Semiparametric Estimation of Latent Variable Asset Pricing Models

Jeroen Dalderop*

August 24, 2021

Abstract

This paper studies semiparametric identification and estimation of the stochastic discount factor in consumption-based asset pricing models with latent state variables. We model consumption, dividends, and a multiplicative discount factor component via unknown functions of Markovian states describing aggregate output growth. For the case of affine state dynamics and polynomial approximation of the measurement and pricing equations, we provide rank conditions for identification and tractable algorithms for filtering, smoothing, and likelihood estimation. Empirically, we find sizable nonlinearities and interactions in the impacts of expected growth and volatility on the price-dividend ratio and the discount factor.

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

JEL Codes: C14, C58, G12

*Department of Economics, University of Notre Dame, 3060 Jenkins-Nanovic Halls, Notre Dame, IN 46556, USA. Email address: jdalderop@nd.edu. This paper has been previously circulated as "Estimating Policy Functions Implicit in Asset Prices". I am grateful to Yacine Aït-Sahalia, Xiaohong Chen, Drew Creal, Dennis Kristensen, 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 consumption-based asset pricing models with moderately risk averse households have difficulty reconciling episodes of highly volatile asset prices with relatively smooth fluctuations in macroeconomic fundamentals. Model extensions in which fundamentals and preferences are driven by persistent yet 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, heightened economic uncertainty during the 1950s and the early 1980s did not trigger excessive stock market volatility, nor did large swings in stock prices around the 2001 dot-com bubble coincide with significant volatility in economic growth. Such episodes suggest an important role for nonlinear state-dependence in fundamentals, preferences, or both.

This paper aims to understand how asset prices depend on state variables that describe aggregate growth dynamics, and whether such dependence works through consumption, cash flows, or time preferences. Therefore we study the identification and estimation of a class of asset pricing models in which consumption and dividends may depend nonlinearly on latent Markovian state variables. In particular, we model the cointegration residuals of consumption and dividends relative to output via unknown functions of state variables describing the conditional distribution of output growth, such as persistent components in its mean and volatility. While unobserved by the econometrician, the state variables' link to an observed aggregrate growth series makes it possible to identify their transition parameters, as well as the shape of the expected consumption, dividend, and pricing functions, under general hidden Markov and stationarity assumptions.

Subsequently, we show that the 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 state-dependent marginal utility function and fixed discount parameter. These take the form of the unique positive

[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).

solution to an eigenfunction problem similar to those studied in Christensen (2017) and Escanciano et al. (2020), but now featuring unobserved state variables.

To avoid the curse of dimensionality of fully nonparametric models, we impose some parametric structure on the stochastic discount factor and the distribution of the state variables. In particular, the stochastic discount factor is assumed to be multiplicatively separable in the power marginal utility of consumption and an unknown function of the state variables that allows for state-dependent time preferences. The latter component can be interpreted as a 'taste shifter' or linked to more structural models. The resulting semiparametric stochastic discount factor decomposes into a non-stationary component depending on the level of output and a stationary state-dependent component as in Hansen and Scheinkman (2009). Parametric models for the state variables allow increasing their dimension and analytic characterization of their dynamic properties. In particular, for affine state variables (Duffie et al., 2000), the framework parsimoniously generalizes the class of affine equilibrium asset pricing models (Eraker and Shaliastovich, 2008) towards nonlinear consumption and dividend dynamics. Their nonlinear state-dependence endogenously generates variation in the mean and volatility of their growth rates, instead of modeling these with additional exogenous state variables. Similarly, autocorrelation in consumption and dividend growth rates derives from that of the stationary state variables, as well as from variable-specific transitory deviations from their long-run cointegration relations.

The framework is highly tractable when the unknown functions are approximated by orthogonal polynomials as in Chen (2007). In particular, expected growth rates of consumption, dividends, and asset prices can be expressed in closed-form as polynomials of affine state variables. Moreover, the identification argument can be expressed in terms of rank conditions on the smoothed moments of the states. We study sieve maximum likelihood estimation of the parameters of the state variable dynamics and the measurement and pricing equations. The estimates are computed using a sequential Monte Carlo variant of the EM-algorithm which analytically solves the maximization step for the approximating polynomial coefficients in terms of simulated smoothed moments of the state variables. We estimate the multiplicative stochastic discount component in a second stage method-of-moments step, by integrating out the latent variables of the Euler equation using their filtered distribution. This 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 long-run risk and stochastic volatility of aggregate output growth on equity valuation ratios. We estimate the model using quarterly data on postwar U.S. macroeconomic variables, S&P 500 stock market index prices and dividends, and 3-month Treasury Bill rates. 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 to some extent the 2008 financial crisis. The frequency and duration of high volatility periods declines steadily over the sampling period, reaching its lows during the high growth 1990s. The consumption-output share mainly responds to expected growth, while the dividend-output residual is convexly increasing in volatility. High expected growth and low growth volatility lift the expected price-dividend ratio to at least one standard deviations above its mean, but only when combined. Meanwhile, return volatility peaks around median levels of growth volatility, suggesting a trade-off between the size and price-sensitivity of economic shocks. The state-dependence in the price-dividend ratio is only partially explained by that of consumption and dividends, as evidenced by the stochastic discount function increasing in expected volatility and decreasing in expected growth. This suggests state-dependent preferences play an important role in relating asset prices to future economic growth and volatility.

Related Literature. The paper is at the intersection of the literatures on nonparametric identification and estimation of stochastic discount factor models and of nonlinear dynamic latent variables models. Empirically, it contributes to the measurement of long-run risks and volatility shocks in macro-financial models.

Gallant and Tauchen (1989), Chapman (1997), Chen and Ludvigson (2009), and others, estimate the stochastic discount factor semi- or nonparametrically based on conditional moment restrictions in the form of Fredholm Type I integral equations, which are common in nonparametric instrumental variables studies. Meanwhile, Hansen and Scheinkman (2009), Chen et al. (2014), Christensen (2017), and Escanciano et al. (2020) formulate the Euler equation as a Fredholm Type II integral equation, establishing identification based on the Kreĭn-Rutman theorem. Similarly, Ross (2015) uses the finite-dimensional Perron-

Frobenius theorem to recover discrete-state option-implied probabilities. While Hansen and Scheinkman (2009) and Christensen (2017) identify the eigenfunction of the long-term valuation operator of a given stochastic discount factor, Escanciano et al. (2020) nonparametrically identify and estimate the marginal utility function of consumption. Our paper extends the latter object of interest to include unobserved state variables, covering an extended class of asset pricing models. Meanwhile, we consider a semiparametric power utility formulation with a multiplicative correction term, which includes the multiplicative habit formulation identified in Chen et al. (2014). This formulation implies that consumption only enters the stochastic discount factor through its growth rate, whose plausible stationarity allows consistent estimation based on long time series instead of large cross-sections of households. Given the joint distribution of our partially observed state vector, we apply existing identification results for Euler equations with fully observed state variables.

The identification of nonlinear dynamic latent variable models has been primarily studied for large cross-sections and panel data, such as Hu and Shum (2012) and Arellano et al. (2017), respectively. These papers focus on individual-specific state variables instead of common state variables. Gagliardini and Gourieroux (2014) and Andersen et al. (2019) extract nonlinear common factors from a large number of cross-sectional units with unpredictable and independent errors. In our 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 exogenous measurement and pricing errors. In finance, latent variables are often dealt with by inverting observations, such as in affine models for the term structure (Piazzesi, 2010), option prices (Pan, 2002; Ait-Sahalia and Kimmel, 2010), and price-dividend ratios (Constantinides and Ghosh, 2011; Jagannathan and Marakani, 2015). In nonlinear and multi-state models, the inverse mapping may not be unique. Alternatively, direct proxies for the state variables could be used, such as estimating current volatility based on high-frequency realized volatility (Andersen et al., 2003) or the option-implied VIX measure (Berger et al., 2020). However stock market volatility does not translate one-to-one into the volatility of economic fundamentals, and may be affected by time-varying risk aversion. Realized variation of low frequency macroeconomic series suffers from non-vanishing measurement error, while 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 absence of reliable proxies, state variables may still be accurately filtered using forward-looking asset prices, which this paper focuses on.

Finally, most empirical studies on the Bansal and Yaron (2004) long-run risk model and its extensions focus on calibration or method-of-moments estimation. Instead, Schorfheide et al. (2018) and Fulop et al. (2021) develop likelihood-based Bayesian methods that allow filtering the long-run risk and volatility components over time using asset price information. Fulop et al. (2021) do so using a similar polynomial approximation as ours of the log price-dividend ratio to incorporate relevant higher order effects (Pohl et al., 2018). We provide a frequentist alternative, using the EM-algorithm to estimate the polynomial coefficients directly, instead of using collocation methods.

Organization. The remainder of this paper is organized as follows. Section 2 introduces the model assumptions and the asset pricing Euler equations. 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 analysis. 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 variable x_t only.

2.1 Aggregate growth

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

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

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.

Example. (Long-run risk model with stochastic volatility) Our baseline model is the discrete-time model with two latent states $s_t = (x_t, \sigma_t^2)$, persistent growth x_t and conditional variance σ_t^2, described by

$$\begin{aligned} \Delta y_{t+1} &= \mu + x_t + \sigma_t \eta_{y,t+1} \\ x_{t+1} &= \rho_x x_t + \phi_x \sigma_t \eta_{x,t+1}, \end{aligned} \tag{1}$$

where $\eta_{y,t+1}$ and $\eta_{x,t+1}$ are i.i.d.$N(0,1)$. The conditional variance σ_t^2 follows an autoregressive Gamma process, described by

$$\sigma_t^2 \sim Gamma(\phi_\sigma + z_t, c), \qquad z_t \sim Poisson\left(\frac{\nu \sigma_{t-1}^2}{c}\right).$$

This process, introduced by Gourieroux and Jasiak (2006), is the discrete-time analogue of the continuous-time Cox-Ingersoll-Ross process. This formulation ensures that the variance is positive, and that its conditional moments are available in closed form. In particular, its conditional mean and variance equal $E\left(\sigma_{t+1}^2 \mid \sigma_t^2\right) = \nu \sigma_t^2 + (1-\nu)\bar{\sigma}^2$ and $\mathrm{Var}\left(\sigma_{t+1}^2 \mid \sigma_t^2\right) = \frac{(1-\nu)\bar{\sigma}^2}{\phi_\sigma}(2\nu\sigma_t^2 + (1-\nu)\bar{\sigma}^2)$, respectively.

2.2 Consumption and dividend policy

In general optimal consumption choice depends on all sources of wealth and all investment opportunities. When the primary interest is in understanding the response of consumption to changing economic growth, a flexible reduced form approach is to model consumption relative to output via an unspecified function $\psi^c(\cdot)$ of the latent states. Together with linear dependence on its lag, and an unexplained shock ε_t^c, this yields the semiparametric additive formulation for the log consumption-to-output ratio:

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

The specification of consumption relative to output guarantees their long run cointegration relation, 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, aggregate corporate dividends per unit of output or consumption is flexibly modeled as a nonparametric function of the state $\psi^d(\cdot)$ plus error component ε_t^d. Suppose a portfolio of equities is traded at the price P_t and pays a stochastic dividend level D_t per share. Dividends can be seen as a leveraged claim on consumption, which implies a cointegration relation between $\log D_t$ and $\log C_t$ (Menzly et al., 2004), and thus between $\log D_t$ and $\log Y_t$. With cointegration parameter λ, the logarithmic residual is modeled analogous to the consumption share by the semiparametric additive specification

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

Alternatively the dividend-to-consumption ratio could be modeled via the cointegration residual $d_t - \lambda c_t$, as any pair of ratios of output, consumption, and dividends, pins down the remaining one.

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 process

$$m_t = R m_{t-1} + \psi(s_t) + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0, \Sigma_\varepsilon), \tag{2}$$

with $\varepsilon_t = (\varepsilon_t^c, \varepsilon_t^d)$ the combined error term, and Σ_ε its covariance matrix.[2] The latent states are assumed to be strongly exogenous, in the sense that $\mathbb{E}(\varepsilon_t \mid (s_{t+j})_{j=-\infty}^{\infty}) = 0$. The identifying assumptions, formalized in Assumption 1, require future augmented states $\mathcal{S}_{t+1} = (\Delta y_{t+1}, s_{t+1})$ to be independent of current and past measurements of m_t, controlling for s_t. This 'no feedback' assumption allows future growth $(\Delta y_{t+h})_{h\geq 1}$ to serve as instruments for s_t. The assumption intuitively implies that consumption and dividend shocks can only affect future output growth *through* the latent states s_t, which in our example contain the persistent growth component. It does not rule out contemporaneous

[2]Our identification results allow for conditional heteroskedastic errors through $\Sigma_{\varepsilon,t} = \Sigma_\varepsilon(s_t)$, but our empirical results cover the homoskedastic case.

correlation between the measurement innovations ε_t and Δy_t. In particular, in our long-run risk example, ε_t is allowed to be correlated with the transitory output growth shock $\eta_{y,t}$. Such correlation is plausible given that m_t is measured relative to y_t, and thus affected by any measurement error in the latter.

