Estimating Policy Functions Implicit in Asset Prices

Jeroen Dalderop*

April 2018

(Link to latest version)

Abstract

I propose a semiparametric asset pricing model to measure how consumption and dividend policies depend on unobserved state variables, such as economic uncertainty and risk aversion. Under a flexible specification of the stochastic discount factor, the state variables are recovered from cross-sections of asset prices and volatility proxies, and the shape of the policy functions is identified from the pricing functions. The model leads to closed-form price-dividend ratios under polynomial approximations of the unknown functions and affine state variable dynamics. In the empirical application uncertainty and risk aversion are separately identified from size-sorted stock portfolios exploiting the heterogeneous impact of uncertainty on dividend policies. I find an asymmetric and convex response in consumption ($-$) and dividend growth ($+$) towards uncertainty shocks, which together with moderate uncertainty aversion, can generate large leverage effects and divergence between macroeconomic and stock market volatility.

Keywords: Asset Prices, Economic Uncertainty, Risk Aversion, Stock Market Volatility, Nonlinear Panel Data, Latent Variables, Sieve Maximum Likelihood

JEL Codes: C14, G12

*Faculty of Economics, University of Cambridge, CB3 9DD, Cambridge, United Kingdom. Email address: jwpd2@cam.ac.uk. I wish to thank Oliver Linton for his guidance and support, as well as Yacine Aït-Sahalia, Debopam Bhattacharya, Xiaohong Chen, Hamish Low, Ralph Koijen, Alexei Onatski, Richard Smith, Jasmine Xiao, and seminar participants at Cambridge, Erasmus University, Federal Reserve Board of Governors, LSE, Notre Dame, Penn State, Princeton, Queen Mary, University of Amsterdam, and Wisconsin, for helpful comments. Financial support by CERF is gratefully acknowledged. Any errors are my own.

1 Introduction

Standard asset pricing models with moderately risk averse households have difficulty reconciling episodes of highly volatile asset prices with relatively smooth fluctuations in macroeconomic fundamentals. Over the last two decades, models in which fundamentals and preferences are jointly affected by unobserved state variables have made substantial progress in rationalizing the level and dynamics of asset price volatility.[1] Many of these models build on the assumption that variables such as consumption and dividend growth, as well as the marginal utility function of a representative agent, depend linearly on the state variables. This leads to highly tractable log-linear stochastic discount factors and valuation ratios. As a result, the variance of asset returns is proportional to that of the state variables. However, sudden drops in asset prices such as the 1987 Black Monday crash did not coincide with significantly increased production growth volatility, nor has the recent upswing in political and economic uncertainty triggered much movement on financial markets. Such observations suggest an important role for nonlinear specifications of the state-dependency in aggregate growth rates and preferences.

This paper investigates whether nonlinear dependence of consumption and dividend choice on state variables such as uncertainty and risk aversion helps to reconcile the dynamics of asset prices and economic fundamentals. In particular, I develop a general class of Markovian asset pricing models in which the preferred level of consumption and dividends per unit of output are functions of state variables that are observed by the household and firms, respectively, but unobserved by the econometrician. This unobservability makes it impossible to directly measure the shape of the consumption and dividend policy functions. However, given the dynamics of the state variables and their influence on the stochastic discount factor, the policy functions determine how asset prices are affected by the state variables. As a result, when the state variables can be uniquely recovered from asset prices, the policy functions are identified from the state variables' impact on asset prices.

The robustness of this approach depends on correct specification of the stochastic discount factor and the distribution of the state variables. To allow for general state-dependent preferences of the representative agent, the stochastic discount factor is constructed by multiplying standard power utility over consumption with an unspecified function of the state variables. In several structural models this additional component takes a specific functional form of the state variables, which is typically log-linear. Its shape is therefore of interest in itself in evaluating

[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).

commonly made linearity assumptions. The model dynamics are fully determined after specifying the distribution of the state variables. Since asset prices depend on the distribution of state variables at possibly any horizon, it is useful to provide some parametric structure to be able to extrapolate the transition density across maturities of interest. In particular, the dynamics of the state variables are specified as continuous time affine processes following Duffie et al. (2000). The resulting semiparametric framework generalizes affine equilibrium asset pricing models in which consumption and dividend growth are linear in the state variables; see Eraker and Shaliastovich (2008) for an overview. The framework combines the tractability of affine models with the generality of models that include unknown components.

The nonlinear measurement equation between observables and unobservables can be related to optimal consumption and dividend policy in fully structural models, or to partial equilibrium models that include components whose distribution or functional form is unknown. The optimal consumption and dividend ratios endogenously generate nonlinear variation in consumption and dividend growth. Still, they are consistent with a notion of optimizing behaviour, which would be more difficult to achieve when directly modelling consumption and dividend growth as exogenous processes. Similarly the framework generates autocorrelation in consumption and dividend growth through the properties of the stationary state variables, but allows for further autocorrelation induced by transitory deviations from the optimal levels.

The framework is highly tractable when the unknown functions are approximated by orthogonal polynomials as in Chen (2007). In particular, closed-form expressions for the price-dividend ratio and its volatility can be derived in terms of ratios of polynomials of the affine state variables. The pricing formulas can be rapidly evaluated and prevent the need for computationally intensive simulation or numerical approximation methods. I study maximum likelihood estimation of the structural parameters and the polynomial coefficients for the general setting where the measurement equations contain unknown components, but the transition density of common state variables is known. The state variables can be recovered from cross-sections of asset prices when the pricing errors are at most weakly dependent, so that they can be averaged out in the cross-section. This gives rise to a fixed-effects estimator that evaluates the measurement and transition densities at the recovered state variables, which is shown to be consistent yet finite-sample biased when the cross-section and time series grow to infinity at the same rate. The structural functions are nonparametrically identified from the conditional mean of the asset prices and macroeconomic ratios, which motivates a quasi-maximum likelihood approach that is robust against the distribution of the errors. For continuous-time Gaussian noise processes, the measurement density simplifies to a generalized least squares criterion. I introduce a novel pairwise concentration procedure that incorporates knowledge of the transition density by penalizing state variables from adjacent cross-sections that are far apart. This is implemented using a closed-form transition density approximation as in Filipović et al. (2013). When the cross-sections become large, the pairwise concentrated estimator is asymptotically equivalent to the maximum likelihood estimator, at low computational cost. The structural functions and parameters are consistently estimated when the number of time periods and cross-sectional observations goes to infinity. The coefficients are asymptotically normally distributed when the approximating error is correctly specified, which provides the basis for performing standard parametric inference.

The framework is illustrated by analyzing the impact of time-varying economic uncertainty and risk aversion. In the benchmark model aggregate productivity or output growth follows a stochastic volatility process whose drift depends on the volatility level. The risk aversion shocks depend on unexpected output shocks, as in models with habit formation, but also contain pure preference shocks as in Bekaert et al. (2009). The resulting stochastic discount factor decomposes into a permanent component that depends on the level of output and a transitory component that depends on the stationary state variables as in Hansen and Scheinkman (2009). The latter allows for general volatility-dependent discount factors, in line with common specifications based on recursive preferences. The model is estimated using data on U.S. output and consumption, the price-dividend ratio and realized variance of the S&P 500 stock market index, and a panel of price-dividend ratios of sorted portfolios starting in 1929 and running until 2016. Economic uncertainty proxies based on the monthly Industrial Production Index are included in the measurement equation and have a penalization effect on the implied uncertainty similar to that for financial volatility in Andersen et al. (2015).

The equilibrium dividend-consumption ratio is shown to be an increasing and convex function of economic uncertainty, after controlling for risk aversion. The sign of the slope is consistent with consumption and dividend growth being negatively and positively correlated, respectively, with changes in the uncertainty proxies, while the convexity is consistent with a fixed adjustment cost for changing policy. The stochastic discount factor is moderately increasing in uncertainty, as opposed to a U-shaped pattern that is obtained when the dividend-consumption ratio is restricted to be linear. The price-dividend ratio is most steeply declining for moderate levels of economic uncertainty, for which return variance is dominated by variation in the price-dividend ratio as opposed to dividend variation. Furthermore large firm dividend policy responds less to uncertainty shocks than that of small firms. Uncertainty aversion therefore creates a wedge between large and small firm valuation ratios, which is larger when risk aversion is high, thus allowing to separate risk aversion from uncertainty. The recovered state variables suggest uncertainty and risk aversion have different dynamics, especially in crisis periods. Low risk aversion played an important role in the build-up of the dotcom-bubble, while both risk aversion and uncertainty peaked during the global financial crisis. Overall the nonlinear response functions, together with moderate uncertainty aversion, are able to reconcile large drops in asset prices in response to uncertainty shocks, the non-monotonic relation between economic uncertainty and stock market volatility, and the dynamics of variance risk premia, with reasonable size and persistence of the underlying shocks.

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

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

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

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

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

2 Setting

This section describes a general class of models for which results are derived. The specific examples are the basis of the empirical section. Throughout let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_t$ be the information filtration process satisfying standard regularity conditions.

2.1 Technology

Let $\mathcal{S}_t = (Y_t, s_t) \subset \mathcal{S} \subseteq \mathbb{R}^{D+1}$ be a Markovian state vector consisting of an output or productivity process Y_t and a D-dimensional state variable s_t that describes its drift, variance, and/or jump intensity, and possibly that of other relevant variables in the economy, as described by the stochastic differential equation

$$\begin{aligned} d\log Y_t &= \mu^y(s_t)dt + \sigma^y(s_t)dW_t \\ ds_t &= \mu^s(s_t)dt + \sigma^s(s_t)dB_t, \end{aligned} \tag{1}$$

where (B_t, W_t) is a standard Brownian motion in $\mathbb{R}^{D+1}$. In particular, the level of the output process Y_t does not affect the distribution of the increments of the state vector. As a consequence, mean-reversion is ruled out and the output process is non-stationary. On the other hand the state variables are stationary under a mean-reversion specification of the drift. This implies output growth $\log \frac{Y_{t+\tau}}{Y_t}$ over any horizon $\tau > 0$ is stationary and its conditional distribution only depends on s_t.

The drift $\mu(s_t) = (\mu^y(s_t), \mu^s(s_t)) = K_0 + K_1 s_t$, and volatility $\sigma(s_t) = (\sigma^y(s_t), \sigma^s(s_t))$ with $\sigma(s_t)\sigma(s_t)' = H_0 + \sum_{j=1}^{D} H_{1d} s_{tj}$ are assumed to be linear in the state variables s_t, so that the process is affine in the sense of Duffie et al. (2000).[2] The affine class contains many models commonly used for stock returns as it can incorporate stochastic volatility and leverage in a tractable fashion. In this paper it is used to compute exact expressions of conditional moments, and for the closed-form approximation of the transition density following Filipović et al. (2013).

The baseline model for output growth is a version of the Heston stochastic volatility model with correlation ρ between the output and volatility Brownian innovations and a volatility-dependent

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

drift:

$$d\log Y_t = \left(\mu - \frac{1}{2}V_t - \lambda V_t\right)dt + \sqrt{V}_t dW_t$$

$$dV_t = \kappa\left(\theta - V_t\right)dt + \omega\sqrt{V}_t dB_t.$$

A positive value of the parameter λ corresponds to an endogenous growth hypothesis where output uncertainty reduces expected growth. The mean reversion parameter κ assures that V_t is stationary around its unconditional mean θ.

2.2 Stochastic discount factor

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

$$U(\mathcal{S}_t) = E\left(\int_t^\infty e^{-\delta(r-t)}u(C_r, s_r)dr \mid \mathcal{S}_t\right),$$

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

$$u(t, C_t, s_t) = e^{-\delta t}v(C_t - X_t; \gamma)H(s_t),$$

with X_t the consumption reference level, $v(\cdot)$ the isoelastic utility function

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

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

Time-varying risk aversion is described by the local curvature

$$\frac{C_t u_{cc}(C_t, s_t)}{u_c(C_t, s_t)} = \frac{C_t}{C_t - X_t} \equiv Q_t,$$

where Q_t is the inverse of the consumption surplus ratio $(C_t - X_t)/C_t$. The consumption reference level X_t is modelled implicitly via the specification for Q_t

$$dQ_t = \kappa^q(\theta^q - Q_t)dt + \sqrt{Q_t}\eta dB_t^q + r_y\sqrt{V}_t dW_t, \qquad Cov(dB_t^q, dW_t) = 0. \tag{2}$$

Under this specification time-varying risk aversion is driven by productivity shocks dW_t and pure preference shocks dB_t^q, and is mean-reverting around the level θ_q. In the baseline model the state discount function $H_t = H(V_t)$ is a general uncertainty aversion index, that is included to control for the impact of uncertainty shocks on preferences. This allows for a direct impact on stochastic discounting beyond the appearance of uncertainty shocks in the specification for Q_t. Formally, let the pricing kernel process $\zeta_t = e^{-\delta t}u_c(C_t, s_t)$ be proportional to the marginal utility of a unit of consumption. Then the stochastic discount factor or marginal rate of substitution over states is given by

$$M_{t,t+\tau} = \frac{\zeta_{t+\tau}}{\zeta_t} = e^{-\delta\tau}\left(\frac{C_{t+\tau} - X_{t+\tau}}{C_t - X_t}\right)^{-\gamma}\frac{H(V_{t+\tau})}{H(V_t)} = e^{-\delta\tau}\left(\frac{C_{t+\tau}}{C_t}\right)^{-\gamma}\left(\frac{Q_{t+\tau}}{Q_t}\right)^{\gamma}\frac{H(V_{t+\tau})}{H(V_t)}.$$

This semiparametric formulation specifies the marginal rate of substitution over states to be the product of exponential time-discounting, a power of consumption growth relative to the habit, and an unknown uncertainty aversion component.

2.3 Consumption and dividend policy

In general the optimal consumption choice depends on all sources of wealth and income and all possible investment opportunities. When the primary interest is in understanding the response of consumption to changing economic circumstances, a flexible reduced form approach is to model the consumption-to-output ratio via some unknown function $\psi^{cy}(\cdot)$ of the latent states and an unexplained stationary component z_t^c:

$$\frac{C_t}{Y_t} = \psi^{cy}(s_t) + z_t^c, \qquad E\left(z_t^c \mid s_t\right) = 0. \tag{3}$$

The specification of consumption as a ratio of output guarantees that in the long run consumption and output cannot drift apart, while the stationary state variables and error component allow for general transitory fluctuations. This specification is the continuous-time analogue of cointegration of the discretely observed realization of the process.

Suppose furthermore an index of equities is traded at price P_t which pays a continuous stochastic dividend stream D_t. Analogue to consumption, the optimal dividend policy as a share of consumption is modelled as a general function of the state $\psi^{dc}(\cdot)$ plus a stationary unexplained component z_t^d:

$$\frac{D_t}{C_t} = \psi^{dc}(s_t) + z_t^d, \qquad E(z_t^d \mid s_t) = 0 \tag{4}$$

Alternatively the dividend-to-output ratio could be modelled first, as any pair of ratios of output, consumption, and dividends, pins down the remaining one. The parametrization of dividends-to-consumption is most convenient for asset pricing, and is also used in Menzly et al. (2004), among others.

The specification of the consumption-output and dividend-consumption ratios leads to potentially nonlinear dynamics of consumption and dividend growth:

$$\begin{aligned} d\log C_t &= d\log Y_t + d\log\psi^{cy}(s_t) \\ d\log D_t &= d\log C_t + d\log\psi^{dc}(s_t), \end{aligned}$$

up to noise processes that depend on $z_t^M = (z_t^c, z_t^d)$. This decomposes consumption and dividend growth into output growth, changes in the optimal ratios driven by the state variables, and unexplained transitory variation. The common affine approach is to model $\psi(s_t)$ as exponential affine functions of the state variables such that consumption and dividend growth are linear in the state variable (Eraker and Shaliastovich, 2008). More general specifications of $\psi(s_t)$ allow for convex or concave relations, or for interaction terms between the state variables.

2.4 Asset prices

From the Euler equation for optimal consumption and investment it follows that the equilibrium asset price satisfies

$$P_t = E\left(\int_0^\infty e^{-\delta\tau}\left(\frac{C_{t+\tau}Q_t}{C_tQ_{t+\tau}}\right)^{-\gamma}\frac{H(V_{t+\tau})}{H(V_t)}D_{t+\tau}d\tau \mid \mathcal{S}_t\right). \tag{5}$$

Under joint Markovian dynamics of (s_t, z_t^d) the asset's price-dividend ratio $\phi_t = \frac{P_t}{D_t} = \phi(s_t, z_t^d)$ is a function of the current state s_t and the transitory deviation in the dividend-consumption ratio z_t^d alone. For the case of log utility $\gamma = 1$ consumption and dividend growth cancel out and the price-dividend ratio is given by:

$$\phi(s_t, z_t^d) = E_t\left(\int_0^\infty e^{-\delta\tau}\frac{Q_{t+\tau}H(V_{t+\tau})}{Q_tH(V_t)}\frac{\psi^{dc}(s_{t+\tau}) + z_{t+\tau}^d}{\psi^{dc}(s_t) + z_t^d}d\tau\right) \tag{6}$$

In the general power utility case the nonstationary component $\left(\frac{Y_{t+\tau}}{Y_t}\right)^{-\gamma}$ appears in the stochastic discount factor. In this case the Girsanov change-of-measure formula can be applied to $e^{-\gamma \log Y_t}$ to write the price-dividend ratio in terms of expectations of functions of stationary affine state variables under a new probability measure. Details are given in the appendix.

