Semiparametric Estimation of Latent Variable Asset Pricing Models

Jeroen Dalderop*

April 12, 2020

Abstract

This paper studies semiparametric identification and estimation of the stochastic discount factor in consumption-based asset pricing models with latent state variables. The measurement equations for consumption and dividend shares are specified non-parametrically 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. 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 dividend shares of output as functions of Markovian state variables describing the conditional distribution of 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 now featuring unobserved state variables.

The robustness of this approach depends on correct specification of the arguments

[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).

of the stochastic discount factor and the distribution of the state variables. To allow for flexible state-dependent preferences of a representative agent, the stochastic discount factor is assumed to be the multiple of the power marginal utility of consumption and an unspecified function of the state variables. This semiparametric formulation nests several structural models, such as those based on recursive preferences, in which the multiplicative component is a restricted parametrized function of the state variables. 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). By imposing some parametric structure on the state variable dynamics, we can increasing their dimensionality, and extrapolate their distribution across maturities of interest. In particular, when state variables are specified as continuous time affine processes (Duffie et al., 2000), the resulting framework is a tractable generalization of the class of affine equilibrium asset pricing models (Eraker and Shaliastovich, 2008) towards arbitrary nonlinear 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 to 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.

For estimation, the framework is highly tractable when the unknown functions are approximated by orthogonal polynomials as in Chen (2007). In particular, conditional moments of growth rates of consumption, dividends, and asset returns can be expressed in closed-form as polynomials of the affine state variables. We study maximum likelihood estimation of the parameters of the state variable dynamics and the coefficients of the approximated 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. This motivates using a variant of the EM-algorithm which reduces the maximization step to estimating the polynomial coefficients by regression. Finally, we discuss the estimation of the stochastic discount function in a second stage method-of-moments step, which avoids numerical solution of the nonlinear pricing function and does not require specifying the dynamic distribution of unexplained price variation. This is achieved by integrating out the latent variables of the conditional Euler equation, which leads to a feasible eigenvector problem that is linear in the approximation coefficients of the stochastic discount function and the filtered moments of the state variables.

The empirical application illustrates the methodology by analyzing the impact of time-variation in the conditional mean and volatility of aggregate output growth on consumption shares and equity valuation ratios. We estimate the model using data on postwar U.S. macroeconomic variables, and prices and dividends on the S&P 500 stock market index. We also consider high-frequency measures of return volatility as well as growth volatility proxies based on the monthly Industrial Production Index, to have a penalization effect on the filtered economic volatility state similar to that for financial volatility in Andersen et al. (2015). We find periods of high 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. Both the consumption-output and dividend-consumption share show a negative interaction of low expected growth and high growth volatility. Vice versa, the combined effect of high expected growth and low growth volatility lifts the expected price-dividend ratio above two standard deviations from its mean. This state-dependence in the price-dividend ratio is only partially explained by that of the consumption and dividend shares, as evidenced by the stochastic discount function reaching its highest levels when expected growth is low and expected volatility is high. This suggests state-dependent preferences have to be accounted for when using asset prices to predict future growth and uncertainty.

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

Nonlinear dynamic latent variable models have been primarily studied for microeconomic panel data, 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. While linear common factor models are widely applied to macroeconomic and financial data, their nonlinear variants are less well studied. Some exceptions are Gagliardini and Gourieroux (2014) and Andersen et al. (2019), who extract nonlinear common factors from a large number of cross-sectional units with unpredictable and independent errors. In this paper, the latent factor dynamics are identified through their relation to a low-dimensional growth series observed over many time periods, while allowing for serially correlated measurement error in the asset prices. In semiparametric conditional moment-based models, Schennach (2014) and Gallant et al. (2017) integrate out the latent variables 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, such as in 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. Cross-sectional dispersion measures based on firm level data (Bloom, 2009) require correctly specifying 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.

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 application. 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 full information filtration satisfying standard regularity conditions. The superscript notation $\mathcal{F}_t^x$ refers to the history $(x_t, x_{t-1}, \ldots)$ 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 instantaneous utility function that decomposes as

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

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 $\phi(\cdot)$ 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 $\zeta_t = \beta^t C_t^{-\gamma} \phi(s_t)$ 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

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

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 log 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 a 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 ia modeled analogous to the consumption share by the semiparametric additive specification

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

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 measurement 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

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

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

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

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, ϕ) 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

$$\frac{P_t}{D_t} = E\left(\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{\phi(s_{t+1})}{\phi(s_t)} \frac{D_{t+1}}{D_t} \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \mid \mathcal{F}_t\right) \tag{3}$$

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

$$\pi(s_t, m_t) = E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{\phi(s_{t+1})}{\phi(s_t)} \frac{D_{t+1}}{D_t}\left(1 + \pi(s_{t+1}, m_{t+1})\right) \mid s_t, m_t\right). \quad (4)$$

In turn, this implies the eigenproblem characterization of the stochastic discount function

$$\frac{1}{\beta}\phi(s_t) = E\left(\phi(s_{t+1})\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{D_{t+1}}{D_t} \frac{1 + \pi(s_{t+1}, m_{t+1})}{\pi(s_t, m_t)} \mid s_t\right). \quad (5)$$

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. In particular, it does not depend on the distribution or dynamic properties of the unexplained variation in the price-dividend ratio.

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 $\mu(s_t)$ and covariance matrix $\mathcal{C}(s_t)$ 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[3]

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

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

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

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

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

[3]Modeling volatility as an affine state variable, instead of the variance as in the Heston (1993) model, provides computational advantages by reducing the order of approximating functions of the state variables.

