{-1}\ln\left(\frac{\bar{G}}{1-\bar{G}}\right)\right).$$

We verify whether or not the approximate closed form given above is close to the true solution. For the true solution, we use a numerical method to compute the conditional expectation, and Panel (b) of Figure A.1 displays the result. According to the comparison, the approximate solution is slightly upward biased for the lower range of w and downward biased for the higher values of w, but it is clear that the overall quality of approximation is very good. Using this, we can compute $\mathbb{E}_t\left(f_{t+1}\right)$ as

$$\mathbb{E}_t\left(f_{t+1}\right) = \alpha + \beta\left(\zeta_0^G + \rho_\omega G_t\right) \tag{A.4}$$

$$= \zeta_0 + \rho_\omega f_t. \tag{A.5}$$

where $\zeta_0 = \beta\zeta_0^G + \alpha\left(1-\rho_\omega\right)$.

Now for the conditional variance, we can show the following relation using Stein's lemma:

$$\begin{aligned}
Var_t\left(f_{t+1}\right) &= \beta^2 Var_t\left(G_{t+1}\right) && \text{(A.6)}\\
&= \left[\beta \mathbb{E}_t\left(\frac{\partial G_{t+1}}{\partial w_{t+1}}\right)\right]^2 && \text{(A.7)}\\
&= \beta^2\lambda^2\left[\mathbb{E}_t\left(G_{t+1}\right) - \mathbb{E}_t\left(G_{t+1}^2\right)\right]^2 && \text{(A.8)}\\
&= \beta^2\lambda^2\left[\left[\mathbb{E}_t\left(G_{t+1}\right)\right]^2 - 2\mathbb{E}_t\left(G_{t+1}\right)\mathbb{E}_t\left(G_{t+1}^2\right) + \left[\mathbb{E}_t\left(G_{t+1}^2\right)\right]^2\right] && \text{(A.9)}
\end{aligned}$$

In addition, from the definition of the variance, we have

$$\begin{aligned}
Var_t\left(f_{t+1}\right) &= \beta^2 Var_t\left(G_{t+1}\right) && \text{(A.10)}\\
&= \beta^2\left[\mathbb{E}_t\left(G_{t+1}^2\right) - \left(\mathbb{E}_t\left(G_{t+1}\right)\right)^2\right]. && \text{(A.11)}
\end{aligned}$$

Thus, putting equations (A.9) and (A.11) together, we have

$$\lambda^2\left[\mathbb{E}_t\left(G_{t+1}^2\right)\right]^2 - \left(1 + 2\lambda^2\mathbb{E}_t\left(G_{t+1}\right)\right)\mathbb{E}_t\left(G_{t+1}^2\right) + \left(1+\lambda^2\right)\left[\mathbb{E}_t\left(G_{t+1}\right)\right]^2 = 0$$

$$\mathbb{E}_t\left(G_{t+1}^2\right) = \frac{\left(1 + 2\lambda^2\mathbb{E}_t\left(G_{t+1}\right)\right) \pm \sqrt{\left(1 + 2\lambda^2\mathbb{E}_t\left(G_{t+1}\right)\right)^2 - 4\lambda^2\left(1+\lambda^2\right)\left[\mathbb{E}_t\left(G_{t+1}\right)\right]^2}}{2\lambda^2}$$

This leads to the following equation.

$$\begin{aligned}
Var_t(G_{t+1}) &= \mathbb{E}_t\left(G_{t+1}^2\right) - \left[\mathbb{E}_t\left(G_{t+1}\right)\right]^2 && \text{(A.12)}\\
&= \left[\sqrt{\frac{1}{4\lambda^2} + \mathbb{E}_t\left(G_{t+1}\right) - \left[\mathbb{E}_t\left(G_{t+1}\right)\right]^2} - \frac{1}{2\lambda}\right]^2. && \text{(A.13)}
\end{aligned}$$

The above equation is the key ingredient to derive equation (2.14), and we use this to numerically evaluate the Euler equation for our estimation and simulation.

Finally, recall that, in Section 2.1.2, we illustrated asset return dynamics of this model using an approximate formula of the equity premium. To derive this formula, we approximate the conditional variance of G_t using the relation, $\mathbb{E}_t\left(G_{t+1}\right) = \zeta_0^G + \rho_\omega G_t$.

$$Var_t(G_{t+1}) \approx \zeta_1 + \zeta_2 G_t,$$

where

$$\begin{aligned}
\zeta_1 = &\left(\sqrt{\frac{1}{4\lambda^2} + \zeta_0^G + \rho_\omega \bar{G} - \left[\zeta_0^G + \rho_\omega \bar{G}\right]^2} - \frac{1}{2\lambda}\right)^2 \\
&- \frac{\left(\sqrt{\frac{1}{4\lambda^2} + \zeta_0^G + \rho_\omega \bar{G} - \left[\zeta_0^G + \rho_\omega \bar{G}\right]^2} - \frac{1}{2\lambda}\right)\left(1 - 2\left[\zeta_0^G + \rho_\omega \bar{G}\right]\right)\left(\zeta_0^G + \rho_\omega \bar{G}\right)}{\sqrt{\frac{1}{4\lambda^2} + \zeta_0^G + \rho_\omega \bar{G} - \left[\zeta_0^G + \rho_\omega \bar{G}\right]^2}} \\
\zeta_2 = &\frac{\left(\sqrt{\frac{1}{4\lambda^2} + \zeta_0^G + \rho_\omega \bar{G} - \left[\zeta_0^G + \rho_\omega \bar{G}\right]^2} - \frac{1}{2}\right)\left(1 - 2\left[\zeta_0^G + \rho_\omega \bar{G}\right]\right)}{\sqrt{\frac{1}{4\lambda^2} + \zeta_0^G + \rho_\omega \bar{G} - \left[\zeta_0^G + \rho_\omega \bar{G}\right]^2}}.
\end{aligned}$$

Then, the conditional variance of f_{t+1} is computed by $\beta^2 Var_t(G_{t+1})$, and denoted

$$\begin{aligned}
Var_t\left(f_{t+1}\right) &= \lambda^2\beta^2\left(\zeta_1 + \zeta_2 G_t\right) \\
&= \xi_1 + \xi_2 f_t,
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= \lambda^2\beta^2\zeta_1, \\
\xi_2 &= \lambda^2\beta^2\zeta_2
\end{aligned}$$

Again in order to check the accuracy of the approximation, we compare the true and approximate solution for the conditional volatility with the estimated parameters. Panel (c) of Figure A.1 shows that the approximated conditional variance of G captures well the true variance.