The conditional moments of affine processes can be computed exactly by solving a first-order linear matrix differential equation. This method was introduced for univariate processes by Zhou (2003), and has been generalized to the multivariate setting by Cuchiero et al. (2012) and Filipović et al. (2016). Define the generator $\mathcal{A}$ of the process S_t as the operator which for a function $f : \mathbb{R}^{D+1} \to \mathbb{R}$ returns

$$\mathcal{A}f = \lim_{\tau \to 0} \frac{1}{\tau} \left(E_t(f(S_{t+\tau})) - f(S_t) \right).$$

Then for affine diffusion processes of the type (1)[3]

$$\mathcal{A}f(S) = (K_0 + K_1 S)^T \nabla f(S) + \frac{1}{2} \left(\mathrm{Tr}(\nabla^2 f H_0) + \sum_{j=1}^{D} \mathrm{Tr}(\nabla^2 f H_{1,j}) s_j \right). \tag{7}$$

Let $|l| = l_1 + \cdots + l_{D+1}$ denote the length of a multi-index $l \in \mathbb{N}^{D+1}$, let $S^l = \Pi_i S_i^{l_i}$ be a mixed polynomial of degree $|l|$, and let $\mathrm{Pol}_L = \{f : \mathcal{S} \subseteq \mathbb{R}^{D+1} \to \mathbb{R} : \exists a, f = \sum_{0 \leq |l| \leq L} a_l S^l\}$ be the vector space of mixed polynomials of maximum degree L. Then it follows from (7) that for any $f_l \in \mathrm{Pol}_L$, the generator $\mathcal{A} f_l \in \mathrm{Pol}_L$ as well. Applying the canonical basis $B_L = \{S^l : |l| \leq L\}$ of Pol_L to $\mathcal{A}$ and collecting the coefficients as

$$\mathcal{A} S^{l_i} = \sum_j a_{ij} S^{l_j} \tag{8}$$

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

$$E(S_{t+\tau}^l \mid S_t^l = S) = S^l + E\left(\int_t^{t+\tau} \mathcal{A} S_s^l ds \right),$$

[3] This property extends to process with quadratic variance specification (Zhou, 2003; Cheng and Scaillet, 2007), and to models with interaction terms such as the habit process (2) whose conditional volatility is $\sqrt{Q_t V_t}$.

which leads to a matrix differential equation with solution

$$E(S_{t+\tau}^{l} \mid S_t = S) = e^{\tau A_L} S^{l}. \tag{9}$$

Suppose the consumption and dividend policy functions are approximated by L-degree polynomial expansions

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

where the inner products use the stacked column vectors of mixed polynomials and their coefficients up to degree L using lexicographic ordering. Orthogonal polynomials such as the Hermite or Chebyshev polynomials are spanned by elementary polynomials and can be represented in this way. Similarly approximate the state-dependent preference function by a K-degree polynomial

$$H_K(s) = \sum_{|l|=0}^{K} e_l s^l = \underline{e} \cdot \underline{s}^K.$$

Define the product of two polynomials $(H_K \cdot \psi_L^{dc})(s) = \sum_{|l|=0}^{M} g_l s^l = \underline{g} \cdot \underline{s}^M$ with coefficients $g_l = \sum_{|m|=0}^{|l|} d_m e_{m-l}$ and $M = K + L$. From the conditional moments (9) it follows that the time $t + \tau$ contribution to the approximated price-dividend ratio equals

$$\begin{aligned} E_t\left(e^{-\delta\tau} H_K(s_{t+\tau})\psi_L^{dc}(s_{t+\tau})\right) &= \underline{g}^T e^{-\delta\tau} E_t\left(\underline{s_{t+\tau}}^M\right) \\ &= \underline{g}^T e^{-\delta\tau} e^{\tau A_M} \underline{s_t}^M. \end{aligned}$$

This implies that the approximated price-dividend ratio becomes a ratio of polynomials given by

$$\begin{aligned} \phi_M(s_t) &= \int_0^\infty e^{-\delta\tau} \frac{E_t\left(H_K(s_{t+\tau})\psi_L^{dc}(s_{t+\tau})\right)}{H_K(s_t)\psi_L^{dc}(s_t)} d\tau \\ &= \frac{\underline{g}^T \int_0^\infty e^{-\delta\tau} e^{\tau A_M} d\tau \underline{s_t}^M}{\underline{g}^T \underline{s_t}^M} \\ &= \frac{\underline{g}^T Q_M \underline{s_t}^M}{\underline{g}^T \underline{s_t}^M}. \end{aligned} \tag{10}$$

The matrix Q_M is the Laplace transform of the matrix exponential of A_M and can be solved in closed form as

$$Q_M = \int_0^\infty e^{-\delta\tau} e^{\tau A_M} d\tau = (\delta I - A_M)^{-1} \tag{11}$$

provided that $\delta I - A_M$ is non-singular.[4] The latter typically holds when $\delta > 0$, but it might be near-singular when A_M is not full rank. In the univariate case A_M is full rank when the diagonal terms of the lower-triangular matrix A_M are all non-zero. For affine diffusions this will be satisfied when the drift is strictly affine. In the multivariate case the diagonal blocks of the blockwise lower-triangular matrix A_M need to be non-singular, which is equivalent to the non-singularity of the matrix K_1 in the drift. This holds for example when the state variables can be ordered to appear in a triangular fashion in the drift of the other state variables.

The rational-polynomial formula (10) is an exact expression for the approximated model. This differs from the more common situation where the input functions are known, but approximation methods are needed to compute the price-dividend ratio as a solution to an integral equation (Wachter, 2005; Calin et al., 2005). Similar equations arise in stochastic growth models; for an overview see Taylor and Uhlig (1990). With unrestricted coefficients the rational-polynomial formula could be used as a Padé expansion, which is known to have good approximation and extrapolation properties (Judd, 1996). The analytic solution overcomes the need to compute expectations via Fourier analysis or Monte Carlo simulation, which is particularly attractive for the purpose of estimation. Exact expressions for asset prices under non-Gaussian distributions and/or non-CRRA preferences are rare. An exception is the linearity generating processes of Gabaix (2007), who modifies an autoregressive process such that the price-dividend ratio becomes linear in the state variable. More often the log-linear approximation method of Campbell and Shiller (1988) is used, which could obscure the impact of higher order moments.

Figure 1 illustrates the impact of nonlinearity in the dividend-consumption ratio as a function of output uncertainty on the price-dividend ratio. In particular, it shows that the second order term in the dividend-consumption ratio expansion $\frac{D_t}{C_t} = 1 + 0.1V_t + c_2V_t^2$ controls the steepness of the decline of the price-dividend ratio in economic uncertainty. The slope of the price-dividend ratio has major implications for its variance, since when the slope is nearly flat the variance of the price-dividend ratio will be near zero regardless of the magnitude of economic uncertainty.

Figure 2 shows that time-varying risk aversion and uncertainty have a similar negative impact on the valuation ratio. Low price-dividend ratios can therefore not be unambiguously interpreted as either high uncertainty or high risk aversion. The lower is economic uncertainty, the steeper is the decline in the price-dividend ratio when risk aversion increases, and vice versa. More observations are therefore needed to disentangle the two effects.

However, Figure 3 shows that when risk aversion amplifies the impact of economic uncertainty

[4] A similar expression obtains for singular $\delta I - A_M$ in terms of its Jordan form UJU^{-1}, yielding $Q_M = U \int_0^\infty e^{-\delta\tau} e^{\tau J} d\tau U^{-1}$.

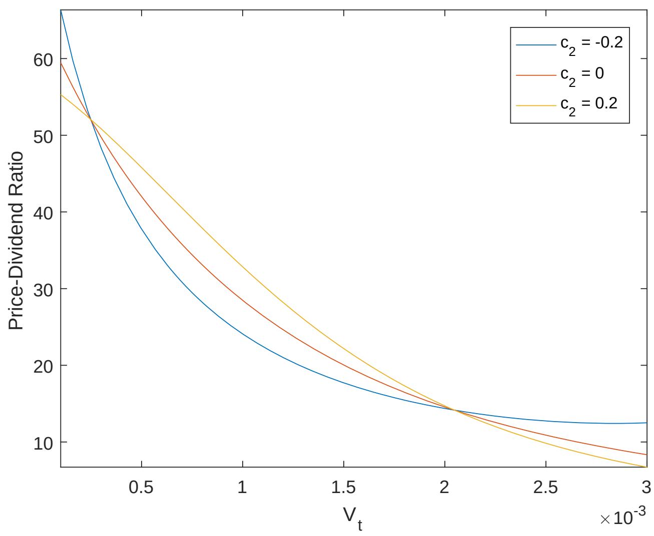


Figure 1: Theoretical Price-Dividend Ratio versus Output Growth Volatility, for Dividend-Consumption ratio set as $\frac{D_t}{C_t} = 1 + 0.1V_t + c_2V_t^2$ for varying c_2. Transition parameters are set as in Table 3; habit Q_t is in the steady state.

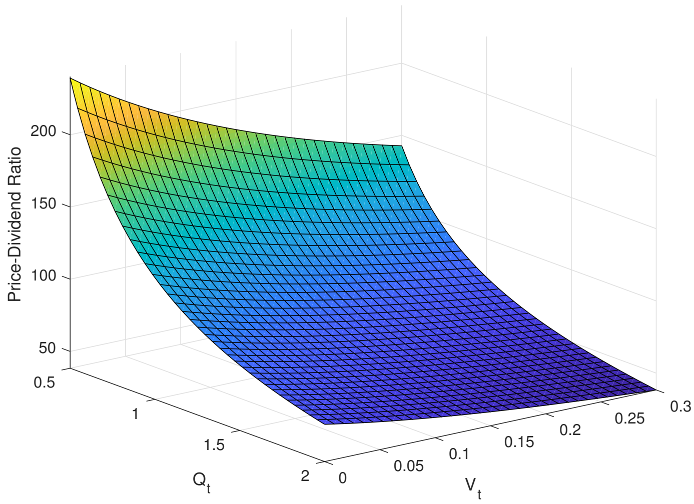
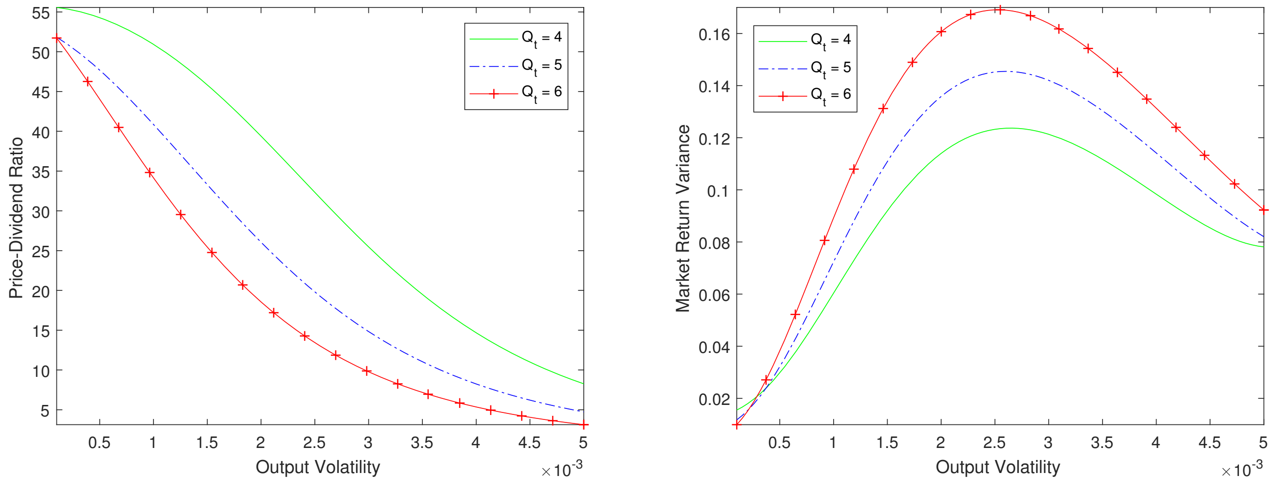


Figure 2: Theoretical Price-Dividend Ratio versus Risk Aversion and Output Growth Volatility. Dividend-Consumption ratio is set as $\frac{D_t}{C_t} = 1 + 0.1V_t$. Transition parameters are set as in Table 3.

on the dividend-consumption ratio, higher levels of risk aversion imply more steeply declining price-dividend ratios for any level of output growth volatility. This in turn implies that the market return variance is monotonically increasing in risk aversion, which it does not necessarily as a function of output growth volatility, as will be illustrated in 4.



(a) Price-Dividend Ratios.

(b) Stock Market Return Variance.

Figure 3: Theoretical Price-Dividend Ratio and Stock Market Return Variance versus Output Growth Volatility for varying levels of risk aversion. Dividend-Consumption ratio is set as $\frac{D_t}{C_t} = 1 + 0.1V_t + 0.2V_t^2 + 0.1Q_tV_t$. Transition parameters are set as in Table 3.

2.5 Risk-free rate and expected excess returns

The pricing equation (5) implies that in equilibrium the expected return on dividend-paying assets should satisfy

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

In particular, if a risk-free security is traded, then its deterministic rate of return r_t^f should satisfy

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

which is shown in the appendix to be increasing in the discount rate δ and expected consumption growth, and decreasing in the inverse consumption surplus Q_t. The impact of output growth volatility on the risk-free rate depends on the shape of the uncertainty aversion function $H(V_t)$ and whether volatility is above or below its long-run mean. The innovations of the pricing kernel follow

$$\frac{d\zeta_t}{\zeta_t} = -r_t^f dt - \lambda^Y(s_t)\sqrt{V_t}dW_t - \lambda^V(s_t)\omega\sqrt{V_t}dB_t - \lambda^Q(s_t)\eta\sqrt{Q_t}dB_t^q,$$

where the prices of growth, uncertainty, and discount rate risks, respectively, are given by

$$\begin{aligned}
\lambda^Y(s_t) &= 1 + \left(\frac{\psi_q^{cy}(s_t)}{\psi^{cy}(s_t)} - \frac{1}{Q_t}\right) r_y \sqrt{Q_t} \\
\lambda^V(s_t) &= \frac{\psi_v^{cy}(s_t)}{\psi^{cy}(s_t)} - \frac{H'(V_t)}{H(V_t)} \\
\lambda^Q(s_t) &= \frac{\psi_q^{cy}(s_t)}{\psi^{cy}(s_t)} - \frac{1}{Q_t}.
\end{aligned} \tag{12}$$

The expected excess return on the aggregate stock is

$$\frac{E_t(dP_t) + D_t dt}{P_t} - r_t^f dt = -E_t\left(\frac{d\zeta_t}{\zeta_t}\frac{dP_t}{P_t}\right) = \lambda^Y(s_t)\beta^Y(s_t) + \lambda^V(s_t)\beta^V(s_t) + \lambda^Q(s_t)\beta^Q(s_t),$$

in terms of the asset's risk exposures $\beta^X(s_t) = E_t\left(dX_t \frac{dP_t}{P_t}\right)$ with respect to $X \in \{\log Y_t, V_t, Q_t\}$.

The prices of growth, uncertainty, and discount rate risks would be constant over time if the consumption policy function ψ^{cy} and the uncertainty aversion function $H(V_t)$ are both log-linear, and the inverse consumption surplus is constant. Instead, the model generates time-varying prices of risks not only from the time-varying consumption reference level, as in habit formation models, but also from the nonlinearity in the consumption policy function and uncertainty aversion. Likewise, the risk exposures depend nonlinearly on the state variables unless the price-dividend ratio is log-linear. Together, these flexible functional forms can generate rich dynamics of growth and variance risk premia.