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

$$dm_t = (R - I)m_t dt + d\psi(s_t) + \sigma^m dW_t^m. \tag{6}$$

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 solve a first-order linear matrix differential equation arising from the polynomial-preserving property of the infinitesimal generator (Zhou, 2003; Cuchiero et al., 2012). Its solution shows the conditional moments of the state variable at horizon τ are polynomials in the current state variable:

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

where $\bar{\mathcal{S}}^L$ is a column vector that stacks monomials up to degree L in lexicographic order, and the coefficients of the matrix A_L are functions of the transition parameters which can be computed symbolically using standard software. The Appendix derives (7) and shows the matrix A_L for the baseline model with order $L = 2$ in Table 3.

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 an inner product. 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

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

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 polynomial 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

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

With the pricing kernel process written as $\zeta_t = e^{-\delta t - \gamma c_t + \tilde{\phi}(s_t)}$ with $\tilde{\phi}(s_t) = \log\phi(s_t)$, 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 $\lambda^\zeta(s_t) = (\gamma, \gamma\psi_s^c(s_t) + \tilde{\phi}_s(s_t))$ 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 $\tilde{\phi}_s(s_t)$.

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 $\lambda^x(s_t) = \gamma\psi_x^c(s_t) + \tilde{\phi}_x(s_t)$ and $\lambda^v(s_t) = \gamma\psi_v^c(s_t) + \tilde{\phi}_v(s_t)$, 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 ϕ 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 generate rich dynamics of growth and volatility risk premia.

2.7 Asset Return Volatility

The framework allows investigating the link between the volatility of asset returns and the volatility of latent state variables describing macroeconomic fundamentals. In particular, the model-based price volatility can be compared with high-frequency measures of realized volatility.

Following our affine-polynomial approximation, suppose the log price-dividend ratio takes the form $\log \frac{P_t}{D_t} = \pi_L^p(s_t) + \alpha \cdot m_t$. Variation in the log return can then be decomposed into variation in the price-dividend ratio, the consumption and dividend shares m_t, and output growth:

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

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

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

where $\lambda^p(s_t) = (\lambda, \pi_s^p(s_t) + \alpha^* \cdot \psi_s^m(s_t))$ are the return's loadings on output growth and the state variables. The quadratic variation of the log return follows by Itô's Lemma as

$$d\langle \log P\rangle_t = \alpha^* \cdot \sigma_m \sigma_m' \alpha^* + \lambda_s^p(s_t)' \mathcal{C}(s_t) \lambda_s^p(s_t). \tag{10}$$

When the pricing and policy functions are polynomials, so are the gradients $\lambda_s^p(s_t)$, and (10) yields 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 π, the preference parameters (β, γ) and stochastic discount function ϕ, and the parameters of the latent variable distribution θ_s. The functional parameters are combined into $h = (\psi, \pi, \phi)$, 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 (6) and (9). 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 interaction 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. Multi-period 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 does not require 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). Lagged state-dependence of the prices is not ruled out to allow for Markovian serially correlated pricing error components. 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

$$
\begin{aligned}
m_t &= Rm_{t-1} + \psi(s_t) + \varepsilon_t \\
p_t &= \tilde{\pi}(s_t, m_t) + \eta_t, \qquad\qquad \eta_t = \rho_p \eta_{t-1} + \omega_t \\
\mathcal{S}_{t+1} &\sim f(\mathcal{S}_{t+1} \mid s_t)
\end{aligned}
\tag{11}
$$

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 error η_t allows for the presence of persistent deviations from the expected price 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 ϕ 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 $(\Delta y_{t+1},\ldots,\Delta y_T)$. The no feedback condition implies the conditional independence $m_t \mid s_t, \mathcal{F}^y_{t+1:T} \sim m_t \mid s_t$, so that

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

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

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

where $f(s_{t+1} \mid m_t, \mathcal{F}^y_{t+1:T}) = \frac{f(m_t, s_{t+1}|\mathcal{F}^y_{t+1:T})}{f(m_t|\mathcal{F}^y_{t+1:T})} = \frac{\int f(m_t|s_t) f(s_{t+1}|s_t,\mathcal{F}^y_{t+1:T}) f(s_t|\mathcal{F}^y_{t+1:T}) ds_t}{f(m_t|\mathcal{F}^y_{t+1:T})}$ 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 case of a polynomial measurement equation, the completeness assumptions reduce to rank conditions involving the smoothed moments of the state variables. In particular, let $s^l_{t|T} = E(s^l_t \mid \mathcal{F}^y_T)$ be the smoothed conditional l-th moment of the state s_t given the full sample of growth realizations, and let the vector $\bar{s}^L_{t|T}$ stack the smoothed moments up to the order L. Consider the univariate specification $m_t = \rho m_{t-1} + c'_L \bar{s}^L_t + \varepsilon_t$, where ε_t is independent of $\mathcal{F}^y_T$. The latter implies $\varepsilon_t \perp s^l_{t|T}$ for any $l = 0,\ldots,L$, which yields the linear system of $L+1$ equations

$$E\left(\bar{s}^{L\,\prime}_{t|T} m_t\right) = \rho E\left(\bar{s}^{L\,\prime}_{t|T} m_{t-1}\right) + c'_L E\left(\bar{s}^L_t \bar{s}^{L\,\prime}_{t|T}\right). \tag{12}$$