A.2 Pricing a Consumption Claim

In this subsection, we derive the price of a consumption claim that pays aggregate consumption (C_t) as dividend each period t. Using equations (2.1) to (2.11), we can solve for the price of the consumption claim by conjecturing $z_{c,t}$ as equation (2.12). Plugging $z_{c,t}$ into the Euler equation (2.1), we can show

$$1 = E_t \left[\exp \left\{ \begin{array}{c} \chi \log \delta - \frac{\chi \mu_c}{\psi} + \chi \left(\phi_{0,c} + \phi_{1,c} A_{0,c} - A_{0,c} + \mu_c \right) \\ +\chi \left(1 - \frac{1}{\psi} + \phi_1 A_{1,c} \rho_c - A_{1,c} \right) v_{c,t} - \chi A_{2,c} f(\lambda_c \omega_t) \\ + \left(\chi - \frac{\chi}{\psi} \right) \sqrt{f(\lambda \omega_t)} \varepsilon_{c,t+1} + \chi \phi_1 A_{1,c} \varphi_c \sqrt{f(\lambda_c \omega_t)} \eta_{c,t+1} \\ +\chi \phi_{1,c} A_{2,c} f(\lambda_c \omega_{t+1}) \end{array} \right\} \right]$$

Taking a logarithmic transformation to the equation above, we can use the method of undetermined coefficients to identify $A_{0,c}$, $A_{1,c}$, and $A_{2,c}$.

Solving for $A_{1,c}$, we collect the terms related to $v_{c,t}$ to obtain

$$A_{1,c} = \frac{1-\gamma}{\chi(1-\phi_{1,c}\rho_c)}.$$

For $A_{2,c}$, terms related to f_t are

$$\xi_2 \left(\chi \phi_{1,c} \right)^2 \left(A_{2,c} \right)^2 - 2\chi \left(1 - \phi_{1,c} \rho_\omega \right) A_{2,c} + \left[(1-\gamma)^2 + \left(\frac{(1-\gamma)\,\phi_{1,c}\varphi_c}{(1-\phi_{1,c}\rho_c)} \right)^2 \right] = 0$$

Then, we compute $A_{2,c}$ as

$$A_{2,c} = \frac{(1-\phi_{1,c}\rho_\omega) + \sqrt{(1-\phi_{1,c}\rho_\omega)^2 - \xi_2 (\phi_{1,c})^2 \left[(1-\gamma)^2 + \left(\frac{(1-\gamma)\phi_{1,c}\varphi_c}{(1-\phi_{1,c}\rho_c)}\right)^2\right]}}{\chi (\phi_{1,c})^2 \xi_2}$$

Similarly for $A_{0,c}$, we have

$$A_{0,c} = \frac{\log\delta + \phi_{0,c} + \left(1-\frac{1}{\psi}\right)\mu_c + \phi_{1,c}A_{2,c}\zeta_0 + \frac{\chi}{2}(\phi_{1,c}A_{2,c})^2\xi_1}{(1-\phi_{1,c})}.$$

A.3 Equity Premium

Now we compute the equity premium of the model. The basic derivation is the same as that of the price of the consumption claim, except for the additional process of dividend d. We first have a conjectured form of z_m as

$$z_{m,t} = A_{0,d} + A_{1,d}v_{d,t} + A_{1,c}^m v_{c,t} + A_{2,d}f(\lambda_d\omega_t).$$

Because $A_{0,c}$, $A_{1,c}$, and $A_{2,c}$ were determined in the previous case, we only need to pin down the coefficients associated with z_m, which are $\left(A_{0,d}, A_{1,d}, A_{1,c}^m, A_{2,d}\right)$. First, it is easy to show that the following relations hold.

$$A_{1,d} = \frac{1}{1-\phi_{1,m}\rho_d},$$
$$A_{1,c}^m = \frac{-\frac{1}{\psi}}{1-\phi_{1,m}\rho_c}.$$

Second, for all the terms related to volatility, we need to compute the following part:

$$E_t \left[\begin{array}{c} (\chi - 1)\left(\phi_{1,c} A_{2,c} f(\lambda_c \omega_{t+1}) - A_{2,c} f(\lambda_c \omega_t)\right) \\ + \left(A_{1,d} \varphi_d \sqrt{f(\lambda_d \omega_t)} \eta_{d,t+1} + A_{2,d} f(\lambda_d \omega_{t+1}) \right) - A_{2,d} f(\lambda_d \omega_t) \end{array} \right]$$
$$+ \frac{1}{2} Var_t \left[\begin{array}{c} -\gamma \sqrt{f(\lambda_c \omega_t)} \varepsilon_{c,t+1} + \sqrt{f(\lambda_d \omega_t)} \varepsilon_{d,t+1} \\ +\phi_{1,m} A_{1,c}^m \varphi_c \sqrt{f(\lambda_c \omega_t)} \eta_{c,t+1} + (\chi - 1)\phi_{1,c} A_{1,c} \varphi_c \sqrt{f(\lambda_c \omega_t)} \eta_{c,t+1} \\ +(\chi - 1)\left(\phi_{1,c} A_{2,c} f(\lambda_c \omega_{t+1})\right) \\ +A_{1,d} \varphi_d \sqrt{f(\lambda_d \omega_t)} \eta_{d,t+1} + A_{2,d} f(\lambda_d \omega_{t+1}) \end{array} \right].$$

The first component, or the conditional expectation, can be re-expressed as follows:

$$E_t \left[\begin{array}{c} (\chi - 1)\left(\phi_{1,c} A_{2,c} f(\lambda_c \omega_{t+1}) - A_{2,c} f(\lambda_c \omega_t)\right) \\ +A_{2,d} f(\lambda_d \omega_{t+1}) - A_{2,d} f(\lambda_d \omega_t) \end{array} \right]$$
$$= A_{2,c} \left[(\chi - 1)\phi_{1,c} E_t \left[f(\lambda_c \omega_{t+1}) \right] - f(\lambda_c \omega_t) \right] + A_{2,d} \left[E_t f(\lambda_d \omega_{t+1}) - f(\lambda_d \omega_t) \right]$$