2.6 Individual stock prices

The general pricing formula can be straightforwardly extended to individual stocks. Suppose the equities of N firms are traded at prices (P_{it}) and pay continuous dividends (D_{it}). To separate a common component and heterogeneity in firm-level dividend policy it is natural to model the shares of total dividend $\frac{D_{it}}{D_t}$, an approach also taken in Menzly et al. (2004). If the dividend shares are stationary then in the long run no single firm dominates aggregate dividends. Suppose that the optimal dividend shares are modelled by individual-specific functions of the aggregate state $\psi_i^d(\cdot)$ plus stationary individual-specific unexplained components z_{it}^d:

$$\frac{D_{it}}{D_t} = \alpha_i + {\beta_i}^T s_t + z_{it}^d, \qquad E(z_{it}^d \mid s_t, z_t^d) = 0. \tag{13}$$

Various extensions can be thought of to model the functions (ψ_i) in a flexible way. In particular nonlinear specifications of $\psi_i^d(s_t)$ could be assumed in the same fashion as the nonlinear response in aggregate dividends D_t. Since the aggregate nonlinearity already implies nonlinear individual dividend-consumption ratios, for parsimony I restrict to the linear specification for individual dividend shares. The error terms z_{it}^d are allowed to be correlated with the aggregate error term z_t^d. A positive correlation suggests the asset pays out less when the aggregate dividend-consumption ratio is low, which makes it less valuable to an investor seeking to diversify its income. Furthermore observed time-invariant characteristics X_i, such as book-to-market or leverage ratios, can be used to explicitly describe parameter heterogeneity.[5] The individual price-dividend ratios are given by

$$\phi^i(s_t, z_{it}^d, z_t^d) = E_t\left(\int_0^\infty e^{-\delta\tau}\frac{Q_{t+\tau}H(V_{t+\tau})}{Q_tH(V_t)}\frac{\psi^{dc}(s_{t+\tau}) + z_{t+\tau}^d}{\psi^{dc}(s_t) + z_t^d}\frac{\alpha_i + \beta_i{}^T s_{t+\tau} + z_{it+\tau}^d}{\alpha_i + \beta_i{}^T s_t + z_{it}^d}d\tau\right). \quad (14)$$

Using the same steps as for the aggregate price-dividend ratio, and abstracting from the noise processes z_t^d, the price-dividend ratio of an asset i has the rational-polynomial expression

$$\phi_M^i(s_t) = \frac{\underline{g_i}^T Q_M \underline{s_t}^M}{\underline{g_i}^T \underline{s_t}^M}, \quad (15)$$

where now g_i are the coefficients of the product of polynomials $(H_K \cdot \psi_L^{dc} \cdot \psi_{i,L}^d)(s)$.

2.7 Asset Price Volatility

Variation in the log return can be decomposed into variation in the log price-dividend ratio ϕ_M, the log dividend-output ratio ψ_L^{dy}, and log output $\log Y_t$:

$$d\log P_t = d\log\phi_M + d\log\psi_L^{dy} + d\log Y_t.$$

The price-dividend ratio and the dividend-output ratio are driven by the latent variables s_t which generate covariation among the components. From Itô's Lemma it follows that the relative size of each component depends on its first derivatives with respond to the state variables. In particular, the variability of the price-dividend ratio follows from its quadratic variation increments

$$d\langle\phi_M(s_t)\rangle_t = \nabla\phi_M(s_t)^T \mathrm{Cov}([dV_t\ dQ_t]^T)\nabla\phi_M(s_t),$$

[5] Including time-varying characteristics would complicate computing asset prices, as their distribution would have to be explicitly modelled. With sorted stock portfolios the time-variation in the sorting variable is naturally controlled for.

with $\nabla\phi_M(s_t)$ the gradient of the price-dividend ratio (see the Appendix). The rational polynomial expression is preserved under differentiation, with the gradient given by

$$\nabla\phi_M(s) = \frac{\nabla\underline{s}^{M^T}\left(G^T Q_M^T - Q_M G\right)\underline{s}^M}{\underline{s}^{M^T} G \underline{s}^M},$$

where $G = \underline{g}\underline{g}^T$ is the outer product of the coefficient vector. The same formula applies to the dividend-output ratio. Figure 4 shows the theoretical variance decomposition of the log market

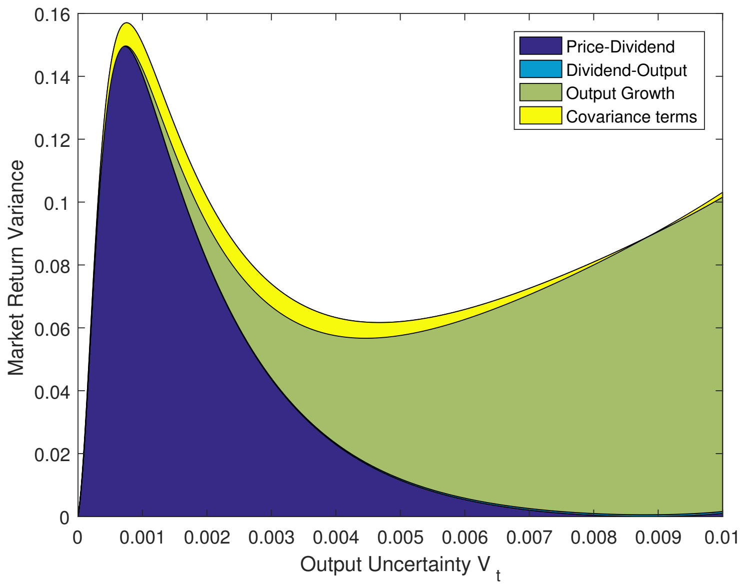


Figure 4: Theoretical variance decomposition of the log market return into variation in the log price-dividend ratio, log dividend-output ratio, and output growth, under quadratic expansion of the policy functions ψ and the pricing kernel H. Parameter values are set as the estimates reported in Table 3; habit Q_t is in the steady state.

return into its components using quadratic approximations. The graph shows that total return variance is non-monotonic in the underlying economic uncertainty. This explains why financial markets can be volatile even at low levels of economic uncertainty, or why stock market variance can be relatively low under high economic uncertainty. At low levels of uncertainty the total return variance is dominated by the large negative slope of the price-dividend ratio, whereas at large levels of uncertainty variation in fundamentals is the main component. Variation in the log dividend-output ratio is of a lower magnitude, so that variation in dividend growth is roughly proportional to output growth. In the empirical section the variation of log returns and dividends is estimated from daily observations over a long time span, and compared against their theoretical counterpart.

3 Estimation

This section discusses the estimation of the policy functions $\psi = \left(\psi^{cy}, \psi^{dc}\right)^T$, the discount parameter δ and uncertainty aversion index H, and the parameters of the latent variable distribution θ_s. The functional parameters are combined into $h = (\psi, H)$, the finite-dimensional into $\theta = (\delta, \theta_s)$, and both functional and finite-dimensional parameters into $\vartheta = (\theta, h)$.

3.1 State space formulation

The vector $M_t = \left(\frac{C_t}{Y_t}, \frac{D_t}{C_t}\right)^T$ contains the aggregate responses whose conditional mean is modelled by polynomial approximation. The N-dimensional vector $P_t = (P_{it})$ contains asset prices and realized volatility measures whose theoretical values are stacked in the vector-valued function $\mathbf{g}$. The theoretical values are functions, derived in Section 2, of the common state s_t and the characteristics $X_t = (X_{it})$ that describe cross-sectional heterogeneity. The system of observations is represented as

$$M_t = \psi(s_t) + z_t^M \tag{16}$$

$$P_t = \mathbf{g}(s_t, z_t^M, X_t, \vartheta) + z_t^P, \tag{17}$$

with $z_t = (z_t^M, z_t^P)$ the combined error terms. Combine the time-t observations into the vector $\mathcal{Y}_t = (Y_t, M_t, P_t)$, where output Y_t is the observed part of the state variable $\mathcal{S}_t$. Let $\mathcal{F}_t^{\mathcal{Y}} = (\mathcal{Y}_t, \mathcal{Y}_{t-1}, \dots)$ and $\mathcal{F}_t^s = (s_t, s_{t-1}, \dots)$ denote the histories of the observed and unobserved components, respectively. The following assumptions describe the interaction between the states and the errors:

Assumption 1.

a) (z_t, s_t) *are jointly stationary and contemporaneously exogenous:* $E\left(z_t \mid s_t\right) = 0$.

b) The joint process is first-order Markov:

$$(\mathcal{Y}_{t+1}, \mathcal{S}_{t+1}) \mid (\mathcal{F}_t^{\mathcal{Y}}, \mathcal{F}_t^s) \sim (\mathcal{Y}_{t+1}, \mathcal{S}_{t+1}) \mid (\mathcal{Y}_t, s_t).$$

c) There is no feedback from the errors to the states:

$$\mathcal{S}_{t+1} \mid (z_t, \mathcal{S}_t) \sim \mathcal{S}_{t+1} \mid \mathcal{S}_t.$$

The stationarity assumption of the states and errors is equivalent to the cointegration of output, consumption, and dividends, that appear as ratios in M_t. This cointegration relation is at the heart of several studies of return predictability (Lettau and Ludvigson, 2001; Bansal et al., 2007). The presence of state variables in the policy functions generalizes the constant cointegration relation and allows for more general dynamic error-correction forms. The joint first-order Markov assumption of observables and states is equivalent to the joint first-order Markov property of the states and the errors z_t. In practice higher order dependence can be allowed for by including further lags in the state vector. The no feedback assumption means that the state variables are themselves Markovian, and are not caused in the sense of Granger (1969) by errors in the observables. This allows for an interpretation of exogenous variation in the state variables that generates endogenous responses in the observations. Together with a notion of weak cross-sectional dependence between the errors z_t this means the state variables describe the systematic variation in the observables. The stronger assumption of a hidden Markov model, which rules out dependence of the error components on past states, is not needed. In particular, the observations do not need to be Markovian themselves, and can depend on shocks at all lags.

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

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

The time $t+1$ contribution to the likelihood function $\mathcal{L}_{TN}(\vartheta) = \prod_{t=1}^{T-1} \mathcal{L}_{t+1}(\vartheta)$ is given by

$$\begin{aligned}
\mathcal{L}_{t+1}(\vartheta) &= f(\mathcal{Y}_{t+1} \mid \mathcal{F}_t^{\mathcal{Y}}; \vartheta) \\
&= \int f\left(\mathcal{Y}_{t+1}, s_{t+1} \mid \mathcal{F}_t^{\mathcal{Y}}; \vartheta\right) ds_{t+1} \\
&= \iint f\left(\mathcal{Y}_{t+1} \mid s_{t+1}, \mathcal{Y}_t, s_t; \vartheta\right) f\left(s_{t+1} \mid s_t; \vartheta\right) f\left(s_t \mid \mathcal{F}_t^{\mathcal{Y}}; \vartheta\right) ds_{t+1} ds_t.
\end{aligned}$$

The normalized log likelihood is written as $\ell_{NT}(\vartheta) = \frac{1}{T}\sum_{t=1}^{T-1} \ell_{t+1}(\vartheta)$ with contributions $\ell_{t+1}(\vartheta) = \frac{1}{N} \log \mathcal{L}_{t+1}(\vartheta)$. The likelihood contributions are integrals over the future and current state variables, where the integrands are the product of the measurement, transition, and updating densities.

The population parameters of interest are given by

$$(\theta_0, h_0) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}} \lim_{N \to \infty} E\left(\ell_{t+1}(\theta, h)\right), \tag{18}$$

and the maximum likelihood estimator by

$$(\hat{\theta}, \hat{h}) = \arg \max_{\theta\in\Theta, h\in\mathcal{H}} \frac{1}{T} \sum_{t=1}^{T-1} \ell_{t+1}(\theta, h), \tag{19}$$

with Θ and $\mathcal{H}$ parameter spaces of finite and infinite dimension, respectively.

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

3.2 Cross-sectional state recovery

The data set takes the form of a panel consisting of N asset prices observed over T time periods. The approach in this paper is to extract the unobserved dynamic state variables from large cross-sections, and evaluate the joint likelihood at the estimated states. While avoiding a computationally intensive simulation procedure, this plug-in approach may lead to biases in the estimated parameters, when the number of cross-sectional units N is not of a larger order than the number of time periods T (e.g. Hahn and Kuersteiner, 2011; Fernández-Val and Weidner, 2016). This is the time series analogue of the incidental parameters problem for individual-specific effects. The bias may be avoided if the number of units N grows faster than the number of time periods T, although this would affect the identification and convergence rate of parameters describing the state dynamics (Gagliardini and Gourieroux, 2014).

While the dynamic state variables are modelled as random, a fixed-effects type estimator is generally consistent as long as N and T go to infinity. When the number of cross-sectional units grows to infinity, the updating density of the dynamic state variables becomes concentrated around its mode, subject to identification conditions. If the true parameter ϑ_0 were known, the prevailing state s_t could be recovered from a single cross-section as long as the cross-sectional errors satisfy a suitable law of large numbers. To formalize this, define the expected limited information conditional likelihood

$$\bar{\ell}_t(\vartheta, s) = \operatorname*{plim}_{N\to\infty} \frac{1}{N} E_{\vartheta_0} \left(\log f_\vartheta \left(\mathcal{Y}_t \mid s\right) \mid s_t\right). \tag{20}$$

Also define the population implied state

$$\bar{s}_t(\vartheta) := \bar{s}(\vartheta, s_t) := \arg\max_s \ \bar{\ell}_t(\vartheta, s).$$

For any parameter value ϑ the implied state maximizes the expected log likelihood function conditional on the prevailing state s_t. At the true parameter ϑ_0 it satisfies $\bar{s}(\vartheta_0, s) = s$ by properties of the Kullback-Leibler divergence. Define the implied state estimator

$$\hat{s}_t(\vartheta) = \arg\max_s \ \log f_\vartheta \left(\mathcal{Y}_t \mid s\right). \tag{21}$$

The implied state estimator is the standard fixed effect estimator when the realized states are treated as parameters. Write the limited information updating density as

$$f_\vartheta \left(s_t \mid \mathcal{Y}_t\right) = \frac{f_\vartheta \left(\mathcal{Y}_t \mid s_t\right) f_\vartheta \left(s_t\right)}{f_\vartheta \left(\mathcal{Y}_t\right)},$$

and denote its mode

$$\tilde{s}_t(\vartheta) = \arg\max_s \ \log f_\vartheta \left(s \mid \mathcal{Y}_t\right).$$

The measurement density $\log f_\vartheta \left(\mathcal{Y}_t \mid s_t\right)$ grows with the number of observations, unlike the unconditional state density $f_\vartheta \left(s_t\right)$ which can be interpreted as a prior. Therefore the maximzers $\hat{s}_t$ and $\tilde{s}_t$ of the measurement and updating density are asymptotically equivalent, provided the unconditional state density $f_\vartheta \left(s\right)$ is uniformly bounded in s. For the state estimators to converge to their population counterpart, the measurement density should become more informative about the state when the number of cross-sectional units increases:

Assumption 2. *For every s_t, when $N \to \infty$*

$$\sup_{\vartheta\in\Theta\times\mathcal{H}, s\in\mathcal{S}} \left| \frac{1}{N} \log f_\vartheta \left(\mathcal{Y}_t \mid s\right) - \bar{\ell}_t(\vartheta, s) \right| \xrightarrow{p} 0$$

Assumption 2 assures that the cross-sectional limit (20) is well defined. For i.i.d. data, it is implied by compactness of Θ and continuity and boundedness of f_ϑ. More generally, it requires the measurement errors to satisfy a uniform (weak) law of large numbers to prevent the measurement density from growing faster than the number of observations. This allows for weak forms of cross-sectional dependence, which the next subsection discusses in more detail. If Assumption 2 holds, and a consistent estimator for ϑ_0 is available, the state variables can be consistently estimated from a large cross-section.

The maximizer of the concentrated likelihood

$$\ell^c_{TN}(\vartheta) = \frac{1}{T}\sum_{t=1}^{T} \log f_\vartheta\left(\mathcal{Y}_t \mid \hat{s}_t(\vartheta)\right) \tag{22}$$

is the analogue of the fixed-effects estimator in panel data models with individual effects, and is therefore generally consistent when $N \to \infty$. The related concentrated nonlinear least square estimator has been applied to option pricing (Andersen et al., 2015; **?**) and term structure models (Andreasen and Christensen, 2015), and can consistently estimate the state variables even when their transition density is unspecified.[6] However, unlike independent cross-sectional units, the state variables create a natural dependency between different time periods, which is ignored by the estimation based on (22). The parameters of the transition density may therefore be inefficiently estimated or not even identified, which would carry over to the estimation of the states and the structural parameters. To incorporate the knowledge about the transition density of the state variables, the limited information maximum likelihood can be extended to condition on the one-period lagged cross-section:

$$\ell_{NT,1}(\vartheta) = \frac{1}{NT}\sum_{t=1}^{T-1} \log f_\vartheta\left(\mathcal{Y}_{t+1} \mid \mathcal{Y}_t\right). \tag{23}$$

The state variables can now be concentrated pairwise, taking account of the transition density. In particular, the augmented time $t+1$ log likelihood including one lag is

$$\ell_{t+1|t}(\vartheta; s', s) = \frac{1}{N}\left(\log f_\theta\left(\mathcal{Y}_{t+1} \mid \mathcal{Y}_t, s', s\right) + \log f_\theta\left(s' \mid s\right) + \log f_\theta\left(s \mid \mathcal{Y}_t\right)\right),$$

and the pairwise concentrated states are

$$\left(\hat{s}'_{t+1}, \hat{s}_t\right)(\vartheta) = \arg\max_{(s',s)} \; \ell_{t+1|t}(\vartheta; s', s).$$

The pairwise concentrated log likelihood function is

$$\ell^{c,pair}_{TN,1}(\vartheta) = \frac{1}{T}\sum_{t=1}^{T-1} \ell^{c,pair}_{t+1}(\vartheta), \qquad \ell^{c,pair}_{t+1}(\vartheta) = \ell_{t+1|t}\left(\vartheta; \hat{s}'_{t+1}(\vartheta), \hat{s}_t(\vartheta)\right),$$

[6] Andersen et al. (2015) apply this estimator to large cross-sections of option prices, and show that when the risk-neutral distribution is correctly specified, a single cross-section suffices for consistent parameter estimation.

with corresponding estimator

$$(\hat{\theta}^{c,pair}, \hat{h}^{c,pair}) = \arg \max_{\theta\in\Theta, h\in\mathcal{H}} \ell^{c,pair}_{TN,1}(\theta, h). \tag{24}$$

This procedure is a hybrid version of simultaneously concentrating out all state variables, and doing so only locally for every point in time. The full simultaneous concentration procedure defines an optimization over TD state variables and is computationally intractable for large T. The local concentration procedure involves simple optimizations but ignores the dependence between the state variables. In finite samples this can lead to large distances between adjacent state vectors, which could result in estimation bias when evaluating the transition density at the concentrated states. The pairwise procedure automatically takes into account the transition density of the states, while solving only $2D$ dimensional optimization problems. For every time period it yields two estimates of the prevailing state, $\hat{s}'_t(\theta, h)$ and $\hat{s}_t(\theta, h)$. The former uses current and one-period lagged observations, while the latter uses current and one-period ahead observations. Under Assumption 2 and a sufficiently large cross-section, the difference between the estimators disappears, which can be used for diagnostic analysis.