The process reduces to a standard first-order vector autoregression when $\psi(s_t)$ is constant, in which case consumption and dividends relative to output do not respond to s_t. For example, Bansal et al. (2007) include the cointegration residuals of consumption and dividends in a vector autoregression with other stationary variables. Furthermore, it includes models without direct lag dependence, such that R is a zero matrix, but where the cointegration residuals depend on latent state variables. For example, Schorfheide et al. (2018) model the consumption-output and dividend-consumption residuals using the linear function $\psi(x_t) = \mu + \psi x_t$, where x_t is a persistent growth component. Such models would typically allow for serial correlation in the error term. Finally, our specification can be related to state-space models expressed in terms of growth rates, by representing (2) as

$$\begin{aligned}\Delta c_{t+1} &= \Delta y_{t+1} + (R_{cc} - 1, R_{cd})^T m_t + \psi^c(s_{t+1}) + \varepsilon^c_{t+1} \\ \Delta d_{t+1} &= \lambda \Delta y_{t+1} + (R_{dc}, R_{dd} - 1)^T m_t + \psi^d(s_{t+1}) + \varepsilon^d_{t+1}.\end{aligned} \tag{3}$$

When $R = I$, output, consumption, and dividends are not subject to cointegration relations. When Δy_{t+1} follows (1) and $\psi(\cdot)$ is constant, this yields the baseline long-run risk model from Bansal and Yaron (2004), in which consumption and dividend growth depend linearly on a common persistent component x_t.

An advantage of formulation (2) is that it does not introduce additional state variables for consumption and dividend growth. Instead, time-variation in their mean and volatility derives from those of output growth. This parsimony may appear restrictive, compared to for example Schorfheide et al. (2018) who allow for three separate stochastic volatility processes. However, allowing for nonlinear state dependence through $\psi(s_t)$ might capture variation in conditional means that would otherwise show up as conditionally heteroskedastic, non-Gaussian error terms.

2.3 Stochastic discount factor

Suppose there is an infinitely-lived representative agent whose consumption and investment choices maximize its life time expected utility U_t given by

$$U_t = E\left(\sum_{s=t}^{\infty} \beta^{s-t} u(C_s, s_s) \mid \mathcal{F}_t\right),$$

where β is a fixed discount parameter, and $u(\cdot)$ is a state-dependent instantaneous utility function with the multiplicative decomposition

$$u(C_t, s_t) = v(C_t; \gamma)\phi(s_t), \tag{4}$$

where $v(\cdot)$ is 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 aim to better explain equity risk premia. The component $\phi(s_t)$ can be directly interpreted as a taste shifter, describing how the marginal utility of consumption changes with the state of the economy. Since economic theory may not prescribe how variables such as the expected growth and its volatility affect such time preferences, it is desirable to not restrict the functional form $\phi(\cdot)$. For example, evidence from option markets suggests marginal utility may be U-shaped with respect to financial market volatility (Song and Xiu, 2016). It remains to be seen whether similar nonlinear time preferences exist for measures of economic uncertainty.

Specification (4) also covers more structural models of the form $u(C_t, M_t) = \frac{C_t^{1-\gamma}}{1-\gamma} M_t$, where the multiplicative component M_t can be written in terms of the Markovian state variables s_t. For example, the utility over wealth models in Bakshi and Chen (1996) specify $M_t = W_t^{\lambda}$ for some parameter λ, where W_t measures either absolute or relative wealth. This is covered by (4) as long as W_t, or the wealth-consumption ratio $\frac{W_t}{C_t}$, is Markovian in s_t, as is common when wealth is defined as the value of a claim on future aggregate consumption.

Models with reference utility imply $M_t = Q_t^\gamma$, where Q_t is the inverse consumption surplus ratio relative to a reference level, which may be determined by s_t. Habit models specify Q_t in terms of lagged consumption growth Δc_t, which in our specification depends in part on s_t, or which could be added into $\psi(s_t, \Delta c_t)$ as an extension discussion below. Models with incomplete markets or private information imply $M_t = E\left((C_t^i/C_t)^{-\gamma} \mid s_t\right)$ and $M_t = E\left((C_t^i/C_t)^{\gamma} \mid s_t\right)^{-1}$, respectively, where C_t^i is consumption by each ex-ante identical consumer i, see e.g. Hansen and Renault (2010). Models with stockholder consumption can be described by $M_t = (C_t^s/C_t)^{-\gamma}$, where stockholder consumption C_t^s has been found to covary more with long-run growth than aggregate consumption (Malloy et al., 2009). In this case, $\phi(x_t)$ could serve as a correction factor when using the latter series. For certain types of subjective beliefs $\phi(s_t)$ may represent probability overweighting of possible outcomes of the state variable. Finally, besides economic interpretation, $\phi(s_t)$ could be used to detect statistical misspecification of the transitory component of the stochastic discount factor, and thereby aid the search for appropriate structural models.

Under this semiparametric specification, the pricing kernel $\zeta_t = \beta^t C_t^{-\gamma}\phi(s_t)$ is the product of a deterministic time-discount factor, a non-stationary 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.

Extensions. The arguments of $\phi(\cdot)$ could be extended to include m_t or other stationary observed variables z_t. For identification of $\phi(\cdot)$, the augmented state vector (s_t, m_t, z_t) would have to satisfy the Markov assumptions in Assumption 1. For example, utility of wealth models may motivate the semiparametric form $\phi(s_t, m_t) = \tilde{\phi}(s_t)e^{\beta^T m_t}$, after conjecturing that the log wealth-consumption ratio is linear in m_t. For habit models featuring a finite number of lags of consumption growth, $\phi(s_t, z_t)$ with $z_t = \Delta c_t$ or its further lags may be chosen. Theoretical and empirical results for related habit models are available in Chen and Ludvigson (2009) and Escanciano et al. (2020). Finally, models where M_t is an unobserved Markovian time preference shock as in Albuquerque et al. (2016) may be included with $z_t = r_t^f$, provided there is a one-to-one mapping between the risk-free rate $r_t^f = \pi^f(s_t, m_t, M_t)$ and the time preference shock, given (s_t, m_t).

2.4 Euler equation for asset prices

In rational expectations equilibrium models, the cum-dividend return R_{t+1}^d on any traded equity price satisfies the Euler equation

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

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

$$\frac{1}{\beta}\phi(s_t) = E\left(\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \phi(s_{t+1}) R_{t+1}^d \mid s_t\right), \tag{5}$$

which can be recognized as a Type-II Fredholm integral equation. Using infinite-dimensional versions of the Perron-Frobenious theorem, Christensen (2017) and Escanciano et al. (2020) 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 $E\left(\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{t+1}^d \mid s_{t+1}, s_t\right)$ given both the current and next period state variables. In terms of price-dividend ratios, the Euler equation reads

$$\frac{P_t}{D_t} = E\left(\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{\phi(s_{t+1})}{\phi(s_t)} \frac{D_{t+1}}{D_t} \left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \mid \mathcal{F}_t\right) \tag{6}$$

When the extended state vector $(s_t, m_t) \in \mathcal{F}_t$, the Markovian consumption and dividend dynamics imply that the price-dividend ratio equals a function $\pi(s_t, m_t)$ which satisfies the recursive relation

$$\pi(s_t, m_t) = E\left(\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \frac{\phi(s_{t+1})}{\phi(s_t)} \frac{D_{t+1}}{D_t} \left(1 + \pi(s_{t+1}, m_{t+1})\right) \mid s_t, m_t\right). \tag{7}$$

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

$$\frac{1}{\beta}\phi(s_t) = E\left(\phi(s_{t+1})\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{D_{t+1}}{D_t}\frac{1+\pi(s_{t+1},m_{t+1})}{\pi(s_t,m_t)} \mid s_t\right). \tag{8}$$

Under the dynamics (2), the logarithm of the consumption-discounted cash flow $\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{D_{t+1}}{D_t}$ can be decomposed as

$$-\gamma\Delta c_{t+1} + \Delta d_{t+1} = (\lambda-\gamma)\Delta y_{t+1} + (-\gamma, 1)^T(m_{t+1}-m_t).$$

The integral equation (8) can therefore be stated in terms of the price-dividend function $\pi(s_t, m_t)$ and the Markovian density $f(\mathcal{S}_{t+1}, m_{t+1} \mid s_t, m_t)$. The latter decomposes into the product of the parametric state density $f(\mathcal{S}_{t+1} \mid s_t)$ and the semiparametric measurement density $f(m_{t+1} \mid s_{t+1}, m_t) = f_\varepsilon(m_{t+1} - \psi(s_{t+1}) - (R-I)m_t)$, both of which can be identified without using asset prices.

Finally, the price of a risk-free bond with one-period maturity is given by

$$\begin{aligned}
P_t^f &= E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\frac{\phi(s_{t+1})}{\phi(s_t)} \mid s_t, m_t\right) \\
&= E\left(\beta\exp\left\{-\gamma\Delta y_{t+1} - \gamma(\psi^c(s_{t+1}) + \varepsilon_{t+1})\right\}\frac{\phi(s_{t+1})}{\phi(s_t)} \mid s_t\right)\exp\left(-\gamma(R_{cc}-1, R_{cd})^T m_t\right) \\
&\equiv \pi^f(s_t)\exp(\alpha_f^T m_t).
\end{aligned}$$

Hence the risk-free rate $r_t^f = \log P_t^f = \log\pi^f(s_t) + \alpha_f^T m_t$ takes a partially linear form.

2.5 Generalizing affine models

By suitably specifying the nonlinear functions $\psi(s_t)$ and latent state dynamics $f(s_{t+1} \mid s_t)$, it is possible to generalize commonly-used affine models for consumption and dividend growth, while retaining some of their tractability.

In particular, suppose the consumption and dividend policy functions are approximated by L-degree polynomial expansions:

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

where $\bar{s}^L$ is a column vector that stacks monomials up to degree L in lexicographic order. Orthogonal polynomials such as the Hermite or Chebyshev polynomials are spanned by elementary polynomials and can be represented in this way.

Affine models are often used to describe non-Gaussian dynamics, as they can incorporate features such as stochastic volatility and leverage effects in a tractable fashion. Discrete-time affine models can be characterized by their exponential-affine conditional Laplace transforms or moment-generating functions:

$$E(e^{u^T s_{t+1}} \mid s_t) = e^{a(u)+b(u)^T s_t},$$

for some coefficient functions $a(\cdot)$ and $b(\cdot)$ that satisfy $a(0) = b(0) = 0$. The conditional mixed moments up to orders $0 \leq j_1, j_2 \leq L$ under such models can be computed as

$$E\left(s_{t+1,i_1}^{j_1} s_{t+1,i_2}^{j_2} \mid s_t\right) = \frac{\partial^{j_1}\partial^{j_2}}{\partial^{j_1} u_{i_1} \partial^{j_2} u_{i_2}} e^{a(u)+b(u)^T s_t}\Big|_{u=0} = Q_{L,j_1 j_2 \cdot}^T \bar{s}_t{}^L, \tag{9}$$

where the rows in the $L^D \times L^D$ matrix Q_L collect coefficients for each of the mixed monomials in $\bar{s}_t{}^L$. As a result, expectations of the state-dependent components in consumption and dividends take the polynomial forms

$$E\left(\psi_L^c(s_{t+1}) \mid s_t\right) = c^T Q_L \bar{s}_t^L, \qquad E\left(\psi_L^d(s_{t+1}) \mid s_t\right) = d^T Q_L \bar{s}_t^L.$$

Example. For the special case of autoregressive Gamma processes, used for modeling stochastic variance σ_t^2 in our application, Gourieroux and Jasiak (2006) show that there exist orthogonal polynomials $\Psi_j(\cdot)$, for any order $j = 0, 1, \ldots$, such that

$$E\left(\Psi_j(\sigma_{t+1}^2) \mid \sigma_t^2\right) = \nu^j \Psi_j(\sigma_t^2),$$

for the scalar persistence parameter ν. The polynomials take the form of scaled generalized Laguerre polynomials, and form a convenient choice of basis functions for $\psi_L^c(\cdot)$ and $\psi_L^d(\cdot)$.

When the functions $\psi_L(s_t)$ are linear in affine state variables ($L = 1$), consumption and dividend growth are themselves affine. More general nonlinear specifications of $\psi_L(s_t)$ ($L \geq 2$) allow for convex or concave relations, for interaction terms between the state variables, or for higher order effects.

2.6 Continuous-time formulation

Our framework can also be used to generalize continuous-time affine models, which are widely used for derivative pricing and risk management. Following Duffie et al. (2000), the continuous-time affine specification for (dy_t, ds_t) requires its drift and covariance matrix to be affine functions of the state variables s_t.[3]

Example. The continuous time counterpart of our baseline long-run risk model with stochastic expected growth x_t and volatility v_t is described by

$$\begin{aligned} dy_t &= (\mu + x_t)dt + \sigma_t dW_t^y \\ dx_t &= -\kappa^x x_t dt + \omega^x \sigma_t dW_t^x \\ d\sigma_t^2 &= \kappa^\sigma \left(\sigma_t^2 - \bar{\sigma}^2\right) dt + \omega^\sigma \sigma_t dW_t^\sigma, \end{aligned} \tag{10}$$

where $(W_t^y, W_t^x, W_t^\sigma)$ are uncorrelated standard Brownian motions. This process is obtained as the limit when the time between observations $\tau \to 0$. The discrete and continuous time parameters are related as $\rho_x = \exp(-\kappa^x \tau)$, $\nu = \exp(-\kappa^\sigma \tau)$, $\phi_s = 2\kappa^\sigma \bar{\sigma}^2 / \omega_\sigma^2$, and $c = \frac{1}{2}\omega_\sigma^2 (1 - \exp(-\kappa^\sigma \tau))/\kappa^\sigma$. Positive values for the mean reversion parameters κ^x and κ^σ assure that $s_t = (x_t, \sigma_t^2)$ is stationary around its unconditional mean $(\mu, \bar{\sigma}^2)$. Moreover, the model could be extended to allow growth and volatility innovations to be correlated.