Similarly for the conditional variance term, we have

$$Var_t \left[\begin{array}{c} -\gamma \sqrt{f(\lambda_c \omega_t)} \varepsilon_{c,t+1} + \sqrt{f(\lambda_d \omega_t)} \varepsilon_{d,t+1} \\ +\phi_{1,m} A_{1,c}^m \varphi_c \sqrt{f(\lambda_c \omega_t)} \eta_{c,t+1} + (\chi - 1)\phi_{1,c} A_{1,c} \varphi_c \sqrt{f(\lambda_c \omega_t)} \eta_{c,t+1} \\ +(\chi - 1)\left(\phi_{1,c} A_{2,c} f(\lambda_c \omega_{t+1})\right) \\ +A_{1,d} \varphi_d \sqrt{f(\lambda_d \omega_t)} \eta_{d,t+1} + A_{2,d} f(\lambda_d \omega_{t+1}) \end{array} \right]$$
$$= \left(\gamma^2 + \left(\phi_{1,m} A_{1,c}^m \varphi_c + (\chi - 1)\phi_{1,c} A_{1,c} \varphi_c \right)^2 \right) f(\lambda_c \omega_t) + \left(1 + (A_{1,d} \varphi_d)^2 \right) f(\lambda_d \omega_t)$$
$$+ Var_t \left((\chi - 1)\phi_{1,c} A_{2,c} f(\lambda_c \omega_{t+1}) + A_{2,d} f(\lambda_d \omega_{t+1}) \right).$$

Now, we approximate $f_d(\lambda_d w_t)$ in terms of $f_{c,t}$ at $f_{d,0}$

$$
\begin{aligned}
f_d(\lambda_d w_t) &\approx \alpha_d + \frac{\beta_d \left(f_{d,0} - \alpha_c\right)^{\lambda} \exp\left(\lambda \kappa_c - \kappa_d\right)}{\left(f_{d,0} - \alpha_c\right)^{\lambda} \exp\left(\lambda \kappa_c - \kappa_d\right) + \left(\alpha_c + \beta_c - f_{d,0}\right)^{\lambda}} \\
&\quad + \frac{\lambda \beta_c \beta_d \left(f_{d,0} - \alpha_c\right)^{\lambda - 1} \left(\alpha_c + \beta_c - f_{d,0}\right)^{\lambda - 1} \exp\left(\lambda \kappa_c - \kappa_d\right)}{\left[\left(f_{c,0} - \alpha_c\right)^{\lambda} \exp\left(\lambda \kappa_c - \kappa_d\right) + \left(\alpha_c + \beta_c - f_{d,0}\right)^{\lambda}\right]^2} \left(f_{c,t} - f_{d,0}\right) \\
&= \pi_0 + \pi_d \left(f_{c,t} - f_{d,0}\right) = \left(\pi_0 - \pi_d f_{d,0}\right) + \pi_d f_{c,t},
\end{aligned}
$$

where $\lambda = \lambda_d / \lambda_c$ and π_0 and π_d are defined from the above equation.

With some algebra, we can show that the term purely related to $f(\lambda \omega_t)$ is zero. That is,

$$
\left[
\begin{array}{c}
\gamma^2 + \left(\phi_{1,m} A_{1,c}^m \varphi_c + (\chi - 1) \phi_{1,c} A_{1,c} \varphi_c\right)^2 \\
+ \pi_d \left(1 + \left(A_{1,d} \varphi_d\right)^2\right) + \xi_2 \left((\chi - 1) \phi_{1,c} A_{2,c} + A_{2,d} \pi_d\right)^2
\end{array}
\right] = 0
$$

$$
A_{2,d} = \frac{\xi_2 (1 - \chi) \phi_{1,c} A_{2,c} \pi_d - \sqrt{\left(\xi_2 (\chi - 1) \phi_{1,c} A_{2,c} \pi_d\right)^2 - \xi_2 \pi_d^2 \left[
\begin{array}{c}
\xi_2 \left((\chi - 1) \phi_{1,c} A_{2,c}\right)^2 \\
- \pi_d \left(1 + \left(A_{1,d} \varphi_d\right)^2\right) - \gamma^2 \\
- \left(\phi_{1,m} A_{1,c}^m \varphi_c + (\chi - 1) \phi_{1,c} A_{1,c} \varphi_c\right)^2
\end{array}
\right]}}{\xi_2 \left(\pi_d\right)^2}
$$

For the constant term $A_{0,d}$, the following relation holds.

$$
A_{0,d} = \frac{\left[
\begin{array}{c}
\chi \log \delta + \phi_{0,m} - \frac{\chi}{\psi} \mu_c + (\chi - 1) \left(\phi_{0,c} + \left(\phi_{1,c} - 1\right) A_{0,c} + \mu_c\right) \\
+ \mu_d + A_{2,c} (\chi - 1) \phi_1 \xi^0 \\
+ \frac{1}{2} \left[\left(1 + \left(A_{1,d} \varphi_d\right)^2\right) \pi_{0,c} + \left((\chi - 1) \phi_{1,c} A_{2,c} + A_{2,d} \pi_d\right)^2 \xi_1\right]
\end{array}
\right]}{1 - \phi_{1,m}}.
$$

Innovation to the return $r_{m,t+1}$ is derived as

$$
\begin{aligned}
r_{m,t+1} - E_t\left(r_{m,t+1}\right) &= \phi_{1,m} z_{m,t+1} + g_{d,t+1} - E_t\left(\phi_{1,m} z_{m,t+1} + g_{d,t+1}\right) \\
&= \sqrt{f(\lambda_d \omega_t)}\varepsilon_{d,t+1} + \phi_{1,m} \left(\begin{array}{c} A_{1,d} v_{d,t+1} + A_{1,c}^m v_{c,t+1} + A_{2,d} f(\lambda_c \omega_{t+1}) \\ -E_t\left(A_{1,d} v_{d,t+1} + A_{1,c}^m v_{c,t+1} + A_{2,d} f(\lambda_c \omega_{t+1})\right) \end{array} \right) \\
&= \sqrt{f(\lambda_d \omega_t)}\varepsilon_{d,t+1} + \phi_{1,m} A_{1,d} \varphi_d \sqrt{f(\lambda_d \omega_t)}\eta_{d,t+1} + \phi_{1,m} A_{1,c}^m \varphi_c \sqrt{f(\lambda_c \omega_t)}\eta_{c,t+1} \\
&+ \phi_{1,m} A_{2,d} \sqrt{var_t\left(f_{t+1}\right)} u_{t+1}.
\end{aligned}
$$