By the maximum theorem the pairwise concentrated likelihood (24) is continuous in the parameters, even when the concentrated state variables themselves are not continuous everywhere. Therefore it satisfies the continuity conditions for sieve M-estimation for weakly dependent data in Chen and Shen (1998). The pairwise concentrated likelihood is asymptotically equivalent, denoted $\overset{\text{a.e.}}{=}$, to the integrated limited information likelihood under weak cross-sectional dependence. When moreover past lags are uninformative about the current state given a large enough cross-section, it poses no efficiency loss compared to the full maximum likelihood estimator:

Proposition 1. a) *Under Assumption 2, when $N \to \infty$*

$$\ell^{c,pair}_{NT,1}(\vartheta) \overset{\text{a.e.}}{=} \ell_{NT,1}(\vartheta).$$

b) *Under Assumptions 2 and* $\arg\max_{s\in\mathcal{S}} \operatorname{plim}_{N\to\infty} \frac{1}{N} E_{\vartheta_0}\left(\log f_\vartheta\left(\mathcal{Y}_t \mid s\right) \mid \mathcal{F}^Y_t\right) = \bar{s}_t(\vartheta)$ *a.s.,*

$$\ell^{c,pair}_{TN,1}(\vartheta) \overset{\text{a.e.}}{=} \ell_{TN}(\vartheta).$$

The proof of Proposition 1 uses a Laplace approximation and is given in the appendix. Part a) states that the difference between the concentrated likelihood and the one-lag limited information likelihood disappears when $N \to \infty$, as long as the scale of the updating density shrinks proportionally to the number of observations. Part b) states that if moreover the updating density is maximized at the same state when past lags are added, the difference with the full information likelihood disappears. Under these assumptions the pairwise concentrated likelihood suffers no asymptotic efficiency loss compared to the full maximum likelihood estimator.

3.3 Quasi-maximum likelihood

When the structural parameters are identified from the conditional mean and variance of the measurement equation, their consistent estimation does not require modelling explicitly the (joint) distribution of the errors. This motivates a quasi-maximum likelihood approach following White (1982), which is robust to specific assumptions on the error distribution, only at the cost of a possible loss of efficiency. In particular, I construct quasi-likelihoods based around Gaussian errors, in which case the joint measurement density reduces to a nonlinear least squares criterion. The errors are not restricted to be uncorrelated over time or in the cross-section, as long as they can be averaged out from a cross-section. Since both the macroeconomic ratios and the price-dividend ratios are highly persistent, I use an autoregressive model for their error terms. Since the state variables already introduce non-Gaussian shocks into the observables, the residuals may not be too far from Gaussian, which limits the associated efficiency loss.

Let the $N+2$-dimensional error vector $z_t = (z_t^M, z_t^P)$ representing the transitory deviation from equilibrium follow the continuous time autoregressive Ornstein-Uhlenbeck process with Brownian increments

$$dz_t = -Az_t dt + \Sigma dW_t. \tag{25}$$

This process reverts around its unconditional mean of zero, with the autocorrelation matrix A capturing the speed of mean reversion, and Σ the covariance matrix of the Gaussian shocks in W_t. The continuous time setup allows computing the distribution of the errors over any horizon. The increments for a given horizon τ follow the normal distribution

$$z_{t+\tau} - z_\tau \sim N(\mu^z_{t,t+\tau}, \Omega^z_\tau),$$

whose expectation $\mu^z_{t,t+\tau}$ is linear in the current value with exponentially decaying weight

$$\mu_{t,\tau}(z) = e^{-A\tau} z_t,$$

and with the covariance matrix Ω^z_τ varying with the horizon but independent of the current value. For notational convenience define the autocorrelation matrix $R_\tau = \exp(-A\tau)$. The unconditional distribution of z_t is also Gaussian, and given by

$$z_0 \sim N(0, \Omega^z_\infty).$$

Under a Gaussian error distribution, the one-period ahead measurement density of the macroeconomic ratios and the price-dividend ratios conditional on the current and previous state is

$$f_\vartheta\left(\mathcal{Y}_{t+1} \mid \mathcal{Y}_t, s_{t+1}, s_t\right) = \phi\left(\varepsilon_{t+1}(\vartheta, s_{t+1}, s_t); \Omega_\tau\right)$$

in terms of the multivariate normal density $\phi(\cdot;\sigma^2)$ with mean zero and variance σ^2, and the generalized residuals

$$\varepsilon_{t+1}(\vartheta, s_{t+1}, s_t) = z_{t+1}(\vartheta, s_{t+1}) - Rz_t(\vartheta, s_t) \tag{26}$$

$$z_t(\vartheta, s_t) = \begin{pmatrix} M_t - \psi(s_t) \\ P_t - \mathbf{g}(s_t, z^M_t, X_t, \vartheta) \end{pmatrix}. \tag{27}$$

The limited information likelihood conditional on the current state alone is given by

$$f_\vartheta\left(\mathcal{Y}_t \mid s_t\right) = \phi\left(z_t(\vartheta, s_t); \Omega_\infty\right).$$

Under a large cross-section of asset prices the dimension of z^P_t grows with N, while that of z^M_t is bounded. Assumption 2 thus requires

$$\frac{1}{N}\log f_\vartheta\left(P_t \mid s_t\right) = -\frac{1}{2N}\log(2\pi) - \frac{1}{2N}\log|\Omega^P_\infty| - \frac{1}{2N} z^P_t(\vartheta, s_t)^T {\Omega^P_\infty}^{-1} z^P_t(\theta, s_t) \tag{28}$$

to remain bounded in probability. For independent Gaussian errors this is implied by non-singularity of Ω^p_∞, which leads to a non-central Chi-squared distribution. More generally this restricts the magnitude of the covariance terms in Ω^p_∞. In particular, it should be ruled out that the variance of some linear combinations of the error terms does not disappear, which would violate the law of large numbers. This is guaranteed by the following assumption of weak cross-sectional dependence based on Chudik et al. (2011):

Assumption 3. *For every period $t = 1, \ldots, T$, for any weight vector $\mathbf{w}_t$ such that $\|\mathbf{w}_t\|^2 = O(N^{-1})$ and $\frac{w_{it}}{\|\mathbf{w}_t\|} = O(N^{-\frac{1}{2}})$ when $N \to \infty$,*

$$\mathrm{Var}\left(\mathbf{w}_t^T \mathbf{z}_t \mid s_t, X_t\right) \to 0.$$

For the quasi-maximum likelihood density (28), Assumption 3 implies Assumption 2. Assumption 3 rules out common components in the error terms. For example, strong factor structures of the form $z_{it} = \beta_i X_t + \varepsilon_{it}$, with ε_{it} i.i.d., would not be allowed. In practice this requires taking out the factors by including sufficiently many latent variables or observed regressors. A natural control in the case of many individual price-dividend ratios is to include the error of the aggregate price-dividend ratio, which effectively demeans the individual errors. The assumption does allow for general spatial dependence in the errors, such as α-mixing errors with distance measured by X_{it} (Sarafidis and Wansbeek, 2012). This type of local dependence can be effectively dealt with by cross-sectional kernel smoothing prior to estimation, as in Dalderop (2018). This also allows using observations of nearby dates to smooth out the errors over time, as is useful in dealing with observation errors in option prices that are likely to be less persistent.

The number of parameters in A and Ω_∞^p grows with N^2 if no restrictions are imposed. In practice dimension reductions are required, for example by assuming uncorrelatedness, which leads to the order of N parameters. In this case $z_{it} = e^{-a_i\tau} z_{it-1} + \omega_{\tau,i}\varepsilon_{it}$, with ε_{it} i.i.d. $N(0,1)$. Furthermore when the cross-sectional units have a natural ordering in the variable X_i, it could be assumed that $(a_i, \sigma_i) = (a(X_i), \sigma(X_i))$ for some global functional parameters $(a(\cdot), \sigma(\cdot))$ that can be approximated by a series of basis functions.

3.4 Closed-form transition density approximation

There is an active literature on approximating the transition density of the state variables based on stochastic differential equations of the type (1). The major advantage of such an approach is that it prevents the need for costly simulation of continuous time processes. Starting from Aït-Sahalia (2002), several papers study approximating the log transition density using Hermite polynomials and solving the coefficients from the Kolmogorov forward and backward equations. This approach works particularly well for multivariate diffusions, and for relatively short horizons. In this paper I use a variant of the approximation method in Filipović et al. (2013). This method is based on series expansions in the state space rather than the time horizon, so that the approximation quality does not deteriorate with the horizon.[7] In particular, starting from an auxiliary density $w(\cdot)$ I approximate the likelihood ratio using orthogonal polynomials up to some order J:

$$f^{(J)}(\mathcal{S} \mid s) = w(\mathcal{S} \mid s)\left(1 + \sum_{|l|=1}^{J} c_l H_l(\mathcal{S})\right),$$

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

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

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

The polynomial moments are linear combinations of the known conditional moments (9). In our implementation I choose the auxiliary densities for output $\log Y_t$, its instantaneous volatility V_t, and the risk aversion state Q_t, to be the symmetric Variance-Gamma density, and twice the noncentral Chi-squared distribution, respectively. The former naturally allows for fat tails, while the latter is the exact distribution of processes of the type of Cox et al. (1985). Each distribution has two parameters which are used to match the conditional mean and variance of each variable. The product of the univariate auxiliary densities then creates the trivariate auxiliary density. The approximations using mixed Hermite polynomials to the fourth order are very close to densities obtained via Fourier inversion. All coefficients are computed symbolically, which only needs to be performed once prior to estimation. The marginal densities of the unobserved states are approximated using the corresponding subset of the conditional moments. Their unconditional density is approximated via conditioning on the mean and setting a large horizon such as $\tau = 5$ years.

3.5 Identification

Under the cross-sectional averaging assumption, the population parameters (θ_0, h_0) defined by (18) also minimize the concentrated version of the population likelihood. This section discusses conditions such that the structural parameters ϑ are identified from the expected quasi-maximum likelihood

$$\begin{aligned}\bar{\ell}^{c,pair}(\vartheta) &= \lim_{N\to\infty} E\left(\ell^{c,pair}_{t+1}(\vartheta)\right) \\ &= \lim_{N\to\infty} E_{\vartheta_0}\left(N^{-1}\|z_{t+1}\left(\vartheta, \bar{s}_{t+1}(\vartheta)\right) - Rz_t\left(\vartheta, \bar{s}_t(\vartheta)\right)\|^2_{\Omega_\tau}\right) + E_{\vartheta_0}\left(N^{-1}\|z_t\left(\vartheta, \bar{s}_t(\vartheta)\right)\|^2_{\Omega_\infty}\right).\end{aligned}$$

The conditions are general and apply to other settings where latent variables are backed out from cross-sections of asset prices, as is common in option pricing and term structure applications. Related conditions are in Pastorello et al. (2003), who study identification based on the transition density of the profiled state. We use information from the measurement equations, which makes it possible to identify structural parameters that do not appear in the transition density. The following assumption suffices to identify the structural parameters from the concentrated quasi-maximum likelihood $\bar{\ell}^{c,pair}(\vartheta)$:

Assumption 4. *For every pair $\vartheta \neq \vartheta'$ there exists some s, such that*

$$\psi(\bar{s}(\vartheta, s)) \neq \psi(\bar{s}(\vartheta', s)) \quad \text{or} \quad \mathbf{g}(\bar{s}(\vartheta, s), X_t, \vartheta) \neq \mathbf{g}(\bar{s}(\vartheta', s), X_t, \vartheta').$$

Assumption 4 rules out that different parameters lead to the same prices, which requires that the implied states cannot respond to a change in parameters in such a way that the price remains unchanged. In general, this is a high level invertibility condition, but it has been studied in the context where the state variables are observable. In particular, given the policy functions, the uncertainty aversion is identified from the price-dividend ratio if H is the only function that satisfies the integral equation

$$E\left(\frac{P_t}{D_t} \mid s_t\right) = E_t\left(\int_0^\infty e^{-\delta\tau} \frac{H(s_{t+\tau})}{H(s_t)} \frac{\psi^{dc}(s_{t+\tau})}{\psi^{dc}(s_t)} d\tau\right)$$

Sufficient conditions are studied in Christensen (2017) using uniqueness of positive eigenfunctions, and in Chen and Ludvigson (2009) in terms of completeness of the conditional density weighted by ψ^{dc}. Since H is only identified up to scale, it is normalized by setting the constant coefficient at 1.

3.6 Consistency

Define the product space $\mathit{\Theta} = \Theta \times \mathcal{H}$, where Θ is a finite-dimensional parameter space, and $\mathcal{H} = \prod_m \mathcal{H}_{\psi_m} \times \mathcal{H}_H$ is a Cartesian product of infinite-dimensional parameter spaces for the policy functions ψ_m and the uncertainty aversion function H. Let the spaces $\mathcal{H}_m$ and $\mathcal{H}_H$ 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 $|\lambda| = \sum_{l=1}^d \lambda_1$, and $D^\lambda = \frac{\partial^{|\lambda|}}{\partial y_1^{\lambda_1} \cdots \partial y_d^{\lambda_d}}$ the partial derivative operator, it is given for some positive integers r and p by

$$\|g\|_{r,p} = \left\{\sum_{|\lambda| \leq r} E\left(D^\lambda g(S)\right)^p\right\}^{1/p}. \tag{29}$$

For vector-valued functions define $\|g\|_{r,p} = \sum_{m=1}^K \|g_m\|_{r,p}$. Instead of maximizing $\ell^c_{NT}(\theta)$ 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 and integer p, define the functional space

$$\mathcal{H}^p = \left\{ g : \mathbb{R}^D \mapsto \mathbb{R} : \|g\|_{r,p}^p \leq B \right\} \tag{30}$$

All functions in the compact space $\mathcal{H}^p$ have at least r partial derivatives that are bounded in expectation. For $p = 2$ the polynomials in this space can be conveniently characterized in terms of their coefficients. Let $\underline{p}_L = (p_1(w), ..., p_L(w))$ be a set of basis functions, and consider the finite-dimensional series approximator $g_L(w) = \sum_{l=1}^{L} \gamma_j p_j(w) = \underline{\gamma} \cdot \underline{p}_L(w)$. Define

$$\Lambda_L = \sum_{|\lambda| \leq m} E\left(D^\lambda \underline{p}_L(z) D^\lambda \underline{p}_L(z)^T \right), \tag{31}$$

which implies that $g_L(w) \in \mathcal{H}^2$ if and only if $\gamma^T \Lambda_L \gamma \leq B$ (Newey and Powell, 2003). Therefore the optimization in (24) is redefined over the compact finite-dimensional subspace $\mathcal{H}^2_{L(T)}$:

$$(\hat{\theta}^{c,pair}, \hat{h}^{c,pair}) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}_{L(T)}} \ell_{TN}^{c,pair}(\theta, h),$$

where $\mathcal{H}^2_{L(T)}$

$$\mathcal{H}^2_{L(T)} = \left\{ g(w) = \sum_{l=1}^{L(T)} \gamma_j p_j(w) : \ \gamma^T \Lambda_{L(T)} \gamma \leq B \right\}.$$

Also define the Sobolev sup-norm

$$\|g\|_{r,\infty} = \max_{|\lambda| \leq r} \sup_{z} \left| D^\lambda g(z) \right|. \tag{32}$$

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).