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:

$$E(\bar{\mathcal{S}}_{t+\tau}^L \mid \mathcal{S}_t) = e^{\tau A_L} \bar{\mathcal{S}}_t^L, \tag{11}$$

where the coefficients of the matrix A_L are functions of the transition parameters which can be computed symbolically using standard software.

The functional specification of the consumption and dividend policies leads to a general

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

class of potentially nonlinear dynamics of consumption and dividend growth according to

$$dm_t = (R - I)m_t dt + d\psi(s_t) + \sigma^m dW_t^m. \tag{12}$$

When $\psi(s_t)$ is linear, the consumption and dividend growth dynamics reduce to the benchmark affine models studied in Eraker and Shaliastovich (2008).

The conditional moment formula (11) allows exact computation of expected values of polynomials of the state variables. For example, growth in consumption relative to output at horizon τ 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

$$E\left(\Delta c_{t,t+\tau} - \Delta y_{t,t+\tau} \mid s_t, c_t - y_t\right) = \left(e^{-\rho^c \tau} - 1\right)(c_t - y_t) + \rho^c c^T Q_L(\tau) \bar{s}_t^L, \tag{13}$$

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. Invertibility of $Q_L(\tau)$ implies a one-to-one mapping between the coefficients of the consumption and dividend policy functions and those of their expected growth rates over any horizon τ, which would thus determine their entire term structure.

Finally, the continuous-time formulation allows linking the volatility of asset returns, which can be estimated at high-frequencies, to that of state variables describing macroeconomic fundamentals. 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

$$d \log P_t - E_t(d \log P_t) = \alpha^* \cdot \sigma_m dW_t^m + \lambda^p(s_t) d\mathcal{S}_t,$$

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

$$d\langle \log P \rangle_t = \alpha^* \cdot \sigma_m \sigma_m' \alpha^* + \lambda_s^p(s_t)' d\langle \mathcal{S} \rangle_t \lambda_s^p(s_t). \tag{14}$$

When the pricing and policy functions are polynomials, so are the gradients $\lambda_s^p(s_t)$, and (14) 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 $\psi = (\psi^c, \psi^d)^T$, 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 (10) and (12). Moreover, the relation between instantaneous and cumulative growth (13) could be used to translate between the timing assumptions.

3.1 State space formulation

The measurements $m_t = \left(\log \frac{C_t}{Y_t}, \log \frac{D_t}{Y_t^\lambda}\right)^T$ and normalized prices $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ contain aggregate quantities whose conditional mean dependence on the unobserved state variables s_t is approximated by polynomials. 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:

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

e) The state- and measurement-dependence of prices is contemporaneous:

$$\eta_{t+1} \mid (m_t, p_t, s_t) \sim \eta_{t+1} \mid \eta_t$$

where $\eta_t = p_t - E(p_t \mid m_t, s_t)$ *is the pricing error.*

Stationarity 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 1.*b)* 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 1.*c)* implies that the partially observed $\mathcal{S}_t$ forms a hidden Markov process, and is not caused in the sense of Granger (1969) by the observations (m_t, p_t). 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 their dependence at all leads and lags. The contemporaneous state-dependence of measurements and prices in 1.*d)* and 1.*e)* rules out their direct dependence on past states, which is a timing assumption also made by Hu and Shum (2012) and describes rational forward-looking behavior. Finally, prices p_t are distinguished from measurements m_t by depending on their own lag only through a Markovian pricing error. The distinction is motivated by the presence of habits or frictions in consumption and cash flow choices, while the pricing error is attributed to market sentiments unrelated to fundamentals. However, if the other type of lag dependence is deemed more appropriate, a series of prices could be included in m_t, or a series of measurements in p_t.

Our application focuses on the special case of partially linear measurement equations with Gaussian errors. Combined with the transition density, this case can be summarized by the state space formulation

$$\begin{aligned} m_t &= R_m m_{t-1} + \psi(s_t) + \varepsilon_t, && \varepsilon_t \sim \text{i.i.d.} N(0, \Sigma_\varepsilon) \\ p_t &= \tilde{\pi}(s_t) + \alpha^T m_t + \eta_t, && \\ \eta_t &= R_p \eta_{t-1} + \omega_t, && \omega_t \sim \text{i.i.d.} N(0, \Sigma_\omega) \\ \mathcal{S}_{t+1} &\sim f(\mathcal{S}_{t+1} \mid s_t), && \end{aligned} \tag{15}$$

where ε_t and ω_t are uncorrelated, and independent of s_t. The serially correlated pricing error η_t allows for the presence of persistent deviations from the fundamental price given the state variables and cointegration residuals. This allows for an autoregressive stochastic discount factor component unrelated to fundamentals, as in 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 and pricing functions π under the hidden Markov assumptions. Given these, we study the identification of the stochastic discount function 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 $(m_t, \Delta y_t)$ identifies the first-order Markovian distribution of $(m_t, \Delta y_t, s_t)$. Intuitively, they exploit the conditional independence of past and future observations given the current partially observed state variable. A related argument is used in Arellano et al. (2017) to identify the consumption rule in terms of a persistent earnings component using future observed earnings. Our no feedback assumption makes this strategy possible as future growth realizations are independent of the current measurement given the current latent state. However, while Hu and Shum (2012) and Arellano et al. (2017) use a large number of cross-sectional units to estimate the multi-period densities, we instead use a stationary time series $(m_t, \Delta y_t)$ observed over a large number of periods T.

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 $\mathcal{F}^y_{t+1:t+K}$ denote the future growth realizations $(\Delta y_{t+1}, \dots, \Delta y_{t+K})$ for a finite number of leads K. Under stationarity of $(m_t, \Delta y_t)$, the joint density $f(m_t, \mathcal{F}^y_{t+1:t+K})$ can be consistently estimated by letting $T \to \infty$. The no feedback condition implies the conditional independence $m_t \mid s_t, \mathcal{F}^y_{t+1:t+K} \sim m_t \mid s_t$, so that

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

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

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

where the updated density $f(s_{t+1} \mid m_t, \mathcal{F}^y_{t+1:t+K}) = \frac{f(m_t, s_{t+1} | \mathcal{F}^y_{t+1:t+K})}{f(m_t | \mathcal{F}^y_{t+1:t+K})}$ is identified from the previous step using

$$f(m_t, s_{t+1} \mid \mathcal{F}^y_{t+1:t+K}) = \int f(m_t \mid s_t) f(s_{t+1} \mid s_t, \mathcal{F}^y_{t+1:t+K}) f(s_t \mid \mathcal{F}^y_{t+1:t+K}) ds_t.$$

Provided the updated density is also complete, the conditional density $f(m_{t+1} \mid m_t, s_{t+1})$ and thus its conditional mean are identified.

In case of a polynomial measurement equation, the completeness assumptions reduce to rank conditions involving conditional moments of the state variables given future and/or past growth realizations. For example, let $s^l_{t|T} = E(s^l_t \mid \mathcal{F}^y_T)$ be the smoothed conditional l-th mixed 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 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 $\mathcal{F}^y_T$.[4] The latter implies $E\left(\varepsilon_t s^l_{t|T}\right) = 0$ for any $l = 0, \ldots, L$, which yields the linear system of $L+1$ equations

$$E\left(\bar{s}^{L\,\prime}_{t|T} m_t\right) = \rho E\left(\bar{s}^{L\,\prime}_{t|T} m_{t-1}\right) + c'_L E\left(\bar{s}^L_t \bar{s}^{L\,\prime}_{t|T}\right). \tag{16}$$

The prediction error $\bar{e}^L_t = \bar{s}^L_{t|T} - \bar{s}^L_t$ by construction satisfies $E\left(e^k_t s^l_{t|T}\right) = 0$ 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

$$E\left(\bar{s}^{L\,\prime}_{t|T} m_{t-1}\right) = \rho E\left(\bar{s}^{L\,\prime}_{t|T} m_{t-2}\right) + c'_L E\left(\bar{s}^L_{t-1} \bar{s}^{L\,\prime}_{t|T}\right), \tag{17}$$

where $E\left(\bar{s}^L_{t-1} \bar{s}^{L\,\prime}_{t|T}\right) = E\left(E\left(\bar{s}^L_{t-1} \mid \mathcal{F}^y_T\right) \bar{s}^{L\,\prime}_{t|T}\right) = E\left(\bar{s}^L_{t-1|T} \bar{s}^{L\,\prime}_{t|T}\right)$ by iterated expectations. Together, the $L+2$ parameters in (ρ, c_L) are thus (over)identified by the $2(L+1)$ linear equations given by (16) and (17), provided the outer product matrix $E\left(\bar{s}^L_{t|T} \bar{s}^{L\,\prime}_{t|T}\right)$ is invertible.

This identification strategy can be seen as a two-stage version of the use of instrumental 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 $\pi(s_t, m_t) = E\left(\frac{P_t}{D_t} \mid s_t, m_t\right)$ can be nonparametrically identified from the integral equation

$$E\left(\frac{P_t}{D_t} \mid m_t, \mathcal{F}^y_{t+1:t+K}\right) = \int \pi(s_t, m_t) f(s_t \mid m_t, \mathcal{F}^y_{t+1:t+K}) ds_t, \tag{18}$$

[4] Correlation between ε_t and Δy_t could be allowed for by adding the latter as a linear regressor.

under the hidden Markov and no feedback Assumptions 1.a)-c) and the completeness of $f(s_t \mid m_t, \mathcal{F}^y_{t+1:t+K})$. Moreover, Assumptions 1.d) and 1.e) imply the pricing errors are independent of future and past measurements, respectively, so that subsets of those could be added as conditioning variables.

The special case of a partially log-linear model for univariate prices $p_t = \log(\frac{P_t}{D_t}) = \tilde{\pi}^p_L(s_t) + \alpha^T m_t + \eta_t$ with polynomial approximation $\tilde{\pi}^p_L(s_t) = b'_L \bar{s}^L_t$ and Gaussian errors $\eta_t = \rho_p \eta_{t-1} + \omega_t$ can be analyzed as a two-stage linear regression model. In particular, let $\tilde{s}^l_{t|T} = E(s^l_t \mid \mathcal{F}^{y,m}_T)$ be the smoothed l-th moment of the state s_t given the leads and lags of both the measurements and growth realizations. Since these are independent of the pricing error η_t, the conditional mean of the log price-dividend ratio p_t is given by the regression equation

$$E\left(p_t \mid \tilde{\bar{s}}^L_{t|T}, m_t\right) = b'_L \tilde{\bar{s}}^L_{t|T} + \alpha' m_t.$$

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

$$\begin{aligned} E\left((p_t - \alpha' m_t)^2\right) &= E\left(\left(b'_L \bar{s}^L_t + \eta_t\right)^2\right) \\ &= b'_L E\left(\bar{s}^L_t \bar{s}^{L'}_t\right) b_L + \sigma^2_\eta, \end{aligned} \tag{19}$$

using $E\left(\eta_t \bar{s}^L_t\right) = 0$, where the outer-product matrix $E\left(\bar{s}^L_t \bar{s}^{L'}_t\right)$ is identified from the state transition parameters. The expected price-dividend ratio is then computed as $\pi(s_t, m_t) = e^{b'_L \bar{s}^L_t + \alpha' m_t + \frac{1}{2}\sigma^2_\eta}$.

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 $E\left(\eta_{t+1}\eta_t\right) = \rho_\eta \sigma^2_\eta$ using

$$\begin{aligned} E\left((p_{t+1} - \alpha' m_{t+1})(p_t - \alpha' m_t)\right) &= E\left(\left(b'_L \bar{s}^L_{t+1} + \eta_{t+1}\right)\left(b'_L \bar{s}^L_t + \eta_t\right)\right) \\ &= b'_L Q_L E\left(\bar{s}^L_t \bar{s}^{L'}_t\right) b_L + E\left(\eta_{t+1}\eta_t\right), \end{aligned}$$

where the coefficient matrix Q_L describing the state moment dynamics (9) and the unconditional moment matrix $E\left(\bar{s}^L_t \bar{s}^{L'}_t\right)$ are known given the transition parameter θ_s. Finally, applying the same identification strategy to $E\left((p_{t+j} - \alpha' m_{t+j})(p_t - \alpha' m_t)\right)$ for lead orders $j \geq 2$ could be used to identify the entire autocovariance function of the pricing error η_t

in order to assess the first-order autoregression specification.