Innovation in the stochastic discount factor is

$$
\begin{aligned}
m_{t+1} &= \chi \log \delta - \frac{\chi}{\psi} g_{c,t+1} + (\chi - 1)\, r_{a,t+1} \\
&= \chi \log \delta - \frac{\chi}{\psi} g_{c,t+1} + (\chi - 1)\left(\phi_{0,c} + \phi_{1,c} z_{c,t+1} - z_{c,t} + g_{c,t+1}\right),
\end{aligned}
$$

with $z_{c,t} = A_{0,c} + A_{1,c} v_t + A_{2,c} f(\lambda \omega_t)$. Then, the conditional expectation of m_{t+1} is

$$
\begin{aligned}
E_{t+1}\left(m_{t+1}\right) &= \chi \log \delta + (\chi - 1)\, \phi_{0,c} + (-\gamma)\left(\mu_c + v_{c,t}\right) \\
&+ (\chi - 1) \left(\begin{array}{c} \phi_{1,c}\left(A_{0,c} + A_{1,c} \rho_c v_{c,t} + A_{2,c}\left(\zeta_0 + \rho_\omega f_t\right)\right) \\ -z_{c,t} \end{array} \right).
\end{aligned}
$$

Thus, the innovation to the stochastic discount factor is computed as

$$
m_{t+1} - E_{t+1}(m_{t+1}) = -\gamma \sqrt{f_t} \varepsilon_{c,t+1} - \Lambda_{m,\eta} \sqrt{f_t} \eta_{c,t+1} - \Lambda_{m,u}(t) u_{t+1},
$$

where $\Lambda_{m,\eta} = (1 - \chi)\, \phi_{1,c} \varphi_c A_{1,c}$, and $\Lambda_{m,u}(t) = (1 - \chi)\, \phi_{1,c} A_{2,c} \sqrt{var_t\left(f_{t+1}\right)}$.

Then, the equity premium can be computed as

$$
\begin{aligned}
&E_t\left(r_{m,t+1} - r_{f,t}\right) - \frac{1}{2} Var_t\left(r_{m,t+1}\right) \\
&= -Cov_t\left[\ln m_{t+1}, r_{m,t+1}\right] \\
&= \gamma \sqrt{f_t} \sqrt{f(\lambda_d \omega_t)} \rho + \Lambda_{m,\eta} \phi_{1,m} A_{1,c} \varphi_c f_t + (1 - \chi) \phi_{1,m} \phi_{1,c} A_{2,c} A_{2,d} var_t\left(f_{t+1}\right).
\end{aligned}
$$

APPENDIX B

APPENDIX TO SECTION 3

B.1 Bernanke et al. (2005) Data

Bernanke et al. (2005) (BBE) dataset contains a wide range of 129 macroeconomic variables over the sample period 1959:01 - 2001:07. The original dataset used in Bernanke et al. (2005) contains the 129 variables at quarterly frequency. The original data are then transformed according to the suggestions given in the transformation codes provided (tcodes): For tcode value 1, no transformation is necessary. If tcode is 2, first difference is taken, and if it is 4, the transformed series is logarithm of an original series. Finally, for the value of 5, log difference of the original data is taken. The Monthly BBE data became available and are already transformed. We calculated sample auto-correlations of 129 variables and reported their minimum, mean, median and maximum values.

Table B.1: Sample Auto-correlation of Bernanke et al. (2005) Data

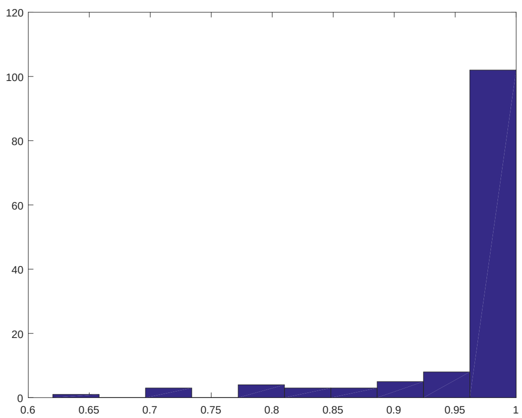
Auto-correlation	Min.	Mean	Median	Max.
Original data (quarterly)	0.62	0.97	0.99	1.00
Transformed data (quarterly)	-0.11	0.64	0.71	1.00
Transformed data (monthly)	-0.41	0.49	0.47	0.99

Notes: Table B.1 reports the minimum, maximum, mean and median values of sample auto-correlation for the Bernanke et al. (2005) (BBE) data variables. The original BBE dataset contains 129 macroeconomic variables.

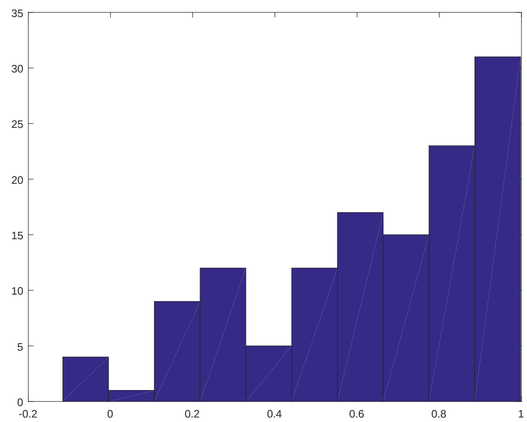
B.2 Boivin et al. (2009) Data

We also try another macroeconomic dataset BGM dataset considered in Boivin et al. (2009). They considered 111 out of 120 variables in the BBE dataset for 1976.Jan-2005.Oct only, but they provide the original as well as the transformed series according to the same transformation codes

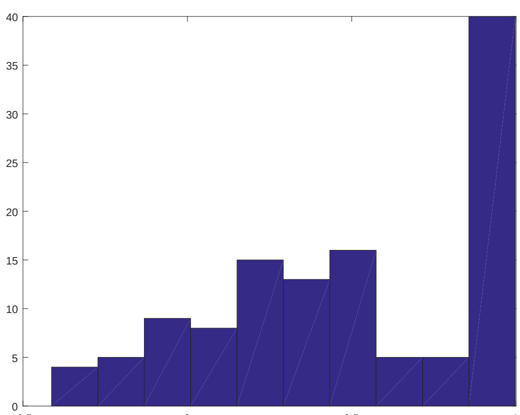
Figure B.1: Histogram of the Bernanke et al. (2005) Data's Sample Auto-correlation



(a) Original data (quarterly)



(b) Transformed data (quarterly)



(c) Transformed data (monthly)

Notes: Figure B.1 displays the histogram plot of the Bernanke et al. (2005) variables' sample auto-correlation. Panel (a) displays the distribution of sample auto-correlation of original BBE variables. Panel (b) and (c) display the histogram plot of transformed quarterly and monthly BBE data, respectively.