The prediction error $\bar{e}^L_t = \bar{s}^L_{t|T} - \bar{s}^L_t$ by construction satisfies $e^k_t \perp s^l_{t|T} \mid \mathcal{F}^y_T$ for each $(k,l) \in \{0,\ldots,L\}$, so that $E\left(\bar{s}^L_t \bar{s}^{L\,\prime}_{t|T}\right) = E\left(\bar{s}^L_{t|T} \bar{s}^{L\,\prime}_{t|T}\right)$. In the special case $\rho = 0$, the coefficient vector c_L could therefore be identified from the regression of m_t on $\bar{s}^L_{t|T}$. More generally, the one-period lagged moments yield the additional $L+1$ linear equations

$$E\left(\bar{s}_{t|T}^{L}{}'m_{t-1}\right) = \rho E\left(\bar{s}_{t|T}^{L}{}'m_{t-2}\right) + c_L' E\left(\bar{s}_{t-1}^{L}\bar{s}_{t|T}^{L}{}'\right), \tag{13}$$

where $E\left(\bar{s}_{t-1}^{L}\bar{s}_{t|T}^{L}{}'\right) = E\left(E\left(\bar{s}_{t-1}^{L} \mid \mathcal{F}_T^y\right)\bar{s}_{t|T}^{L}{}'\right) = E\left(\bar{s}_{t-1|T}^{L}\bar{s}_{t|T}^{L}{}'\right)$ by iterated expectations. Together, the $L+2$ parameters in (ρ, c_L) are thus identified as the solution to the $2(L+1)$ linear equations given by (12) and (13), provided the outer product matrix $E\left(\bar{s}_{t|T}^{L}\bar{s}_{t|T}^{L}{}'\right)$ 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 $\pi(s_t, m_t)$ could be nonparametrically identified from the equation

$$E\left(\frac{P_t}{D_t} \mid m_t, \mathcal{F}_{t+1:T}^y\right) = \int \pi(s_t, m_t) f(s_t \mid m_t, \mathcal{F}_{t+1:T}^y) ds_t, \tag{14}$$

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 $\pi(s_t, m_t) = \pi^p(s_t)e^{\alpha' m_t}$ that arises from the log-linear lag dependence of m_t.

The conditionally Gaussian model combined with the polynomial approximation $\tilde{\pi}_L^p = b_L'\bar{s}_t^L$ gives rise to a two-stage linear regression approach. In particular, let $\tilde{s}_{t|T}^l = E(s_t^l \mid \mathcal{F}_T^{y,m})$ 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 (14) can be written as the regression equation for the log price-dividend ratio p_t

$$E\left(p_t \mid \mathcal{F}_T^{y,m}\right) = b_L'\tilde{\bar{s}}_{t|T}^L + \alpha' m_t.$$

The variance σ_η^2 of the pricing error η_t can be identified from the squared deviations as

$$E\left((p_t - \alpha' m_t)^2\right) = E\left((b_L' \bar{s}_t^L + \eta_t)^2\right) \tag{15}$$
$$= b_L' E\left(\bar{s}_t^L \bar{s}_t^{L'}\right) b_L + \sigma_\eta^2,$$

using $\eta_t \perp s_t$, where the outer-product matrix $E\left(\bar{s}_t^L \bar{s}_t^{L'}\right)$ is identified from the state transition parameters. The expected price-dividend ratio is then computed as $\pi(s_t, m_t) = e^{b_L' \bar{s}_t^L + \alpha' m_t + \frac{1}{2}\sigma_\eta^2}$.

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

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

where the coefficient matrix A_L^s describing the state moment dynamics (21) and the unconditional moment matrix $E\left(\bar{s}_t^L \bar{s}_t^{L'}\right)$ are known given the transition parameter θ_s.

The identification strategy of the pricing function does not restrict the serial dependence in the pricing error η_t. 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 consistently estimated without any assumption on the pricing function and errors. The moment-based estimators are therefore a natural choice as initial values for computing the maximum likelihood estimator.

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 as long as there is unique eigenvalue-eigenfunction pair $(\phi, \frac{1}{\beta})$ that solves (5). 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

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

where

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

Escanciano et al. (2015) and Christensen (2017) show that the following assumption is sufficient for the uniqueness (up to scale) of a positive eigenfunction ϕ 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 $\pi(s_t, m_t)$ 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 ϕ is only identified up to scale, the estimation uses the normalization $E\left(\phi(s_t)^2\right) = 1$.

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 ϕ under the completeness of an expected return-weighted density of the state variables.

3.3 Likelihood formulation

The joint log-likelihood function of the combined observations can be decomposed as

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

The parameter structure 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 π 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

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

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

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

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

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

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

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

The latter limited-information likelihood does not condition on past prices $\mathcal{F}_t^p$, which is not needed to identify the pricing and stochastic discount function. Further efficiency can be achieved by conditioning on past prices $\mathcal{F}_t^p$ when the conditional distribution of the Markovian pricing error is correctly specified. The likelihood will then integrate over two periods of the state variables.

In nonlinear dynamic models it is generally not possible to integrate out the latent variables analytically from the likelihood components, unlike in linear models with Gaussian errors where the updating density $f\left(s_t \mid \mathcal{F}_t^y; \theta_s\right)$ can be computed recursively by the Kalman filter. 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 could 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 approximate 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, when the measurement equations are approximated by polynomials, we develop a variant of the EM algorithm whose maximization step reduces to a linear regression on the filtered moments of the states.

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 we 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 $w(\cdot)$, we 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

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

weighted $\mathcal{L}^2_w$ norm satisfy

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

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

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 $\bar{\mathcal{S}}^L_{t+\tau|t} = E(\bar{\mathcal{S}}^L_{t+\tau} \mid \mathcal{F}_t)$, 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 likelihood 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:

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

where $(s_{i,t+1})_{i=1}^{N^s}$ are simulated data from the auxiliary densities, and $(w_{i,t+1})_{i=1}^{N^s}$ are the normalized sampling weights

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

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

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

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

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

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.