Given the correct specification of the joint Markovian transition density of (m_t, p_t, s_t), the policy and pricing functions can be efficiently estimated using full-information likelihood methods. Still, 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. This robustness motivates using the moment-based estimators as initial values for finding the maximum likelihood estimator via the Expectation-Maximization or similar algorithm.

3.2.3 Identification of the stochastic discount function

Once the transition density and the policy and pricing 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 (8). Let $\mathcal{L}^2 = L^2(\mathbb{P})$ denote the Hilbert space of square integrable functions with the marginal 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

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

Christensen (2017) and Escanciano et al. (2020) 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 that the price-dividend function $\pi(s_t, m_t)$ is positive almost everywhere. Some mild sufficient conditions on $\mathcal{K}$ for $\mathbb{M}$ to be bounded and compact are given in Christensen (2017) and Escanciano et al. (2020). In particular, compactness follows from

$$\iint \mathcal{K}(s_t, s_{t+1})^2 f_s(s_{t+1}) f_s(s_t) ds_{t+1} ds_t < \infty,$$

where $f_s(\cdot)$ is the marginal density of s_t. Since ϕ is only identified up to scale, estimation requires a normalization such as $E\left(\phi(s_t)^2\right) = 1$.

Further restrictions can be formed by adding conditioning variables such as m_t to identify additional arguments in $\phi(\cdot)$ or to achieve overidentification. This approach is taken in Chen and Ludvigson (2009) to help identify $\phi(\cdot)$ 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

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

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

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

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

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

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

$$f(p_{t+1} \mid \mathcal{F}_{t+1}^{m,y}, \mathcal{F}_t^{p}; \vartheta) = \iint f_\eta\left(\eta_{t+1}(s_{t+1}) \mid \eta_t(s_t); \pi, \sigma_\eta^2\right) f\left(s_{t+1}, s_t \mid \mathcal{F}_{t+1}^{m,y}, \mathcal{F}_t^{p}; \vartheta\right) ds_{t+1} ds_t,$$

where $\eta_t(s_t) = p_t - \pi(s_t, m_t)$ are the implied pricing errors.

3.4 Sequential Monte Carlo filtering and smoothing

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. Instead, particle filtering or sequential Monte Carlo simulation can be used to recursively approximate expectations of any nonlinear function of the state vector, see Doucet and Johansen (2009) for an overview.

The filtering density of the latent states satisfies the recursion

$$f(s_{t+1} \mid \mathcal{F}_{t+1}^{p,m,y}) \propto \int f(p_{t+1}, m_{t+1} \mid \mathcal{S}_{t+1}, p_t, m_t, s_t) f(\mathcal{S}_{t+1} \mid s_t) f(s_t \mid \mathcal{F}_t^{p,m,y}) ds_t.$$

This motivates the following recursive algorithm. Let $(s_{i,t})_{i=1}^{N^s}$ be a set of N^s particles drawn from $\mathcal{F}_t^{p,m,y}$ with weights $(w_{i,t})_{i=1}^{N^s}$. First, draw next period's states $(s_{i,t+1})_{i=1}^{N^s}$ from the transition density $f(s_{t+1} \mid s_t; \theta_s)$. Second, compute the updated sampling weights of

$(s_{i,t+1})_{i=1}^{N^s}$ given $(p_{t+1}, m_{t+1}, \Delta y_{t+1})$ as

$$w_{i,t+1} \propto f_\eta \left(\eta_{t+1}(s_{i,t+1}) \mid \eta_t(s_{i,t}); \vartheta\right) f\left(m_{t+1} \mid s_{i,t+1}, m_t; \vartheta\right) f(\Delta y_{t+1} \mid s_{i,t+1}, s_{i,t}; \theta_s), \quad (20)$$

and normalize the weights such that $\sum_{i=1}^{N^s} w_{i,t} = 1$. The updated moments of s_{t+1} given $(p_{t+1}, m_{t+1}, \Delta y_{t+1})$ follow as $\bar{s}^L_{t+1|t+1} = \frac{1}{N^s}\sum_{i=1}^{N^s} w^*_{i,t+1}\bar{s}^L_{i,t+1}$. The predictive likelihood of $(p_{t+1}, m_{t+1}, \Delta y_{t+1})$ is approximated by the simulated average

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

Before proceeding to the the next period, check whether the Effective Sample size $ESS_{t+1} = 1/(\sum_{i=1}^{N^s} w^2_{i,t+1})$ falls below a specified threshold to reduce the risk of particle degeneracy. If so, re-sample the draws $s_{i,t+1}$ according to a multinomial distribution with probabilities $w_{i,t+1}$, and set their weights equal to $1/N^s$.

Alternatively, one could draw from an auxiliary transition density $\omega(s_{t+1} \mid s_t)$ that is easy to simulate from, and multiply the weights (20) by the importance sampling factors $f(s_{i,t+1} \mid s_{i,t}; \theta_s)/\omega(s_{t+1} \mid s_t)$ before normalizing. The auxiliary densities can be chosen to improve the efficiency of the simulations. In particular, the variance of the importance sampling factors is minimized by choosing the updated density $\omega_{t+1}(s_{t+1} \mid s_t) = f(s_{t+1} \mid s_t, p_{t+1}, m_{t+1}, \Delta y_{t+1})$ given next period's observations (e.g. Doucet and Johansen, 2009). For nonlinear models the latter density is typically not available in closed form, but can be approximated as Gaussian using the Unscented Kalman Filter (e.g. Fulop et al., 2021).

The smoothing distribution can be approximated similarly 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 last period's filtered distribution $f(s_T \mid \mathcal{F}_T)$. In particular, the smoothed sampling weights $(w^*_{i,t})_{i=1}^{N^s}$ are recursively computed from the filtered weights $(w_{i,t})_{i=1}^{N^s}$ as

$$w^*_{i,t} = w_{i,t} \sum_{j=1}^{N^s} \frac{f(s_{j,t+1} \mid s_{i,t}) w^*_{j,t+1}}{\sum_{k=1}^{N^s} f(s_{j,t+1} \mid s_{k,t}) w_{k,t}}.$$

Finally, the pairwise smoothed state densities, required by the Expectation-Maximization algorithm, can be computed from the marginal filtering and smoothing distributions using

$$\begin{aligned} f(s_{t+1}, s_t \mid \mathcal{F}_T) &= f(s_{t+1} \mid \mathcal{F}_T) f(s_t \mid s_{t+1}, \mathcal{F}_t) \\ &= \frac{f(s_{t+1} \mid \mathcal{F}_T) f(s_{t+1} \mid s_t)}{f(s_{t+1} \mid \mathcal{F}_t)} f(s_t \mid \mathcal{F}_t). \end{aligned}$$

The particle filter approximates this as

$$f(s_{i,t+1}, s_{a(i),t} \mid \mathcal{F}_T) = w_{a(i),t} \frac{w^*_{i,t+1} f(s_{i,t+1} \mid s_{a(i),t})}{\sum_{k=1}^{N^s} f(s_{i,t+1} \mid s_{k,t}) w_{k,t}},$$

where $a(i)$ is the index of the 'ancestor' of the i-th particle of the next period.

3.5 Expectation-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 estimated by the method-of-moments based on the conditional moments of the states, in line with the identification argument in Section 3.2. This motivates the following variant of the Expectation-Maximization (EM) algorithm, in which the M-step of optimizing the expected log-likelihood given the observations reduces to a linear regression. Related iterative algorithms in which the M-step is performed analytically are available for linear Gaussian models (Watson and Engle, 1983), finite mixture models (Arcidiacono and Jones, 2003), and conditional quantile models (Arellano et al., 2017).

E-step. Let $\vartheta = (\theta_s, \vartheta_m, \vartheta_p)$, where $\vartheta_m = (\psi_L, R_m, \Sigma_\varepsilon)$ and $\vartheta_p = (\pi_L, \alpha, R_p, \Sigma_\omega)$ combine the polynomial coefficients and finite-dimensional parameters of the error distributions for the measurement and prices, respectively. The Expectation-step involves computing the expected augmented-state log-likelihood over ϑ given some initial values $\tilde{\vartheta}$, defined as

$$\begin{aligned} Q(\vartheta, \tilde{\vartheta}) &= E_{\tilde{\vartheta}} \left(\log f(p_{2:T}, m_{2:T}, \Delta y_{2:T}, s_{1:T}; \vartheta) \mid \mathcal{F}_T^{p,m,y} \right) \\ &\equiv Q_p(\vartheta_p, \tilde{\vartheta}) + Q_m(\vartheta_m, \tilde{\vartheta}) + Q_y(\theta_s, \tilde{\vartheta}), \end{aligned}$$

where the price, measurement, and growth components equal

$$Q_p(\vartheta_p, \tilde{\vartheta}) = E_{\tilde{\vartheta}}\left(\sum_{t=2}^{T} \log f_\omega\left(\eta_t(s_t) - R_p\eta_{t-1}(s_{t-1})\right) \mid \mathcal{F}_T^{p,m,y}\right)$$

$$Q_m(\vartheta_m, \tilde{\vartheta}) = E_{\tilde{\vartheta}}\left(\sum_{t=2}^{T} \log f_\varepsilon\left(m_t - \psi_L(s_t) - R_m m_{t-1}\right) \mid \mathcal{F}_T^{p,m,y}\right)$$

$$Q_y(\theta_s, \tilde{\vartheta}) = E_{\tilde{\vartheta}}\left(\log f(s_1;\theta_s) + \sum_{t=2}^{T} \log f(\mathcal{S}_t \mid s_{t-1};\theta_s) \mid \mathcal{F}_T^{p,m,y}\right),$$

where $\eta_t(s_t) = p_t - \pi_L(s_t) - \alpha^T m_t$ are the implied pricing errors given parameters (π_L, α).[5] The expectations under the smoothing distribution $f_{\tilde{\vartheta}}(s_{1:T} \mid \mathcal{F}_T^{p,m,y})$ can be approximated as weighted averages using the simulated particles $(w_{it}^*, s_{it})_{i,t}$.

M-step. The three-way decomposition indicates that $Q(\vartheta, \tilde{\vartheta})$ can be maximized component-wise using corresponding subsets of parameters. Let $\tilde{\bar{s}}_{t|T}^L = E_{\tilde{\vartheta}}(\bar{s}_t^L \mid \mathcal{F}_T^{p,m,y})$ and $\tilde{V}_{t|T}^L = \mathrm{Var}_{\tilde{\vartheta}}(\bar{s}_t^L \mid \mathcal{F}_T^{p,m,y})$ denote the smoothed means and variances, respectively, of the polynomials $\bar{s}_t^L$ given the initial parameter values. For Gaussian measurement errors ε_t, $Q_m(\cdot)$ is maximized over c_L and R_m as

$$\begin{pmatrix}\hat{c}_L & \hat{R}_m\end{pmatrix}' = \begin{pmatrix}\sum_{t=2}^T \tilde{\bar{s}}_{t|T}^L \tilde{\bar{s}}_{t|T}^{L\,\prime} + \tilde{V}_{t|T}^L & \sum_{t=2}^T \tilde{\bar{s}}_{t|T}^L m_{t-1} \\ \sum_{t=2}^T m_{t-1}' \tilde{\bar{s}}_{t|T}^{L\,\prime} & \sum_{t=2}^T m_{t-1}' m_{t-1}\end{pmatrix}^{-1} \begin{pmatrix}\sum_{t=2}^T \tilde{\bar{s}}_{t|T}^L m_t \\ \sum_{t=2}^T m_{t-1} m_t\end{pmatrix}.$$

For Gaussian pricing error innovations ω_t, $Q_p(\cdot)$ can be maximized over b_L and α given the initial serial correlation parameters $\tilde{R}_p$ as

$$\begin{pmatrix}\hat{b}_L & \hat{\alpha}\end{pmatrix}' = \begin{pmatrix}\sum_{t=2}^T \tilde{\bar{s}}_{t|T}^L \tilde{\bar{s}}_{t|T}^{L\,\prime} + \tilde{V}_{t|T}^L & \sum_{t=2}^T \tilde{\bar{s}}_{t|T}^L m_t \\ \sum_{t=2}^T m_t' \tilde{\bar{s}}_{t|T}^{L\,\prime} & \sum_{t=2}^T m_t' m_t\end{pmatrix}^{-1} \begin{pmatrix}\sum_{t=2}^T \tilde{\bar{s}}_{t|T}^L \tilde{p}_t \\ \sum_{t=2}^T m_t \tilde{p}_t\end{pmatrix}$$

in terms of the prices $\tilde{p}_t = p_t - \tilde{R}_p\tilde{\eta}_{t-1}$ adjusted for serial correlation, where $\tilde{\eta}_t = p_t - \tilde{b}_L' \tilde{\bar{s}}_{t|T}^L - \tilde{\alpha}' m_t$ are the lagged pricing errors given the initial parameters. The transition density parameter estimate $\hat{\theta}_s$ can be found using gradient-descent methods, as the simulated $Q(\theta_s, \tilde{\vartheta})$ is continuous in θ_s as long as the transition density is.

[5] The first observed growth realization is $\Delta y_2 = y_2 - y_1$. The augmented-state likelihoods of the first prices $f_\eta(\eta_1(s_1))$ and measurements $f(m_1, s_1)$ could be taken into account. The former is unconditionally Gaussian, while the latter can be approximated by simulation.