Let $\hat{\Sigma} := (\hat{\Sigma}_M, \hat{\Sigma}_P)$ be initial estimators of the covariance matrices $\Sigma := (\Sigma_M, \Sigma_P)$, and $\hat{R} = e^{-\hat{A}}$ be initial estimators of the autocorrelation parameters. Plugging $\hat{\Sigma}$ and $\hat{R}$ into (3.6) yields the feasible quasi-maximum likelihood objective

$$\hat{\ell}_{TN}^{c,pair}(\theta, h) = \ell_{TN}^{c,pair}(\theta, h; \hat{\Sigma}, \hat{R}). \tag{33}$$

This is the objective of two-step estimators which do not involve simultaneous maximization over possibly high-dimensional covariance matrices.

Consider the following set of assumptions:

Assumption 5.

a) The parameter space $\varTheta = \Theta \times \mathcal{H}$ is compact, and the population quasi-maximum likelihood $\bar{\ell}^c(\theta, h)$is uniquely maximized at the interior point $\vartheta_0 = (\theta_0, h_0)$.

b) $\hat{\Sigma}_N$ and Σ_N are positive definite, Σ_N is bounded, and $\hat{\Sigma}_N - \Sigma_N \xrightarrow{a.s.} 0$; R is bounded, and $\hat{R}_N - R_N \xrightarrow{a.s.} 0$.

c) (X_t, z_t) is a strong mixing stationary process, with $E\left(\|z_t\|^2\right) < \infty$

d) The pricing functions satisfy for every $i = 1, ..., N$,

$$|g_i(s_t, X_{it}, \vartheta) - g_i(s_t, X_t, \tilde{\vartheta})| \leq b(s_t, X_t)\|\vartheta - \tilde{\vartheta}\|^v$$

for some $v > 0$ with $E\left(b(s_t, X_t)^2\right) < \infty$, and $\operatorname{Var}\left(g_i(s_t, X_t, \vartheta_0)\right) < \infty$. The transition density satisfies

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

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

Theorem 1. *Under Assumptions 1-5, the maximizer $(\hat{\theta}, \hat{h}_L)$ of (33) satisfies*

$$\hat{\theta} \xrightarrow{p} \theta_0,$$
$$\|\hat{h}_L - h_0\|_{r,\infty} \xrightarrow{p} 0,$$

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

3.7 Inference

Ackerberg et al. (2012) show that when the unknown functions are approximated by the sieve method, then for M-estimators inference based on assuming the approximating order is correct is numerically equivalent to inference taking into account the presence of the infinite-dimensional parameters. This provides a basis for performing standard parametric-style inference that is relatively simple to implement. While the numerical equivalence of standard errors only considers the parametric part of the model, its application to the sieve coefficients is justified when the approximation up to a certain order is correct. This is a parametric specification hypothesis that can be tested for sequentially. This section therefore derives the asymptotic distribution of the parameters and sieve coefficients under the assumption that the order of the polynomial is correctly specified.

Let $c_L = (c_\psi, c_H)$ denote the coefficients of the L degree polynomial approximations to ψ and H, and let $\vartheta_L = (\theta, c_L)$ be the combined parameters. Let ∂_ϑ and $\partial_{\vartheta\vartheta'}$ be the first and second partial derivative operator with respect to ϑ, respectively. The score vector and the Hessian for observation t are then given by $\partial_{\vartheta_L}\ell_t^c(\theta_L)$ and $\partial_{\vartheta\vartheta'}\ell_t^c(\vartheta)$.

The following additional regularity conditions are required for the asymptotic normality of $\hat{\vartheta}_L$:

Assumption 6.

a) $E\left(\sup_\alpha \left\|\partial_{\alpha\alpha'} \log f(s' \mid s; \alpha)\|t\right) < \infty$, $E\left(\left\|\partial_\alpha \log f(s' \mid s; \alpha_0)\right\|^4\right) < \infty$, *and for* $i = 1, \ldots, N$ $E\left(\sup_\vartheta \left\|\partial_{\vartheta\vartheta'} g_i(s_t, X_t, \vartheta_0)\right\|\right) < \infty$ *and* $E\left(\left\|\partial_\vartheta g_i(s_t, X_t, \vartheta_0)\right\|^4\right) < \infty$

b) $E(\partial_{\vartheta\vartheta'}\ell_t^c(\vartheta_0))$ *is non-singular*

c) $E(\varepsilon_{it} \mid \mathcal{F}_{t-1}) = 0$, *and*

$$E\left(\left(\mathbf{w}_t^T \varepsilon_t\right)^4 \mid s_t, X_t\right)^{\frac{1}{2}} = O_p\left(\frac{1}{N}\right)$$

Theorem 2. *Let Assumptions 1 - 6 hold. When* $N, T \to \infty$, *and* $\frac{T}{N} \to \kappa$ for some $0 < \kappa < \infty$, *then*

$$\sqrt{NT}(\hat{\vartheta} - \vartheta_0) \xrightarrow{d} \mathcal{H}_0^{-1} \times \mathcal{N}\left(\kappa E(\bar{B}_t), \mathcal{V}_0\right),$$

where

$$\begin{aligned}
\mathcal{V}_0 &= \lim_{N\to\infty} \frac{1}{N} E\left(\partial_\vartheta \varepsilon_t(\vartheta_0)^T \Omega_N^{-1} \Omega_{tN} \Omega_N^{-1} \partial_\vartheta \varepsilon_t(\vartheta_0)\right) \\
\Omega_{tN} &= E\left(\varepsilon_t \varepsilon_t^T \mid s_t, s_{t-1}\right) \\
\mathcal{H}_0 &= E\left(\partial_{\vartheta\vartheta'} \bar{\ell}(\vartheta_0, s_{p,t}(\vartheta_0))\right) + E\left(\partial_{\vartheta s_p'} \bar{\ell}(\vartheta_0, s_{p,t}(\vartheta_0)) \partial_{\vartheta'} s_{p,t}(\vartheta_0)\right).
\end{aligned}$$

The expression for the bias term $\bar{B}_t$ is given in the appendix. If the cross-section is large compared to the time series, then κ is close to zero and the bias is small, and vice versa for a relatively small cross-section. Under conditional homoskedasticity $\mathcal{V}_0$ simplifies to

$$\lim_{N\to\infty} \frac{1}{N} E\left(\partial_\vartheta \varepsilon_t(\vartheta_0)^T \Omega_N^{-1} \partial_\vartheta \varepsilon_t(\vartheta_0)\right).$$

Sample counterparts $\hat{\mathcal{V}}$ and $\hat{\mathcal{H}}$ based on the outer product of the gradient and the Hessian matrix can be used to estimate the standard errors of the parameters, and to construct Wald or Lagrange-Multiplier test statistics for hypotheses on the coefficients.

4 Empirical Results

4.1 Data

Output and consumption data are obtained from the Bureau of Economic Analysis of the U.S. Department of Commerce. 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, and scaled to match the average total real consumption expenditure. The sample combines annual data starting from January 1929 with quarterly data starting from January 1947 and ending in December 2016. Monthly observations of the Industrial Production Index starting from 1929 obtained from the Federal Reserve are used to construct a proxy for economic uncertainty.

Aggregate stock market prices and dividends are based on the S&P 500 index obtained from CRSP. The individual price-dividend ratios are constructed from 10 Size and 10 Book-to-Market portfolios obtained from Kenneth French's Data Library. All prices and dividends are expressed in real terms using the price index for U.S. gross domestic product. Dividends per unit are computed from the difference in value-weighted returns with (R^d_{t+1}) and without (R^x_{t+1}) dividends:

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

Price-dividend ratios are then computed as

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

and dividends from

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

The initial aggregate dividend $D^m_1 = C_1$ is normalized to aggregate consumption, and the initial portfolio dividends normalized by their relative market share $D_{i1} = \frac{M_{i1}}{\sum_j M_{j1}} D^m_1$. The constructed dividend series are equivalent to reinvesting intermediate cash payments in the underlying stock (Cochrane, 1992).

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

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

with $\overline{\Delta ip}_t$ the rolling window annual mean. The realized stock market variance (RV) is similarly constructed from daily log returns R_{t+1} with demeaning at the quarterly frequency as[8]

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

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

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

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

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

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

4.2 Economic Uncertainty and Stock Market Volatility

Figure 5 shows the historical series of the measure of economic uncertainty (REV) and the measure of financial market volatility (RV). While both measures of variation peaked during the Great Depression and the Great Recession, there are several well-known episodes during which they diverged. The first postwar decades saw substantial economic uncertainty but historically calm financial markets. On the other hand, during the LTCM collapse of 1998 stock market volatility peaked while production growth remained largely unchanged. In recent years, stock market volatility has plummeted while economic uncertainty remains relatively high. Figure 6

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

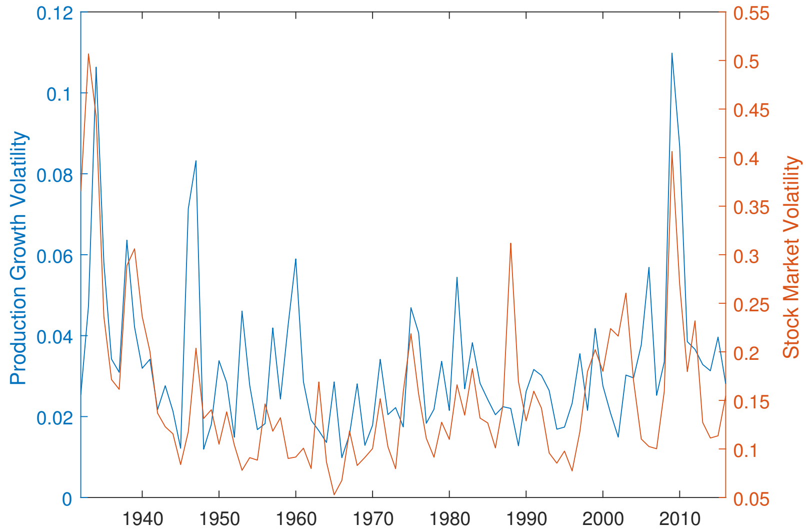


Figure 5: Realized Variance of Industrial Production growth and of S&P 500 returns from 1930-2016.

compares REV with the dividend-consumption ratio. It displays substantial comovement between the series, with the dividend-consumption ratio starting high during the uncertain 1930s, reaching historical lows during the post war recovery period, and steadily rising again during and after the Great Recession. This suggests output uncertainty affects not just the scale but also the level of consumption and dividend growth. In particular, dividends must grow faster than consumption when uncertainty increases in order to explain the behaviour of the dividend-consumption ratio. To understand the different channels via which economic and financial market uncertainty affect asset prices, Figures 7 and 8 plot the annual percentage change in REV and RV, respectively, against the annual log return on output, consumption, dividends, and the S&P 500 Index from 1930 to 2016. Figure 7 suggests a negative relation between uncertainty shocks and output and consumption growth, in line with the evidence in Bloom (2009) and Nakamura et al. (2017). The market return also goes down contemporaneously when uncertainty increases in line with the well known leverage effect. Dividend growth, on the other hand, does not go down and even is slightly convexly increasing in changes to uncertainty, in line with rebalancing of the dividend-consumption ratio. Figure 8 shows the impact of changes in financial market volatility on the same response variables. While the direction of the responses is the same as for changes in economic uncertainty, the realized stock market variance correlates much stronger with dividend growth and the market return and much weaker with consumption and output growth. In particular, dividend growth is pronounced convexly increasing and the market return convexly decreasing in

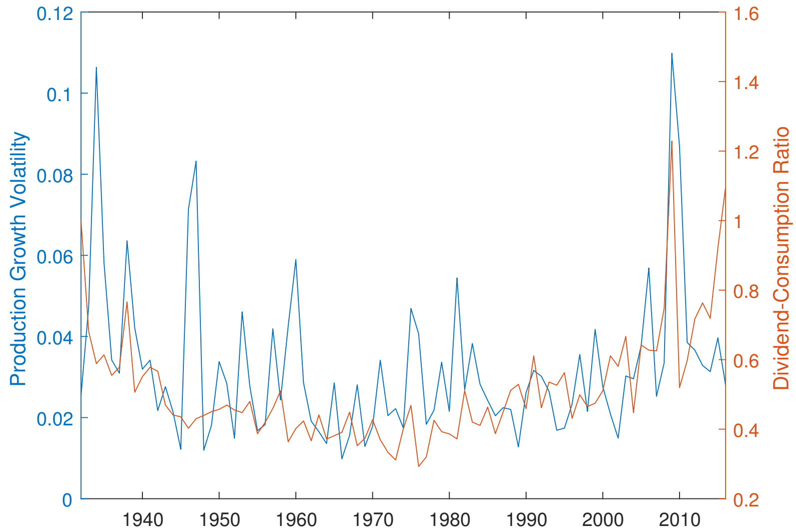


Figure 6: Realized Variance of Industrial Production growth versus Dividend-Consumption ratio ($DC_{1930} = 1$) from 1930-2016.

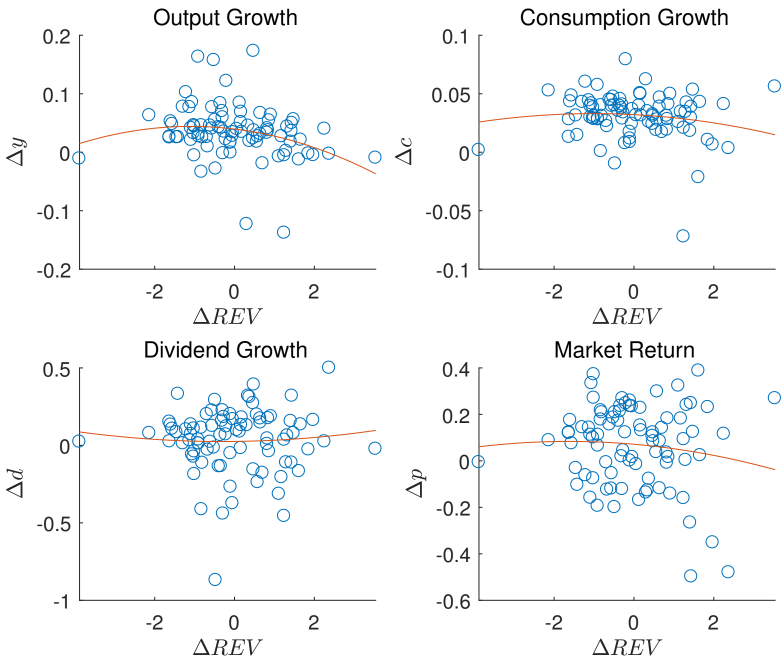


Figure 7: Annual change in log Realized Economic Variance (REV) of Industrial Production growth versus annual log return on Output, Consumption, Dividends, and the S&P 500 Index from 1930-2016. Fitted line corresponds to a quadratic fit.

changes in the realized variance. This provides further evidence against a simple linear relation between economic and financial uncertainty, which would not explain the different impact they have on fundamentals. Figures 14 and 15 in the Appendix show that the same patterns can be

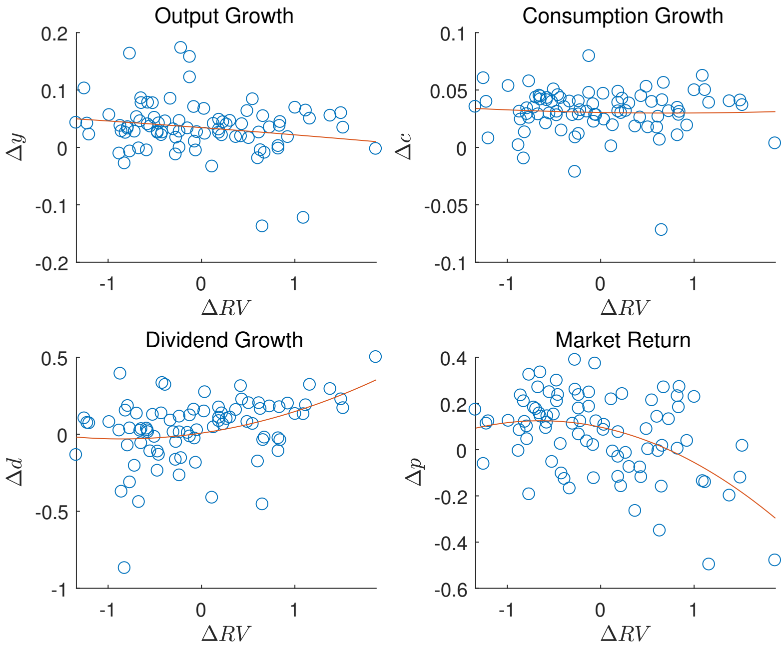


Figure 8: Annual change in log Realized Variance (RV) versus Annual log return on Output, Consumption, Dividends, and the S&P 500 Index from 1930-2016. Fitted line corresponds to a quadratic fit.

observed when quarterly instead of annual data is used.