The smoothing step can be performed using a similar sampling approximation based on the backward recursive relation

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

starting from the last period's updated moments $\bar{s}^L_{T|T}$. In particular, the smoothed sampling weights can be recursively computed as proportional to

$$w^\dagger_{i,t} \propto w^*_{i,t} \frac{1}{N_s} \sum_{i=1}^{N^s} f(s_{i,t+1} \mid s_{i,t}) \frac{w^\dagger_{i,t+1}}{w^*_{i,t+1}}$$

and normalizing their sum to 1. Like the particle filter, the smoothing step requires storing each periods simulated 'particles' and weights. Therefore, for large-scale parameter optimization, we consider an iterative algorithm that avoids the smoothing step.

3.6 Implementation of simulated likelihood evaluation

The closed-form transition density is approximated by setting the auxiliary densities for output growth Δy_t, its persistent mean component 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 prior to estimation. The unconditional density is approximated as the transition density over a horizon of $\tau = 5$ years, starting from the unconditional means.

The simulated maximum likelihood is computed 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 variable, 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.

3.7 Iterative maximization 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 the method-of-moments involving the estimated moments of the states, in line with the identification argument for polynomial policy and pricing functions in section 3.2. This motivates using the following iterative algorithm, which resembles the Expectation-Maximization (EM) algorithm but replaces the M-step by a linear regression instead of optimizing the expected log-likelihood given observations. The algorithm is computationally attractive is it avoids computing moments of the order $2L$, and works with the filtered instead of the smoothed moments of the state variables. Related iterative algorithms that simplify the maximization step are in Arcidiacono and Jones (2003) and Arellano and Bonhomme (2011).

Starting from the initial parameter estimates $\tilde{\vartheta}_L = (\tilde{\theta}_s, \tilde{h}_L)$, running the above likelihood evaluation yields as a by-product the filtered moments $\dot{\bar{s}}^L_{t|t} = E(\bar{s}^L_t \mid \Delta y_t, \mathcal{F}^{m,y}_{t-1}; \tilde{\vartheta}_L)$ given the growth realization but not the other measurements. For the maximization step, we estimate the polynomial coefficients $\hat{c}_L$ and lag parameters $\hat{R}$ by the least-squares regression of m_t on $(m_{t-1}, \dot{\bar{s}}^L_{t|t})$. This yields consistent estimates since its conditional mean given available information equals

$$E\left(m_t \mid \Delta y_t, \mathcal{F}^{m,y}_{t-1}\right) = Rm_{t-1} + c'_L \dot{\bar{s}}^L_{t|t},$$

since the lagged value m_{t-1} is in the information set used to estimate $\dot{\bar{s}}^L_{t|t}$. Similarly, we estimate the coefficients $\hat{b}_L$ of the pricing function by polynomial regression of p_t on m_t and the filtered states $\tilde{\bar{s}}^L_{t|t} = E(\bar{s}^L_t \mid \mathcal{F}^{m,y}_t; \tilde{\vartheta}_L)$ after including $(m_t, \Delta y_t)$:

$$E\left(p_t \mid \mathcal{F}^{y,m}_t\right) = b'_L \tilde{\bar{s}}^L_t + \alpha' m_t.$$

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

$$\hat{\Sigma}_\varepsilon = \frac{1}{T-1} \sum_{t=2}^{T} m_t (m_t - \hat{R}m_{t-1} + \hat{c}'_L \tilde{\bar{s}}^L_{t|t})',$$

using the orthogonality conditions $\varepsilon_t \perp (s_t, m_{t-1})$ and $m_t \perp \tilde{\bar{s}}^L_{t|t} - \bar{s}^L_t$ by definition of prediction error. Similarly, the variance of the pricing errors σ_η can be consistently estimated as

$$\hat{\sigma}_\eta = \frac{1}{T} \sum_{t=1}^{T} (p_t - \alpha' m_t) \left(p_t - \alpha' m_t - b'_L \hat{\bar{s}}^L_t\right)$$

where $\hat{\bar{s}}^L_{t|t} = E(\bar{s}^L_t \mid p_t, \mathcal{F}^{m,y}_t; \tilde{\vartheta}_L)$ are the filtered states after additional updating with p_t, using the orthogonality conditions $\eta_t \perp (s_t, m_t)$ and $p_t \perp \hat{\bar{s}}^L_{t|t} - \bar{s}^L_t$.

After the algorithm converges, we perform joint parameter optimization starting from the found local optimum to compute the MLE estimate and information matrix for inference, following the approach recommended in Watson and Engle (1983).

3.8 Consistency

The population parameters of interest are given by

$$(\theta_0, h_0) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}} \lim_{T \to \infty} \ell_T(\theta, h), \tag{16}$$

and the maximum likelihood estimator by

$$(\hat{\theta}, \hat{h}) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}} \ell_T(\theta, h), \tag{17}$$

where Θ is a finite-dimensional parameter space, and $\mathcal{H} = \prod_{m=1}^{K} \mathcal{H}_{\psi_m} \times \mathcal{H}_\pi$ is a Cartesian product of infinite-dimensional parameter spaces for the policy functions ψ_m, $m = 1, \ldots, K$, and the pricing function π. Also define the product space $\mathit{\Theta} = \Theta \times \mathcal{H}$. Let the spaces $\mathcal{H}_m$ and $\mathcal{H}_\pi$ 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, this norm 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 $\mathcal{H}_L \subseteq \mathcal{H}_{L+1} \subseteq \ldots \subseteq \mathcal{H}$ 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 $\underline{p}_L = (p_1(w), \ldots, p_L(w))$ 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

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

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 (17) is redefined over the compact finite-dimensional subspace $\mathcal{H}_{L(T)}$:

$$(\hat{\theta}, \hat{h}_L) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}_{L(T)}} \ell_T(\theta, h), \tag{18}$$

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 $\operatorname{Var}\left(\log f\left(\mathcal{S}_{t+1} | s_t; \theta_{s0}\right)\right) < \infty$.