The error covariance matrices can be consistently estimated by sample averages, avoiding numerical optimization over its parameters. In particular, the covariance matrix of the errors Σ_ε can be consistently estimated as

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

using the orthogonality conditions $\varepsilon_t \perp (s_t, m_{t-1})$ and $m_t \perp \tilde{\bar{s}}_{t|T}^L - \bar{s}_t^L$ by definition of prediction error. Similarly, the auto-covariance matrices of the pricing errors $\Sigma_\eta(j) = \mathrm{Var}(\eta_t, \eta_{t-j})$ can be consistently estimated as

$$\hat{\Sigma}_\eta(j) = \frac{1}{T}\sum_{t=1}^{T} (p_t - \hat{\alpha}' m_t)\left(p_t - \hat{\alpha}' m_t - b_L' \tilde{\bar{s}}_{t|T}^L\right)$$

using the orthogonality conditions $\eta_t \perp (s_t, m_t)$ and $p_t \perp \tilde{\bar{s}}_{t|T}^L - \bar{s}_t^L$. These imply the first-order auto-regression coefficient matrix $R_p = \Sigma_\eta(1)\Sigma_\eta(0)^{-1}$ and innovation covariance matrix $\Sigma_\omega = \Sigma_\eta(0) - R_p \Sigma_\eta(0) R_p'$.

We repeat the E- and M-steps until convergence to a local optimum. Afterwards, we perform parameter inference based on the scores of $Q(\vartheta, \tilde{\vartheta})$ around the local optimum $\vartheta = \hat{\vartheta}$, where they equal the scores of the log-likelihood $\ell_T(\vartheta)$. While the former is continuous in the parameters given the simulated state distribution, simulating the latter may result in discontinuities.

3.6 Consistency

The population parameters of interest are given by

$$(\theta_0, h_0) = \arg\max_{\theta\in\Theta, h\in\mathcal{H}} \lim_{T\to\infty} \ell_T(\theta, h), \tag{21}$$

and the maximum likelihood estimator by

$$(\hat{\theta}, \hat{h}) = \arg\max_{\theta\in\Theta, h\in\mathcal{H}} \ell_T(\theta, h), \tag{22}$$

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 $(\psi_m)_{m=1}^K$ 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

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

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

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

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

$$(\hat{\theta}, \hat{h}_L) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}_{L(T)}} \ell_T(\theta, h), \tag{23}$$

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

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

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 $\varTheta = \Theta \times \mathcal{H}$ 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

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

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

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

Theorem 1. *Under Assumptions 3, the maximizer $(\hat{\theta}, \hat{h}_L)$ of (23) 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 $L^{D+1}/T \to 0$.

3.7 Conditional method of moments estimation of ϕ

The recursive pricing equation (7) 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 method-of-moments procedure for estimating ϕ based on the filtered moments of s_t given the observations. This estimation method could therefore be performed within the M-step of the EM-algorithm, or 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 (5). When the stochastic discount function is approximated by the polynomial $\phi_L(s_t) = e_L' \bar{s}_t^L$, its projection on the conditioning information is a polynomial in the filtered moments $\hat{\bar{s}}_{t|t}^L = E_{\hat{\vartheta}}(\bar{s}_t^L \mid \mathcal{F}_t^{m,y,p})$ of the states:

$$E\left(\phi_L(s_t) \mid \mathcal{F}_t^{m,y,p}\right) = e_L' \hat{\bar{s}}_{t|t}^L.$$

By the Law of iterated expectations, the Euler equation (5) therefore gives rise to the following conditional moment restrictions stated in terms of the filtered moments:

$$E\left(\frac{1}{\beta} e_L' \hat{\bar{s}}_{t|t}^L - \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} e_L' \hat{\bar{s}}_{t+1|t+1}^L R_{t+1} \mid (m_t, p_t, \hat{\bar{s}}_{t|t}^L)\right) = 0.$$

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, s_t). When the filtered state density $f(s_t \mid \mathcal{F}_t^{m,y,p})$ can be summarized in terms of its L moments, there is no loss of information from conditioning down on $(m_t, p_t, \hat{\bar{s}}_{t|t}^L)$. The resulting unconditional moments can be represented as an eigenproblem in the coefficient vector e_L:

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

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 (6) 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 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. Dividends per share are computed from the difference in value-weighted returns with (R_{t+1}^d) and without (R_{t+1}^x) dividends:

$$\frac{D_{t+1}}{P_t} = R_{t+1}^d - R_{t+1}^x.$$

These dividends are then aggregated at the annual frequency to diminish seasonal effects. Price-dividend ratios are computed as

$$\frac{P_{t+1}}{D_{t+1}} = \frac{R_{t+1}^x}{R_{t+1}^d - R_{t+1}^x}.$$

The 3-month U.S. Treasury Bill rate is used to measure the risk-free rate. Stock prices, dividends, and the risk-free rate are expressed in real terms using the price index for U.S.

gross domestic product.

Let ip_t denote the log observed industrial production in month t, and let its increment be $\Delta ip_t = ip_t - ip_{t-1}$. Define the Realized Economic Variance (REV) as a quarterly proxy for the variance of output growth computed as

$$REV_t = \sum_{m=1}^{3} (\Delta ip_{t+1-m} - \overline{\Delta ip}_t)^2,$$

with $\overline{\Delta ip}_t$ the rolling window quarterly mean. The Realized stock market Variance (RV) is similarly constructed from daily log returns R_{t+1} after centering with the quarterly mean $\overline{R}_t$ as[6]

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

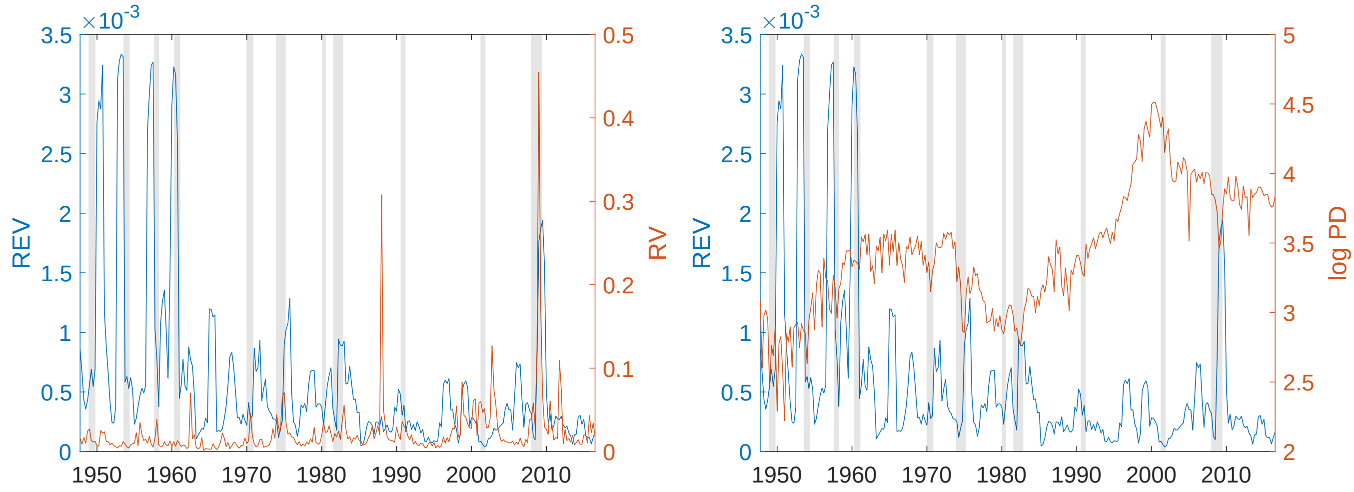
where n_t is the number of trading days in quarter t.

4.2 Economic Uncertainty and Stock Market Volatility

Figure 1a) shows the quarterly measures of economic (REV) and financial market variation (RV). Most episodes of high economic volatility occurred during the first half of the sample ending in the early 1980s, while most episodes of high financial market turbulence occurred thereafter. In particular, the 1950s saw substantial economic uncertainty but historically calm financial markets. On the other hand, stock market volatility spiked during the 1987 Black Monday crash and in the aftermath of the 1998 LTCM collapse, while industrial production growth remained largely unchanged. Moreover, while spikes in economic volatility tend to happen during recessions, jumps in financial markets have regularly occurred during expansions. The notable exception to these patterns was the 2008 financial crisis and subsequent Great Recession, during which both measures peaked. To understand how economic uncertainty affects asset prices, the right panel of Figure 1b) compares REV with the log price-dividend ratio. The two series display a clear inverse relation, with the 1950s seeing peak economic uncertainty and record low valuation ratios, while the peak valuations of the dot-com bubble coincided with minimal economic uncertainty. However, economic uncertainty cannot explain the decline in valuations leading

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

up to the early 1980s recession or their rise during the expansionary 1990s, suggesting economic growth itself may be needed to explain asset prices.



(a) Economic and financial market variance.

(b) Economic uncertainty and price-dividend ratio.

Figure 1: Realized Variance of Industrial Production growth versus (a) Realized Variance of S&P 500 returns and (b) log S&P 500 Price-Dividend ratio, quarterly data from 1947Q3-2016Q2.

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. 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 decreases contemporaneously with 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, which could be explained by reduced corporate investment. Figure 3 shows responses to changes in financial market volatility are in the same direction as for changes in economic volatility, but that stock market variance correlates stronger with dividend growth and market returns 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.

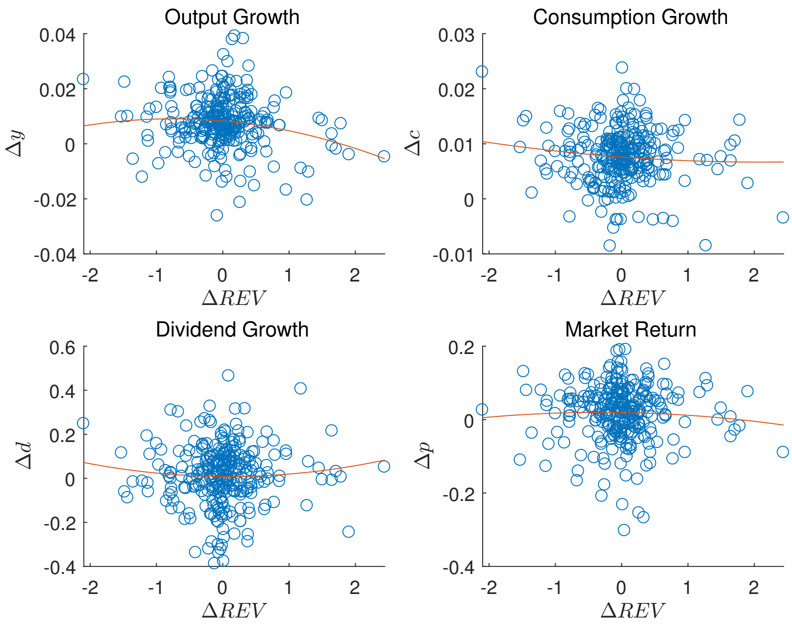


Figure 2: Quarterly changes in log Realized Economic Variance (REV) of Industrial Production growth versus log return on Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016. Fitted line shows the quadratic least squares fit.

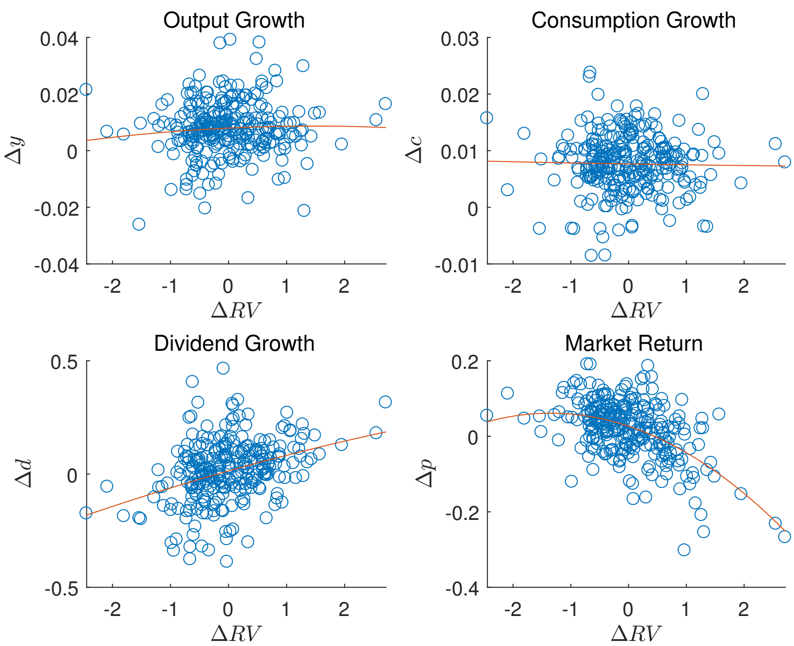


Figure 3: Quarterly changes in log Realized Variance (RV) versus log return Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016. Fitted line shows the quadratic least squares fit.