4.3 Cross-sectional heterogeneity

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

$$\frac{D_{it}}{D_t} = \alpha_{0i} + \beta_i{}^T REV_t + z^d_{it}, \qquad E(z^d_{it} \mid REV_t) = 0. \tag{34}$$

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

Table 2: Parameter estimates and standard errors of regression (34) using annual data from 1926-2016 using one-lag feasible generalized least squares.

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

times, whereas smaller firms tend to reduce them. A possible explanation is the real options theory of investment which suggests that large firms invest less when uncertainty is high and can afford to pay out cash to their shareholders. Small firms have less cash reserves to fall back on, which could force them to cut dividends in uncertain times. This heterogeneity in individual dividend-consumption ratios generates heterogeneous impact of the portfolio price-dividend ratios in response to uncertainty shocks. This is illustrated in Figure 9, which compares the price-dividend ratios as a function of output volatility for small and big firms, at the estimated parameters. Larger firm are expected to pay more dividends in uncertain times, which is priced by uncertainty averse investors, making large firm equity relatively expensive. Moreover large firm price-dividend ratios are more sensitive to changes in uncertainty, as their more countercyclical dividend payouts induce larger variations in expected dividend growth, which is reflected in their prices. As a result, the spread between large and small firm price-dividend ratios helps to identify uncertainty beyond the level of the price-dividend ratios.

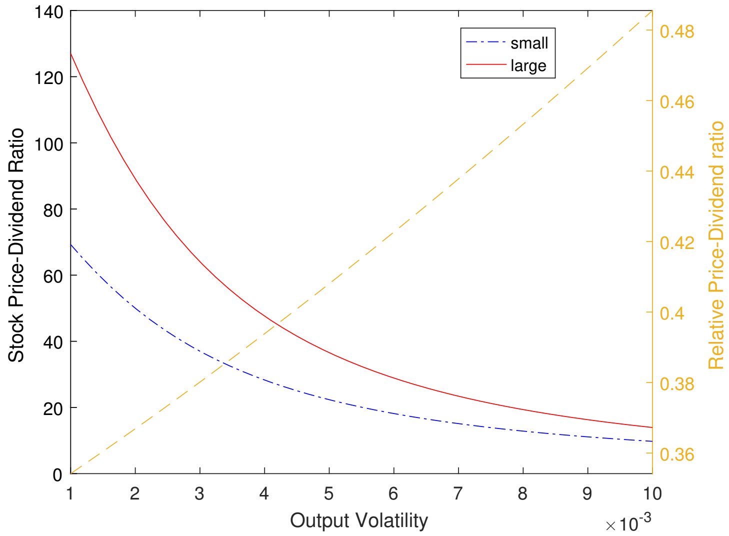
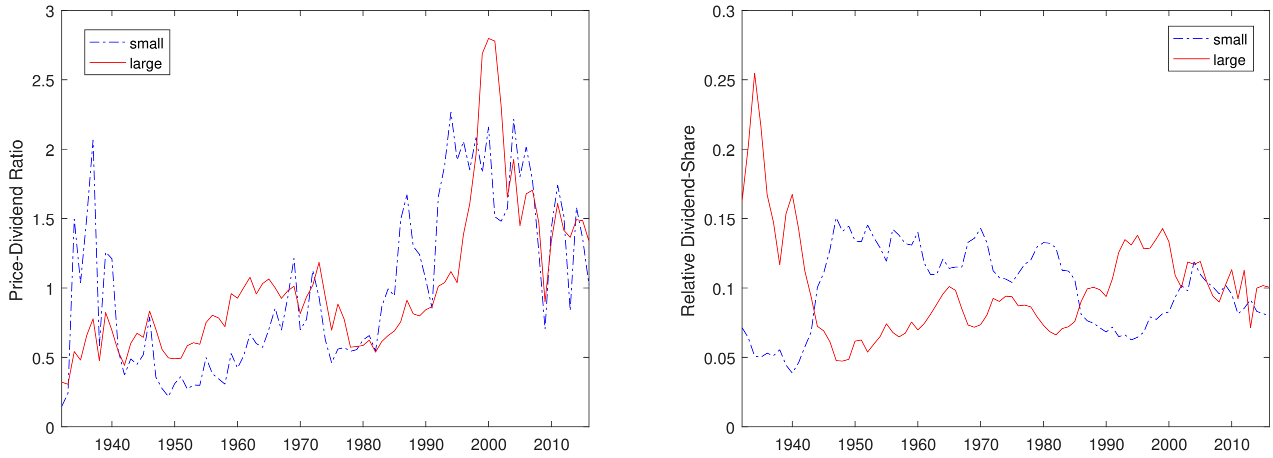


Figure 9: Price-dividend ratios as function of output volatility for the 10% smallest and largest firms, and their ratio, based on estimated parameters.

Figure 10 shows the time series of the price-dividend ratios and the dividend shares of the stock portfolio corresponding to the smallest and largest size decile. The price-dividend ratios of the small firms have a strong negative correlation of -0.52 with its dividend share, suggesting an important role for the mean-reverting error to drive expected dividend growth. The price-dividend ratios of large firms are more smooth, reflecting their ability to smooth dividends. During several episodes, in particular in the 80s and 90s, the price-dividend ratios moved substantially apart even when their dividend shares did not. Small firms were more vulnerable to the increased uncertainty in the aftermath of the 1987 stock market crash, but recovered sharply when markets calmed. The peak in large-firm valuation ratios during the dot-com bubble is difficult to explain by uncertainty alone but can be attributed to high risk appetite.



(a) Price-Dividend Ratios.

(b) Relative Dividend Share.

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

The estimated coefficients are used as initial estimates in the full estimation algorithm, and updated for each set of estimated states. The proxy $\hat{V}_t = REV_t$ for the unobserved economic uncertainty V_t is similarly used to construct initial estimates of the policy functions in (16). When the proxies are consistent estimators of the underlying states, the structural parameters can be consistently estimated albeit typically with a bias due to the measurement error. This approach is particularly useful when using high frequency financial data, for which appropriate limit theory is derived in Li and Xiu (2016). The proxy for economic uncertainty is made up from low frequency observations, so that a direct estimation likely involves some non-negligible biases, but is nevertheless useful to get initial estimates of the parameters and the error distribution. In particular, the quasi-maximum likelihood set out in subsection 3.3 leads to the feasible GLS regressions

$$M_{it} - \hat{R}_i M_{it-1} = \sum_{l=0}^{L} c_l (\hat{V}_t^l - \hat{R}_i \hat{V}_{t-1}^l) + \eta_{it},$$

where $\hat{R}_i$ is an initial estimator of the autocorrelation of the noise process Z_{it}. The next section includes further observations in the form of asset prices that improve estimating the latent states.

4.4 Estimates

This section reports estimated parameters for the feasible quasi-maximum likelihood estimator (33) and a second-order expansion of the policy functions and the pricing kernel. The multiplicative form of the asset price means that conditional moments up to the sixth order are being used in the approximation. The parameters of the error distribution are initialized based on the residuals of an initial estimation based on the proxy REV_t. Under the Ornstein-Uhlenbeck specification these errors are normally distributed and the parameters (A, Σ) in (25) are estimated by maximum likelihood.

Table 3 reports the estimates of the parameter vector $\theta = (\delta, \theta_s)$ that include the discount rate δ and the transition density parameters θ_s with their standard errors. The discount rate of 0.02 is equivalent to a time discount parameter of 0.98 and is accurately estimated. The triple (κ, θ, ω) captures the mean reversion, steady state, and volatility of output growth volatility V_t. The estimates confirm this is a quickly mean-reverting process as κ is large relative to ω. In particular, the Feller condition $2\kappa\theta > \omega^2$ is satisfied, which guarantees the positivity of the volatility process. The large negative correlation parameter $\rho = -0.6$ suggests a pronounced leverage effect between adverse economic shocks and economic uncertainty, akin to the financial leverage effects in stock returns. The negative value for λ suggests that moreover increases in economic uncertainty lead to lower subsequent economic growth. The triple $(\kappa^Q, \theta^Q, \eta)$ captures the mean reversion, steady state, and volatility of the inverse consumption surplus ratio Q_t that describes changing relative risk aversion. The steady state level is reasonable as it corresponds locally to a constant relative risk aversion of about five. The inverse consumption surplus ratio is slightly more volatile than economic uncertainty, but reverts slightly quicker. The impact r_y of output shocks on risk aversion is negative, in line with models of external habit formation that generate countercyclical risk premia. Unlike pure habit models however, the presence of a non-negligible pure preference component proportional to η suggests the fundamental shocks and consumption surplus ratio are only imperfectly correlated; similar findings are in Bekaert et al. (2009). Nonetheless, the hypothesis that $\eta = 0$ cannot be rejected at standard confidence values.

Figure 11 shows the estimated uncertainty aversion index $\hat{H}_L$ for $L = 2$, together with a 95% pointwise confidence intervals. The estimated uncertainty aversion index is increasing and slightly convex. This suggests that uncertainty is a priced state variable beyond its impact on consumption growth. This is related to findings that stock market volatility is priced, which explains the presence of variance risk premia. However, the shape of the aversion against stock

Table 3: Estimates and standard errors (in brackets) of the discount rate and transition density parameters.

Estimates based on mixed frequency data, with annual observations from 1926 to 1946 and quarterly observations from 1947 to 2016.

δ	κ	θ	ω	ρ	μ	λ	κ^Q	θ^Q	η	r_y
0.02	0.38	$3 * 10^{-3}$	0.08	-0.60	0.04	-1.98	0.55	5.45	0.21	-2.86
(0.00)	(0.19)	$(1.7 * 10^{-3})$	(0.10)	(0.12)	(0.01)	(0.73)	(0.67)	(0.03)	(0.84)	(2.25)

market versus macroeconomic volatility need not be the same in line with their nonlinear relation as seen in Figure 4. This could explain the evidence in Song and Xiu (2016) from variance derivatives that investors are primarily concerned about either very high or very low levels of stock market volatility, creating a U-shaped pricing kernel. In my model, the aversion towards low levels of stock market volatility would merely be an artefact of not controlling for economic uncertainty.

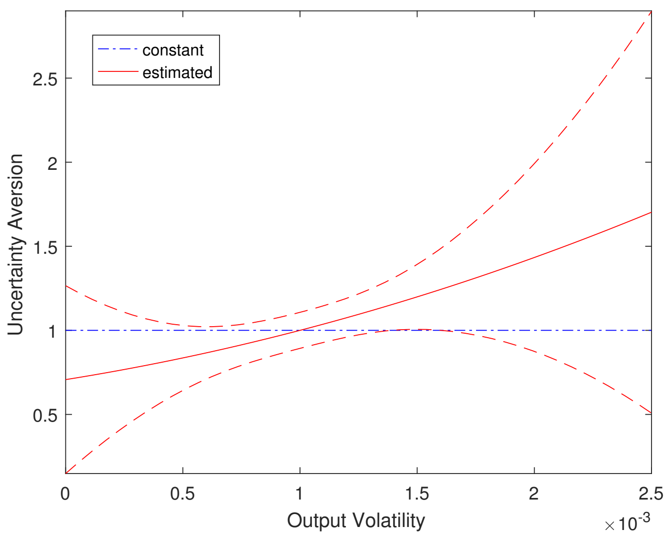


Figure 11: Estimated uncertainty aversion index $\hat{H}_L$ for $L = 2$ against a constant value of one, together with 95% pointwise confidence intervals. Estimates are based on annual observations from 1926 to 1946 and quarterly observations from 1947 to 2016.

Figure 12 shows the estimated states using the pairwise concentration procedure. During most of the post war economic expansion the estimated uncertainty and relative risk aversion are relatively smooth and are moderately negative correlated. Uncertainty has been steadily declining with the exception of heightened levels around the 1980s energy crisis. The risk aversion measure shows strong mean reversion towards a level around 5.5 up to and including the 1980s. However, during the 1990s it started a steady decline, reaching historical lows prior to the burst of the technology bubble in the early 2000s. The period after that risk aversion showed a strong upward trajectory that lasts until the most recent observations. Economic uncertainty did not experience a similar trend but does show an extreme peak in the aftermath of the 2008 financial crisis. A

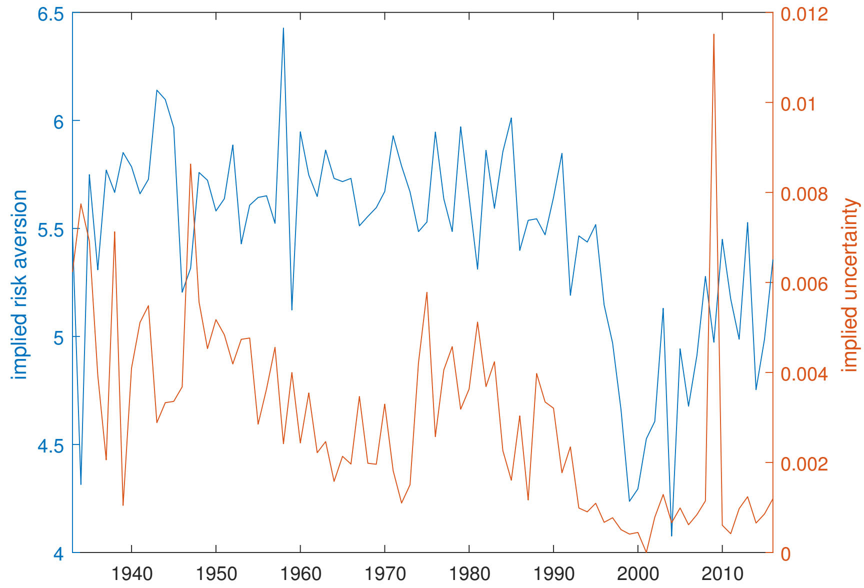


Figure 12: Concentrated states $\hat{s}'_t = (\hat{V}'_t, \hat{Q}'_t)$ over the period 1930-2016. Time t states are estimated using observations from dates $t-1$ and t.

useful measure of the reliability of the pairwise concentrated states is the proximity of the states estimated from the two overlapping pairs of cross-sections. Figure 16 in the Appendix shows that the difference in the implied states based on using the current and lagged cross-section or the current and future cross-section is of a smaller magnitude than the variation of the states over time. This suggests the states are generally well identified from single cross-sections, as adding information from leads or lags does not make a major difference. Figure 13 shows the estimated expected dividend-consumption ratio as a function of the states. It shows a strong convexity, with the dividend-consumption ratio initially barely responsive or even declining in output growth volatility, while strongly increasing when output growth volatility reaches higher levels.

Figure 17 in the Appendix shows the consumption-to-output ratio as a function of the states. It shows a moderate decline in consumption per unit of output when uncertainty or risk aversion increases, with a slightly negative interaction effect. Unlike the dividend-consumption ratio, it is less responsive to the states and does not show strong evidence for nonlinearity.

Figures 18, 19, and 20 in the Appendix show the time-series fitted values of the macroeco-

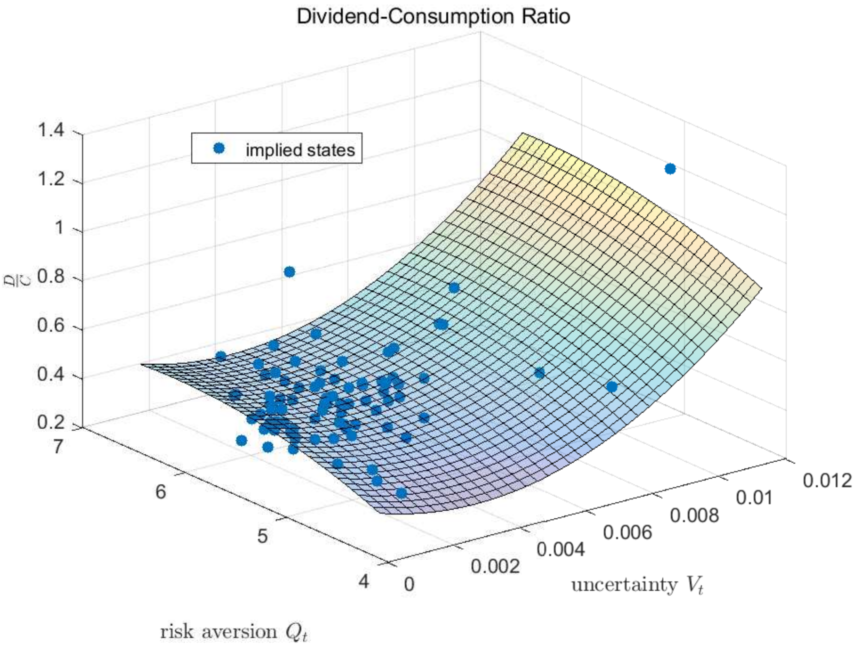


Figure 13: Fitted dividend-consumption ratio as a function of the estimated states using a second-order bivariate expansion. Time t states are estimated using observations from dates $t-1$ and t.

nomic ratios, the price-dividend ratio, and the realized stock market variance, respectively, using the estimated states. The residual processes display clear first-order autocorrelation, but their prediction errors do not, nor are there strong violations of non-Gaussianity apart from during the 1987 and 2008 stock market crashes. The time-series fitted values of the quarterly price-dividend ratio in Figure 19 show that even with only two state variables, the model is able to match the peaks and troughs in the data, including periods of sharp declines such as the 2008 crisis. The dividend-consumption residual, however, is important in explaining, for example, the high price-dividend ratios prior to the dotcom-bubble, which coincided with relatively low dividend payouts. The model cannot fully explain the very low valuation ratios during the 50s and 70s, closely related to the equity premium puzzle, for which institutional issues such as limited access to stock markets may have prevented the cheap stock prices from being corrected.

