Semiparametric Estimation of Latent Variable Asset Pricing Models

Jeroen Dalderop*

January 2020

Abstract

This paper studies semiparametric identification and estimation of stochastic discount factors in consumption-based asset pricing models with latent state variables. The measurement equations for consumption and dividend shares are specified nonparametrically to allow for robust updating of the Markovian states describing the aggregate growth distribution. For the special case of affine state dynamics and polynomial approximation of the measurement equations, we derive rank conditions for identification, tractable filtering algorithms for likelihood estimation, and closed-form expressions for risk premia and return volatility. Empirically, we find sizable nonlinearities and interactions in the impact of shocks to expected growth and volatility on the consumption share and the discount factor, that help explain the divergence between macroeconomic and stock market volatility.

Keywords: Asset Prices, Volatility, Risk Aversion, Latent Variables, Nonlinear Time Series, Sieve Maximum Likelihood

JEL Codes: C14, G12

---

*Department of Economics, University of Notre Dame, 3060 Jenkins-Nanovic Halls, Notre Dame, IN 46556, USA. This paper has been previously circulated as “Estimating Policy Functions Implicit in Asset Prices”. I am grateful to Yacine Aït-Sahalia, Xiaohong Chen, Oliver Linton, Hamish Low, Ralph Koijen, Alexei Onatski, Richard Smith, and seminar participants at Cambridge, Erasmus University, Federal Reserve Board, LSE, Notre Dame, Penn State, Princeton, Queen Mary, University of Amsterdam, and Wisconsin School of Business, and several conferences, for helpful comments. Financial support by CERF is gratefully acknowledged. Any errors are my own.

1 Introduction

Standard asset pricing models with moderately risk averse households have difficulty reconciling episodes of highly volatile asset prices with relatively smooth fluctuations in macroeconomic fundamentals. More advanced models in which fundamentals and preferences are driven by unobserved state variables have made substantial progress in rationalizing the distribution of asset returns.[1] For the sake of tractability, such models commonly assume that variables such as consumption, dividends, and the stochastic discount factor depend log-linearly on the latent state variables. The resulting log-linear pricing formulas imply that the volatility of asset returns is proportional to that of the state variables. However, sudden drops in asset prices such as the 1987 Black Monday crash did not coincide with significantly volatility in economic growth, nor have upswings in economic uncertainty always triggered much movement on financial markets. Such episodes suggest an important role for nonlinear state-dependence in aggregate growth rates and preferences.

This paper considers the identification and estimation of consumption-based asset pricing models featuring nonlinear dependence of aggregate choice variables on latent state variables. While unobserved by the econometrician, the dynamics of the state variables are identified from their link to an observed aggregrate growth process. In particular, we model the expected consumption and dividends shares of output as functions of Markovian state variables describing the conditional distribution of aggregrate output growth, such as persistent components in its mean and volatility. While their unobservability makes it impossible to directly measure the shape of the consumption, dividend, and pricing functions, the latter can be identified under general hidden Markov assumptions.

Subsequently, we show that the forms of the expected consumption, dividend, and pricing functions identify the dependence of the stochastic discount factor on the states. In particular, the Euler equation for optimal consumption and investment pins down a unique stochastic discount function and fixed discount parameter. These take the form of the unique positive solution to an eigenfunction problem similar to those studied in Escanciano et al. (2015) and Christensen (2017), but allowing for the presence of unobserved state variables.

The robustness of this approach depends on correct specification of the arguments of the stochastic discount factor and the distribution of the state variables. To allow for general state-dependent preferences of a representative agent, the stochastic discount factor is assumed to

[1]Prominent examples are models that feature habit formation (Campbell and Cochrane, 1999), long-run risk (Bansal and Yaron, 2004), stochastic volatility (Drechsler and Yaron, 2010), or variable rare disasters (Gabaix, 2012).

be the multiple of standard power utility over consumption and an unspecified function of the state variables. This semiparametric formulation nests several structural models, notably those based on recursive preferences, in which the multiplicative component is a parametrized function of the state variables, and whose functional form assumptions can therefore be tested. The resulting stochastic discount factor decomposes into a permanent component depending on the level of output and a transitory component depending on the state variables as in Hansen and Scheinkman (2009). We impose some parametric structure on the state variables, which enables to increase their dimensionality, and to extrapolate their distribution across maturities of interest. In particular, the dynamics of the state variables are specified as continuous time affine processes following Duffie et al. (2000). The resulting framework is a tractable generalization of affine equilibrium asset pricing models (see Eraker and Shaliastovich, 2008, for an overview), towards arbitrary functional dependence of consumption and dividends on the state variables.

The nonlinear measurement equations specify the observed levels of consumption, dividends, and asset prices in terms of the unobservables states. Their functional form can be related to optimal consumption and dividend policy in fully structural models, or to partial equilibrium models that include unknown distributions or functional forms. Moreover, we can naturally incorporate long-run cointegration relations in the levels of the macroeconomic series. The optimal policy functions endogenously generate nonlinear variation in the growth rates of the series, without introducing additional exogenous state variables that model these explicitly. Similarly, autocorrelation in growth rates derives from the properties of the stationary state variables, as well as from variable-specific transitory deviations from their optimal levels.

The framework is highly tractable when the unknown functions are approximated by orthogonal polynomials as in Chen (2007). In particular, closed-form expressions for expected returns and return volatility are derived in terms of polynomials of the affine state variables. We study maximum likelihood estimation of the coefficients of the parameters of the state variable dynamics and the polynomial approximation to the measurement equations. The affine case yields tractable recursive algorithms for filtering and predicting the moments of the state variables, and allows using closed-form approximations of the transition density. Moreover, the identification argument can be expressed in terms of rank conditions on the filtered moments of the states. As a result, the maximization step of the EM-algorithm we use reduces to estimating the polynomial coefficients by regression. Finally, we show that integrating out the latent variables of the conditional Euler equation leads to a feasible eigenvector problem that is linear in the coefficients of the approximated stochastic discount function.

The methodology is illustrated by analyzing the impact of time-variation in the conditional mean and volatility of aggregate output growth. We estimate the model using data on U.S. output and consumption, the price-dividend ratio and realized variance of the S&P 500 stock market index, and a panel of price-dividend ratios of sorted portfolios starting in 1929 and running until 2016. Economic uncertainty proxies based on the monthly Industrial Production Index are included in the measurement equation and have a penalization effect on the implied economic volatility state similar to that for financial volatility in Andersen et al. (2015). Empirically, we find periods of high estimated growth volatility clustered around episodes such as the post-war years, the 1980s energy crisis, and the 2008 financial crisis. The frequency and duration of high volatility periods is steadily declining over the sampling period, especially during the high growth 1990s. The consumption-output share shows a negative interaction of low expected growth and high volatility. The dividend-consumption ratio increases with volatility, but appears less sensitive to changes in expected growth. The price-dividend function monotonically increases in expected growth, and decreases in expected volatility. These effects interact such that simultaneously high expected growth and low expected volatility further lifts the price-dividend ratio above two standard deviations from its mean. Conversely, the stochastic discount function reaches its highest levels when expected growth is low and expected volatility is high, suggesting investors have higher marginal utility for payoffs in adverse times. As a result state-dependence in consumption and dividend levels only partially explains state-dependence in the price-dividend ratio, suggesting state-dependent preferences have to be accounted for when using asset prices as predictors of future growth and uncertainty.

Related Literature. The paper contributes to the literature on the estimation of nonlinear dynamic latent variables models, on the identification of risk and risk aversion from asset prices, and on computational methods for nonlinear equilibrium asset pricing models.

Nonlinear dynamic panel data methods have been primarily applied in microeconomics, such as Hu and Shum (2012) and Arellano et al. (2017); for an overview see Arellano and Bonhomme (2011). These papers focus on individual-specific state variables instead of common state variables. An exception is Gagliardini and Gourieroux (2014), who extract common factors in a setting where N grows faster than T and the cross-sectional units are identical and independent, unlike in this paper. Schennach (2014) and Gallant et al. (2017) provide methods to integrate out latent variables in conditional moment models via a minimum entropy criterion or Bayesian methods, respectively, which this paper avoids by employing series approximations. Gallant and Tauchen (1989) and Gallant et al. (1993) introduce series approximations to the transition densities, whereas in this paper non-Gaussianity arises from the nonlinear measurement equations. Latent variables have also been dealt with by inverting observations under monotonicity assumptions. This has been used for affine models for the term structure (Piazzesi, 2010) or option prices (Pan, 2002; Ait-Sahalia and Kimmel, 2010). For nonlinear or multivariate models the inverse mapping is generally not unique. Alternatively, direct proxies for the state variables could be used, such as realized volatility measures to estimate the current level of volatility. For example, stock market volatility can be computed from the variation of high-frequency stock returns (Andersen et al., 2003) or option-implied measures such as the VIX (Carr and Wu, 2008). However stock market volatility does not translate one-to-one into the volatility of economic fundamentals, and moreover is affected by time-varying risk aversion. Using the realized variation of low frequency macroeconomic series suffers from backward looking bias. Using cross-sectional dispersion measures based on firm level data (Bloom, 2009) overcomes this, but requires modelling the conditional means and covariance structure (Jurado et al., 2015). For state variables corresponding to time-varying drift, disaster probability, or changing preferences, no obvious proxy is available. In the presence of noisy or unavailable proxies, the state variables can still be recovered from forward-looking asset prices, which this paper focuses on.

Bansal and Viswanathan (1993), Chapman (1997), Chen and Ludvigson (2009), and Escanciano et al. (2015) estimate unknown components of the stochastic discount factor nonparametrically by generalized method-of-moments, but require observed factor proxies in the absence of knowledge about their distribution. Hansen and Scheinkman (2009) and Christensen (2017) study eigenfunction decompositions of the stochastic discount factor and investigate its temporal properties, also requiring the state variables to be observed or estimated in a first step. Aït-Sahalia and Lo (1998) show that the risk-neutral distribution is nonparametrically identified from option prices, but not its decomposition into the physical distribution and pricing kernel. Bollerslev et al. (2009), Garcia et al. (2011), and Dew-Becker et al. (2017) disentangle variance expectations and variance risk premia, but do not empirically link these to risk in fundamentals. Recent papers by Constantinides and Ghosh (2011) and Jagannathan and Marakani (2015) use equity price-dividend ratios to extract long-run risks, but do not extract uncertainty and risk aversion, unlike this paper.

Finally the paper relates to the literature on approximation methods for expectations of nonlinear functions of continuous-time stochastic processes often encountered in derivative pricing. The generalized transform analysis in Bakshi and Madan (2000) and Chen and Joslin (2012) extends the Fourier transform analysis for affine jump-diffusions in Duffie et al. (2000) towards a large class of models with nonlinear components. This is particularly suitable when the multivariate characteristic function of the driving variables is tractable, and the variables appear in single-index form. Whereas Fourier analysis uses expansions in the complex domain, the approach here uses series expansions in the real domain, which does not require knowledge of the characteristic function. Heston and Rossi (2016) establishes the asymptotic equivalence between series approximations of payoff functions and Edgeworth density expansions. Filipović et al. (2016) show that the class of linear-rational models can also be linked to the class of linearity-generating process introduced in Gabaix (2007). The rational-polynomial formulas in this paper nests both approaches, and make use of the attractive extrapolation properties of ratios of polynomial functions.

Organization. The remainder of this paper is organized as follows. Section 2 introduces the model assumptions and derives closed-form asset pricing formulas. Section 3 outlines the estimation procedure and its asymptotic properties. Section 4 discusses the empirical findings. Section 5 concludes.

2 Setting

This section describes a general class of models for which results are derived. The specific examples are the basis of the empirical section. Throughout let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_t$ be the information filtration satisfying standard regularity conditions. The superscript notation $\mathcal{F}_t^x$ refers to the history $(x_t, x_{t-1}, \dots)$ of the series x_t only.

2.1 Aggregate growth

Let Y_t be an aggregate output or productivity process and let s_t be a D-dimensional state variable that describes the conditional mean, variance, or other distributional characteristics of its growth process $\Delta y_{t+1} = \log\left(\frac{Y_{t+1}}{Y_t}\right)$. The state vector $\mathcal{S}_t = (\Delta y_{t+1}, s_{t+1}) \subset \mathcal{S} \subseteq \mathbb{R}^{D+1}$ in terms of the growth process is assumed to Markovian in s_t, defined as

$$f(\Delta y_{t+1}, s_{t+1} \mid \mathcal{F}_t^{\mathcal{S}}) = f(\Delta y_{t+1}, s_{t+1} \mid s_t).$$

In particular, the level of the output process Y_t does not affect the distribution of its future growth. As a consequence, mean-reversion is ruled out and the output process is non-stationary. On the other hand, the state variables s_t are assumed to be jointly stationary. As a result output growth $\log \frac{Y_{t+\tau}}{Y_t}$ is stationary over any horizon $\tau > 0$ and its conditional distribution only depends on s_t.

2.2 Stochastic discount factor

Suppose there is an infinitely-lived representative agent who maximizes its life time utility $U(\cdot)$ given by

$$U(s_t) = E\left(\int_t^{\infty} \beta^{r-t} u(C_r, s_r) dr \mid s_t\right),$$

where β is a fixed discount parameter, and $u(\cdot)$ is a state-dependent instant utility function that decomposes as

$$u(C_t, s_t) = v(C_t; \gamma) H(s_t),$$

with $v(\cdot)$ the isoelastic utility function

$$v(C_t; \gamma) = \begin{cases} \frac{C_t^{1-\gamma}}{1-\gamma} & \gamma \neq 1 \\ \log C_t & \gamma = 1, \end{cases}$$

and $H(\cdot)$ a general function of the state that could be fully or partially unspecified. Such a specification provides additional stochastic discounting in line with extensions of the standard power utility consumption-based model that include further relevant state variables. Commonly used models with habit formation, recursive preferences, or imperfect risk sharing can be written in this form (Hansen and Renault, 2010).