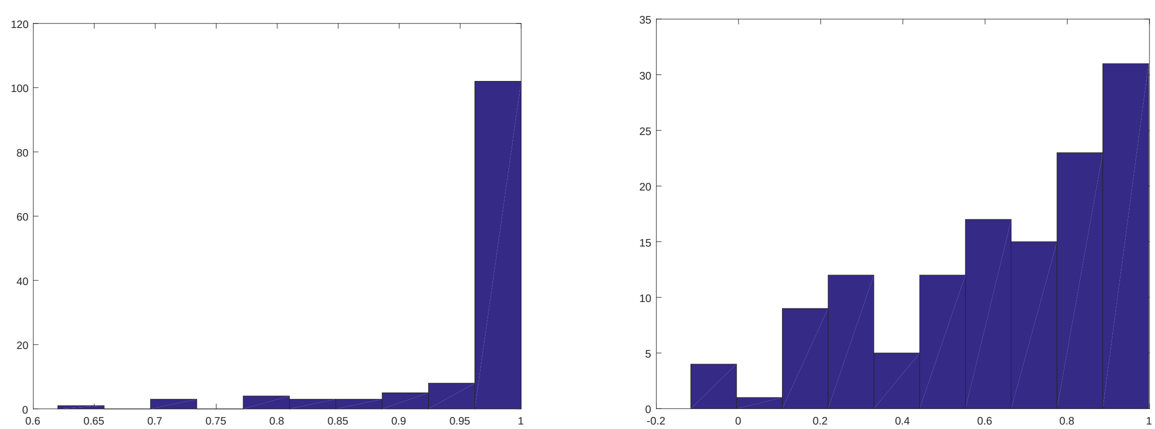
used in BBE. We compute the autocorrelations of the original and transformed variables, and compare them to learn about how their serial dependence structures.

Table B.2: Sample Auto-correlation of Boivin et al. (2009) Data

Auto-correlation	Min	Mean	Median	Max
Monthly Original Data	0.54	0.97	1.00	1.00
Monthly Transformed Data	-0.40	0.52	0.58	1.00

Notes: Table B.2 reports the minimum, maximum, mean and median values of sample auto-correlation for the Boivin et al. (2009) (BGM) data variables. The original BGM dataset contains 111 macroeconomic variables.

Figure B.2: Histogram of the Boivin et al. (2009) Data' Sample Auto-correlation

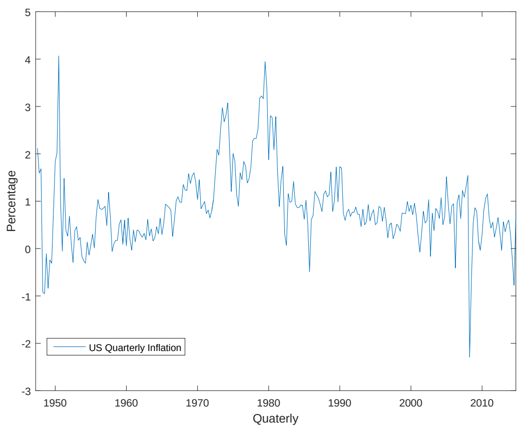


(a) Monthly Data (b) Transformed Monthly Data

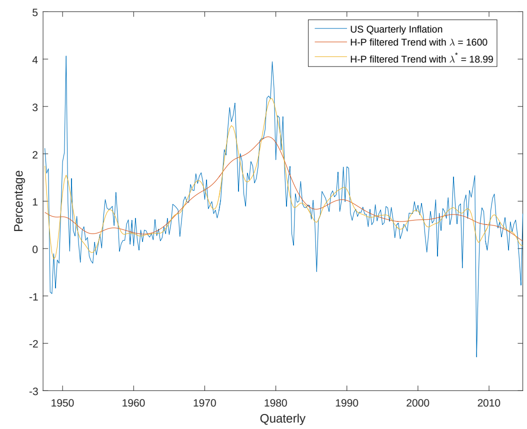
Notes: Figure B.2 displays the histogram plot of the Boivin et al. (2009) variables' sample auto-correlation. Panel (a) displays the distribution of sample auto-correlation of original BGM variables. Panel (b) displays the histogram plot of transformed monthly BGM data.