5 Conclusion

This paper develops a class of nonlinear Markovian asset pricing models in which the dynamics of consumption and dividends per unit of output is described via general policy functions of latent state variables, with a focus on understanding the impact of economic uncertainty and risk aversion on economic outcomes. The state variables are recovered from cross-sections of price-dividend ratios and proxies for real output uncertainty and financial market volatility. The policy functions are estimated to be consistent with asset prices, given a semiparametric specification of the stochastic discount factor. Tractable closed-form expressions for price-dividend ratios and financial volatility are obtained under polynomial approximations of the policy functions and the pricing kernel. The paper establishes the consistency and asymptotic normality of a profile maximum likelihood estimator for the general case where the distribution of observables conditional on unobservables is unknown, but many observations on functions of the same state variables are available. This setting is typical in asset pricing contexts, and can be naturally extended to other types of cross-sectional pricing data such as bonds or derivatives. The paper proposes a practical pairwise profiling procedure that takes into account the persistence of dividend-consumption and price-dividend ratios, and shows that the state variables are accurately estimated from two cross-sections of asset prices. The expected consumption-dividend ratio is found to be an increasing and convex function of economic uncertainty, and the positive impact of uncertainty on dividends is greater for large than for small firms. Together with a moderately uncertainty averse representative investor this leads to steeply declining price-dividend ratios for moderate levels of economic uncertainty, which can explain episodes of large stock market volatility that occurred during periods of moderate economic uncertainty. Low risk aversion played an important role in the build-up of the dotcom bubble and the period preceding the Global Financial Crisis. Finally, the non-monotonic relation between economic uncertainty and stock market volatility helps to explain why variance risk premia are small on average, but peak during highly volatile crisis periods.

References

Ackerberg, D., Chen, X., and Hahn, J. (2012). A practical asymptotic variance estimator for two-step semiparametric estimators. *Review of Economics and Statistics*, 94(2):481–498.

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

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

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

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

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

Andreasen, M. M. and Christensen, B. J. (2015). The sr approach: A new estimation procedure for non-linear and non-gaussian dynamic term structure models. *Journal of Econometrics*, 184(2):420–451.

Andrews, D. W. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric theory*, 4(3):458–467.

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

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

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

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

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

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

Bekaert, G., Engstrom, E., and Xing, Y. (2009). Risk, uncertainty, and asset prices. *Journal of Financial Economics*, 91(1):59–82.

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

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

Calin, O. L., Chen, Y., Cosimano, T. F., and Himonas, A. A. (2005). Solving asset pricing models when the price–dividend function is analytic. *Econometrica*, 73(3):961–982.

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.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.

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

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

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

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

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

Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, pages 289–314.

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

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

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 14(1).

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

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

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407.

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

Dalderop, J. (2018). Nonparametric filtering of conditional state-price densities. Working paper.

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

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

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

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

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

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

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312.

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

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

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

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

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

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

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

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

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

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

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

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

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6):1152–1191.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

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.

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

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

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

Judd, K. L. (1996). Approximation, perturbation, and projection methods in economic analysis. *Handbook of computational economics*, 1:509–585.

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.

Li, J. and Xiu, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica*, 85(4).

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.

Pastorello, S., Patilea, V., and Renault, E. (2003). Iterative and recursive estimation in structural nonadaptive models. *Journal of Business & Economic Statistics*, 21(4):449–509.

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

Sarafidis, V. and Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, 31(5):483–531.

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

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

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

Taylor, J. B. and Uhlig, H. (1990). Solving nonlinear stochastic growth models: a comparison of alternative solution methods. *Journal of Business & Economic Statistics*, 8(1):1–17.

Wachter, J. A. (2005). Solving models with external habit. *Finance Research Letters*, 2(4):210–226.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25.

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

A Appendix

A.1 Proofs.

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

$$\left.\frac{d\mathbb{Q}^T}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \frac{e^{-\log Y_T}}{E_t^{\mathbb{P}}(e^{-\log Y_T})}. \tag{35}$$

The affine property of (y_t, s_t) implies that

$$\tilde{p}^T(t) = E_t^{\mathbb{P}}(e^{-\gamma(y_T - y_t)}) = e^{\alpha^T(t;\gamma)+\beta^T(t;\gamma)\cdot s_t}, \tag{36}$$

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

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

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

$$K_0^{\mathbb{Q}^T}(t) = K_0 + H_0\beta^T(t), \qquad K_1^{\mathbb{Q}^T}(t) = K_1 + H_1\beta^T(t). \tag{37}$$

For the benchmark model with $s_t = (V_t, h_t)$, the differential equations can be explicitly solved as $\beta = (\beta_V, \beta_h)$, where $\beta_h = 0$, and letting $\lambda_1, \lambda_2 = \frac{\kappa+\omega\rho\gamma\pm\sqrt{\Delta}}{(1-\rho^2)\omega^2}$ and $\Delta = (\kappa + \omega\rho\gamma)^2 - (1 -$

$\rho^2)\omega^2(-2\lambda\gamma+\gamma^2)$,

$$\beta_V^T(t)=\frac{\lambda_1\lambda_2(1-e^{-\sqrt{\Delta}\tau})}{\lambda_1-\lambda_2 e^{-\sqrt{\Delta}\tau}} \tag{38}$$

$$\alpha^T(t)=(-\mu\gamma+\lambda_2\kappa\theta)\tau+\kappa\theta\frac{2}{(1-\rho^2)\omega^2}\log\left(\frac{\lambda_1-\lambda_2}{\lambda_1-\lambda_2 e^{-\sqrt{\Delta}\tau}}\right), \tag{39}$$

so that under $\mathbb{Q}^T$ the variables solve the s.d.e.

$$d\log Y_t=\left(\mu-\left(\frac{1}{2}+\lambda+\beta_V^T(t)\right)V_t\right)dt+\sqrt{V}_t dW_t$$
$$dV_t=\kappa\left(\theta-(1-\frac{\omega}{\kappa}\beta_V^T(t))V_t\right)dt+\omega\sqrt{V}_t dB_t.$$

Write

$$E_t^{\mathbb{P}}(e^{-\gamma(y_T-y_t)}s_t^j)=\tilde{p}^T(t)E_t^{\mathbb{Q}^T}(s_t^j), \tag{40}$$

where by the affine property the conditional $\mathbb{Q}^T$-moments are given by

$$E_t^{\mathbb{Q}^T}\left(\underline{s}_T^L \mid s_t=s\right)=\exp\left(\int_t^T A_L(\tau)d\tau\right)\underline{s}^L. \tag{41}$$

provided $A_L(\tau_1)$ and $A_L(\tau_2)$ are commutative for any $\tau_1\neq\tau_2$, otherwise the Magnus series expansion can be used. The time-varying matrix $A_L(\tau)$ has a known expression that involves

$$\int_t^T \beta_V(\tau)\tau=\lambda_1(T-t)-\frac{\lambda_1-\lambda_2}{\sqrt{\Delta}}\log\left(\lambda_1-\lambda_2 e^{-\sqrt{\Delta}(T-t)}\right). \tag{42}$$

Derivations used for the risk-free rate and the expected excess return. The innovations of the pricing kernel under log utility ($\gamma=1$) follow from Itô's Lemma applied to $\zeta_t=e^{-\delta t}\frac{Q_t}{C_t}H(V_t)$:

$$\frac{d\zeta_t}{\zeta_t}=-\delta dt-\frac{dC_t}{C_t}+\frac{dQ_t}{Q_t}+\frac{dH(V_t)}{H(V_t)}+\frac{d\langle C_t,Q_t\rangle}{C_tQ_t}+\frac{d\langle C_t,H(V_t)\rangle}{C_tH(V_t)}+\frac{d\langle Q_t,H(V_t)\rangle}{Q_tH(V_t)}.$$

Therefore

$$r_t^f=-\frac{1}{dt}E\left(\frac{d\zeta_t}{\zeta_t}|s_t\right)=\delta+\frac{1}{dt}E_t\left(\frac{dC_t}{C_t}\right)+\kappa^q\left(\frac{\theta^q}{Q_t}-1\right)+\frac{1}{H(V_t)}\left(\kappa H'(V_t)(\theta-V_t)+\frac{1}{2}H''(V_t)\omega^2V_t\right)$$
$$+\text{ covariance terms.}$$

The innovations of the pricing kernel follow from demeaning the stochastic terms:

$$\frac{d\zeta_t}{\zeta_t} = -r_t^f dt - \frac{dC_t - E_t(dC_t)}{C_t} + \frac{1}{\sqrt{Q_t}}\left(\eta dB_t^q + r_y\sqrt{V}_t dW_t\right) + \frac{H'(V_t)\sqrt{V_t}}{H(V_t)}\omega dB_t,$$

which, after applying Itô's Lemma to the consumption policy function, can be rearranged to yield the prices of risk given in (12).

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

Table 4: Coefficient matrix A mapping the Itô generator of the second-degree moments of $(\log Y_t, V_t, Q_t)$ into itself.

0	0	0	0	0	0	0	0	0	0
μ	0	λ^V	0	0	0	0	0	0	0
$\kappa\theta$	0	-κ	0	0	0	0	0	0	0
$\kappa^q\theta^q$	0	λ^V	-κ^q	0	0	0	0	0	0
0	2μ	1	0	0	$2\lambda^V$	0	0	0	0
0	$\kappa\theta$	$\mu + \omega\rho$	0	0	-κ	0	λ^V	0	0
0	$\kappa^q\theta^q$	$r_y + \omega r_v \rho$	μ	0	λ^V	-κ^q	0	λ^V	0
0	0	$2\kappa\theta + \omega^2$	0	0	0	0	-2κ	0	0
0	0	$\kappa^q\theta^q + \omega^2 r_v/2 + \omega(\omega r_v + r_y\rho)/2 + \omega r_y \rho/2$	$\kappa\theta$	0	0	0	λ^V	- κ - κ^q	0
0	0	$r_y(r_y + \omega r_v\rho) + \omega r_v(\omega r_v + r_y\rho)$	$2\kappa^q\theta^q + \sigma_q^2$	0	0	0	0	$2\lambda^V$	-$2\kappa^q$

Quadratic variation of the price-dividend ratio for multivariate affine diffusions. Itô's Lemma for multivariate diffusions yields that

$$d\phi_M(s_t) = (\nabla\phi_M(s_t))^T\sigma(s_t)dW_t + \mu_t dt,$$

for some drift term μ_t that does not contribute to the spot variance. The increment of the quadratic variation of the approximated price-dividend ratio for affine diffusions is

$$d\langle\phi_M(s_t)\rangle = \nabla\phi_M(s_t)^T\left(H_0 + \sum_{j=1}^{D} H_j s_j\right)\nabla\phi_M(s_t)dt,$$

where $\mathrm{Cov}(ds_t) = H_0 + \sum_{j=1}^{D} H_j s_j$ for affine models.

Proof of Proposition 1. Denote $\nabla_{(S',S)}f = \left(\frac{\partial}{\partial S'}, \frac{\partial}{\partial S}\right)^T$ as the gradient and $H_{(S',S)}f = \nabla_{(S',S)}(\nabla_{(S',S)}f)^T$ as the Hessian of a twice differentiable function f with respect to the future and current state variables. For the below argument drop the dependence of $(\hat{\mathcal{S}}_{t+1}(\theta,h),\hat{\mathcal{S}}_t(\theta,h))$ on the parameter vector (θ,h). Combine the current and future state variables into the vector $\mathcal{S}_{t:t+1}=(\mathcal{S}_{t+1},\mathcal{S}_t)^T$. Define the time-$t$ likelihood conditional on its state

$$l_t(\theta,h,\mathcal{S})=\frac{1}{N}\log f_{\theta,h}\left(\mathcal{Y}_t|\mathcal{S}_t=\mathcal{S}\right),$$

and write Bayes' rule for time-t observations as

$$f_{\theta,h}\left(\mathcal{S}_t|\mathcal{Y}_t\right)=\frac{\exp(Nl_t(\theta,h,\mathcal{S}_t))f_{\theta,h}\left(\mathcal{S}_t\right)}{f_{\theta,h}\left(\mathcal{Y}_t\right)}.$$

Denote the objective function for the choice of state variables as

$$q_{t+1}(S',S;\theta,h)=l_{t+1}(\theta,h,\mathcal{S}',\mathcal{S})+\frac{1}{N}\log f_{\theta,h}\left(\mathcal{Y}_t|\mathcal{S}\right)+\frac{1}{N}\log f_{\theta,h}\left(\mathcal{S}\right).$$

A Laplace approximation of the limited information integrated likelihood $f_{\theta,h}\left(\mathcal{Y}_{t+1}|\mathcal{Y}_t\right)$ around the pair of concentrated state variables shows that

$$\begin{aligned}
f_{\theta,h}\left(\mathcal{Y}_{t+1}|\mathcal{Y}_t\right)&=\iint\exp\left(Nl_{t+1}(\theta,h,\mathcal{S}',\mathcal{S})+\log f_{\theta,h}(\mathcal{S}|\mathcal{Y}_t)\right)d\mathcal{S}'d\mathcal{S}\\
&=\iint\exp\left(Nq_{t+1}(S',S;\theta,h)\right)d\mathcal{S}'d\mathcal{S}\,f_{\theta,h}\left(\mathcal{Y}_t\right)\\
&=\exp\left(Nq_{t+1}(\hat{\mathcal{S}}'_{t+1},\hat{\mathcal{S}}_t;\theta,h)\right)f_{\theta,h}\left(\mathcal{Y}_t\right)\left(1+O_p\left(\frac{1}{N}\right)\right)\\
&\quad\iint\exp\left(\frac{N}{2}(\mathcal{S}_{t:t+1}-\hat{\mathcal{S}}_{t:t+1})^T H_{(S',S)}\Big|_{\hat{\mathcal{S}}_{t:t+1}}q_{t+1}(S',S;\theta,h)(\mathcal{S}_{t:t+1}-\hat{\mathcal{S}}_{t:t+1})\right)d\mathcal{S}'d\mathcal{S}\\
&=\exp\left(Nq_{t+1}(\hat{\mathcal{S}}'_{t+1},\hat{\mathcal{S}}_t;\theta,h)\right)f_{\theta,h}\left(\mathcal{Y}_t\right)\left(\frac{2\pi}{T}\right)^D\\
&\qquad\left|-NH_{(S',S)}\Big|_{\hat{\mathcal{S}}_{t:t+1}}q_{t+1}(S',S;\theta,h)\right|^{-\frac{1}{2}}\left(1+O_p\left(\frac{1}{N}\right)\right),
\end{aligned}$$

with $|\cdot|$ the matrix determinant. It follows that

$$\frac{1}{N}\log f_{\theta,h}\left(\mathcal{Y}_{t+1}|\mathcal{Y}_t\right)=l^c_{t+1}(\theta,h)+\frac{1}{N}\log\left(\frac{2\pi}{N}\right)^D-\frac{1}{2N}\log\left|-H_{(S',S)}\Big|_{\hat{\mathcal{S}}_{t:t+1}}q_{t+1}(S',S;\theta,h)\right|+O_p\left(\frac{1}{N^2}\right),$$

which implies the first part of the proposition.

The second part follows from performing the same Laplace approximation around the pair of states

$$\left(\hat{\mathcal{S}}'_{t+1|t}(\theta,h),\hat{\mathcal{S}}_{t|t}(\theta,h)\right)=\arg\max_{S',S} l_{t+1}(\theta,h,\mathcal{S}',\mathcal{S})+\frac{1}{N}\log f_{\theta,h}\left(\mathcal{Y}_t|\mathcal{S},\mathcal{F}^Y_{t-1}\right)+\frac{1}{N}\log f_{\theta,h}\left(\mathcal{S}|\mathcal{F}^Y_{t-1}\right). \tag{43}$$

From the assumed probability limit of the implied state it follows that $\hat{\mathcal{S}}'_{t+1|t}(\theta,h)\overset{\text{a.e.}}{=}\tilde{\mathcal{S}}'_{t+1}(\theta,h)\overset{\text{a.e.}}{=}\hat{\mathcal{S}}'_{t+1}(\theta,h)$, with $\tilde{\mathcal{S}}'_{t+1}(\theta,h)$ as in (21), as their respective objections functions agree on leading terms. In particular the first part of objective functions (3.2) and (43) coincide under the first-order Markov assumption. Similarly $\hat{\mathcal{S}}'_{t+1|t}(\theta,h)\overset{\text{a.e.}}{=}\tilde{\mathcal{S}}'_{t+1}(\theta,h)\overset{\text{a.e.}}{=}\hat{\mathcal{S}}'_{t+1}(\theta,h)$. Hence the updating densities are evaluated at the same state asymptotically, and the conclusion follows from the continuity of the concentrated likelihoods. □

Proof of Theorem 1. The proof is based on Lemma A1 in Newey and Powell (2003). Let $Q'(\theta)=l^c(\theta)$ and $Q(\theta)=E(l^c_t(\theta))$. This requires that (i) there is unique θ_0 that minimizes $Q'(\theta)$ 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, $Q'(\theta)$ is compact, and $\max_{\theta\in\Theta}|Q'_T(\theta)-Q'(\theta)|\xrightarrow{p}0$.