Under this semiparametric specification, the pricing kernel process $\zeta_t = \beta^t C_t^{-\gamma} H(s_t)$ is the product of a deterministic time-discount factor, a permanent component proportional to the marginal utility of consumption, and a stationary component that allows for general state-dependent preferences. The stochastic discount factor or marginal rate of substitution over states between times t and $t+\tau$ is given by

$$M_{t,t+\tau} = \frac{\zeta_{t+\tau}}{\zeta_t} = \beta^{\tau} \left(\frac{C_{t+\tau}}{C_t}\right)^{-\gamma} \frac{H(s_{t+\tau})}{H(s_t)}.$$

The stochastic discount factor $M_{t,t+\tau}$ is stationary for any fixed horizon τ due to the joint stationarity of consumption growth and the state variables.

2.3 Consumption and dividend policy

In general the optimal consumption choice depends on all sources of wealth and income and all possible investment opportunities. When the primary interest is in understanding the response of consumption to changing economic circumstances, a flexible reduced form approach is to model consumption relative to output via a nonparametric regression function $\psi^c(\cdot)$ of the latent states.

Together with linear dependence on its lag, and an unexplained stationary residual ε_t^c, this yields the semiparametric additive formulation for the log consumption-to-output ratio:

$$c_{t+1} - y_{t+1} = \psi^c(s_{t+1}) + \rho^c(c_t - y_t) + \varepsilon_{t+1}^c, \qquad E\left(\varepsilon_t^c \mid s_t, c_t - y_t\right) = 0.$$

The specification of consumption as a ratio of output guarantees the long run cointegration relation between consumption and output, while the stationary state variables and error component allow for general transitory fluctuations. The inclusion of the lagged value is in line with partial adjustment models for the consumption share towards a target level that changes with the state variables.

Similarly, the aggregate dividend of firms per unit of output or consumption is flexibly modeled as a nonparametric function of the state $\psi^d(\cdot)$ plus a stationary error component ε_t^d. Suppose an index of equities is traded at the price P_t that pays a stochastic dividend level D_t per share. Dividends can be seen as leveraged claim on consumption, which empirically corresponds to a cointegration relation of $\log D_t$ and $\log C_t$ (Menzly et al., 2004). With cointegration parameter λ, the logarithmic residual can be modeled analog to the consumption share as the semiparametric additive specification

$$d_{t+1} - \lambda c_{t+1} = \psi^d(s_{t+1}) + \rho^d\left(d_t - \lambda c_t\right) + \varepsilon_{t+1}^d, \qquad E\left(\varepsilon_t^d \mid s_t, d_t - \lambda c_t\right) = 0.$$

Alternatively the dividend-to-output ratio could be modeled, as any pair of ratios of output, consumption, and dividends, pins down the remaining one. Modeling dividends relative to consumption is the most convenient for our asset price computations.

Combining the cointegration residuals into the observed vector $m_t = (c_t - y_t, d_t - \lambda y_t)$, and allowing for interaction, yields the vector representation

$$m_t = Rm_{t-1} + \psi(s_t) + \varepsilon_t \tag{1}$$

with $\varepsilon_t = (\varepsilon_t^c, \varepsilon_t^d)$ the combined error terms.

2.4 Euler equation for asset prices

In rational expectations equilibrium models, the holding return R_{t+1} on any traded asset satisfies the Euler equation

$$1 = E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{H(s_{t+1})}{H(s_t)}R_{t+1} \mid \mathcal{F}_t\right).$$

When $s_t \in \mathcal{F}_t$, that is, when the latent state variables are in the investor's information set, the Euler equation implies

$$H(s_t) = E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}H(s_{t+1})R_{t+1} \mid s_t\right), \tag{2}$$

which can be recognized as a Type-II Fredholm integral equation. Using infinite-dimensional versions of the Perron-Frobenious theorem, Escanciano et al. (2015) and Christensen (2017) provide conditions for the existence and uniqueness of a positive eigenvalue-eigenfunction pair (β, H) that solves this type of equation.

Computing the solution using the return formulation requires knowledge of the conditional expectation $E\left(\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{t+1} \mid s_{t+1}, s_t\right)$ given both the current and next period state variables. In terms of price-dividend ratios, the Euler equation reads

$$E\left(\frac{P_t}{D_t} \mid s_t\right) = E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{H(s_{t+1})}{H(s_t)}\frac{D_{t+1}}{D_t}\left(1+\frac{P_{t+1}}{D_{t+1}}\right) \mid s_t\right)$$

The Markovian property of the consumption and dividend policies in the extended state vector (s_t, m_t) implies that the expected price-dividend ratio $\phi(s_t, m_t) = E\left(\frac{P_t}{D_t} \mid s_t, m_t\right)$ satisfies the recursive relation

$$\phi(s_t, m_t) = E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{H(s_{t+1})}{H(s_t)}\frac{D_{t+1}}{D_t}\left(1+\phi(s_{t+1}, m_{t+1})\right) \mid s_t, m_t\right). \tag{3}$$

This version of the Euler equation is stated in terms of the contemporaneous projection of the price-dividend ratio on the latent variables and the consumption and dividend shares.

2.5 Generalizing affine models

The class of affine models is widely used to describe non-Gaussian dynamics, as it can incorporate features such as stochastic volatility and leverage effects in a tractable fashion. Following Duffie et al. (2000), the continuous-time affine specification for (dy_t, ds_t) requires its drift $\mu(s_t) =$

$K_0 + K_1 s_t$, and covariance matrix $\mathcal{C}(s_t) = H_0 + \sum_{j=1}^{D} H_{1d} s_{tj}$ to be affine functions of the state variables s_t[2].

Example. Our baseline long-run risk model with stochastic expected growth x_t and volatility v_t is described in continuous time by

$$dy_t = (x_t - \lambda^v v_t)dt + v_t dW_t^y$$

$$dx_t = \kappa^x(\mu - x_t)dt + \omega^x v_t dW_t^x$$

$$dv_t = \kappa^v \left(\theta^v - v_t\right) dt + \omega^v v_t dW_t^v.$$

A positive value for λ^v corresponds to an endogenous growth hypothesis where output uncertainty reduces expected growth. Positive values for the mean reversion parameters κ^x and κ^v assure that $s_t = (x_t, v_t)$ is stationary around its unconditional mean (μ, θ^v). The growth and volatility innovations can have non-zero correlation ρ, while the correlation between growth and expected growth innovations is set to zero to separate persistent and transitory shocks.

The functional specifications of the consumption and dividend policies leads to a general class of potentially nonlinear dynamics of consumption and dividend growth according to

$$dm_t = (R - I)m_t dt + d\psi(s_t) + \sigma^m dW_t^m. \tag{4}$$

When the functions $\psi(s_t)$ are affine in the state variables, this reduces to the benchmark affine model for consumption and dividend growth (Eraker and Shaliastovich, 2008). Nonlinear specifications of $\psi(s_t)$ allow for convex or concave relations, for interaction terms between the state variables, or for higher order effects.

The conditional moments of affine processes can be computed exactly by solving a first-order linear matrix differential equation. This method was introduced for univariate processes by Zhou (2003), and has been generalized to the multivariate setting by Cuchiero et al. (2012) and Filipović et al. (2016). The infinitesimal generator $\mathcal{A}$ of the process S_t, defined as the limit $\mathcal{A}f = \lim_{\tau\to 0} \frac{1}{\tau}\left(E_t(f(S_{t+\tau})) - f(S_t)\right)$ for a function $f : \mathbb{R}^{D+1} \to \mathbb{R}$, for a general affine diffusion process is given by[3]

$$\mathcal{A}f(S) = (K_0 + K_1 S)^T \nabla f(S) + \frac{1}{2}\left(\mathrm{Tr}(\nabla^2 f H_0) + \sum_{j=1}^{D} \mathrm{Tr}(\nabla^2 f H_{1,j}) s_j\right). \tag{5}$$

[2]The affine framework also accommodates discontinuous shocks provided the jump intensity is linear in the state variables.

[3]This property extends to process with quadratic variance specification (Zhou, 2003; Cheng and Scaillet, 2007).

Let $|l| = l_1 + \cdots + l_{D+1}$ denote the length of a multi-index $l \in \mathbb{N}^{D+1}$, let $S^l = \Pi_i S_i^{l_i}$ be a mixed polynomial of degree $|l|$, and let $\mathrm{Pol}_L = \{f : \mathcal{S} \subseteq \mathbb{R}^{D+1} \to \mathbb{R} : \exists a, f = \sum_{0 \le |l| \le L} a_l S^l\}$ be the vector space of mixed polynomials of maximum degree L. For any $f_l \in \mathrm{Pol}_L$, the generator (5) implies that $\mathcal{A} f_l \in \mathrm{Pol}_L$ as well. Applying the canonical basis $B_L = \{S^l : |l| \le L\}$ of Pol_L to $\mathcal{A}$ and collecting the coefficients as

$$\mathcal{A}\mathcal{S}^{l_i} = \sum_{|l^j| \le L} a_{ij} \mathcal{S}^{l_j} \tag{6}$$

leads to a lower triangular matrix $A_L = (a_{ij})$ that by linearity of $\mathcal{A}$ can be used to compute the generator for any polynomial in Pol_L. The coefficients of A can be solved symbolically using standard software. For the baseline model with second-order expansion $L = 2$ its solution is given in Table 4. The conditional moments follow from Dynkin's formula

$$E(\mathcal{S}_{t+\tau}^l \mid \mathcal{S}_t) = \mathcal{S}_t^l + E\left(\int_t^{t+\tau} \mathcal{A}\mathcal{S}_s^l ds\right),$$

which leads to a matrix differential equation with solution

$$E(\bar{\mathcal{S}}_{t+\tau}^L \mid \mathcal{S}_t) = e^{\tau A_L} \bar{\mathcal{S}}_t^L. \tag{7}$$

where $\bar{\mathcal{S}}^L$ is a column vector that stacks monomials up to degree L in lexicographic order.

The conditional moment formula (7) allows exact computation of expected values of polynomials of the state variables. Suppose the consumption and dividend policy functions are approximated by L-degree polynomial expansions

$$\psi_L^c(s) = \sum_{0 \le |l| \le L} c_l s^l = c \cdot \bar{s}^L, \qquad \psi_L^d(s) = \sum_{0 \le |l| \le L} d_l s^l = d \cdot \bar{s}^L,$$

where $\cdot$ indicates inner products. Orthogonal polynomials such as the Hermite or Chebyshev polynomials are spanned by elementary polynomials and can be represented in this way. The growth in consumption relative to output after τ periods equals

$$\Delta c_{t,t+\tau} - \Delta y_{t,t+\tau} = \left(e^{-\rho^c \tau} - 1\right)(c_t - y_t) + \rho^c \int_t^{t+\tau} e^{-\rho^c (t-s)} \psi^c(s_s) ds + \int_t^{t+\tau} e^{-\rho^c (t-s)} dW_s^c.$$

Under the polynomial consumption function approximation, its expected value equals

$$E\left(\Delta c_{t,t+\tau} - \Delta y_{t,t+\tau} \mid s_t, c_t - y_t\right) = \left(e^{-\rho^c \tau} - 1\right)(c_t - y_t) + \rho^c c \cdot Q_L(\tau) \bar{s}_t^L, \tag{8}$$

where the matrix $Q_L(\tau) = \int_0^\tau e^{s(A_L - \rho^c I)} ds$ converges to $Q_L(\infty) = (A_L - \rho^c I)^{-1}$ provided $A_L - \rho^c I$

is invertible. Thus, under the affine-polynomial formulation expected consumption growth is itself a polynomials of the state variables. The same result holds for expected dividend growth. Provided $Q_L(\tau)$ is invertible, this implies a one-to-one mapping between the coefficients of the consumption and dividend policy functions and the coefficients of their expected growth over any horizon τ. In particular, this gives rise to testable overidentifying restrictions.

2.6 Risk-free rate and risk premia

The continuous-time version of the Euler equation for the expected return on dividend-paying assets is

$$0 = \zeta_t D_t dt + E\left(d(\zeta_t P_t) \mid s_t, m_t\right)$$

In particular, the equilibrium rate of return r_t^f on a risk-free asset satisfies

$$r_t^f dt = -E\left(\frac{d\zeta_t}{\zeta_t} \;\Big|\; s_t, m_t\right).$$

With the pricing kernel process written as $\zeta_t = e^{-\delta t - \gamma c_t + h(s_t)}$ with $h(s_t) = \log H(s_t)$, Itô's Lemma yields

$$\frac{d\zeta_t}{\zeta_t} = -\delta dt + (\rho^c - 1)(c_t - y_t)dt - \gamma\sigma_c dW_t^c - \lambda^\zeta(s_t)dS_t + \frac{1}{2}\lambda_s^\zeta(s_t)^T \mathcal{C}(s_t)\lambda_s^\zeta(s_t)dt,$$

where $\lambda^\zeta(s_t) = (\gamma, \gamma\psi_s^c(s_t) + h_s(s_t))$ are the loadings of the pricing kernel increments on $dS_t = (dy_t, ds_t)$. The dependence of the risk-free rate on s_t can take a variety of shapes as determined by the gradients $\psi_s^c(s_t)$ and $h_s(s_t)$.