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 (18) 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* $L^{D+1}/T \to 0$.

3.9 Conditional method of moments estimation of ϕ

The recursive pricing equation (4) pins down the dependence of the expected price-dividend ratio $\pi(s_t; \phi, \psi, \theta)$ on the stochastic discount function ϕ and other structural parameters. When $\pi(s_t; \phi, \psi, \theta)$ can be quickly and accurately computed, ϕ could be efficiently estimated by maximizing the restricted likelihood function. However, in general this requires an additional numerical approximation step, and leads to pricing functions that are no longer linear in parameters. Instead, we consider a second-stage method-of-moments procedure for estimating ϕ after the first-stage unrestricted maximum likelihood estimation of (π, ψ, θ).

The stochastic discount function ϕ is identified as the unique eigenfunction that solves the Euler equation (2). Given that the joint distribution of the endogenous variables $(\Delta c_{t+1}, \Delta d_{t+1}, s_{t+1})$ is identified, ϕ could thus be found via numerical integration and minimizing a distance criterion. This requires a choice of grid points or weighting distribution over the conditioning state s_t. By conditioning on past observations $\mathcal{F}_t^{m,y,p}$, the Euler equation in terms of returns implies the conditional moment

$$0 = E\left(\frac{1}{\beta}\phi(s_t) - \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \phi(s_{t+1}) R_{t+1} \;\middle|\; \mathcal{F}_t^{m,y,p}\right), \tag{19}$$

Under completeness of the filtered state density $f(s_t \mid \mathcal{F}_t^{m,y,p})$, the conditional moment (19) still identifies the unique eigenfunction ϕ solving (2).

By the Law of iterated expectations, the conditional moment can be stated in terms of observables by integrating out the latent state variables given observed endogenous variables. When the stochastic discount function is approximated by the polynomial $\phi_L(s_t) = e \cdot \bar{s}_t^L$, its projection on the conditioning information is a polynomial in the filtered moments of the states:

$$E\left(\phi_L(s_t) \mid \mathcal{F}_t^{m,y,p}\right) = e_L \cdot \hat{\bar{s}}_{t|t}^L.$$

The conditional moment (19) can therefore be stated in terms of the filtered moments as

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

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

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

For 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).

The above procedure can be modified to be robust against the dynamic properties of the pricing error η_t. In this case, we project the Euler equation (3) in terms of price-dividend ratios on the restricted conditioning information $\mathcal{F}_t^{m,y}$ which does not include past prices. Then, the conditional moment can be written in terms of the limited-information filtered moments of the current and next period state vectors given the augmented information set $(\mathcal{F}_t^{m,y}, p_t)$.

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 and dividend growth are computed, respectively, as

$$pd_{t+1} = \frac{P_{t+1}}{D_{t+1}} = \frac{R^x_{t+1}}{R^d_{t+1} - R^x_{t+1}}, \qquad \text{and} \qquad \frac{D_{t+1}}{D_t} = \frac{pd_{t+1}}{pd_t} R^x_{t+1}.$$

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

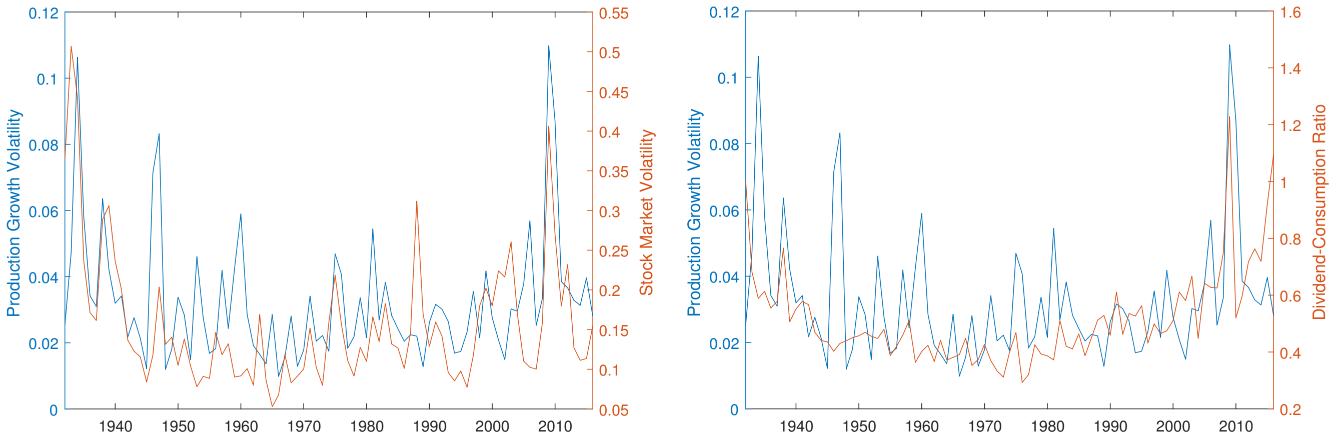
$$REV_t = \sum_{m=1}^{12} (\Delta ip_{t+1-m} - \overline{\Delta ip}_t)^2,$$

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} after demeaning with the quarterly frequency $\overline{R}_t$ as[5]

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

4.2 Economic Uncertainty and Stock Market Volatility

Figure 1 in the left panel shows the annual measures of economic uncertainty (REV) and financial market volatility (RV). Both measures peaked during the Great Depression and the Great Recession, but diverged during other well-known episodes. The first 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. The right panel of



(a) Economic uncertainty and financial volatility. (b) Economic uncertainty and dividend-consumption ratio.