B.3 HP Estimated Trends for Key Macroeconomic Variables

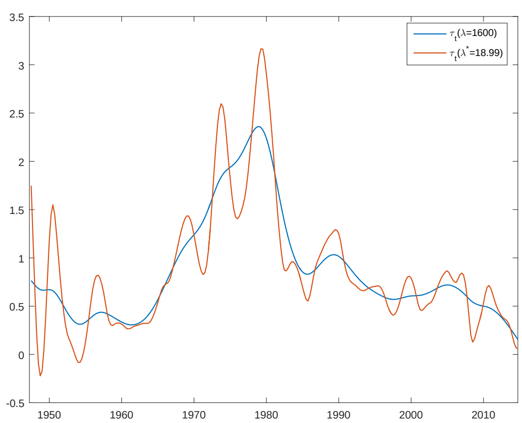
Figure B.3: Quarterly Inflation



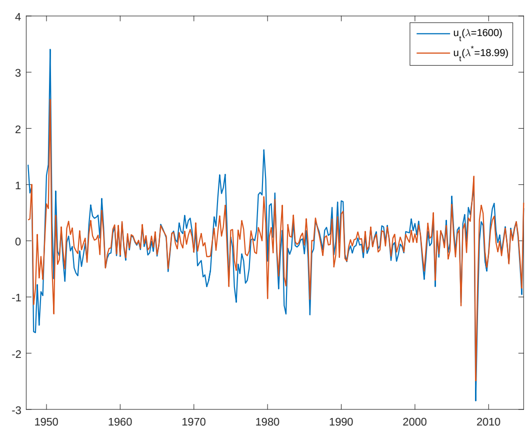
(a) Quarterly CPI



(b) Fitted CPI



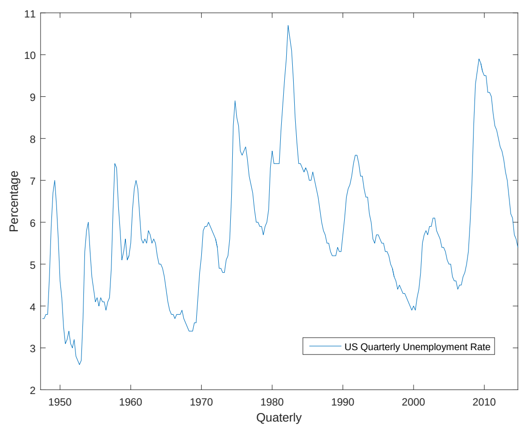
(c) Fitted trends with λ_Q and λ^*



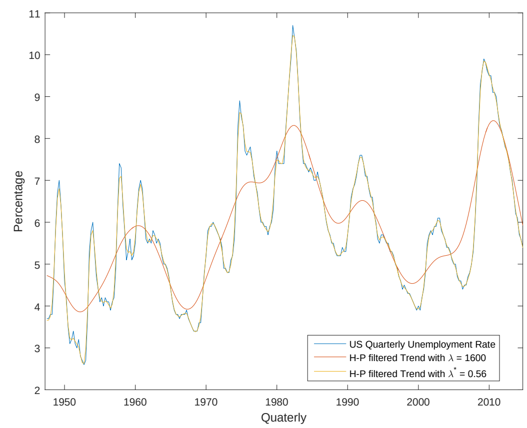
(d) Fitted Residuals: $u_t(\lambda_Q)$ and $u_t(\lambda^*)$

Notes: Figure B.3 reports our empirical result using the Quarterly Inflation. In Panel (a), a time series plot of U.S. Consumer Price Index is displayed. Panel (b) and (c) compare the HP filtered trend with the λ_Q and λ^*. Our optimal smoothing parameter λ^* is found at 18.99. Panel (d) compares the fitted residuals $u_t(\lambda_Q)$ and $u_t(\lambda^*)$ after subtracting the trend components $\tau_t(\lambda_Q)$ and $\tau_t(\lambda^*)$ from data,respectively.

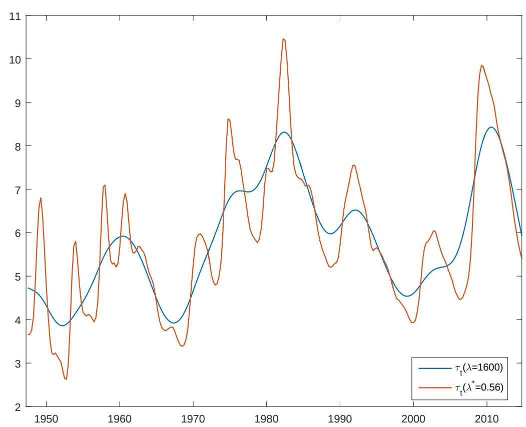
Figure B.4: Quarterly Unemployment Rates



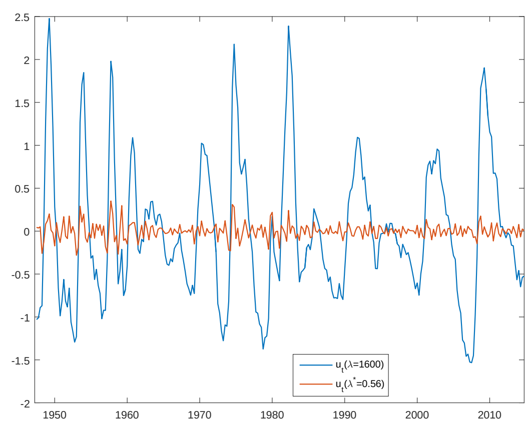
(a) Unemployment rates (quarterly)



(b) Fitted unemployment rates



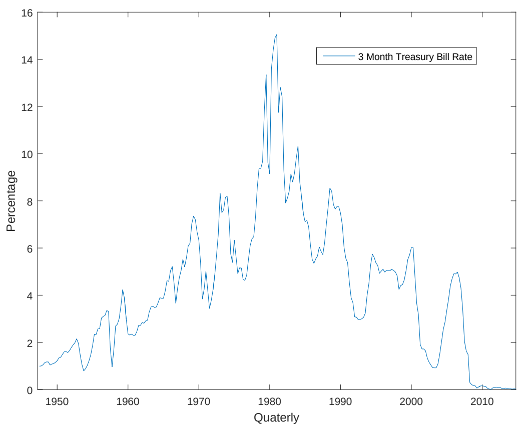
(c) Fitted trends with λ_Q and λ^*



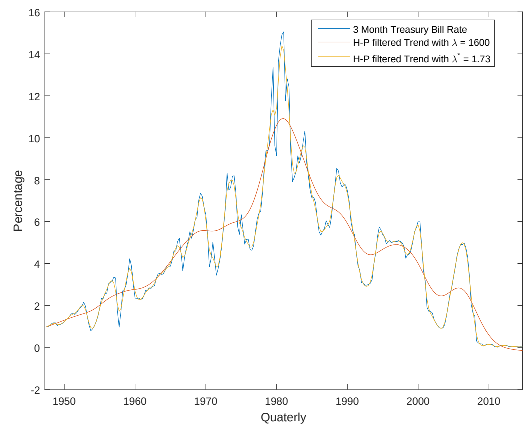
(d) Fitted Residuals: $u_t(\lambda_Q)$ and $u_t(\lambda^*)$

Notes: Figure B.4 reports our empirical result using the quarterly unemployment rate. In Panel (a), a time series plot of U.S. quarterly unemployment rate is displayed. Panel (b) and (c) compare the HP filtered trend with the λ_Q and λ^*. Our optimal smoothing parameter λ^* is found at 0.56. Panel (d) compares the fitted residuals $u_t(\lambda_Q)$ and $u_t(\lambda^*)$ after subtracting the trend components $\tau_t(\lambda_Q)$ and $\tau_t(\lambda^*)$ from data,respectively.