Under the baseline long-run risk model, the innovations to the pricing kernel follow as

$$\frac{d\zeta_t}{\zeta_t} = -r_t^f dt - \gamma\sigma_c dW_t^c - \gamma v_t dW_t^y - \lambda^x(s_t)\omega^x v_t dW_t^x - \lambda^v(s_t)\omega^v v_t dW_t^v,$$

where the loadings on expected growth and volatility risks are $\lambda^x(s_t) = \gamma\psi_x^c(s_t) + h_x(s_t)$ and $\lambda^v(s_t) = \gamma\psi_v^c(s_t) + h_v(s_t)$, respectively.

The expected excess return on the aggregate stock therefore equals

$$\frac{E_t(dP_t) + D_t dt}{P_t} - r_t^f dt = \lambda(s_t)^T \beta(s_t), \tag{9}$$

in terms of the risk prices $\lambda(s_t) = (\gamma, \gamma, \lambda^x(s_t), \lambda^v(s_t))$ and the asset's risk exposures $\beta(s_t) = E_t\left(\frac{dP_t}{P_t} dX_t\right)$ with respect to $dX_t = (d(c_t - y_t), dS_t)$.

The prices of growth, expected growth, and volatility risks are constant over time if the consumption function ψ^c and the stochastic discount function $H(s_t)$ are log-linear. More generally, the model generates time-varying prices of risks from any nonlinearity in the consumption policy and stochastic discount function. Likewise, the risk exposures depend nonlinearly on the state variables unless the price-dividend ratio is log-linear. Together, these flexible functional forms can generate rich dynamics of growth and volatility risk premia.

2.7 Asset Return Volatility

The framework allows investigating the link between the volatilities of asset returns and the hidden state variables. When the latter include the volatilities of macroeconomic fundamentals, these can be contrasted against measures of price volatility. In particular, the continuous-time theoretical volatility can be linked to realized volatility measures based on high frequency data.

Following our affine-polynomial approximation, suppose the log price-dividend ratio takes the form $\log \frac{P_t}{D_t} = \phi^p_L(s_t) + \alpha \cdot m_t$. Then variation in the log return can be decomposed into variation in the polynomials for the price-dividend ratio, the dividend share of output $d_t - \lambda y_t$, and output growth:

$$d \log P_t = d\phi^p(s_t) + \alpha^* \cdot dm_t + \lambda dy_t$$

where $\alpha^* = \alpha + (\lambda, 1)$. Its unexpected innovation is

$$d \log P_t - E_t(d \log P_t) = \alpha^* \cdot \sigma_m dW^m_t + \lambda^p(s_t) dS_t,$$

where $\lambda^p(s_t) = (\lambda, \phi^p_s(s_t) + \alpha^* \cdot \psi^m_s(s_t))$ are the return's loadings on output growth and the state variables. The price-dividend and dividend-output ratios are driven by the latent variables s_t which generate variation among the components. Itô's Lemma links their joint variation to their gradients with respect to the state variables. In particular, the quadratic variation of the log return equals

$$d\langle \log P \rangle_t = \alpha^* \cdot \sigma_m \sigma'_m \alpha^* + \lambda^p_s(s_t)' \mathcal{C}(s_t) \lambda^p_s(s_t).$$

When the pricing and policy functions are approximated by polynomials, the polynomial gradients give rise to an exact formula for the spot variation of returns that can be used for estimation.

3 Estimation

This section discusses the identification and estimation of the policy functions $\psi = \left(\psi^c, \psi^d\right)^T$, the pricing function ϕ, the preference parameters (β, γ) and stochastic discount function H, and the parameters of the latent variable distribution θ_s. The functional parameters are combined into $h = (\psi, \phi, H)$, the finite-dimensional parameters into $\theta = (\beta, \gamma, \theta_s)$, and both types of parameters into $\vartheta = (\theta, h)$.

The results in this section apply to the discrete-time model formulated by (1) and (2). When the frequency of observation is high, the resulting parameters are expected to be close to their continuous-time counterparts in (4) and (9) that underpin the risk premia and volatility calculations. Moreover, the relation between the instantaneous and cumulative growth in (8) could be used to translate between the timing assumptions.

3.1 State space formulation

The measurements $m_t = \left(\log \frac{C_t}{Y_t}, \log \frac{D_t}{C_t^\lambda}\right)^T$ and normalized prices $p_t = \log \frac{P_t}{D_t}$ contain the aggregate quantities whose conditional mean is approximated by a polynomial in the unobserved state variables s_t. The dynamics of the partially observed Markovian state vector $\mathcal{S}_{t+1} = (\Delta y_{t+1}, s_{t+1})$ are defined by its transition density. The following assumptions describe the interactions between the observations and states:

Assumption 1.

a) $(m_t, p_t, \mathcal{S}_t)$ *are jointly stationary*

b) *The joint process is first-order Markov:*

$$(m_{t+1}, p_{t+1}, \mathcal{S}_{t+1}) \mid \mathcal{F}_t^{m,p,y,s} \sim (m_{t+1}, p_{t+1}, \mathcal{S}_{t+1}) \mid (m_t, p_t, s_t)$$

c) *There is no feedback from the measurements and prices to the states:*

$$\mathcal{S}_{t+1} \mid (m_t, p_t, s_t) \sim \mathcal{S}_{t+1} \mid s_t$$

d) *The state-dependence of the measurements is contemporaneous:*

$$m_{t+1} \mid \mathcal{S}_{t+1}, \mathcal{F}_t^{m,y,s} \sim m_{t+1} \mid (s_{t+1}, m_t).$$

The stationarity assumption of the measurements m_t implies the cointegration of the logarithms of output, consumption, and dividends. The resulting mean-reverting behavior of m_t is a well-known source of return predictability (Lettau and Ludvigson, 2001; Bansal et al., 2007). The presence of state variables in the policy functions allows for the flexible modeling of the cointegration residuals. The joint first-order Markov assumption of observables rules out any dependence on past states or errors. In practice higher order dependence can be allowed for by including further lags in the state vector. The no feedback assumption means that the state variables are a hidden Markov process, and are not caused in the sense of Granger (1969) by errors in the observables. This allows for an interpretation of exogenous variation in the state variables generating endogenous responses in the observations. The hidden Markov assumption is much weaker than the requirement that observations are themselves Markovian, as it allows for dependent observations at all leads and lags. Finally, contemporaneous state-dependence of the measurements rules out their direct dependence on past states, which is a timing assumption also made by Hu and Shum (2012). Moreover, measurements and prices are assumed not to depend on growth Δy_{t+1} beyond the latent states s_{t+1}. This dimensionality reduction is motivated by the Markovianity of growth in s_t. The size-effect of Δy_{t+1} is captured via the denominator in the consumption and dividend shares of output.

For the estimation we focus on the special case of linear Gaussian errors in the measurement equation. In particular, combined with the transition density our dynamic assumptions are summarized by the state space formulation

$$\begin{aligned} m_t &= Rm_{t-1} + \psi(s_t) + \varepsilon_t \\ p_t &= \tilde{\phi}(s_t, m_t) + \eta_t, \qquad\qquad \eta_t = \rho_p \eta_{t-1} + \omega_t \\ \mathcal{S}_{t+1} &\sim f(\mathcal{S}_{t+1} \mid s_t) \end{aligned} \tag{10}$$

where $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$ are i.i.d. Gaussian errors of the measurement equation, and η_t are pricing errors which follows an AR(1) process with i.i.d. Gaussian innovations $\omega_t \sim N(0, \Sigma_\omega)$. Moreover we assume that ε_t and ω_t are uncorrelated. The serially correlated pricing errors, as allowed for under Assumption 1, allow for the presence of persistent deviations from their expected value given the state variables describing aggregate growth. In particular, it allows for persistent stochastic discount rate variation unrelated to fundamentals, following Albuquerque et al. (2016) and Schorfheide et al. (2018).

3.2 Identification

The identification of the functional parameters follows a sequential argument. First, we study the identification of the policy functions ψ under the hidden Markovian assumption. Given ψ, we study the identification of the stochastic discount functions H from the conditional Euler equation.

3.2.1 Identification of the policy functions

Under Assumption 1, our semiparametric formulation is a special case of the nonparametric dynamic latent variable models considered in Hu and Shum (2012). Applying their main result yields high-level invertibility conditions under which the four-period joint density of $(m_{t+1}, \Delta y_{t+1})$ identifies the first-order Markovian distribution of observed and unobserved variables. Intuitively, they exploit the conditional independence of past and future observations given the unobserved state variable. A related argument is used in Arellano and Bonhomme (2011) to identify the consumption rule in terms of a persistent earnings component using future observations. Under our no feedback assumption, future growth realizations are independent of the current measurement given next period's latent state.

The identification argument proceeds sequentially. First, we assume that the parameter θ_s of the state transition density $f(\mathcal{S}_{t+1}|s_t;\theta_s)$ are identified from the dynamics of observed growth Δy_{t+1}. For affine models this can be verified from their Laplace transform (Gagliardini and Gouriéroux, 2019). Second, let $\mathcal{F}^y_{t+1:T}$ denote the future growth realizations $(\Delta y_{t+1}, \ldots, \Delta y_T)$. From the conditional independence $m_t \mid s_t, \mathcal{F}^y_{t+1:T} \sim m_t \mid s_t$, it follows that

$$f(m_t \mid \mathcal{F}^y_{t+1:T}) = \int f(m_t \mid s_{t+1}) f(s_{t+1} \mid \mathcal{F}^y_{t+1:T};\theta_s) ds_{t+1}.$$

Hence, provided the density $f(s_{t+1} \mid \mathcal{F}^y_{t+1:T})$ is complete, the density $f(m_t \mid s_{t+1})$ is identified. Finally, let

$$f(m_{t+1} \mid m_t, \mathcal{F}^y_{t+1:T}) = \int f(m_{t+1} \mid m_t, s_{t+1}) f(s_{t+1} \mid m_t, \mathcal{F}^y_{t+1:T};\theta_s) ds_t,$$

where $f(s_{t+1} \mid m_t, \mathcal{F}^y_{t+1:T}) = \frac{f(m_t|s_{t+1})}{f(m_t|\mathcal{F}^y_{t+1:T})}$ is identified from the previous step. Provided the latter density is also complete, the conditional density of $f(m_{t+1} \mid m_t, s_{t+1})$ and thus its conditional mean are identified.

In the polynomial case, the completeness assumptions reduce to rank conditions, which are easier to interpret and verify. In particular, let $\hat{s}^l_t = E(s^l_t \mid \mathcal{F}^y_{t+1:T})$ be the smoothed conditional

l-th moment of the state s_t given the sample of future growth realizations, and denote its error $e_t^l = s_t^l - \hat{s}_t^l$. Consider the univariate specification $m_t = \rho m_{t-1} + c_L' \bar{s}_t^L + \varepsilon_t$, where ε_t is independent of $\mathcal{F}_{t+1:T}^y$. The latter implies $\varepsilon_t \perp \hat{s}_t^l$ for any $l = 0, \ldots, L$, from which we obtain the linear system of $L+1$ equations

$$E\left(\bar{\hat{s}}_t^{L\prime} m_t\right) = \rho E\left(\bar{\hat{s}}_t^{L\prime} m_{t-1}\right) + c_L' E\left(\bar{s}_t^L (\bar{s}_t^L + \bar{e}_t^L)'\right).$$

By definition of prediction error, $e_t^k \perp \hat{s}_t^l \mid \mathcal{F}_T^y$ for each $(k, l) \in \{0, \ldots, L\}$, so that $E\left(\bar{s}_t^L \bar{\hat{s}}_t^{L\prime}\right) = E\left(\bar{\hat{s}}_t^L \bar{\hat{s}}_t^{L\prime}\right)$. To simultaneously identify ρ we add the first autocovariance of m_t given by

$$E(m_t m_{t-1}) = \rho E\left(m_{t-1}^2\right) + c_L' E\left(\bar{\hat{s}}_t^L m_{t-1}\right),$$

using that $E\left(\bar{s}_t^L m_{t-1}\right) = E\left(\bar{\hat{s}}_t^L m_{t-1}\right)$ under the no feedback assumption. Together, the $L+2$ parameters in (ρ, c_L) are thus identified as the population regression coefficients of m_t on $\bar{\hat{s}}_t^L$ and m_{t-1}, provided the outer product matrix $E\left(\left(\bar{\hat{s}}_t^L;\ m_{t-1}\right)\left(\bar{\hat{s}}_t^L;\ m_{t-1}\right)'\right)$ is invertible.

This identification strategy can be seen as a two-stage version of the use of instrument variables in polynomial measurement errors models by Hausman et al. (1991), who solve a linear system involving moments of the measurement error. In our case, the first stage directly estimates the moments of the unobserved regressors, using the path of growth realizations as instrument.

3.2.2 Identification of the pricing function

When the measurement density is known, the expected price-dividend ratio $\phi(s_t, m_t)$ could be nonparametrically identified from the equation

$$E\left(\frac{P_t}{D_t} \mid m_t, \mathcal{F}_{t+1:T}^y\right) = \int \phi(s_t, m_t) f(s_t \mid m_t, \mathcal{F}_{t+1:T}^y) ds_t, \tag{11}$$

under the hidden Markov assumption and the completeness of $f(s_t \mid m_t, \mathcal{F}_{t+1:T}^y)$. This also covers the semiparametric case $\phi(s_t, m_t) = \phi^p(s_t) e^{\alpha' m_t}$ that arises from the log-linear lag dependence of m_t.