Figure 1: Realized Variance of Industrial Production growth versus (a) Realized Variance of S&P 500 returns and (b) Dividend-Consumption ratio from 1930-2016.

Figure 1 compares REV with the dividend-consumption ratio. It displays substantial co-movement 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

[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.

order to explain the behaviour of the dividend-consumption ratio.

To understand how economic and financial market uncertainty affects asset prices, Figures 2 and 3 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 2 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, is actually slightly convexly increasing in changes to uncertainty, in line with rebalancing of the dividend-consumption ratio. Figure 3 shows the responses to changes in financial market volatility are in the same direction as for changes in economic uncertainty, but that realized stock market variance correlates stronger with dividend growth and the market return and weaker with consumption and output growth. In particular, dividend growth is pronouncedly 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, and their impact on fundamentals.

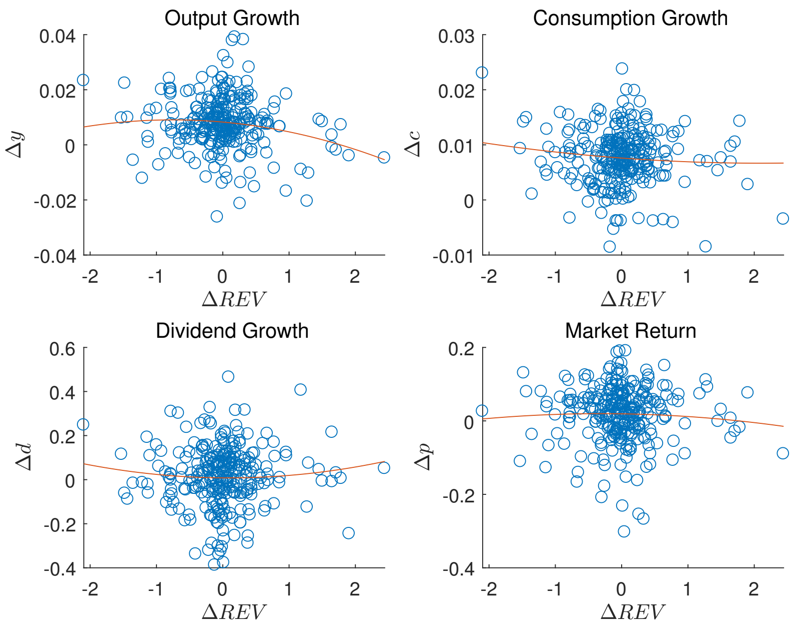


Figure 2: Quarterly changes in log Realized Economic Variance (REV) of Industrial Production growth versus log return on Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016. Fitted line corresponds to a quadratic fit.

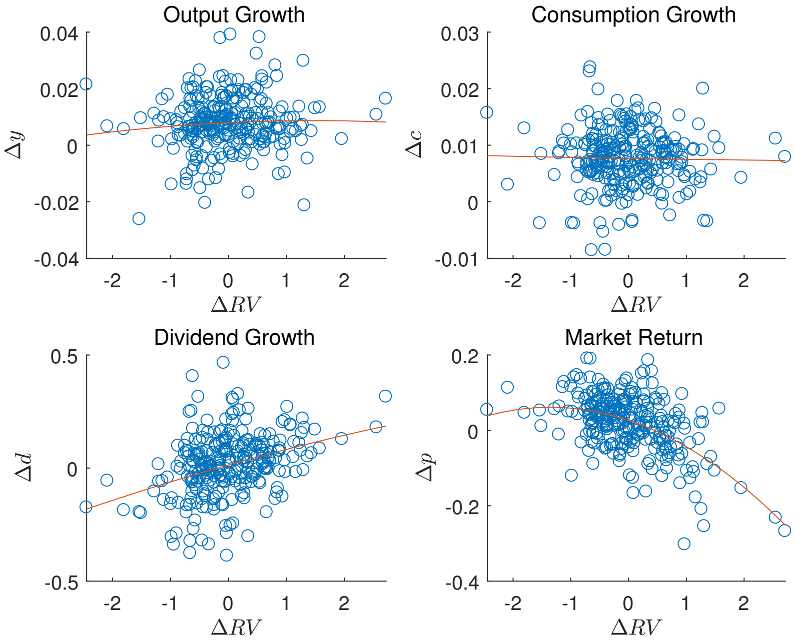


Figure 3: Quarterly changes in log Realized Variance (RV) versus log return Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016. Fitted line corresponds to a quadratic fit.

4.3 Estimates

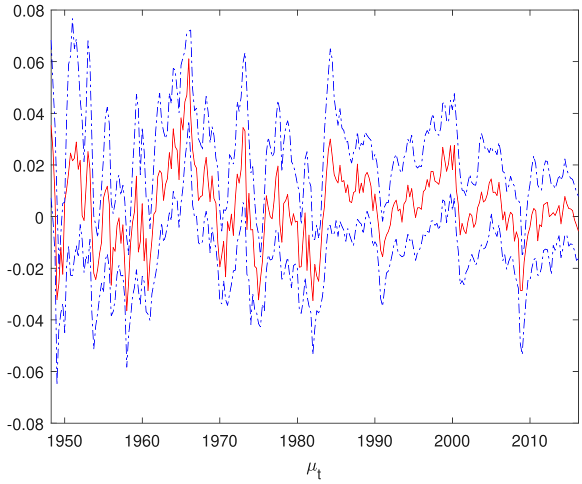
Table 1 reports the simulated maximum likelihood estimates of the transition density parameters θ_s. The mean reversion parameters (κ^x, κ^v) correspond to half-lives of the expected growth and volatility of around 7 and 2 years, respectively. Thus suggests the presence of a highly persistent growth 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 4 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 period, especially during the high growth 1990s.