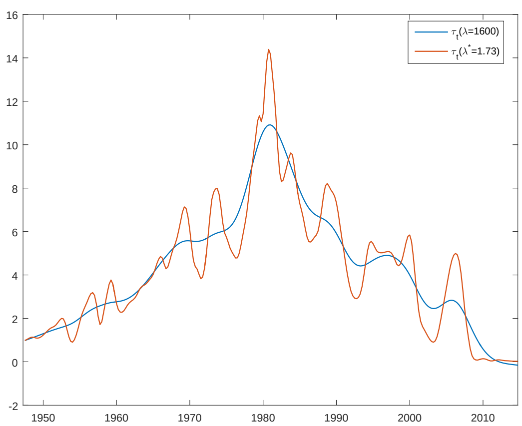
Figure B.5: Quarterly 3 Month T-bill Rate



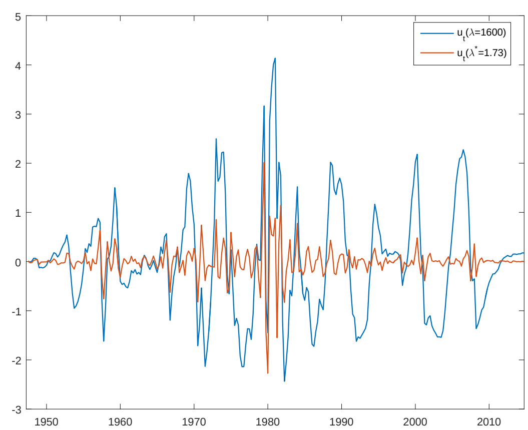
(a) 3 Month T-bill Rate (quarterly)



(b) Fitted short-term interest rate



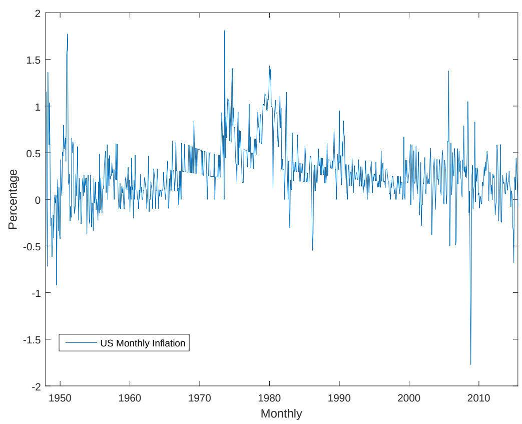
(c) Fitted trends with λ_Q and λ^*



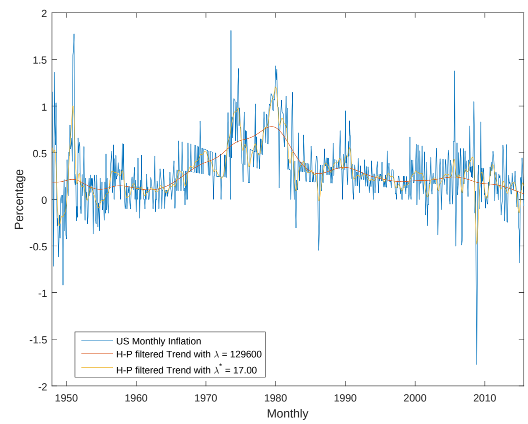
(d) Fitted residuals: $u_t(\lambda_Q)$ and $u_t(\lambda^*)$

Notes: Figure B.5 reports our empirical result using the quarterly 3 month T-bill rate. In Panel (a), a time series plot of quarterly 3 month T-bill is displayed. Panel (b) and (c) compare the HP filtered trend with the λ_Q and λ^*. Our optimal smoothing parameter λ^* is found at 1.73. Panel (d) compares the fitted residuals $u_t(\lambda_Q)$ and $u_t(\lambda^*)$ after subtracting the trend components $\tau_t(\lambda_Q)$ and $\tau_t(\lambda^*)$ from data,respectively.

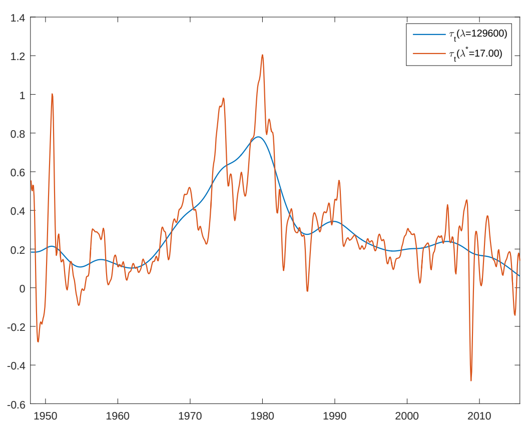
Figure B.6: Monthly Inflation



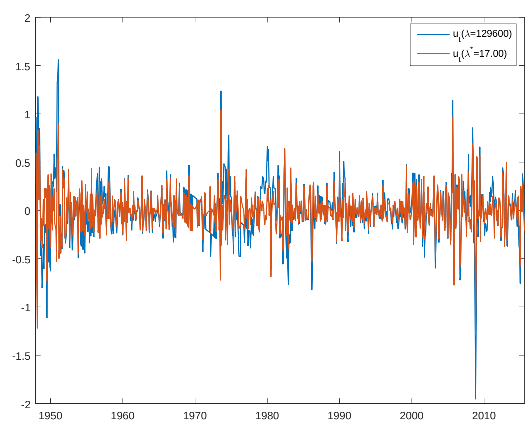
(a) Monthly CPI



(b) Fitted monthly inflation



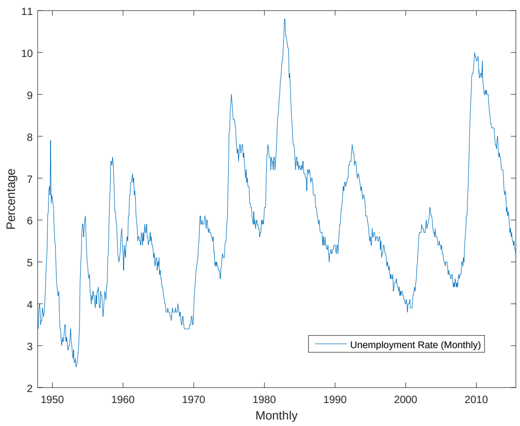
(c) Fitted trends with λ_M and λ^*



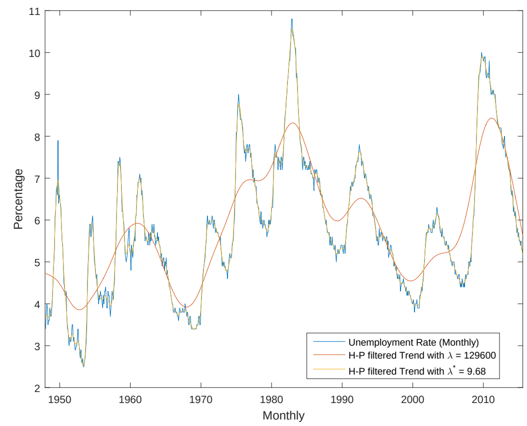
(d) Fitted residuals: $u_t(\lambda_M)$ and $u_t(\lambda^*)$

Notes: Figure B.6 reports our empirical result using the monthly inflation. In Panel (a), a time series plot of monthly CPI is displayed. Panel (b) and (c) compare the HP filtered trend with the λ_M and λ^*. Our optimal smoothing parameter λ^* is found at 17.00. Panel (d) compares the fitted residuals $u_t(\lambda_M)$ and $u_t(\lambda^*)$ after subtracting the trend components $\tau_t(\lambda_M)$ and $\tau_t(\lambda^*)$ from data,respectively.