The conditionally Gaussian model combined with the polynomial approximation $\tilde{\phi}_L^p = b_L' \bar{s}_t^L$ gives rise to a two-stage linear regression approach. In particular, let $\tilde{s}_t^l = E(s_t^l \mid \mathcal{F}_T^{y,m})$ be the smoothed l-th moment of the state s_t, which now additionally conditions on the leads and lags of the measurements and growth realizations which are assumed independent of the pricing error η_t. The conditional moment (11) can be written as the regression equation for the log price-dividend ratio p_t

$$E\left(p_t \mid \mathcal{F}_T^{y,m}\right) = b_L' \bar{\bar{s}}_t^L + \alpha' m_t.$$

The variance σ_η^2 of η_t can be identified from the residuals, using $\eta_t \perp (s_t, m_t, \bar{\bar{s}}_t^L)$, as

$$\begin{aligned} Var(\eta_t) &= E\left(\left(p_t - \alpha' m_t\right)\left(p_t - b_L' \bar{\bar{s}}_t^L - \alpha' m_t\right)\right) \\ &= E\left(\left(b_L' \bar{s}_t^L + \eta_t\right)\left(b_L'\left(\bar{s}_t^L - \bar{\bar{s}}_t^L\right) + \eta_t\right)\right). \end{aligned} \tag{12}$$

The expected price-dividend ratio is then computed as $\phi(s_t, m_t) = e^{b_L' \bar{s}_t^L + \alpha' m_t + \frac{1}{2}\sigma_\eta^2}$.

When the pricing error η_t is independent of s_t at all leads and lags, its autocorrelation ρ_η is identified from the autocovariance $E\left(\eta_{t+1}\eta_t\right) = \rho_\eta \sigma_\eta^2$ using

$$\begin{aligned} E\left(\left(p_{t+1} - \alpha' m_{t+1}\right)\left(p_t - \alpha' m_t\right)\right) &= E\left(\left(b_L' \bar{s}_{t+1}^L + \eta_{t+1}\right)\left(b_L' \bar{s}_t^L + \eta_t\right)\right) \\ &= b_L' e^{A_L^s} E\left(\bar{s}_t^L \bar{s}_t^{L\prime}\right) b_L + E\left(\eta_{t+1}\eta_t\right), \end{aligned}$$

where the coefficient matrix A_L^s describing the state moment dynamics (6) and the unconditional moment matrix $E\left(\bar{s}_t^L \bar{s}_t^{L\prime}\right)$ are known given the transition parameter θ_s.

The identification strategy of the pricing function does not restrict any dependence of p_t on its lag p_{t-1}, which is allowed under the hidden Markov assumption. A more efficient estimation approach may use conditioning information in the full sample of leads and lags of p_t, for estimating the moments of s_t. However, computing these moments requires the correct specification of the joint Markovian transition density of (m_t, p_t, s_t) or the corresponding subset used as conditioning variables. The block diagonal structure of $(m_t, p_t) \mid s_t$ means that the measurement dynamics $(m_t \mid s_t, m_{t-1})$ can be estimated without any assumption on the pricing function and errors.

3.2.3 Identification of the stochastic discount function

Once the transition density and the policy functions are known, the identification of the stochastic discount function proceeds essentially as if the state variables are observable. In particular, the stochastic discount function is identified from the price-dividend function if there is a unique solution H that satisfies the Euler equation (3). Equivalently, there is a unique eigenvalue-eigenfunction pair $(H, \frac{1}{\beta})$ that solves

$$\frac{1}{\beta} H(s_t) = E\left(H(s_{t+1})\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{D_{t+1}}{D_t} \frac{1 + \phi(s_{t+1}, m_{t+1})}{\phi(s_t, m_t)} \mid s_t\right). \tag{13}$$

Let $\mathcal{L}^2 = L^2(\mathbb{P})$ denote the Hilbert space of square integrable functions with the unconditional distribution $\mathbb{P}(s)$ of s_t as measure. Let $\mathbb{M} : \mathcal{L}^2 \to \mathcal{L}^2$ be the linear operator defined by

$$\mathbb{M}H(s_t) = E\left(H(s_{t+1})\mathcal{K}(s_t, s_{t+1}) \mid s_t\right),$$

where

$$\mathcal{K}(s_t, s_{t+1}) = E\left(\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{D_{t+1}}{D_t} \frac{1 + \phi(s_{t+1}, m_{t+1})}{\phi(s_t, m_t)} \mid s_t, s_{t+1}\right).$$

Escanciano et al. (2015) and Christensen (2017) show that the following assumption is sufficient for the uniqueness (up to scale) of a positive eigenfunction H and corresponding positive eigenvalue:

Assumption 2.

a) $\mathbb{M}$ *is bounded and compact*

b) $\mathcal{K}(s_t, s_{t+1})$ *is positive a.e.*

The positivity assumption facilitates the use of an infinite-dimensional extension of the Perron-Frobenious theorem for positive valued matrices. This theorem also underpins the pricing kernel recovery theorem for finite Markov chains in Ross (2015). In our setting, a sufficient condition for the positivity of $\mathcal{K}$ is the positivity of the expected price-dividend ratio $\phi(s_t, m_t)$ almost everywhere. Some mild sufficient conditions on $\mathcal{K}$ for $\mathbb{M}$ to be bounded and compact are given in Escanciano et al. (2015) and Christensen (2017). Since H is only identified up to scale, the estimation uses the normalization $E\left(H(s_t)^2\right) = 1$.

Further restrictions that are potentially overidentifying can be formed by adding conditioning variables, such as m_t. This approach is taken in Chen and Ludvigson (2009), and helps to identify H under the completeness of an expected return-weighted density of the state variables.

3.3 Likelihood formulation

Under Assumption 1 the transition density of the state vector $\mathcal{S}_t = (\Delta y_t, s_t)$ depends only on the current unobserved state:

$$f(\mathcal{S}_{t+1} \mid \mathcal{F}_t) = f(\mathcal{S}_{t+1} \mid s_t).$$

The joint log-likelihood function can be decomposed as

$$\begin{aligned}\ell_T(\vartheta) &= \log f(\mathcal{F}_T^p, \mathcal{F}_T^m, \mathcal{F}_T^y; \vartheta) \\ &= \log f(\mathcal{F}_T^p \mid \mathcal{F}_T^{m,y}; \phi, \psi, \theta_s) + \log f(\mathcal{F}_T^m \mid \mathcal{F}_T^y; \psi, \theta_s) + \log f(\mathcal{F}_T^y; \theta_s).\end{aligned}$$

This parameter structures allows for both joint and sequential estimation procedures. In particular, θ_s can be consistently estimated from the series Δy_t alone, ψ from $(m_t, \Delta y_t)$ given $\hat{\theta}_s$, and ϕ from the full observation vector $(p_t, m_t, \Delta y_t)$ given $(\hat{\theta}_s, \hat{\psi})$. The time $t+1$ contribution to the joint log-likelihood function $\ell_T(\vartheta) = \frac{1}{T-1}\sum_{t=1}^{T-1} l_{t+1}(\vartheta)$ is given by

$$l_{t+1}(\vartheta) = \log f(m_{t+1}, p_{t+1}, \Delta y_{t+1} \mid \mathcal{F}_t; \vartheta)$$

The likelihood components are the predictive likelihood of the growth realization Δy_{t+1} given by

$$f(\Delta y_{t+1} \mid \mathcal{F}_t; \vartheta) = \iint f\left(\Delta y_{t+1}, s_{t+1} \mid s_t; \theta_s\right) f\left(s_t \mid \mathcal{F}_t; \vartheta\right) ds_{t+1} ds_t,$$

the conditional likelihood of the measurements m_t after updating by Δy_{t+1}

$$f(m_{t+1} \mid \Delta y_{t+1}, \mathcal{F}_t; \vartheta) = \int f\left(m_{t+1} \mid s_{t+1}, m_t; \psi\right) f\left(s_{t+1} \mid \Delta y_{t+1}, \mathcal{F}_t; \vartheta\right) ds_{t+1},$$

and the conditional likelihood of the prices p_t after updating by $(m_{t+1}, \Delta y_{t+1})$

$$f(p_{t+1} \mid \mathcal{F}_{t+1}^{m,y}; \vartheta) = \int f\left(p_{t+1} \mid m_{t+1}, s_{t+1}; \phi, \sigma_\eta^2\right) f\left(s_{t+1} \mid \mathcal{F}_{t+1}^{m,y}; \vartheta\right) ds_{t+1},$$

where the last likelihood does not condition on past prices $\mathcal{F}_t^p$ so that the integration involves only one period of the state variables.

The log-likelihood contributions are integrals over the latent state variables, where the integrands are the product of the updating density and the transition or measurement density.

In dynamic models it is generally not possible to integrate out the latent variables analytically from the likelihood. Whereas in linear models with Gaussian errors the updating density $f\left(s_t \mid \mathcal{F}_t^y; \theta_s\right)$ can be computed recursively by the Kalman filter, in nonlinear models exact filtering is rarely possible. In line with Taylor expansion methods of solving equilibrium models (e.g. Schmitt-Grohé and Uribe, 2004), a second order approximation to the measurement equation can be performed to identify parameters corresponding to volatility shocks (Fernández-Villaverde and Rubio-Ramírez, 2007). However, this may cause parameters related to higher order moments to become unidentified. Alternatively, particle filtering in combination with Bayesian updating can be used to numerically compute expectations over the state vector, see Doucet and Johansen (2009) for an overview.

Below, we develop an algorithm to efficiently compute the likelihood function when the predictive moments of the state variables are known in closed form, as in (7). In this case, we can use closed-form polynomial approximations to the likelihood ratio of the transition density relative to an auxiliary density. For well-chosen auxiliary densities, this allows for fast and accurate computation of the likelihood integrals by importance sampling. Moreover, the maximization step of the EM algorithm reduces to a linear regression when the measurement equations are approximated by polynomials.

3.4 Closed-form transition density approximation

There is an active literature on approximating the transition density of continuous time Markovian state variables. The major advantage of such an approach is that it prevents the need for pathwise simulation of continuous time processes. Starting from Aït-Sahalia (2002), several papers study approximating the log transition density using Hermite polynomials and solving the coefficients from the Kolmogorov forward and backward equations. This approach works particularly well for multivariate diffusions, and for relatively short horizons. In this paper I use a variant of the approximation method in Filipović et al. (2013), which is based on a series expansion in the state space rather than the time horizon. Therefore the approximation quality does not deteriorate with the horizon.[4] In particular, starting from an auxiliary density $w(\cdot)$, I approximate the likelihood ratio using orthogonal polynomials up to some order J:

$$f^{(J)}(\mathcal{S} \mid s) = w(\mathcal{S} \mid s) \left(\sum_{|l|=0}^{J} c_l(s) H_l(\mathcal{S}) \right),$$

where l is a multi-index and $H_l(\cdot)$ is the Hermite polynomial of degree l whose coefficients are constructed from the Gram-Schmidt process. The projection coefficients based on the weighted $\mathcal{L}_w^2$ norm satisfy

$$c_l(s) = \left\langle \frac{f^{(J)}}{w}, H_l \right\rangle_{\mathcal{L}_w^2} = E\left(H_l(S_{t+\tau}) \mid s_t = s \right).$$

The polynomial moments are linear combinations of the known conditional moments (7).

[4] See Filipović et al. (2013) for details on the approximation properties when $J \to \infty$.

3.5 Filtering and likelihood evaluation

For affine processes, the matrix differential equation (7) can be used to compute the moments of next period's state variables $\mathcal{S}_{t+1} = (\Delta y_{t+1}, s_{t+1})$ by taking linear combinations of the moments of the current unobserved state variables s_t. In particular, using the notation $\bar{\mathcal{S}}^L_{t|t} = E(\bar{\mathcal{S}}^L_t \mid \mathcal{F}_t)$, the predictive moments are given by

$$\bar{\mathcal{S}}^L_{t+\tau|t} = e^{\tau A_L} I^s \bar{s}^L_{t|t}.$$

where I^s is a selection matrix that sets to zero the monomials in $\bar{\mathcal{S}}^L_{t|t}$ that feature lagged growth Δy_t. Combined with the polynomial approximation of the transition density, this relation allows the closed-form approximation of the predictive density as a function of the filtered moments

$$f(\Delta y_{t+1}, s_{t+1} \mid \mathcal{F}_t; \theta_s) \approx w(\Delta y_{t+1}, s_{t+1}) \left(\sum_{|l|=0}^{J} c_{lt} H_l(\Delta y_{t+1}, s_{t+1}) \right),$$

with $c_{lt} = c_l(\bar{s}^L_{t|t})$ the updated coefficients of the transition density approximation. In particular, the predictive density of observed growth $f_{\theta_s}(\Delta y_{t+1} \mid \mathcal{F}_t)$ follows from the marginal predictive moments of Δy_{t+1}.

The updated moments of s_{t+1} given Δy_{t+1} can be approximated via importance sampling:

$$E(s^l_{t+1} \mid \mathcal{F}_t, \Delta y_{t+1}) \approx \frac{1}{N^s} \sum_{i=1}^{N^s} w_i s^l_i$$

where $(s_i)_{i=1}^{N^s}$ are simulated data from the auxiliary densities, and $(w_i)_{i=1}^{N^s}$ are the sampling weights

$$w_i = \frac{\sum_{|l|=0}^{J} c_{lt} H_l(\Delta y_{t+1}, s_i)}{\sum_{i=1}^{N^s} \sum_{|l|=0}^{J} c_{lt} H_l(\Delta y_{t+1}, s_i)}.$$

The predictive likelihood of m_{t+1} can then be computed from the simulated states as

$$f(m_{t+1} \mid \Delta y_{t+1}, \mathcal{F}_t; \vartheta) = \frac{1}{N^s} \sum_{i=1}^{N^s} w_i f\left(m_{t+1} \mid s_i, m_t; \vartheta\right).$$

The updated moments of s_{t+1} given (m_{t+1}, p_{t+1}) follow as $\bar{s}^L_{t+1|t+1} = \frac{1}{N^s} \sum_{i=1}^{N^s} w^*_i \bar{s}_i$ using the updated sampling weights

$$w^*_i = \frac{w_i f\left(m_{t+1}, p_{t+1} \mid s_i, m_t; \vartheta\right)}{\frac{1}{N^s} \sum_{i=1}^{N^s} w_i f\left(m_{t+1}, p_{t+1} \mid s_i, m_t; \vartheta\right)}.$$

The updated moments of the states $\bar{s}^L_{t+1|t+1}$ are sufficient information to perform the prediction and updating steps in the next period.