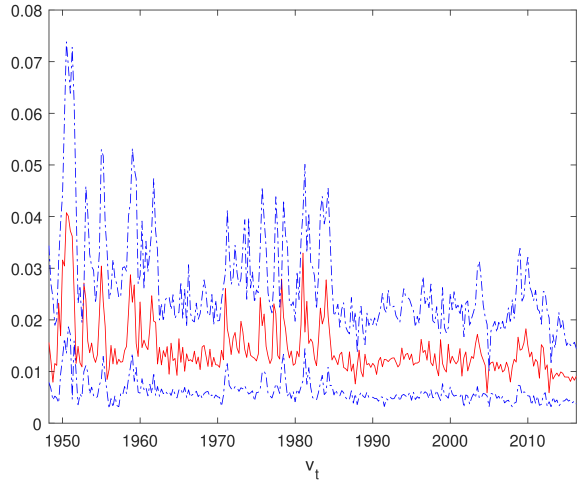
Table 1: Simulated maximum likelihood estimates of the transition density parameters θ_s for the long-run risk model with geometric stochastic volatility.

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)



(a) Filtered expected growth.

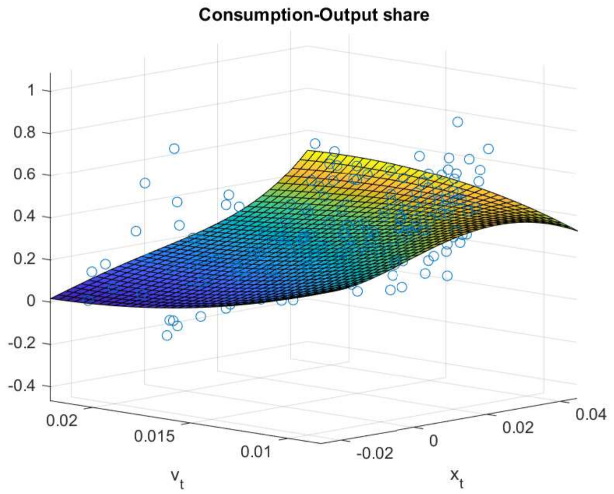


(b) Filtered expected volatility.

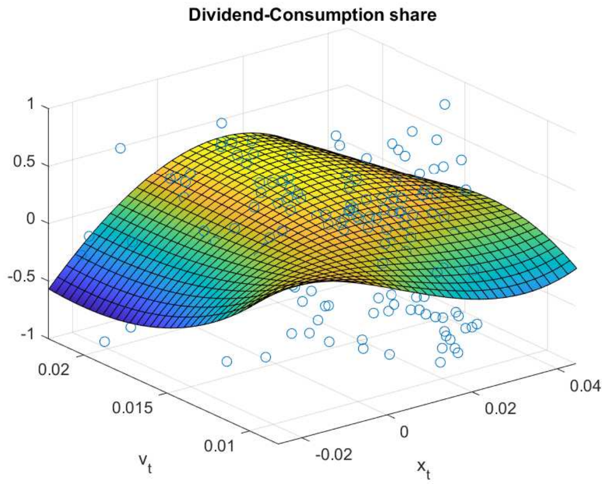
Figure 4: Filtered conditional mean of expected growth x_t and expected volatility v_t of output growth, using quarterly observations. Dashed lines are 95% confidence intervals from simulated filtered distributions.

Figure 5 shows the estimated policy functions ψ^c and ψ^d of the log consumption-output and dividend-consumption ratio. The consumption-share is generally increasing in expected growth and decreasing in growth volatility, with some amplifying interaction between the two effects. The dividend-consumption share appears relatively flat for median levels of expected and volatility of growth, but decreases sharply when expected growth is well below its median, especially when growth volatility is high.

Table 2 shows the estimated parameters $(\hat{\Sigma}_\varepsilon, \hat{R})$ of the error distribution. The consumption-share has less volatile innovations but is more persistent than the dividend-consumption share. The two series display negative contemporaneous error correlation and negative cross-autoregressive effects, which could be explained by measurement error in the consumption level featuring in the numerator and denominator, respectively. The table also



(a) Consumption-output response function.



(b) Dividend-consumption response function.

Figure 5: Estimated policy response functions ψ 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. Blue circles show filtered means of state variables. Vertical axis shows standard deviations from the mean.