4.3 Estimates

Table 1 reports the simulated maximum likelihood estimates of the transition density parameters θ_s for the long-run risk model (1) with autoregressive Gamma stochastic volatility. The estimates are for three datasets created by consecutively adding the measurements $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$ and prices $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ to the growth observations, using the partially linear specification (15) with $L = 4$ order polynomials (ψ_L, π_L). The annual dividend-output cointegration parameter is estimated by ordinary least squares as $\hat{\lambda} = 1.65$. When m_t and p_t are used for estimation, the proxies $\log REV_t$ and $\log RV_t$, respectively, are added to the measurement equation using the polynomial formulation $V_{t+1} = \psi_L^v(s_t) + \eta_{t+1}^v$, where η_t^v are serially correlated errors.

Differences in parameter estimates across datasets are generally less than two standard errors, suggesting no major misspecification of the measurement and pricing equations. Moreover, standard errors fall when adding measurements and prices, confirming their state-dependence helps to identify the transition density parameters. The values of the mean reversion parameters (ρ_x, ν) in the range 0.95-0.98 correspond to half-lives of the expected growth and variance components between 4 to 8 years, suggesting both are highly persistent.

Figure 4 shows the smoothed means of the persistent growth and variance components using the simulated filtering and smoothing algorithm, for the three subsets of observations. The addition of both the measurements and prices leads to more volatile paths for the smoothed means with narrower confidence intervals, thus yielding more precise estimates. All three datasets display sudden drops in the smoothed growth component during recessions, with varying degrees of severity. Moreover, all three datasets show clusters of high output growth volatility around episodes such as the post-war years, the 1980s energy crisis, and to a lesser extent the 2008 financial crisis. The frequency and scale of high volatility periods has been steadily declining over the sampling period, reaching its lows during the late 1980s and the 1990s.

Figure 5a) shows the estimated response functions $\psi^c(s_t)$ and $\psi^d(s_t)$ of the log consumption-output and dividend-output cointegration residuals based on the simulation maximum likelihood of $(\Delta y_t, m_t)$. The reported estimates are for the specification $m_t = R_m m_{t-1} + \psi_L(x_t, \sigma_t^2) + a_y \Delta y_t + \varepsilon_t$, thus allowing a contemporaneous effect of Δy_t on the cointegration

Table 1: Simulated maximum likelihood estimates of the transition density parameters θ_s for the long-run risk with autoregressive Gamma stochastic volatility model, based on quarterly observations from 1947 to 2016 on subsets of Δy_t, $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$ and $p_t = (\log \frac{P_t}{D_t}, r_t^f)$. Estimates and standard errors (in brackets) based on the EM algorithm, using $N^s = 5,000$ simulated particles and $L = 4$.

(a) Based on (Δy_t) only.

μ	ρ_x	ϕ_x	ϕ_v	$\bar{\sigma}$	ν
0.008	0.952	0.160	1.011	0.007	0.958
(0.000)	(0.051)	(0.039)	(0.006)	(0.002)	(0.003)

(b) Based on $(\Delta y_t, m_t)$.

μ	ρ_x	ϕ_x	ϕ_v	$\bar{\sigma}$	ν
0.007	0.969	0.133	1.011	0.008	0.979
(0.000)	(0.032)	(0.010)	(0.002)	(0.001)	(0.002)

(c) Based on $(\Delta y_t, m_t, p_t)$.

μ	ρ_x	ϕ_x	ϕ_v	$\bar{\sigma}$	ν
0.008	0.972	0.131	1.011	0.007	0.974
(0.000)	(0.015)	(0.009)	(0.002)	(0.001)	(0.002)

residuals. The consumption-output residual is moderately increasing in expected growth, but does not appear strongly related to growth volatility, in line with the scatterplots in Figure 2. Meanwhile, the dividend-output residual increases in volatility levels above the median, yet is U-shaped in volatility when x_t is low, such as during recessions. Figure 5b) shows the responses are qualitatively similar when p_t is included in the simulated maximum likelihood estimation and the latent states are more accurately estimated.

To relate the response functions to specific episodes, Figure 6 plots the growth rates in consumption and dividends relative to output against their conditional means under the smoothing distribution of the state variables. The consumption share of output tends to peak during recessions, as consumption does not immediately respond to output declines. However, controlling for this contemporaneous 'denominator' effect, the state-dependent component $\hat{\psi}^c(\cdot)$ actually falls during the low-growth, high volatility early 1980s and Great Recession, in order to rise again during their recoveries. Dividends help up relatively well throughout the economically uncertain late 1950s and early 1980s, explaining why $\hat{\psi}^d(\cdot)$ is positive for large σ_t^2 and all levels of x_t. Meanwhile, the U-shape of $\hat{\psi}^d(\cdot)$ in σ_t^2 for low levels of x_t rationalizes dividends during the below average growth 2000s, which suddenly increased around 2005 when economic volatility dropped, yet fell continuously when economic uncertainty returned during the Great Recession.

Table 2 shows the estimated parameters $(\hat{R}_m, \hat{a}_y, \hat{\Sigma}_\varepsilon)$ of the measurement equation. It includes regression estimates of the model with $\psi^m(\cdot)$ constant as a benchmark to quantitatively assess the role of state-dependence. Both series are highly persistent, but not strongly mutually correlated. The consumption-output residual has more volatile innovations than the annually aggregated dividend-output residual. Moreover, a larger fraction of the dividend-output residual innovations is estimated to be due to the latent states, regardless of whether asset prices are used to estimate them. However, the consumption-output residual depends more strongly on contemporaneous output growth. The parameter estimates indicate that consumption growth is expected to increase by around 0.6 per unit of output growth, keeping the mean and variance of next period's growth constant.

Table 2: Estimated autoregression matrix $\hat{R}_m$, growth regression coefficient $\hat{a}_y$, and error covariance matrix $\hat{\Sigma}_\varepsilon$ of the quarterly measurements $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$ based on (a) least squares regression of m_t on $(m_{t-1}, \Delta y_t)$, and (b-c) SML estimates with two sets of observations. Measurements are centered and standardized to unit variance.

(a) Initial estimates.

$\hat{R}_m$	0.957	-0.023
	0.014	1.003
$\hat{a}_y$	-0.361	-0.025
$\hat{\Sigma}_\varepsilon$	0.042	–
	-0.001	0.027

(b) Based on $(\Delta y_t, m_t)$.

$\hat{R}_m$	0.949	-0.027
	0.032	0.846
$\hat{a}_y$	-0.403	-0.034
$\hat{\Sigma}_\varepsilon$	0.021	–
	0.002	0.007

(c) Based on $(\Delta y_t, m_t, p_t)$.

$\hat{R}_m$	0.960	-0.031
	0.036	0.846
$\hat{a}_y$	-0.411	-0.032
$\hat{\Sigma}_\varepsilon$	0.028	–
	0.002	0.010

Figure 7 shows the estimated pricing functions $\hat{\pi}_L$ and variance function $\hat{\psi}^v_L$. The estimated price-dividend ratio and risk-free rate functions in $\hat{\pi}_L$ are controlled for the consumption and dividend cointegration residuals m_t. The price-dividend function appears to increase monotonically in expected growth, and to decrease monotonically in the variance. Both effects interact on the upside such that combining high expected growth and low volatility lifts the expected price-dividend ratio to at least one standard deviations above its mean. In contrast, when expected growth is low, volatility dropping below its median is not expected to lift the price-dividend ratio. Intuitively, during recessions investors may actually prefer some volatility for the growth rate to revert to its mean faster. The risk-free rate function increases fairly linearly with expected growth for low and moderate levels of uncertainty. However, it also peaks when low expected growth combines with high uncertainty, which may be explained by contractionary monetary policy when the latter is due to inflationary spirals. Finally, the estimated realized return variance function reveals its nonlinear relation to the variance of economic growth. It appears to trade-off the effect of more volatile economic shocks against higher sensitivity to shocks when the price-dividend ratio is high due to low uncertainty. High valuation ratios can explain the highly volatile returns during 'good' times when growth is high and uncertainty is low. Overall, return volatility peaks during recessions with moderate rather than high levels of uncertainty, when there are still downside risks to valuations.

Figure 8 shows the time series fit of the prices and realized variances. The model is able to explain drops in the price-dividend ratio during highly uncertain recessions such as the early 1980s and the Great Recession. However, the sustained rise in valuation from the mid 1980s until the dot-com bubble cannot be explained by the smoothed state variables, which remained relatively stable during this period. Instead, the estimated undervaluation during the 1980s coincides with higher than expected risk-free rates, whereas the dot-com bubble is treated as a pure pricing error unrelated to fundamentals. The model matches the high risk-free rates during the early 1980s contractionary monetary policy as well as the low rates since the Great Recession. The pricing error parameter estimates in Table 3 confirm the negative relation between the unexplained components of the price-dividend ratio and risk-free rates, as well as their much reduced scale after allowing for state-dependence. Finally, the relatively low financial market volatility during the economically uncertain 1950s can be explained by the low sensitivity of the price-dividend ratio when growth volatility is high. In contrast, valuations were more sensitive to shocks during the moderate levels of economic uncertainty around the Great Recession. Meanwhile, short-lived bumps in return volatility outside recessions such as the 1987 Black Monday crash are treated as pure variance shocks.

Figure 9 shows the estimated stochastic discount factor residual function $\hat{\phi}_L$, and the profiled GMM criterion in terms of the parameters (β, γ). 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 volatility is high, and its lowest values in the opposite cases. Moreover, it remains elevated at low volatility levels for low expected growth. These findings indicate that state-dependence in consumption and dividends can-

Table 3: Estimated regression coefficients $\hat{\alpha}$ of the prices on measurements, and autocovariance matrices $\hat{\Sigma}_\eta(j)$ of order $j = 0, 1$ of the pricing error η_t, based on (a) least squares regression of $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ on $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$, and (b) SML estimates based on $(\Delta y_t, m_t, p_t)$. Prices and measurements are centered and standardized to unit variance.

(a) Initial estimates.

$\hat{\alpha}$	-0.120	-0.138
	0.207	-0.006
$\hat{\Sigma}_\eta(0)$	0.974	–
	-0.033	0.957
$\hat{\Sigma}_\eta(1)$	0.953	–
	-0.043	0.762

(b) SML estimates.

$\hat{\alpha}$	0.104	0.036
	0.227	0.036
$\hat{\Sigma}_\eta(0)$	0.675	–
	-0.074	0.296
$\hat{\Sigma}_\eta(1)$	0.647	–
	-0.078	0.183

not fully explain the state-dependence of the price-dividend ratio. Instead, our estimates suggests rationalizing the latter through the discount factor, giving higher marginal utility for payoffs during low growth and/or uncertain times. In particular, it supports stochastic discount factor models which induce a negative risk premium for output growth volatility. The profiled GMM-criterion is minimized by the discount and risk aversion parameters $(\hat{\beta}, \hat{\gamma}) = (0.9975, 7.25)$, implying an annual discount rate of around 1%. These moderate values suggest that power utility over consumption may not be unreasonable once additional state-dependent discounting is allowed for. However, identification of the risk aversion parameter is relatively weak, as values in the range $5 < \gamma < 9$ yield qualitatively similar stochastic discount functions.

5 Conclusion

This paper develops a class of nonlinear Markovian asset pricing models in which the dynamics of consumption and dividend relative to output are described via general functions of latent state variables describing persistent components in the aggregate growth distribution. We study the identification and estimation of a semiparametric specification of the stochastic discount factor by formulating the Euler equation as an eigenfunction problem. We establish the consistency of a sieve maximum likelihood estimator for the unknown functions under high-level assumptions. For affine state variables and polynomial approximations of the measurement and pricing functions, we derive closed-form expressions for expected growth and financial volatility, and a tractable simulation-based algorithm for filtering, smoothing, and parameter optimization. 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 can be rationalized by that of consumption and dividends. Instead, the evidence supports models in which investors have moderate relative risk aversion but higher marginal utility for payoffs in times of low expected growth and high volatility. Finally, the steeply declining price-dividend ratios for moderate levels of growth volatility help explain bursts of stock market volatility during periods of moderate economic uncertainty.

References

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.

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.

Bakshi, G. S. and Chen, Z. (1996). The spirit of capitalism and stock-market prices. *The American Economic Review*, pages 133–157.

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

Berger, D., Dew-Becker, I., and Giglio, S. (2020). Uncertainty shocks as second-moment news shocks. *The Review of Economic Studies*, 87(1):40–76.

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

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.

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., Chernozhukov, V., Lee, S., and Newey, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, 82(2):785–809.

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.

Christensen, T. M. (2017). Nonparametric stochastic discount factor decomposition. *Econometrica*, 85(5):1501–1536.

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.

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. (2020). Nonparametric euler equation identification and estimation. *Econometric Theory*, pages 1–41.

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

Fulop, A., Heng, J., Li, J., and Liu, H. (2021). Bayesian estimation of long-run risk models using sequential monte carlo. *Journal of Econometrics*.

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

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

Gourieroux, C. and Jasiak, J. (2006). Autoregressive gamma processes. *Journal of forecasting*, 25(2):129–152.

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.

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.

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

Malloy, C. J., Moskowitz, T. J., and Vissing-Jørgensen, A. (2009). Long-run stockholder consumption risk and asset returns. *The Journal of Finance*, 64(6):2427–2479.