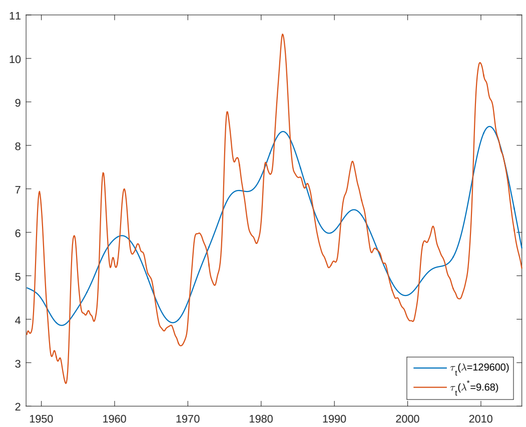
Figure B.7: Monthly Unemployment Rate



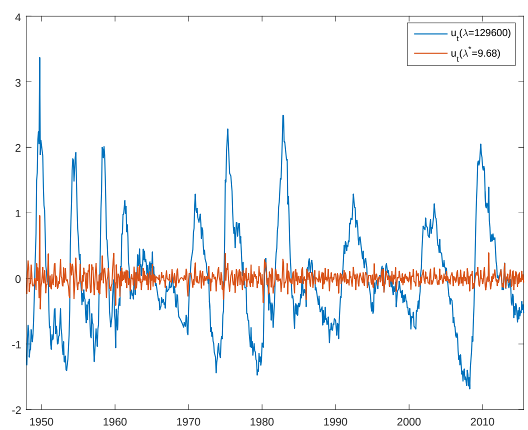
(a) Unemployment rates (monthly)



(b) Fitted monthly unemployment rates



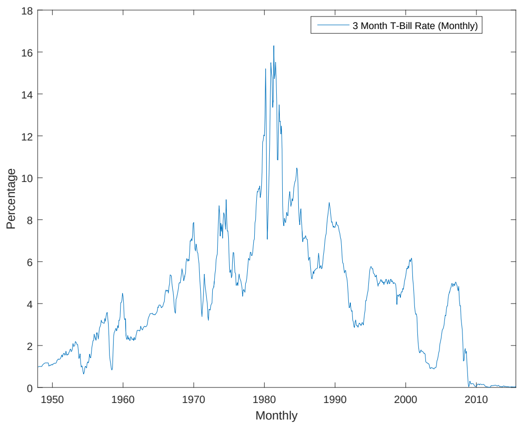
(c) Fitted trends with λ_M and λ^*



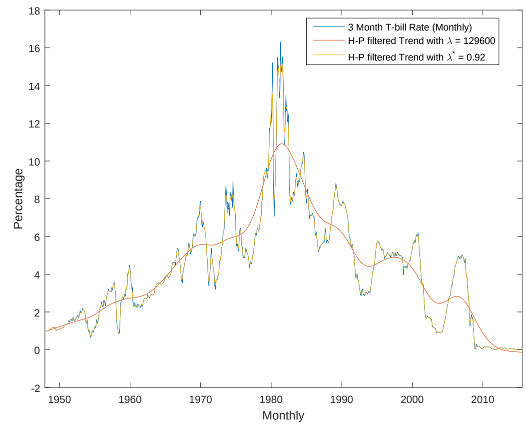
(d) Fitted residuals: $u_t(\lambda_M)$ and $u_t(\lambda^*)$

Notes: Figure B.7 reports our empirical result using the monthly unemployment rate. In Panel (a), a time series plot of monthly unemployment rate is displayed. Panel (b) and (c) compare the HP filtered trend with the λ_M and λ^*. Our optimal smoothing parameter λ^* is found at 9.68. Panel (d) compares the fitted residuals $u_t(\lambda_M)$ and $u_t(\lambda^*)$ after subtracting the trend components $\tau_t(\lambda_M)$ and $\tau_t(\lambda^*)$ from data,respectively.

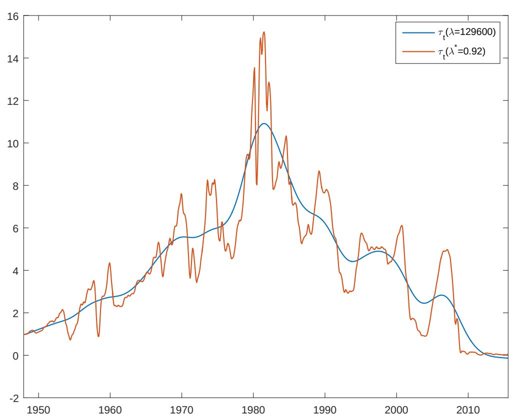
Figure B.8: Monthly 3 Month T-bill Rate



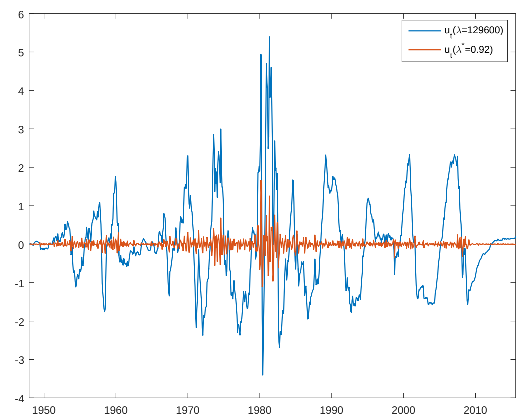
(a) 3 month T-bill rate (monthly)



(b) Fitted monthly interest rate



(c) Fitted trends with λ_M and λ^*



(d) Fitted residuals: $u_t(\lambda_M)$ and $u_t(\lambda^*)$

Notes: Figure B.8 reports our empirical result using the monthly short-term interest rate. In Panel (a), a time series plot of monthly 3 month T-bill rate is displayed. Panel (b) and (c) compare the HP filtered trend with the λ_M and λ^*. Our optimal smoothing parameter λ^* is found at 0.92. Panel (d) compares the fitted residuals $u_t(\lambda_M)$ and $u_t(\lambda^*)$ after subtracting the trend components $\tau_t(\lambda_M)$ and $\tau_t(\lambda^*)$ from data,respectively.