For the smoothing step, we perform a similar sampling approximation based on the backward recursive relation

$$f(s_t \mid \mathcal{F}_T) = f(s_t \mid \mathcal{F}_t) \int \frac{f(s_{t+1} \mid s_t)}{f(s_{t+1} \mid \mathcal{F}_t)} f(s_{t+1} \mid \mathcal{F}_T) ds_{t+1},$$

starting from the last period's updated moments $\bar{s}^L_{T|T}$.

3.6 EM-algorithm

Global maximization of the approximated likelihood function is computationally unattractive when the parameter space is large-dimensional, as is the case when approximating functional parameters. However, when the measurement equations are approximated by polynomials, their coefficients can be found by linear regression on the estimated moments of the states, in line with the identification argument for polynomial policy and pricing functions in section 3.2. This motivates the use of the following version of the Expectation-Maximization (EM) algorithm. Starting from the initial parameter estimates $\tilde{\vartheta}_L = (\tilde{\theta}_s, \tilde{h}_L)$, running the above likelihood evaluation yields the filtered moments $\tilde{\bar{s}}^L_{t|t} = E(\bar{s}^L_t \mid \Delta y_t, \mathcal{F}^{m,y}_{t-1}; \tilde{\vartheta}_L)$ based on observing the growth realization but not the other measurements. For the maximization step, estimate the polynomial coefficients $\hat{c}_L$ and lag parameters $\hat{R}$ by the least-squares regression of m_t on $(m_{t-1}, \tilde{\bar{s}}^L_{t|t})$. This yields consistent estimates since its conditional mean given available information reduces to

$$E\left(m_t \mid m_{t-1}, \mathcal{F}_t\right) = R m_{t-1} + c'_L \tilde{\bar{s}}^L_{t|t},$$

provided the lagged value m_{t-1} is in the information set used to estimate $\tilde{\bar{s}}^L_{t|t}$. Similarly, we estimate the coefficients $\hat{b}_L$ of the pricing function by polynomial regression of p_t on $\hat{\bar{s}}^L_{t|t}$ and m_t, where $\hat{\bar{s}}^L_{t|t}$ are the filtered states after observing $(m_t, \Delta y_t)$ and past observations.

The covariance matrix of the errors Σ_ε can be consistently estimated as

$$\hat{\Sigma}_\varepsilon = E\left(m_t(m_t - \hat{R} m_{t-1} + \hat{c}'_L \tilde{\bar{s}}^L_{t|t})'\right),$$

using the orthogonality $\varepsilon_t \perp (s_t, m_{t-1}, \tilde{\bar{s}}^L_{t|t})$, following the same steps as in (12).

After the EM-algorithm converges, joint parameter optimization starting from the EM-estimate can be performed to compute the MLE estimate and information matrix for inference, following Watson and Engle (1983).

3.7 Consistency

The population parameters of interest are given by

$$(\theta_0, h_0) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}} \lim_{T \to \infty} \ell_T(\theta, h), \tag{14}$$

and the maximum likelihood estimator by

$$(\hat{\theta}, \hat{h}) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}} \ell_T(\theta, h), \tag{15}$$

where Θ is a finite-dimensional parameter space, and $\mathcal{H} = \prod_{m=1}^{K} \mathcal{H}_{\psi_m} \times \mathcal{H}_\phi \times \mathcal{H}_H$ is a Cartesian product of infinite-dimensional parameter spaces for the policy functions ψ_m, $m = 1, \ldots, K$, and the pricing function ϕ. Also define the product space $\mathit{\Theta} = \Theta \times \mathcal{H}$. Let the spaces $\mathcal{H}_m$ and $\mathcal{H}_\phi$ be equipped with the weighted Sobolev norm $\| \cdot \|$, which sums the expectations of the partial derivatives of a function. In particular, for λ a $D \times 1$ vector of non-negative integers such that $|\lambda| = \sum_{s=1}^{D} \lambda_s$, and $D^\lambda = \frac{\partial^{|\lambda|}}{\partial y_1^{\lambda_1} \cdots \partial y_D^{\lambda_D}}$ the partial derivative operator, it is given for some positive integers r and p by

$$\|g\|_{r,p} = \left\{ \sum_{|\lambda| \le r} E \left(D^\lambda g(S) \right)^p \right\}^{1/p}.$$

For vector-valued functions define $\|g\|_{r,p} = \sum_{m=1}^{K} \|g_m\|_{r,p}$. Instead of maximizing $\ell_T(\theta)$ over the infinite dimensional functional space $\mathcal{H}$, the method of sieves (Chen, 2007) controls the complexity of the model in relation to the sample size by minimizing over approximating finite-dimensional spaces $\mathcal{H}_L \subseteq \mathcal{H}_{L+1} \subseteq ... \subseteq \mathcal{H}$ which become dense in $\mathcal{H}$. For some positive constant B, define $\mathcal{H}$ as the compact functional space

$$\mathcal{H} = \left\{ g : \mathbb{R}^D \mapsto \mathbb{R} : \|g\|_{r,2}^2 \le B \right\}$$

All functions in $\mathcal{H}$ have at least r partial derivatives that are bounded in squared expectation. The polynomials in this space can be conveniently characterized in terms of their coefficients. Let $\underline{p}_L = (p_1(w), ..., p_L(w))$ be a set of basis functions, and consider the finite-dimensional series approximator $g_L(w) = \sum_{l=1}^{L} \gamma_j p_j(w) = \underline{\gamma} \cdot \underline{p}_L(w)$. Define

$$\Lambda_L = \sum_{|\lambda| \le r} E \left(D^\lambda \underline{p}_L(z) D^\lambda \underline{p}_L(z)^T \right),$$

which implies that $g_L(w) \in \mathcal{H}$ if and only if $\gamma^T \Lambda_L \gamma \leq B$ (Newey and Powell, 2003). Therefore the optimization in (15) is redefined over the compact finite-dimensional subspace $\mathcal{H}_{L(T)}$:

$$(\hat{\theta}, \hat{h}_L) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}_{L(T)}} \ell_T(\theta, h), \tag{16}$$

where $\mathcal{H}_{L(T)}$

$$\mathcal{H}_{L(T)} = \left\{ g(w) = \sum_{l=1}^{L(T)} \gamma_j p_j(w) : \; \gamma^T \Lambda_{L(T)} \gamma \leq B \right\}.$$

Also define the Sobolev sup-norm

$$\|g\|_{r,\infty} = \max_{|\lambda| \leq r} \sup_z \left| D^\lambda g(z) \right|.$$

Then the closure $\bar{\mathcal{H}}$ of $\mathcal{H}$ with respect to the norm $\|g\|_{r,\infty}$ is compact (Gallant and Nychka, 1987; Newey and Powell, 2003).

Consider the following set of assumptions:

Assumption 3.

a) *The parameter space $\mathit{\Theta} = \Theta \times \mathcal{H}$ is compact, and the population log-likelihood is uniquely maximized at the interior point $\vartheta_0 = (\theta_0, h_0)$.*

b) *(m_t, p_t) is a strong mixing stationary process, with $E\left(\|m_t\|^2\right) < \infty$ and $E\left(\|p_t\|^2\right) < \infty$*

c) *The transition density satisfies*

$$\left| \log f\left(\mathcal{S} \mid s;\; \theta_s\right) - \log f\left(\mathcal{S} \mid s;\; \tilde{\theta}_s\right) \right| \leq c(\mathcal{S}, s) \|\theta_s - \tilde{\theta}_s\|^u$$

for some $u > 0$ with $E\left(c(\mathcal{S}_{t+1}, s_t)^2\right) < \infty$, and $\mathrm{Var}\left(\log f\left(\mathcal{S}_{t+1} | s_t; \theta_{s0}\right)\right) < \infty$.

Under these conditions, the following consistency result applies when both the sample size and approximation order increase:

Theorem 1. *Under Assumptions 3, the maximizer $(\hat{\theta}, \hat{h}_L)$ of (16) satisfies*

$$\hat{\theta} \xrightarrow{p} \theta_0,$$

$$\|\hat{h}_L - h_0\|_{r,\infty} \xrightarrow{p} 0,$$

when $T \to \infty$, $L \to \infty$, and $L^{D+1}/T \to 0$.

3.8 Filtered latent variable conditional method of moments estimation of H

The stochastic discount function H is identified as the unique eigenfunction that solves (2). Given the joint distribution of the endogenous variables $(\Delta c_{t+1}, \Delta d_{t+1}, s_{t+1})$ is identified, H could therefore be found via numerical integration and solution.

By conditioning on past observations, and applying the Law of Iterated Expectation, the Euler equation implies the conditional moment

$$0 = E\left(\frac{1}{\beta}H(s_t) - \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} E\left(H(s_{t+1}) \mid \mathcal{F}_{t+1}\right)R_{t+1} \mid \mathcal{F}_t\right), \tag{17}$$

using the fact that $(\Delta c_{t+1}, R_{t+1})$ are included in the information set $\mathcal{F}_{t+1} = (\mathcal{F}_t, \Delta y_{t+1}, m_{t+1}, p_{t+1})$ where $\mathcal{F}_t = \mathcal{F}_t^{y,m,p}$. Under completeness of the state density $f(s_t \mid \mathcal{F}_t)$, the conditional moment (17) identifies the unique eigenfunction H solving (13).

When the stochastic discount function is approximated by the polynomial $H_L(s_t) = e \cdot \bar{s}_t^L$, its projection on observables is a polynomial in the filtered moments of the states:

$$E\left(H_L(s_t) \mid \mathcal{F}_t\right) = e_L \cdot \tilde{\bar{s}}_{t|t}^L.$$

The conditional moment (17) can therefore be stated in terms of observables only as

$$0 = E\left(\frac{1}{\beta}e_L \cdot \tilde{\bar{s}}_{t|t}^L - \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} e_L \cdot \tilde{\bar{s}}_{t+1|t+1}^L R_{t+1} \mid \mathcal{F}_t\right).$$

We convert the latter into unconditional moments using the finite-dimensional instrument vector $(m_t, p_t, \tilde{\bar{s}}_{t|t}^L)$. The Markovian assumptions imply that the distribution of $(\Delta c_{t+1}, R_{t+1}, s_{t+1})$ only depends on $\mathcal{F}_t$ through (m_t, p_t) and the filtered state density $f(s_t \mid \mathcal{F}_t)$. When the latter density is summarized in terms of its L moments, the loss of efficiency from conditioning down on $(m_t, p_t, \tilde{\bar{s}}_{t|t}^L)$ vanishes when L increases. The resulting unconditional moments can be represented as an eigenproblem in the coefficient vector e_L:

$$\begin{aligned}
0 =& E\left((m_t, p_t, \tilde{\bar{s}}_{t|t}^L)^T \left(\frac{1}{\beta}\tilde{\bar{s}}_{t|t}^L - C_{t,t+1}^{-\gamma} R_{t+1} \tilde{\bar{s}}_{t+1|t+1}^L\right) \cdot e_L\right) \\
\Leftrightarrow & E\left((m_t, p_t)^T C_{t,t+1}^{-\gamma} R_{t+1} \tilde{\bar{s}}_{t+1|t+1}^{L\,\prime}\right) e_L = \frac{1}{\beta} E\left((m_t, p_t)^T \tilde{\bar{s}}_{t|t}^L\right) e_L, \\
& E\left(\tilde{\bar{s}}_{t|t}^L C_{t,t+1}^{-\gamma} R_{t+1} \tilde{\bar{s}}_{t+1|t+1}^{L\,\prime}\right) e_L = \frac{1}{\beta} E\left(\tilde{\bar{s}}_{t|t}^L \tilde{\bar{s}}_{t|t}^{L\,\prime}\right) e_L.
\end{aligned}$$

For the estimation we replace the unconditional moments by their empirical averages. Instead of direct GMM estimation of the parameters (β, γ, e_L), we profile the risk aversion parameter γ and solve for $(\beta(\gamma), e_L(\gamma))$ as the eigenvalue-eigenvector of the lower $L+1$ equations, recognizing the particular structure of the problem. The parameter γ is then set using the moments obtained by instrumenting with (m_t, p_t).

4 Empirical Results

4.1 Data

Aggregate output and consumption data are obtained from the U.S. Bureau of Economic Analysis. We consider both annual data from 1930 until 2016 and quarterly data from January 1947 until December 2016. Output is measured by U.S. real gross domestic product in 1992 chained dollars. Consumption is measured as the real expenditure on nondurables and service, excluding shoes and clothing, scaled to match the average total real consumption expenditure. Monthly observations of the Industrial Production Index are obtained from the Federal Reserve to construct initial proxies for economic uncertainty.