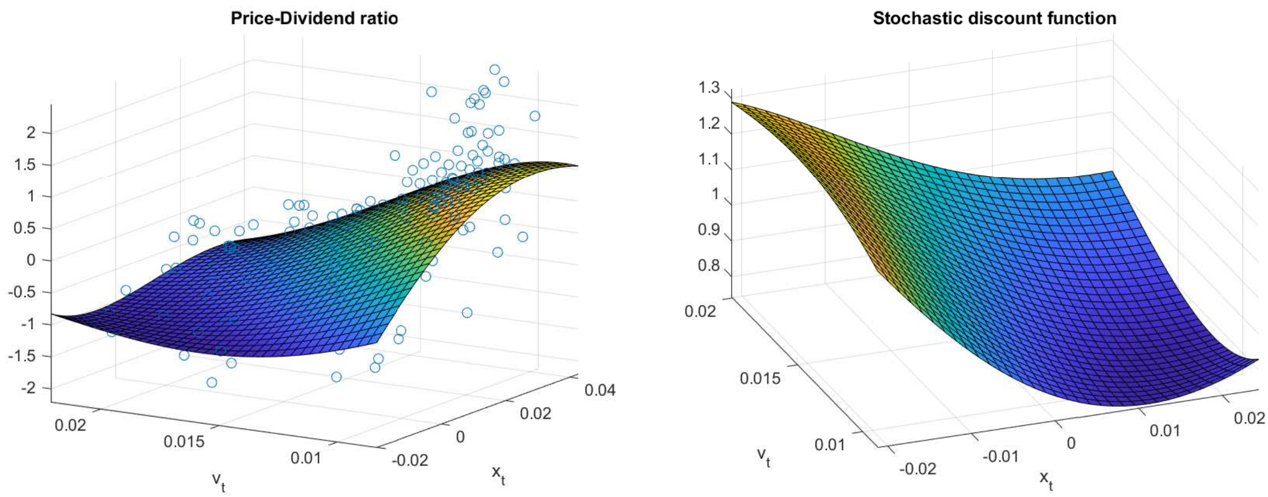
shows the estimate of the log dividend-consumption cointegration parameter λ^{dc}, which is found to exceed one.

Table 2: Simulated maximum likelihood estimates of the error parameters.

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

σ_ε^c	σ_ε^{cd}	σ_ε^d	R_{cc}	R_{cd}	R_{dc}	R_{dd}	λ^{dc}
0.160	-0.016	0.631	0.945	-0.061	-0.051	0.752	1.411
(0.001)	(0.001)	(0.028)	(0.002)	(0.002)	(0.002)	(0.005)	(0.007)

Figure 6 shows in the left panel the estimated price-dividend ratio π 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 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.



(a) Log price-dividend ratio function.

(b) Stochastic discount function.

Figure 6: Estimated 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. Blue circles show filtered means of state variables. Vertical axis shows standard deviations from the mean.

Figure 7 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.

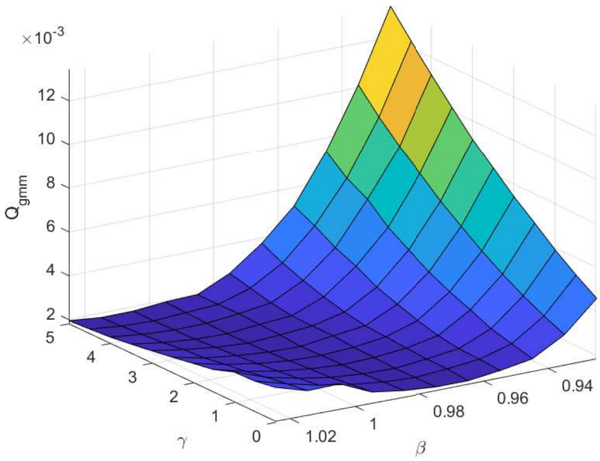


Figure 7: GMM-criterion as a function of the discount and risk aversion parameters (β, γ), where the stochastic discount function approximation coefficients are profiled as $e_L(\beta, \gamma)$.

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. We establish the consistency of a sieve maximum likelihood estimator for the general case where the measurement equation 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 price-dividend ratio is found to be increasing in expected growth, decreasing in growth volatility, with both effects interacting, and showing stronger state-dependence than could be explained by the expected consumption and dividend shares alone. Instead, this could be reconciled by investors with moderate risk aversion but high marginal utility for payoffs in times of low expected growth and high volatility. Finally, the steeply declining price-dividend ratios for moderate levels of economic uncertainty can explain episodes of large stock market volatility occurring 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., Fusari, N., and Todorov, V. (2015). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Andersen, T. G., Fusari, N., Todorov, V., and Varneskov, R. T. (2019). Unified inference for nonlinear factor models from panels with fixed and large time span. *Journal of econometrics*, 212(1):4–25.

Arcidiacono, P. and Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946.

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. *Annual Reviews of Economics*, 3(1):395–424.

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. *The Journal of Finance*, 52(4):1383–1410.

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*, 85(5):1501–1536.

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

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*, 201(2):198–211.

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

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica*, pages 871–907.

Gallant, A. R. and Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica*, 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. (1993). A closed-form solution for options with stochastic volatility and applications to bond and currency options. *The Review of Financial Studies*.

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.

Appendix

Derivation of recursive conditional moment formula. Let S_t be an affine process with drift $\mu(s_t) = K_0 + K_1 s_t$, and covariance matrix $\mathcal{C}(s_t) = H_0 + \sum_{j=1}^{D} H_{1d} s_{tj}$ affine functions of the state variables $s_t \in S_t$. 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[6]

$$\mathcal{A}f(S) = (K_0 + K_1 S)^T \nabla f(S) + \frac{1}{2}\left(\mathrm{Tr}(\nabla^2 f H_0) + \sum_{j=1}^{D} \mathrm{Tr}(\nabla^2 f H_{1,j}) s_j\right). \tag{20}$$

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 (20) 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

$$\mathcal{A}\mathcal{S}^{l_i} = \sum_{|l^j| \le L} a_{ij} \mathcal{S}^{l_j} \tag{21}$$

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_L can be solved symbolically using standard software. For the baseline model with second-order expansion $L = 2$ its solution is given in Table 3. 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 given by (7).

Example of the coefficient matrix for computing conditional moments in the baseline

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

model. Table 3 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.

Table 3: 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	$\omega_v^2 - 2\kappa^v$

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 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. $\square$