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.

Pohl, W., Schmedders, K., and Wilms, O. (2018). Higher order effects in asset pricing models with long-run risks. *The Journal of Finance*, 73(3):1061–1111.

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

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.

Song, Z. and Xiu, D. (2016). A tale of two option markets: Pricing kernels and volatility risk. *Journal of Econometrics*, 190(1):176–196.

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

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

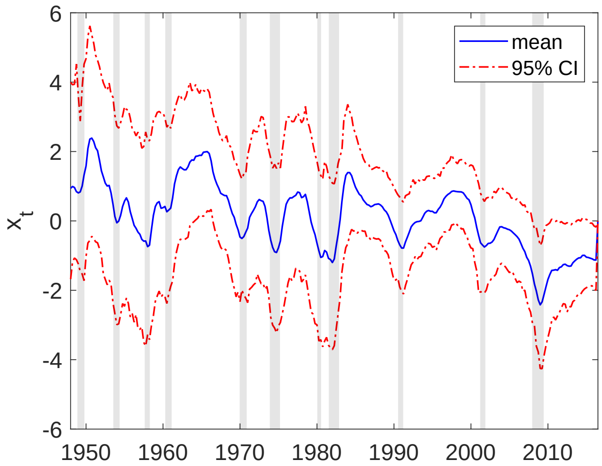
A Appendix

A.1 Proofs

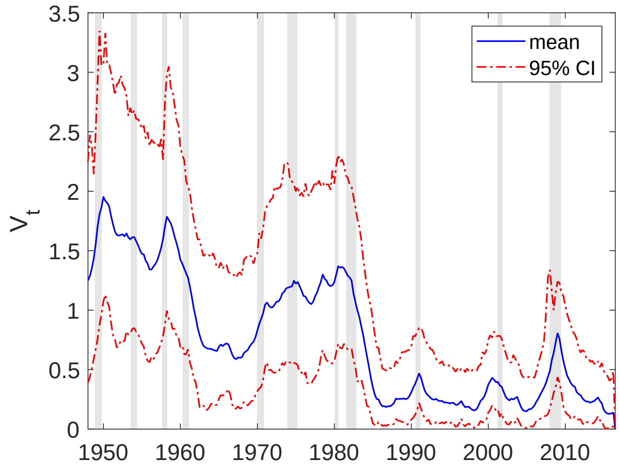
Proof of Theorem 1. The proof is based on Lemma A1 in Newey and Powell (2003). Let $Q_T(\theta) = \ell_T(\theta)$ and $Q(\theta) = E(l_t(\theta))$. This requires that (i) there is unique θ_0 that minimizes $Q_T(\theta)$ on Θ, (ii) Θ_T are compact subsets of Θ such that for any $\theta \in \Theta$ there 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 $\mathcal{H}$.

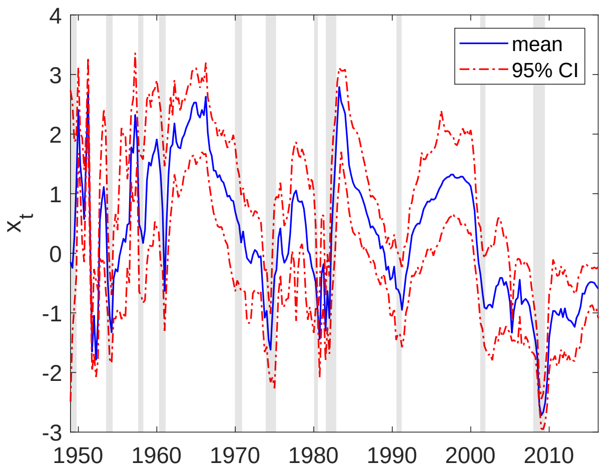
For (iii), continuity of $Q_T(\theta)$ follows from continuity of the policy and pricing functions and the transition density. The remaining conditions of continuity of $Q(\theta)$ and uniform convergence follow from Lemma A2 in Newey and Powell (2003). This requires pointwise convergence $Q_T(\theta) - Q(\theta) \xrightarrow{p} 0$ 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. $\square$



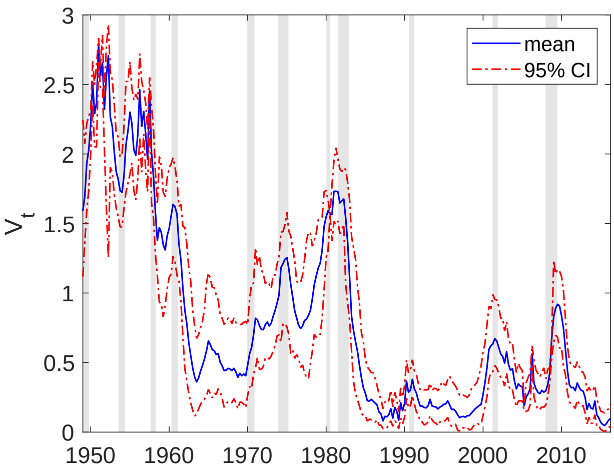
(a) Smoothed expected growth using (Δy_t).



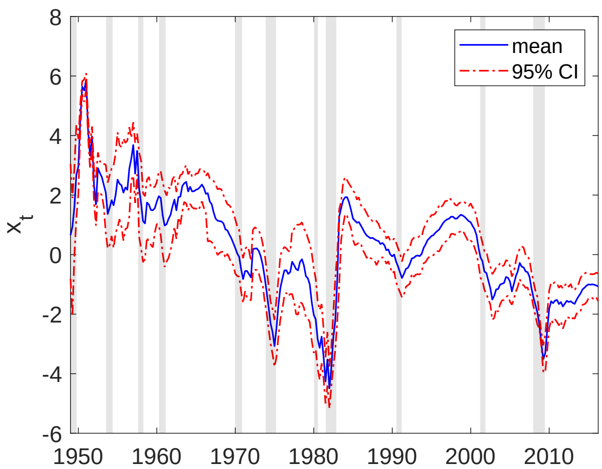
(b) Smoothed variances using (Δy_t).



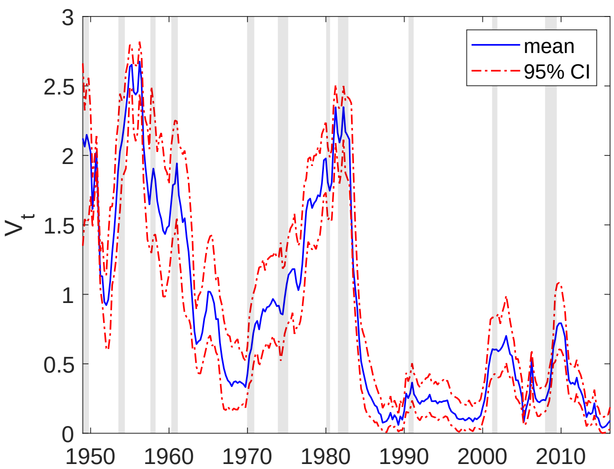
(c) Smoothed expected growth using $(\Delta y_t, m_t)$.



(d) Smoothed variances using $(\Delta y_t, m_t)$.

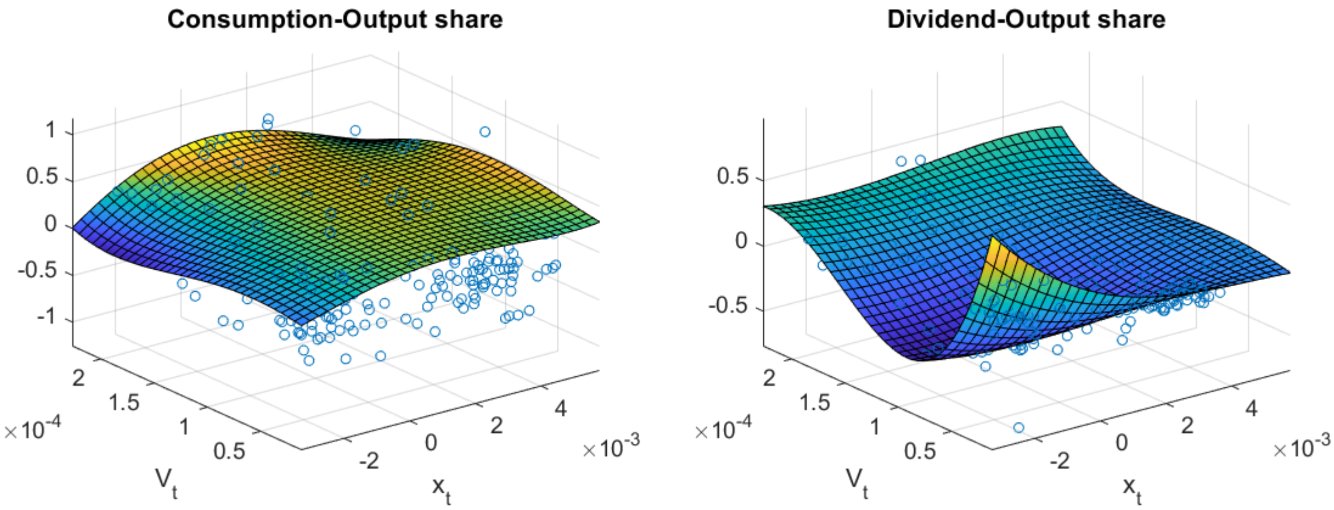


(e) Smoothed expected growth using $(\Delta y_t, m_t, p_t)$.

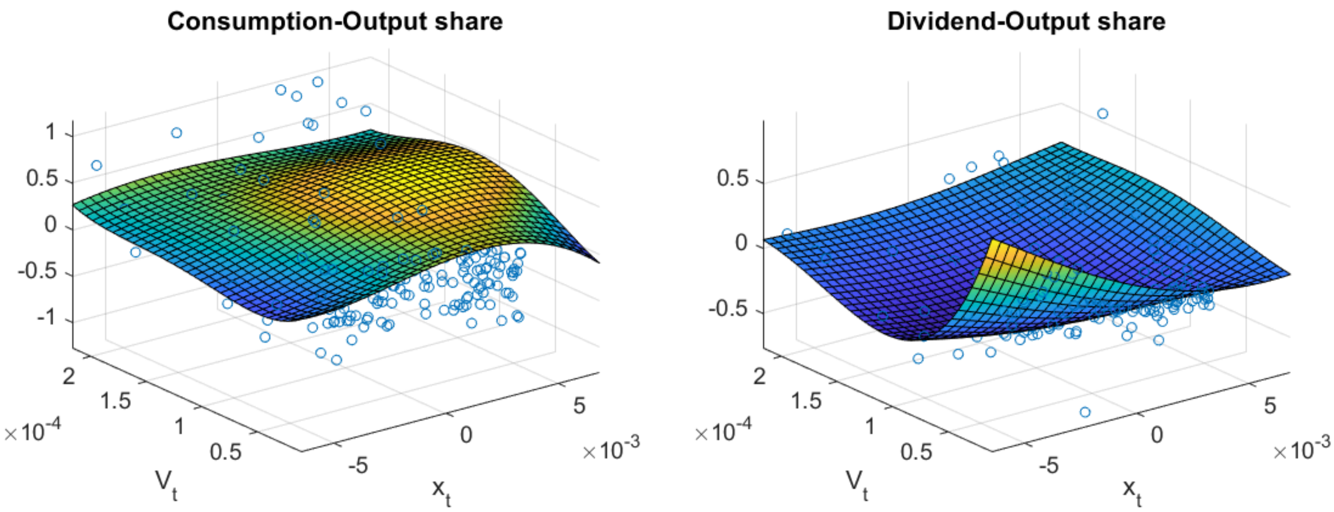


(f) Smoothed variances using $(\Delta y_t, m_t, p_t)$.

Figure 4: Smoothed conditional means x_t (left panels, in annualized percentages) and variances V_t (right panels) of the long run risk with autoregressive Gamma stochastic volatility model, based on quarterly observations of Δy_t, $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$ and $p_t = (\log \frac{P_t}{D_t}, r_t^f)$. Dashed lines represent 95% confidence intervals based on simulated particles. Grey shades indicate NBER recession dates.

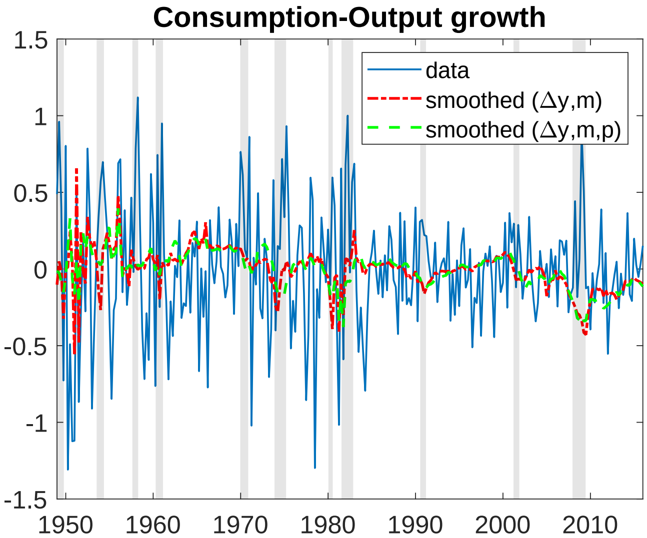


(a) Estimated response functions using $(\Delta y_t, m_t)$.