Stock market prices and dividends are based on the S&P 500 index obtained from the CRSP database. All prices and dividends are expressed in real terms using the price index for U.S. gross domestic product. Dividends per share are computed from the difference in value-weighted returns with (R^d_{t+1}) and without (R^x_{t+1}) dividends:

$$\frac{D_{t+1}}{P_t} = R^d_{t+1} - R^x_{t+1}.$$

Price-dividend ratios are then computed as

$$pd_{t+1} = \frac{P_{t+1}}{D_{t+1}} = \frac{P_{t+1}}{P_t}\frac{P_t}{D_{t+1}} = \frac{R^x_{t+1}}{R^d_{t+1} - R^x_{t+1}},$$

and dividend growth as

$$\frac{D_{t+1}}{D_t} = \frac{pd_{t+1}}{pd_t} R^x_{t+1}.$$

The initial aggregate dividend $D_1 = C_1$ is normalized to aggregate consumption. The constructed dividend series are equivalent to reinvesting intermediate cash payments in the underlying stock (Cochrane, 1992).

Let ip_t denote the log observed industrial production in month t, and let its increment be $\Delta ip_t = ip_t - ip_{t-1}$. The underlying volatility of output growth can be estimated using the annualized Realized Economic Variance (REV) measure

$$REV_t = \sum_{m=1}^{12} (\Delta ip_{t+1-m} - \overline{\Delta ip_t})^2,$$

with $\overline{\Delta ip_t}$ the rolling window annual mean. The realized stock market variance (RV) is similarly constructed from daily log returns R_{t+1} with demeaning at the quarterly frequency as[5]

$$RV_t = \sum_{d=1}^{252} (\Delta R_{t+1-d} - \overline{R}_t)^2,$$

with $\overline{R}_t$ the rolling window quarterly mean.

Table 1 contains descriptive statistics of the aggregate series. Output and consumption have on average grown at a comparable pace of 3-4%. Both series are fairly persistent, with output growth about twice as volatile. The S&P 500 market return earned on average over 6%, and is highly volatile but not persistent. The consumption-to-output ratio is more smooth and more persistent than the dividend-consumption ratio. The price-dividend ratio is relatively volatile yet highly persistent. Realized stock market variance is about ten times higher than that of realized productivity growth, and about five times as volatile. The two variation measures feature similar autocorrelation, but the REV is estimated using lower frequency (monthly) data which could lead to underestimate its autocorrelation.

Table 1: Sample mean, standard deviation and first-order autocorrelation (ACF(1)) of selected annual series from 1930-2016.

Real output growth ($\Delta \log Y$), real consumption growth ($\Delta \log C$), and the real market return ($\Delta \log P$) are in percentages. The ratios are in levels with the dividend-consumption ratio (D/C) normalized to start at unity. Realized Variance (RV) and Realized Economic Variance (REV) are multiplied by 100.

	$\Delta \log Y$	$\Delta \log C$	$\Delta \log P$	C/Y	D/C	P/D	RV	REV
Mean	3.45	3.09	6.54	0.69	0.51	32.97	3.01	0.33
Std Dev	4.57	1.99	17.77	0.07	0.16	16.06	4.00	0.92
ACF(1)	0.44	0.36	-0.13	0.85	0.58	0.90	0.50	0.42

4.2 Economic Uncertainty and Stock Market Volatility

Figure 1 shows the annual series of the measure of economic uncertainty (REV) and the measure of financial market volatility (RV). While both measures peaked during the Great Depression and the Great Recession, there are several well-known episodes during which they diverged. The first

[5]Cum-dividend returns are used to control for price changes due to anticipated payments. At the index level the difference compared to using ex-dividends returns is negligible.

postwar decades saw substantial economic uncertainty but historically calm financial markets. On the other hand, during the LTCM collapse of 1998 stock market volatility peaked while production growth remained largely unchanged. In recent years, stock market volatility has plummeted while economic uncertainty remained relatively high. Figure 2 compares REV with the dividend-

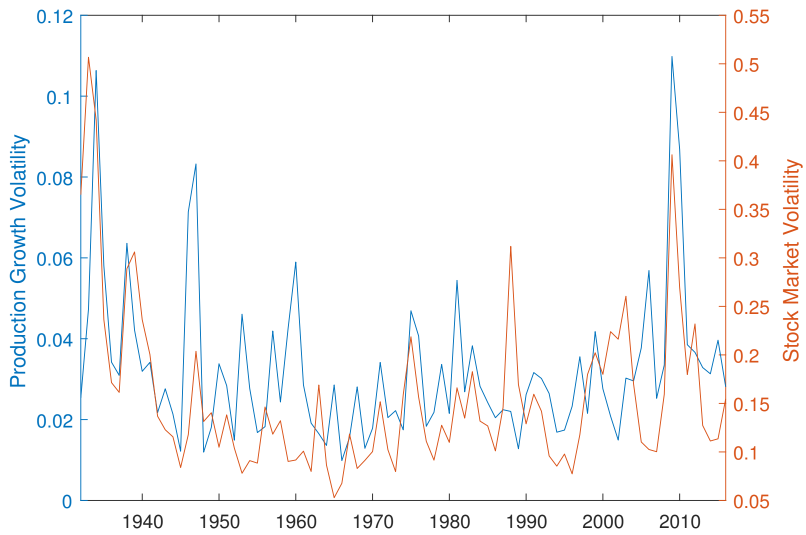


Figure 1: Realized Variance of Industrial Production growth and of S&P 500 returns from 1930-2016.

consumption ratio. It displays substantial comovement between the series, with the dividend-consumption ratio starting high during the uncertain 1930s, reaching historical lows during the post war recovery period, and steadily rising again during and after the Great Recession. This suggests output uncertainty affects not just the scale but also the level of consumption and dividend growth. In particular, dividends must grow faster than consumption when uncertainty increases in order to explain the behaviour of the dividend-consumption ratio. To understand the different channels via which economic and financial market uncertainty affect asset prices, Figures 3 and 4 plot the quarterly growth rates in REV and RV, respectively, against the growth rates of output, consumption, dividends, and the S&P 500 Index from 1947 to 2016. Figure 3 suggests a negative relation between uncertainty shocks and output and consumption growth, in line with the evidence in Bloom (2009) and Nakamura et al. (2017). The market return also goes down contemporaneously when uncertainty increases in line with the well known leverage effect. Dividend growth, on the other hand, does not go down and even is slightly convexly increasing in changes to uncertainty, in line with rebalancing of the dividend-consumption ratio. Figure 4 shows the impact of changes in financial market volatility on the same response variables.

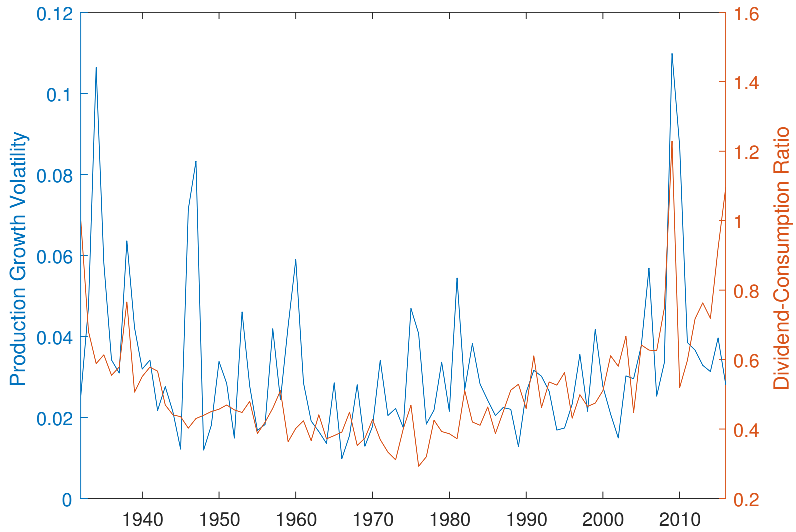


Figure 2: Realized Variance of Industrial Production growth versus Dividend-Consumption ratio ($DC_{1930} = 1$) from 1930-2016.

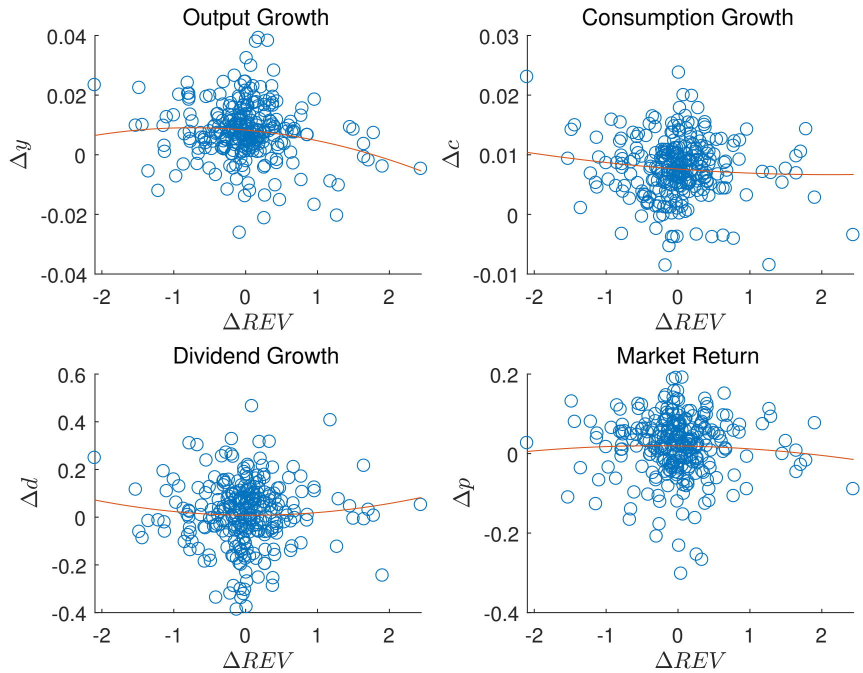


Figure 3: Quarterly changes in log Realized Economic Variance (REV) of Industrial Production growth versus log return on Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016. Fitted line corresponds to a quadratic fit.

While the direction of the responses is the same as for changes in economic uncertainty, the realized stock market variance correlates much stronger with dividend growth and the market return and much weaker with consumption and output growth. In particular, dividend growth is pronounced increasing and the market return convexly decreasing in changes in the realized variance. This provides further evidence against a simple linear relation between economic and financial uncertainty, which would not explain the different impact they have on fundamentals.

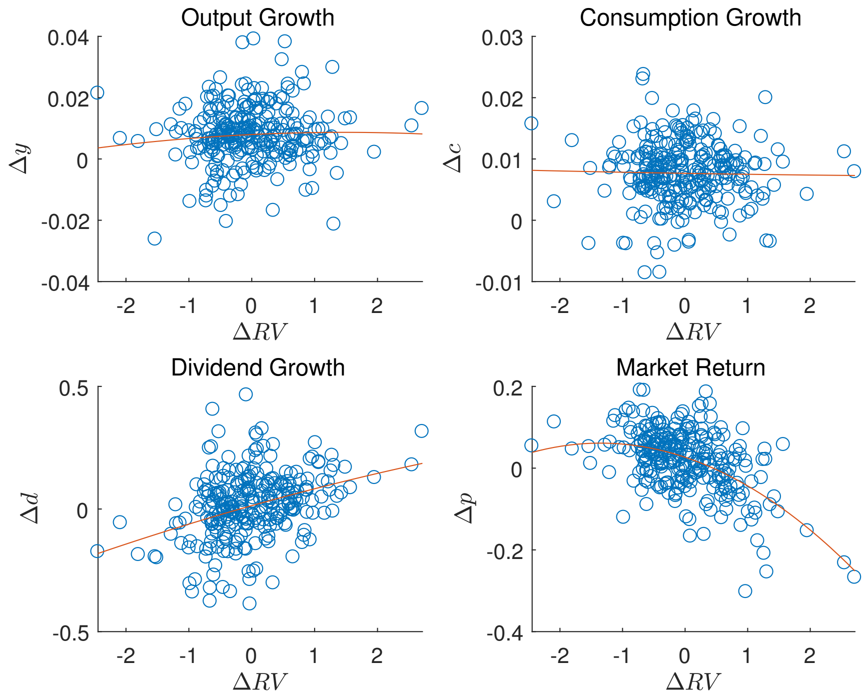


Figure 4: Quarterly changes in log Realized Variance (RV) versus log return Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016. Fitted line corresponds to a quadratic fit.

4.3 Cross-sectional heterogeneity

Table 2 shows the heterogeneous impact of increases in uncertainty on the dividend share of small and large firms based on the regression

$$\frac{D_{it}}{D_t} = \alpha_{0i} + \beta_i{}^T REV_t + z_{it}^d, \qquad E(z_{it}^d \mid REV_t) = 0. \tag{18}$$

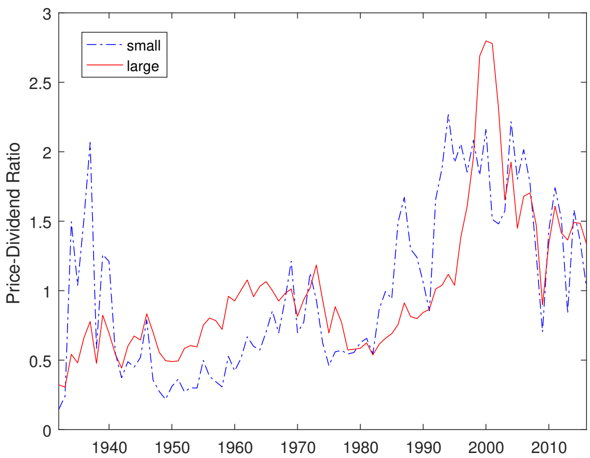
The estimated coefficients show that larger firms tend to increase their dividend share in uncertain

Table 2: Parameter estimates and standard errors of regression (18) using annual data from 1926-2016 using one-lag feasible generalized least squares.

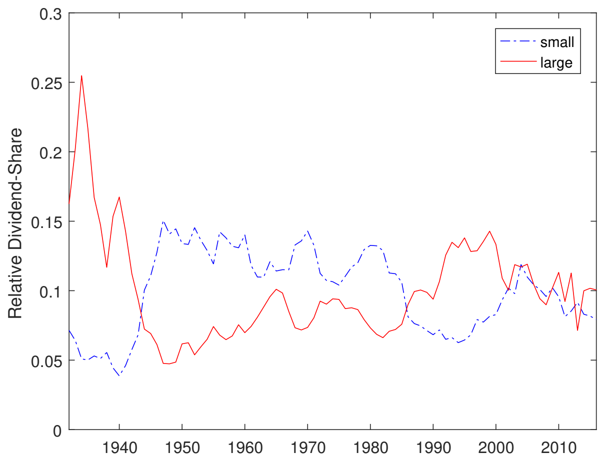
Size decile	1	2	3	4	5	6	7	8	9	10
$\hat{\beta}_i$	-2.67	-2.24	-1.35	-1.58	0.77	0.47	2.04	1.13	1.58	1.85
	(4.15)	(1.18)	(0.56)	(0.73)	(0.85)	(1.08)	(0.81)	(0.94)	(0.71)	(0.70)

times, whereas smaller firms tend to reduce them. A possible explanation is the real options theory of investment which suggests that large firms invest less when uncertainty is high and can afford to pay out cash to their shareholders. Small firms have less cash reserves to fall back on, which could force them to cut dividends in uncertain times. This heterogeneity in individual dividend-consumption ratios generates heterogeneous impact of the portfolio price-dividend ratios in response to uncertainty shocks.

Figure 5 shows the time series of the price-dividend ratios and the dividend shares of the stock portfolio corresponding to the smallest and largest size decile. The price-dividend ratios of the small firms have a strong negative correlation of -0.52 with its dividend share, suggesting an important role for the mean-reverting error to drive expected dividend growth. The price-dividend ratios of large firms are more smooth, reflecting their ability to smooth dividends. During several episodes, in particular in the 80s and 90s, the price-dividend ratios moved substantially apart even when their dividend shares did not. Small firms were more vulnerable to the increased uncertainty in the aftermath of the 1987 stock market crash, but recovered sharply when markets calmed. The peak in large-firm valuation ratios during the dot-com bubble is difficult to explain by uncertainty alone but can be attributed to high risk appetite.



(a) Price-Dividend Ratios.



(b) Relative Dividend Share.

Figure 5: Annualized Price-Dividend ratios and Dividend Share of the stock portfolios of the 10% smallest and largest U.S. exchange-traded firms by market capitalization over the period 1930-2016. Price-Dividend ratios are normalized by their sample average, Dividend shares are normalized to have mean 0.1.

4.4 Implementation

The closed-form transition density is approximated by setting the auxiliary densities for output growth Δy_t, its persistent components x_t, and its volatility v_t, as the symmetric Variance-Gamma density for the first two series, and the Gamma distribution for the latter. The Variance-Gamma is a mixture distribution that parsimoniously generates fat tails, while the Gamma distribution ensures that the volatility process is positive. Each distribution has two parameters which are used to match the conditional variance and kurtosis of the variables after centering around their conditional mean. The product of the univariate auxiliary densities creates the trivariate auxiliary density. The approximations using mixed Hermite polynomials to the fourth order are very close to densities obtained via Fourier inversion. All coefficients are computed symbolically, which only needs to be performed once prior to estimation. Their unconditional density is approximated as the transition density over a horizon of $\tau = 5$ years, starting from the unconditional means.

The simulated maximum likelihood estimation is implemented by sampling from the Bilateral Gamma and the Gamma distribution for the unbounded and positive variables, respectively. The Bilateral Gamma is a reparametrization of the symmetric Variance-Gamma distribution (Küchler and Tappe, 2008). It is the distribution of the sum of a positive and a negative Gamma random variables, which allows for easy simulation. The simulated moment-updating steps are performed using $N_s = 10,000$ draws in each period, starting from a fixed seed.

4.5 Estimates

Table 3 report the simulated maximum likelihood estimates of the transition density parameters θ_s. The mean reversion parameters (κ^x, κ^v) correspond to half-lives of the expected growth and volatility of around 7 and 2 years, respectively. Thus suggests the presence of a highly persistent growth component, while the volatility component is somewhat more transitory. Still, the persistent growth component is itself highly variable, as its standard deviation ω^x is much larger than its mean-reversion parameter κ^x. The negative correlation parameter $\rho = -0.28$ suggests a pronounced leverage effect between adverse economic shocks and economic uncertainty, akin to the financial leverage effects in stock returns. Moreover, the volatility-in-mean parameter λ^v indicates that volatility tends to reduce expected growth.

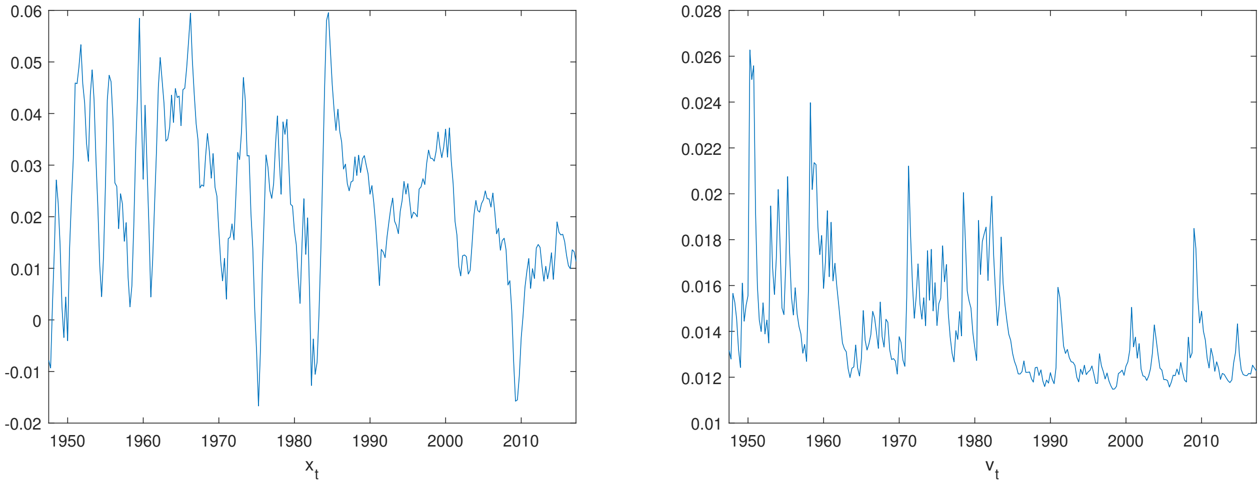
Figure 6 shows the conditional means of the state variables describing expected growth and volatility using the simulated moment-filtering algorithm. Periods of high volatility are clustered around episodes such as the post-war years, the 1980s energy crisis, and the 2008 financial crisis. The frequency and duration of high volatility periods has been steadily declining over the sampling

Table 3: Simulated maximum likelihood estimates of the transition density parameters θ_s for the long-run risk model with geometric stochastic volatility.

Estimates based on quarterly output growth observations from 1947 to 2016. Standard errors according to the Hessian obtained by numerical differentiation.

μ	κ^x	ω^x	κ^v	θ^v	ω^v	ρ	λ^v
0.026	0.088	0.88	0.25	0.021	0.70	-0.28	0.26
(0.013)	(0.026)	(0.08)	(0.009)	(0.002)	(0.011)	(0.025)	(0.020)

period, especially during the high growth 1990s.



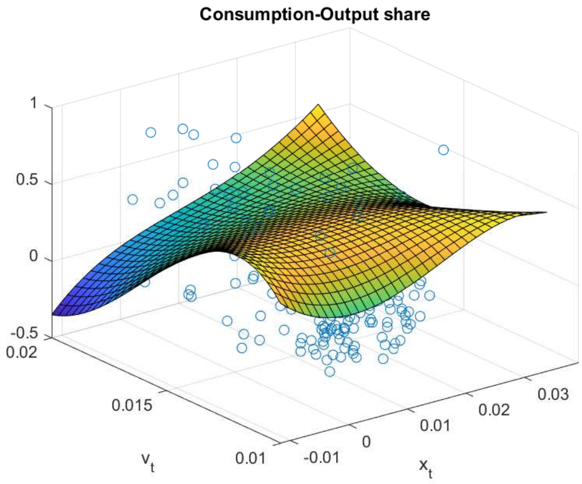
(a) Filtered expected growth.

(b) Filtered expected volatility.

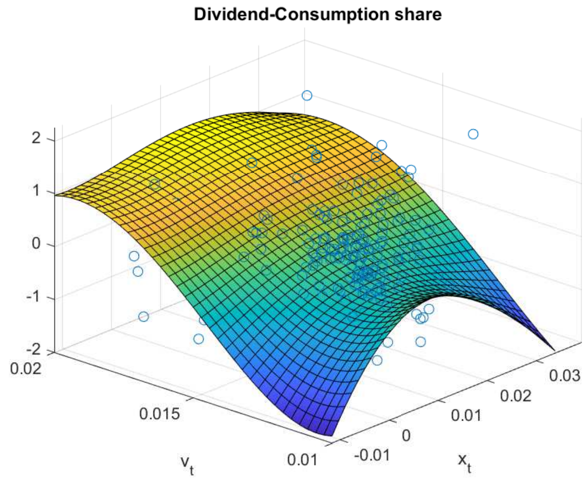
Figure 6: Filtered conditional mean of expected growth x_t and expected volatility v_t of output growth, using quarterly observations.

Figure 7 shows the estimated policy functions ψ^c and ψ^c of the log consumption-output and dividend-consumption ratio as a function of the state variables. The consumption-output share shows a negative interaction of low expected growth and high volatility, while absent one of these the function is relatively flat. The dividend-consumption ratio strongly increases with volatility, reflecting the decrease in consumption, and appears relatively robust to changes in expected growth.

Figure 8 shows in the left panel the estimated price-dividend ratio ϕ as a function of the state variables, after controlling for the consumption and dividend ratios. The price-dividend function appears to increase monotonically in expected growth, and to decrease monotonically in expected volatility. Both effects appear to interact such that simultaneously high expected growth and low expected volatility further lifts the price-dividend ratio to levels exceeding two standard deviations above the its mean. The right panel shows the estimated stochastic discount function



(a) Consumption-output response function.

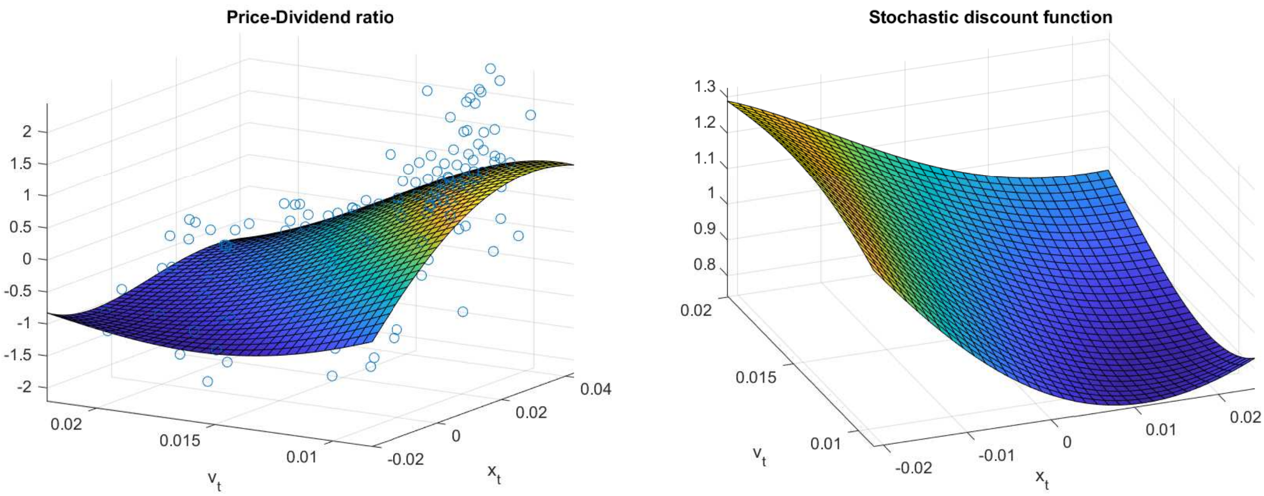


(b) Dividends-consumption response function.

Figure 7: Fitted policy functions for log consumption and dividend shares of output as a function of expected growth x_t and expected volatility v_t, using quarterly observations and a $L = 4$ order expansion. Vertical axis shows standard deviations from the mean.

that minimizes the GMM criterion in terms of the filtered states. The stochastic discount function tends to move inversely to the price-dividend ratio. In particular, it reaches its highest levels when expected growth is low and expected volatility is high. This variation in the stochastic discount function indicates that state-dependence in the consumption and dividend levels cannot fully explain state-dependence in the price-dividend ratio. Instead, our estimates suggests rationalizing the latter by state-dependent preferences through the discount factor. Economically, the evidence suggests that the marginal investor has relatively high marginal utility for payoffs in adverse times as defined by low growth and high volatility.