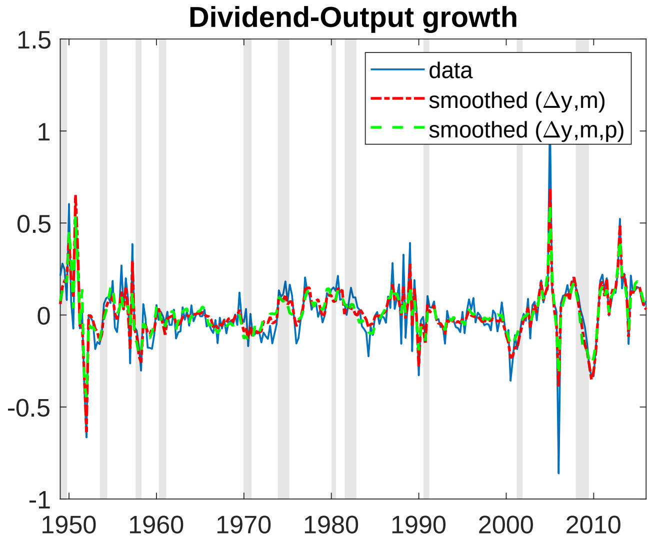


(b) Estimated response functions using $(\Delta y_t, m_t, p_t)$.

Figure 5: Estimated policy response functions ψ for log consumption (left panels) and dividend (right panels) relative to output as a function of the conditional mean x_t and variance V_t of output growth, using quarterly observations and a $L = 4$ order expansion. Blue circles plot the partial differences $m_t - \hat{R}_m m_{t-1}$ against the smoothed means of state variables. Vertical axis measures standard deviations $\sqrt{\mathrm{Var}(m_t)}$ from the mean.

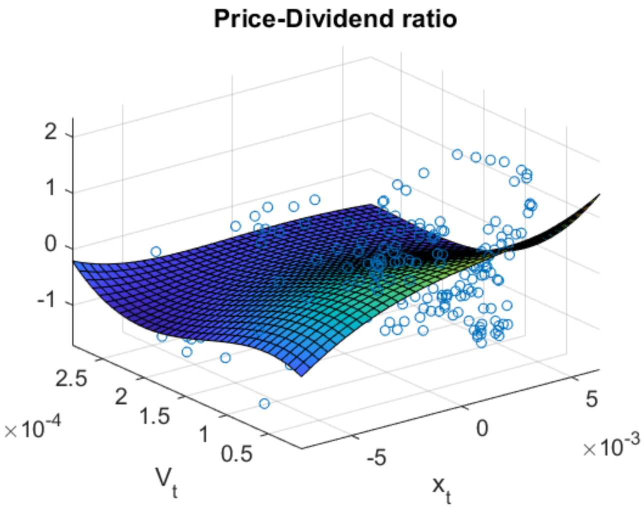


(a) $\Delta(c_t - y_t)$ against smoothed values.

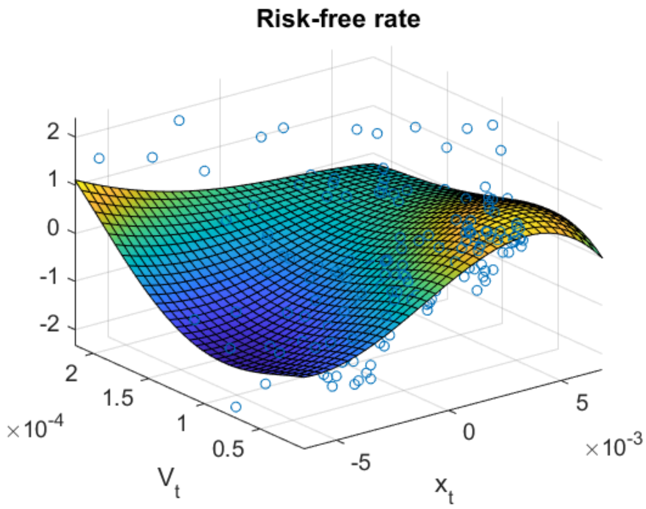


(b) $\Delta(d_t - \hat{\lambda} y_t)$ against smoothed values.

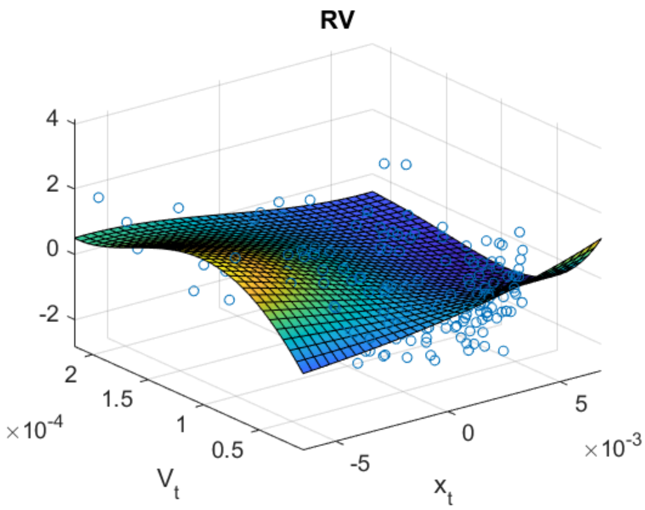
Figure 6: Growth rates in quarterly consumption (left) and annual dividends (right) relative to output, against smoothed values $(\hat{R}_m - I)m_{t-1} + \hat{c}_L' \hat{\hat{s}}_{t|T}^L$ according to (3), using $L = 4$ order joint conditional moments of state variables (x_t, σ_t^2) given quarterly observations of $(\Delta y_t, m_t)$ or $(\Delta y_t, m_t, p_t)$. Vertical axis shows standard deviations from the mean.



(a) Log price-dividend ratio function.

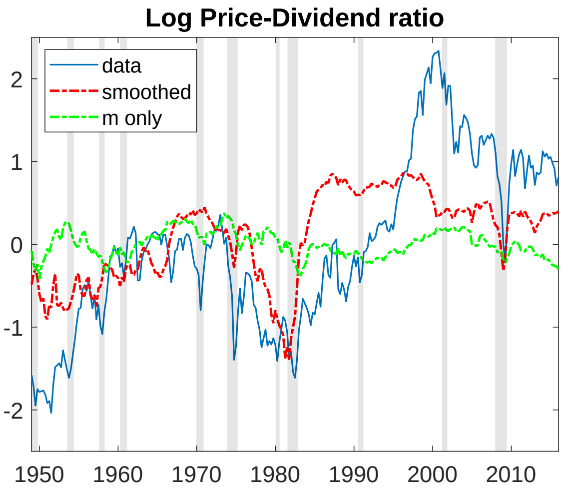


(b) Risk-free rate function.

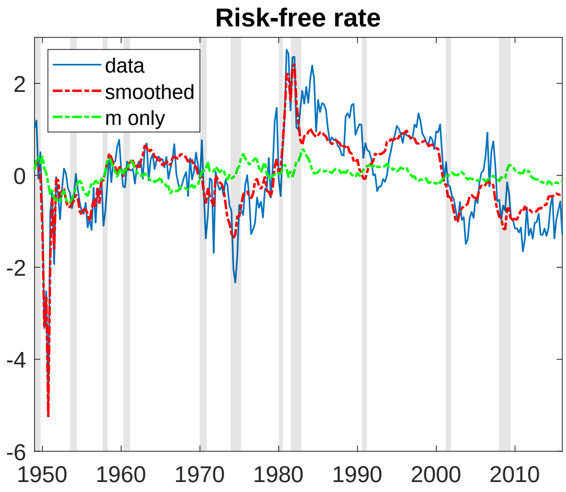


(c) Log realized variance.

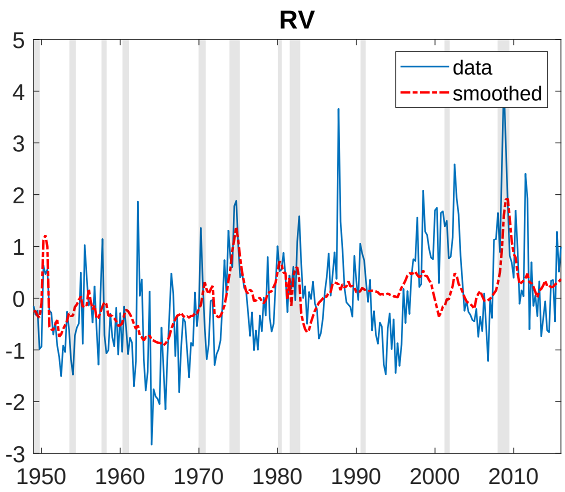
Figure 7: Simulated maximum likelihood estimates of pricing and variance functions as a function of expected growth x_t and variance V_t, using quarterly observations and $L = 4$ order expansion. Blue circles show smoothed means of state variables. Vertical axes show standard deviations from the mean.



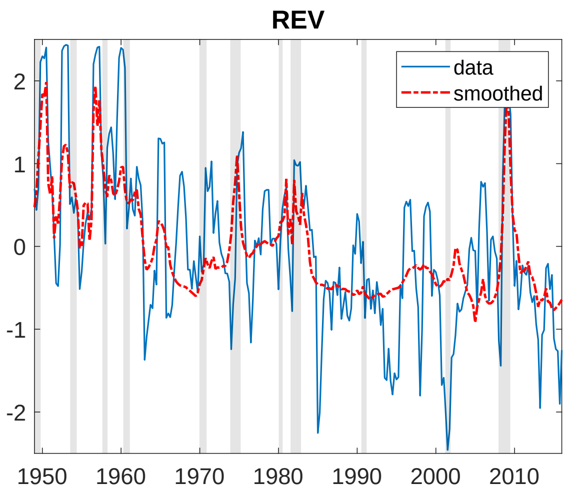
(a) Log price-dividend ratio.



(b) Risk-free rate.

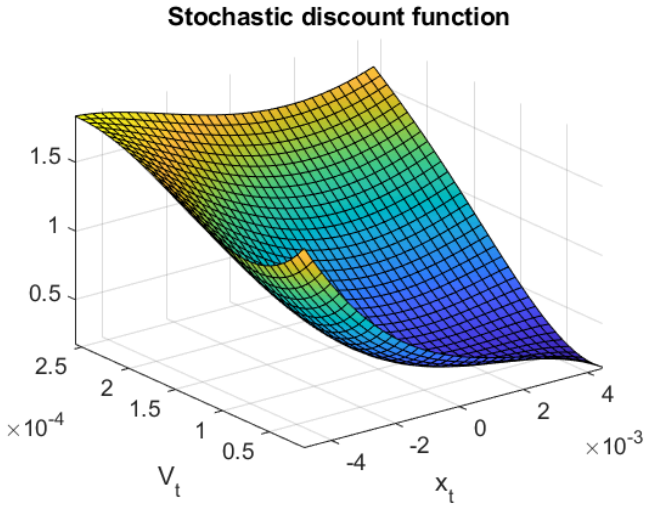


(c) Log realized return variance.

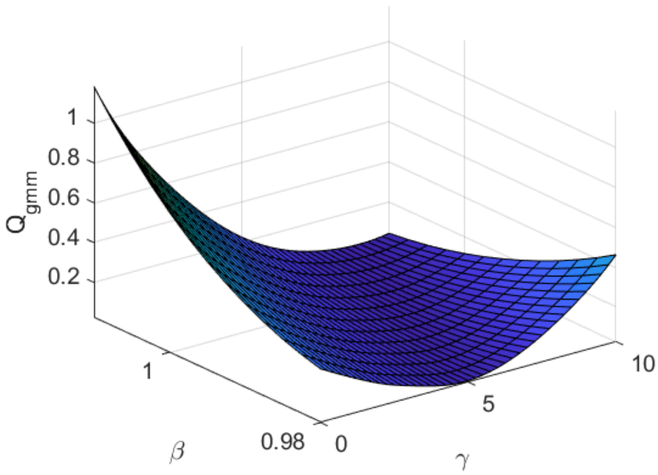


(d) Log realized economic variance.

Figure 8: Time series of prices and realized variance proxies against smoothed values $\hat{b}_L' \hat{\hat{s}}_{t|T}^L + \hat{\alpha}' m_t$ using $L = 4$ order joint conditional moments of state variables (x_t, σ_t^2) given quarterly observations of $(\Delta y_t, m_t, p_t)$. Plots with prices include the least squares fitted values $\tilde{p}_t = \tilde{\alpha}' m_t$ based on measurements only. Vertical axis shows standard deviations from the mean.



(a) Stochastic discount function.



(b) Profiled GMM criterion.

Figure 9: Stochastic discount function minimizing the GMM criterion in terms of the filtered states, and GMM-criterion as a function of the quarterly discount parameter β and risk aversion parameter γ, after profiling the stochastic discount function approximation $\hat{\phi}_L(\cdot;\beta,\gamma)$. The criterion is minimized at $(\hat{\beta},\hat{\gamma}) = (0.9975, 7.25)$.