Figure 9 shows the GMM-criterion as a function of the discount and risk aversion parameters (β, γ), after profiling the approximation coefficients $e_L(\beta, \gamma)$ of the stochastic discount function. The minimizing values of the parameters equal $(\hat{\beta}, \hat{\gamma}) = (0.97, 1.09)$. The latter suggests that the common assumption of logarithmic utility over consumption is not unreasonable once additional variation in the discount factor is allowed. However, identification of the risk aversion parameter is relatively weak, as values up to $\gamma = 3$ yield qualitatively similar stochastic discount functions.



(a) Log price-dividend ratio function.

(b) Stochastic discount function.

Figure 8: Fitted price-dividend ratio and stochastic discount factor as a function of expected growth x_t and expected volatility v_t, using quarterly observations and a $L = 4$ order expansion. Vertical axis shows standard deviations from the mean.

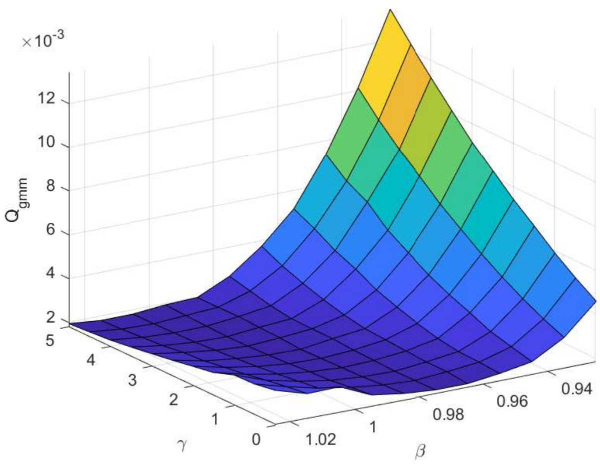


Figure 9: GMM-criterion as a function of the discount and risk aversion parameters (β, γ), where the stochastic discount function approximation coefficients are profiled as $e_L(\beta, \gamma)$.

5 Conclusion

This paper develops a class of nonlinear Markovian asset pricing models in which the dynamics of consumption and dividend shares of output are described via general policy functions of latent state variables describing persistent components in the aggregate growth distribution. Tractable closed-form expressions for expected returns and financial volatility are obtained under polynomial approximations of the policy functions. The paper establishes the consistency of a sieve maximum likelihood estimator for the general case where the measurement distribution of observables is unknown. Moreover, we study the identification and estimation of a semiparametric specification of the stochastic discount factor by formulating the Euler equation as an eigenfunction problem. The expected dividend-consumption ratio is found to be increasing of economic volatility, but not enough to explain the decline in the expected price-dividend ratio when volatility increases. Instead, the latter could be explained by investors with moderate risk aversion but with higher marginal utility for payoffs in times of low expected growth and high volatility. The steeply declining price-dividend ratios for moderate levels of economic uncertainty can explain episodes of large stock market volatility that occurred during periods of moderate economic uncertainty.

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262.

Ait-Sahalia, Y. and Kimmel, R. L. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. *Journal of Financial Economics*, 98(1):113–144.

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53(2):499–547.

Albuquerque, R., Eichenbaum, M., Luo, V. X., and Rebelo, S. (2016). Valuation risk and asset pricing. *The Journal of Finance*, 71(6):2861–2904.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Bondarenko, O., Todorov, V., and Tauchen, G. (2015). The fine structure of equity-index option dynamics. *Journal of Econometrics*, 187(2):532–546.

Arellano, M., Blundell, R., and Bonhomme, S. (2017). Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica*, 85(3):693–734.

Arellano, M. and Bonhomme, S. (2011). Nonlinear panel data analysis.

Bakshi, G. and Madan, D. (2000). Spanning and derivative-security valuation. *Journal of Financial Economics*, 55(2):205–238.

Bansal, R., Dittmar, R., and Kiku, D. (2007). Cointegration and consumption risks in asset returns. *The Review of Financial Studies*, 22(3):1343–1375.

Bansal, R. and Viswanathan, S. (1993). No arbitrage and arbitrage pricing: A new approach. *The Journal of Finance*, 48(4):1231–1262.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3):623–685.

Bollerslev, T., Tauchen, G., and Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies*, 22(11):4463–4492.

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(2):205–251.

Carr, P. and Wu, L. (2008). Variance risk premiums. *The Review of Financial Studies*, 22(3):1311–1341.

Chapman, D. A. (1997). Approximating the asset pricing kernel. *Journal of Finance*, pages 1383–1410.

Chen, H. and Joslin, S. (2012). Generalized transform analysis of affine processes and applications in finance. *The Review of Financial Studies*, 25(7):2225–2256.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of econometrics*, 6:5549–5632.

Chen, X. and Ludvigson, S. C. (2009). Land of addicts? an empirical investigation of habit-based asset pricing models. *Journal of Applied Econometrics*, 24(7):1057–1093.

Cheng, P. and Scaillet, O. (2007). Linear-quadratic jump-diffusion modeling. *Mathematical Finance*, 17(4):575–598.

Christensen, T. M. (2017). Nonparametric stochastic discount factor decomposition. *Econometrica.*

Cochrane, J. H. (1992). Explaining the variance of price–dividend ratios. *The Review of Financial Studies*, 5(2):243–280.

Constantinides, G. M. and Ghosh, A. (2011). Asset pricing tests with long-run risks in consumption growth. *The Review of Asset Pricing Studies*, 1(1):96–136.

Cuchiero, C., Keller-Ressel, M., and Teichmann, J. (2012). Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740.

Dew-Becker, I., Giglio, S., Le, A., and Rodriguez, M. (2017). The price of variance risk. *Journal of Financial Economics*, 123(2):225–250.

Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3.

Drechsler, I. and Yaron, A. (2010). What's vol got to do with it. *The Review of Financial Studies*, 24(1):1–45.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Eraker, B. and Shaliastovich, I. (2008). An equilibrium guide to designing affine pricing models. *Mathematical Finance*, 18(4):519–543.

Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., Srisuma, S., et al. (2015). Nonparametric euler equation identification and estimation. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4):1059–1087.

Filipović, D., Larsson, M., and Trolle, A. B. (2016). On the relation between linearity-generating processes and linear-rational models.

Filipović, D., Mayerhofer, E., and Schneider, P. (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111.

Gabaix, X. (2007). Linearity-generating processes: A modelling tool yielding closed forms for asset prices. Technical report, National Bureau of Economic Research.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly journal of economics*, 127(2):645–700.

Gagliardini, P. and Gourieroux, C. (2014). Efficiency in large dynamic panel models with common factors. *Econometric Theory*, 30(5):961–1020.

Gagliardini, P. and Gouriéroux, C. (2019). Identification by laplace transforms in nonlinear time series and panel models with unobserved stochastic dynamic effects. *Journal of Econometrics*, 208(2):613–637.

Gallant, A. R., Giacomini, R., and Ragusa, G. (2017). Bayesian estimation of state space models using moment conditions. *Journal of Econometrics*.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, pages 363–390.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica: Journal of the Econometric Society*, pages 871–907.

Gallant, A. R. and Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica: Journal of the Econometric Society*, pages 1091–1120.

Garcia, R., Lewis, M.-A., Pastorello, S., and Renault, É. (2011). Estimation of objective and risk-neutral distributions based on moments of integrated volatility. *Journal of Econometrics*, 160(1):22–32.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Hansen, L. P. and Renault, E. (2010). Pricing kernels. *Encyclopedia of Quantitative Finance*.

Hansen, L. P. and Scheinkman, J. A. (2009). Long-term risk: An operator approach. *Econometrica*, 77(1):177–234.

Hausman, J. A., Newey, W. K., Ichimura, H., and Powell, J. L. (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics*, 50(3):273–295.

Heston, S. L. and Rossi, A. G. (2016). A spanning series approach to options. *The Review of Asset Pricing Studies*, 7(1):2–42.

Hu, Y. and Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics*, 171(1):32–44.

Jagannathan, R. and Marakani, S. (2015). Price-dividend ratio factor proxies for long-run risks. *The Review of Asset Pricing Studies*, 5(1):1–47.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *The American Economic Review*, 105(3):1177–1216.

Küchler, U. and Tappe, S. (2008). Bilateral gamma distributions and processes in financial mathematics. *Stochastic Processes and their Applications*, 118(2):261 – 283.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849.

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

Nakamura, E., Sergeyev, D., and Steinsson, J. (2017). Growth-rate and uncertainty shocks in consumption: Cross-country evidence. *American Economic Journal: Macroeconomics*, 9(1):1–39.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, pages 1565–1578.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of financial economics*, 63(1):3–50.

Piazzesi, M. (2010). Affine term structure models. *Handbook of financial econometrics*, 1:691–766.

Ross, S. (2015). The recovery theorem. *The Journal of Finance*, 70(2):615–648.

Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica*, 82(1):345–385.

Schmitt-Grohé, S. and Uribe, M. (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of economic dynamics and control*, 28(4):755–775.

Schorfheide, F., Song, D., and Yaron, A. (2018). Identifying long-run risks: A bayesian mixed-frequency approach. *Econometrica*, 86(2):617–654.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400.

Zhou, H. (2003). Itô conditional moment generator and the estimation of short-rate processes. *Journal of Financial Econometrics*, 1(2):250–271.

A Appendix

A.1 Proofs.

Change-of-measure formula for price-dividend ratio.
Define the time-T forward measure $\mathbb{Q}^T$ by

$$\left.\frac{d\mathbb{Q}^T}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \frac{e^{-\log Y_T}}{E_t^{\mathbb{P}}(e^{-\log Y_T})}. \tag{19}$$

The affine property of (y_t, s_t) implies that

$$\tilde{p}^T(t) = E_t^{\mathbb{P}}(e^{-\gamma(y_T - y_t)}) = e^{\alpha^T(t;\gamma)+\beta^T(t;\gamma)\cdot s_t}, \tag{20}$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ solve the differential equations

$$\begin{aligned}
\dot{\beta}(t) &= -K_1(t)^T\beta(t) - \frac{1}{2}\beta(t)^T H_1(t)\beta(t) \\
\dot{\alpha}(t) &= -K_0(t)^T\beta(t) - \frac{1}{2}\beta(t)^T H_0(t)\beta(t)
\end{aligned}$$

with boundary conditions $\alpha^T(T) = \beta^T(T) = 0$. Under the forward measure the drift of the state variables changes to (Duffie et al., 2000, Prop. 5)

$$K_0^{\mathbb{Q}^T}(t) = K_0 + H_0\beta^T(t), \qquad K_1^{\mathbb{Q}^T}(t) = K_1 + H_1\beta^T(t). \tag{21}$$

Example of the coefficient matrix for computing conditional moments in the baseline model. Table 4 lists the coefficient matrix used for computing conditional moments in the baseline model with stochastic volatility V_t of output growth and time-varying risk aversion Q_t.

Proof of Theorem 1. The proof is based on Lemma A1 in Newey and Powell (2003). Let $Q_T(\theta) = \ell_T(\theta)$ and $Q(\theta) = E(l_t(\theta))$. This requires that (i) there is unique θ_0 that minimizes $Q_T(\theta)$ on Θ, (ii) Θ_T are compact subsets of Θ such that for any $\theta \in \Theta$ there exists a $\tilde{\theta}_T \in \Theta_T$ such that $\tilde{\theta}_T \xrightarrow{p} \theta$, and (iii) $Q_T(\theta)$ and $Q(\theta)$ are continuous, $Q_T(\theta)$ is compact, and $\max_{\theta\in\Theta} |Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0$.

The identification condition (i) follows from subsection 3.2. The compact subset condition in (ii) holds by construction of $\mathcal{H}_T$ and $\mathcal{H}$. Moreover for any $\theta \in \Theta$ we can find a series approximator $\theta_T \in \Theta_T$ that satisfies $\|\theta_T - \theta\| \to 0$ as by construction the approximating spaces $\mathcal{H}_T$ are dense in $\mathcal{H}$.

For (iii), continuity of $Q_T(\theta)$ follows from continuity of the policy and pricing functions and the

Table 4: Coefficient matrix A_2 mapping the Itô generator of the second-degree moments of $(d\log Y_t, d\mu_t, dv_t)$ into itself.

0	0	0	0	0	0	0	0	0	0
μ	0	1	λ^v	0	0	0	0	0	0
0	0	$-\kappa^\mu$	0	0	0	0	0	0	0
$\kappa^v\theta^v$	0	0	$-\kappa^v$	0	0	0	0	0	0
0	2μ	0	0	0	2	$2\lambda^v$	0	0	1
0	0	μ	0	0	$-\kappa^\mu$	0	1	λ^v	0
0	$\kappa^v\theta^v$	0	μ	0	0	$-\kappa^v$	0	1	$\lambda^v + \omega^v\rho$
0	0	0	0	0	0	0	$-2\kappa^\mu$	0	ω_μ^2
0	0	$\kappa^v\theta^v$	0	0	0	0	0	$- \kappa^v - \kappa^\mu$	0
0	0	0	$2\kappa^v\theta^v$	0	0	0	0	0	$\omega_v^2 - 2\kappa^v$

transition density. The remaining conditions of continuity of $Q(\theta)$ and uniform convergence follow from Lemma A2 in Newey and Powell (2003). This requires pointwise convergence $Q_T(\theta) - Q(\theta) \xrightarrow{p} 0$ as well as the stochastic equicontinuity condition that there is a $v > 0$ and $B_n = O_p(1)$ such that for all $\theta, \tilde{\theta} \in \Theta$, $\|Q_T(\theta) - Q_T(\tilde{\theta})\| \leq B_n \|\theta - \tilde{\theta}\|^v$. Pointwise convergence follows from the weak law of large numbers due to the stationarity and mixing conditions. Stochastic equicontinuity follows from the Lipschitz condition on the transition density. $\square$