The identification condition (i) follows from subsection 3.5, together with the positive definiteness of Σ_t. 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 pricing function and the transition density. Their continuity carries over to the profiled $l^c(\theta)$ by the maximum theorem. 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 condition in Assumption 5. Stochastic equicontinuity follows from Assumption 5. □

Proof of Theorem 2. Introduce the notation:

$$U_t(\vartheta)=\partial_\vartheta\ell_t(\vartheta,s_{p,t}(\vartheta))$$

$$\mathcal{S}_t(\vartheta)=\partial_{s_p}\ell_t(\vartheta,s_{p,t}(\vartheta))$$

$$\mathcal{H}_t(\vartheta)=\partial_{s_ps_p}\ell_t(\vartheta,s_{p,t}(\vartheta))$$

Furthermore let $U_t^{s_p}(\vartheta)$ and $U_t^{s_p s_p}(\vartheta)$ denote the first and second derivative of $U_t(\vartheta)$ with respect to s_p. The parameter argument is suppressed when evaluated at $\vartheta = \vartheta_0$.

A mean value expansion around the parameter ϑ_0 yields

$$\partial_\vartheta \ell^c(\vartheta_0) + \partial_{\vartheta\vartheta'} \ell^c(\tilde{\vartheta})(\hat{\vartheta} - \vartheta_0) = 0. \tag{44}$$

for some $\tilde{\vartheta}$ that lies in between ϑ_0 and $\hat{\vartheta}$. Rearranging yields

$$\sqrt{NT}(\hat{\vartheta} - \vartheta_0) = -\left(\partial_{\vartheta\vartheta'} \ell^c(\tilde{\vartheta})\right)^{-1} \sqrt{NT} \partial_\vartheta \ell^c(\vartheta_0). \tag{45}$$

The uniform convergence $\sup_\vartheta \left| \frac{1}{T} \sum_{t=1}^{T-1} \partial_{\theta\theta'} \ell_{t+1}^c(\tilde{\theta}) - E(\partial_{\vartheta\vartheta'} \ell^c(\vartheta_0)) \right| \xrightarrow{p} 0$ follows from stationarity, mixing, and the bounded supremum Assumption 6a.). Here

$$\begin{aligned} E\left(\partial_{\vartheta\vartheta'} \ell^c(\vartheta_0)\right) &= E\left(d_\vartheta \partial_\vartheta \ell(\vartheta_0, s_{p,t}(\vartheta_0))\right) \\ &= E\left(\partial_{\vartheta\vartheta'} \ell(\vartheta_0, s_{p,t}(\vartheta_0))\right) + E\left(\partial_{\vartheta s_p'} \ell(\vartheta_0, s_{p,t}(\vartheta_0)) \partial_{\vartheta'} s_{p,t}(\vartheta_0)\right), \end{aligned}$$

where d_ϑ denotes the total derivative with respect to ϑ.

Let $\hat{s}_{t+1}^p(\vartheta) = (\hat{s}_{t+1}'(\vartheta), \hat{s}_t(\vartheta))$, which satisfies for every t the estimating equation

$$0 = \partial_{s_p} \ell_{t+1}(\vartheta, \hat{s}_{p,t}(\vartheta)).$$

Similarly, a mean value expansion of $\hat{s}_{p,t}(\vartheta)$ around $s_{p,t}(\vartheta)$ yields

$$\sqrt{N}\left(\hat{s}_{p,t}(\vartheta) - s_{p,t}(\vartheta)\right) = -\left(\partial_{s_p s_p} \ell_{t+1}(\vartheta, \tilde{s}_{p,t}(\vartheta))\right)^{-1} \sqrt{N} \mathcal{S}_t. \tag{46}$$

for some $\tilde{s}_{p,t}(\vartheta)$ that lies in between $\hat{s}_{p,t}(\vartheta)$ and $s_{p,t}(\vartheta)$ for every ϑ. Write

$$\partial_\vartheta \ell^c(\vartheta) = \frac{1}{T} \sum_{t=1}^{T-1} \partial_\vartheta \ell_{t+1}^c(\vartheta) = \frac{1}{T} \sum_{t=1}^{T-1} \partial_\vartheta \ell_{t+1}(\vartheta, \hat{s}_{p,t}(\vartheta)), \tag{47}$$

using the envelope theorem. A second-order Taylor expansion around the true states $s_{p,t}(\vartheta_0) = s_{p,t}$ yields for $l = 1, \ldots, \dim \vartheta$

$$\partial_{\vartheta_l} \ell_{t+1}(\vartheta_0, \hat{s}_{p,t}(\vartheta_0)) = U_{tl} + U_{tl}^{s_p T}(\hat{s}_{p,t}(\vartheta_0) - s_{p,t}) + \frac{1}{2}(\hat{s}_{p,t}(\vartheta_0) - s_{p,t})^T U_{tl}^{s_p s_p^T}(\hat{s}_{p,t}(\vartheta_0) - s_{p,t}) + R_{3t}. \tag{48}$$

Plugging in the expansion (48) into (47) and rescaling yields

$$\sqrt{NT}\partial_\vartheta \ell^c(\vartheta_0) = V + B + R_3, \tag{49}$$

with leading variance term

$$V = \sqrt{\frac{N}{T}} \sum_{t=1}^{T-1} U_{tN},$$

a bias term whose l-th element equals

$$\begin{aligned} B_l &= \sqrt{\frac{N}{T}} \sum_{t=1}^{T-1} \left(\frac{1}{2} \mathcal{S}_{t+1}^T \mathcal{H}_{t+1}^{-1} U_{tl}^{s_p s_p^T} - U_{tl}^{s_p^T} \right) \mathcal{H}_{t+1}^{-1} \mathcal{S}_{t+1} \\ &= \frac{1}{T} \sqrt{\frac{T}{N}} \sum_{t=1}^{T-1} \left(\frac{1}{2} \frac{\mathcal{S}_{t+1}^T}{\sqrt{N}} \mathcal{H}_{t+1}^{-1} U_{tl}^{s_p s_p^T} - \frac{U_{tl}^{s_p^T}}{\sqrt{N}} \right) \mathcal{H}_{t+1}^{-1} \frac{\mathcal{S}_{t+1}}{\sqrt{N}} \end{aligned}$$

and the remainder $R_3 = \sqrt{\frac{N}{T}} \sum_{t=1}^{T-1} R_{t3} = o_p(1)$. Denote

$$\begin{aligned} \bar{\mathcal{H}}_{t+1} &= \operatorname*{plim}_{N\to\infty} \mathcal{H}_{t+1}, & \bar{I}_t &= \operatorname*{plim}_{N\to\infty} N E_t \left(\mathcal{S}_{t+1} \mathcal{S}_{t+1}^T \right), \\ \bar{U}_{tl}^{s_p s_p^T} &= \operatorname*{plim}_{N\to\infty} U_{tl}^{s_p s_p^T}, & \bar{I}_t^u &= \operatorname*{plim}_{N\to\infty} N E_t \left(U_t^{s_p} \mathcal{S}_{t+1}^T \right), \end{aligned}$$

where the subscript t denotes conditioning on s_{pt}. The summands of B_l are then recognized as quadratic forms with probability limit

$$\bar{B}_{tl} = \operatorname{tr}\left(\bar{\mathcal{H}}_{t+1}^{-1} \left(\frac{1}{2} \bar{U}_{tl}^{s_p s_p^T} \bar{\mathcal{H}}_{t+1}^{-1} \bar{I}_t - \bar{I}_t^u \right) \right). \tag{50}$$

It follows that $B = \kappa E(\bar{B}_t) + O_p\left(\frac{1}{T}\right)$.

The summands of the variance term are given by

$$U_{tN} = \frac{1}{N} \left(-2 \left(\partial_\vartheta \hat{\varepsilon}_t(\vartheta_0) \right)^T \hat{\Sigma} \hat{\varepsilon}_t(\vartheta_0) + \partial_\vartheta \log f\left(s_t | s_{t-1}; \vartheta_0 \right) \right) \tag{51}$$

Using $\hat{\varepsilon}_t - \varepsilon_t = (\hat{A} - A)\varepsilon_{t-1}$, write

$$\begin{aligned} \hat{\varepsilon}_t^T \hat{\Omega}^{-1} \hat{\varepsilon}_t - \varepsilon_t^T \Omega^{-1} \varepsilon_t &= (\hat{A} - A)^T \varepsilon_{t-1}^T \hat{\Omega} (\hat{A} - A) \varepsilon_{t-1} \\ &\quad + 2\varepsilon_t^T \hat{\Omega}^{-1} (\hat{A} - A) \varepsilon_{t-1} \\ &\quad + \varepsilon_{t-1}^T (\hat{\Omega}^{-1} - \Omega^{-1}) \varepsilon_t \\ &= o_p(N) \end{aligned}$$

by the triangle inequality and the consistency of $\hat{A}$ and $\hat{\Omega}$. The limiting distribution of the score vector is therefore determined by

$$U_{tN}^{*} = \frac{1}{\sqrt{N}}\left(-2\left(\partial_{\vartheta}\varepsilon_t(\vartheta_0)\right)^T \Omega^{-1}\varepsilon_t(\vartheta_0) + \partial_{\vartheta}\log f\left(s_t|s_{t-1};\vartheta_0\right)\right) =: \xi_{tN,1} + \xi_{tN,2} \tag{52}$$

where $\xi_{tN,1}$ is a linear combination of the prediction errors, and $\xi_{tN,2}$ is the score of the transition density. From Corollary 3.1 in Hall and Heyde (1980) it follows that $V \overset{\text{a.e.}}{=} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{tN}^{*} \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_0)$ provided the following two conditions hold:

i) for all $\varepsilon > 0$, $\frac{1}{T}\sum_{t=1}^{T} E_{t-1}\left(\|U_{tN}^{*}\|^2 1_{\|U_{tN}^{*}\| \geq \epsilon\sqrt{T}}\right) \xrightarrow{p} 0$

ii) $\frac{1}{T}\sum_{t=1}^{T} E_{t-1}\left(\|U_{tN}^{*}\|^2\right) \xrightarrow{p} \mathcal{V}_0$

To prove the Lindeberg condition i), note

$$E_{t-1}\left(\|U_{tN}^{*}\|^2 1_{\|U_{tN}^{*}\| \geq \varepsilon\sqrt{T}}\right) \leq E_{t-1}\left(\|U_{tN}^{*}\|^4\right)^{\frac{1}{2}} \mathbb{P}_{t-1}\left(\|U_{tN}^{*}\| \geq \varepsilon\sqrt{T}\right)^{\frac{1}{2}},$$

by the Cauchy-Schwartz inequality, where

$$\mathbb{P}_{t-1}\left(\|U_{tN}^{*}\| \geq \varepsilon\sqrt{T}\right) \leq \mathbb{P}_{t-1}\left(\|\xi_{tN,1}\| \geq \frac{1}{2}\varepsilon\sqrt{T}\right) + \mathbb{P}_{t-1}\left(\|\xi_{tN,2}\| \geq \frac{1}{2}\varepsilon\sqrt{T}\right),$$

The first probability satisfies

$$\mathbb{P}_{t-1}\left(\|\xi_{tN,1}\| \geq \frac{1}{2}\varepsilon\sqrt{T}\right) = \mathbb{P}_{t-1}\left(\left\|\frac{1}{\sqrt{N}}\xi_{tN,1}\right\| \geq \frac{1}{2}\varepsilon\sqrt{\frac{T}{N}}\right) \leq \frac{4N}{\varepsilon^2 T}\mathrm{tr}\left(\mathrm{Var}\left(\frac{1}{\sqrt{N}}\xi_{tN,1}\right)\right)$$

by the Chebyshev inequality for random vectors, where

$$\begin{aligned}\mathrm{tr}\left(\mathrm{Var}\left(\frac{1}{\sqrt{N}}\xi_{tN,1}\right)\right) &= 4\sum_{j=1}^{\dim\vartheta} E\left(\left(\frac{1}{N}\partial_{\vartheta_j}\varepsilon_t(\vartheta_0)^T\Omega^{-1}\varepsilon_t\right)^2\right)\\ &=: 4\sum_{j=1}^{\dim\vartheta} E\left(\mathrm{Var}\left(w_{tN,j}^T\varepsilon_t \mid s_{pt}\right)\right) \to 0\end{aligned}$$

since $w_{tN,j} = \frac{1}{N}\Omega^{-1}\partial_{\vartheta_j}\varepsilon_t(\vartheta_0) = \frac{1}{N}\Omega\partial_{\vartheta_j}\left(z_{t+1}(\vartheta, s_{t+1}) - Rz_t(\vartheta, s_t)\right)$ satisfies the granularity Assumption 3 using boundedness of the weighting matrix Ω and Assumption 5d) on the pricing function. Similarly, the second probability satisfies

$$T\mathbb{P}\left(\|\xi_{tN,2}\| \geq \frac{1}{2}\varepsilon\sqrt{T}\right) \leq \frac{4E\left(\|\partial_{\vartheta}\log f\left(s_t|s_{t-1};\vartheta_0\right)\|^2\right)}{N\varepsilon^2} \to 0,$$

using Assumption 5d) on the transition density. Meanwhile

$$E_{t-1}\left(\|U_{tN}^*\|^4\right) \leq 8\left(E_{t-1}\left(\|\xi_{tN,1}\|^4\right) + E_{t-1}\left(\|\xi_{tN,2}\|^4\right)\right)$$

by the triangle inequality, where

$$\begin{aligned}
E_{t-1}\left(\|\xi_{tN,1}\|^4\right) &= \sum_{i=1}^{\dim\vartheta}\sum_{j=1}^{\dim\vartheta} E_{t-1}\left(\left(\frac{1}{\sqrt{N}}\partial_{\vartheta_i}\varepsilon_t(\vartheta_0)^T\Omega^{-1}\varepsilon_t\right)^2\left(\frac{1}{\sqrt{N}}\partial_{\vartheta_j}\varepsilon_t(\vartheta_0)^T\Omega^{-1}\varepsilon_t\right)^2\right) \\
&\leq \sum_{i=1}^{\dim\vartheta}\sum_{j=1}^{\dim\vartheta} E_{t-1}\left(\left(\frac{1}{\sqrt{N}}\partial_{\vartheta_i}\varepsilon_t(\vartheta_0)^T\Omega^{-1}\varepsilon_t\right)^4\right)^{\frac{1}{2}} E_{t-1}\left(\left(\frac{1}{\sqrt{N}}\partial_{\vartheta_j}\varepsilon_t(\vartheta_0)^T\Omega^{-1}\varepsilon_t\right)^4\right)^{\frac{1}{2}} \\
&= O_p(1),
\end{aligned}$$

by the Cauchy-Schartz inequality and Assumption 6c), and

$$E_{t-1}\left(\|\xi_{tN,2}\|^4\right) = \frac{1}{N^2}E_{t-1}\left(\|\partial_\vartheta \log f\left(s_t|s_{t-1};\vartheta_0\right)\|^4\right) \to 0,$$

by Assumption 6a). It follows that $E_{t-1}\left(\|U_{tN}^*\|^4\right) = O_p(1)$, while $\mathbb{P}_{t-1}\left(\|U_{tN}^*\| \geq \varepsilon\sqrt{T}\right) = o_p(1)$, which together implies i).

For ii), an application of the weak law of large numbers for mixingale arrays in Andrews (1988, Theorem 2) to the stationary strong mixing array $U_{tN}^* {U_{tN}^*}^T$ implies that

$$\frac{1}{T}\sum_{t=1}^{T} U_{tN}^* {U_{tN}^*}^T \xrightarrow{p} \mathcal{V}_0,$$

where

$$\begin{aligned}
\mathcal{V}_0 &= \lim_{N\to\infty} E\left(U_{tN}^* {U_{tN}^*}^T\right) = \lim_{N\to\infty} E\left(\xi_{tN,1}{\xi_{tN,1}}^T\right) \\
&= \lim_{N\to\infty}\frac{1}{N}E\left(\partial_\vartheta\varepsilon_t(\vartheta_0)^T\Omega^{-1}\varepsilon_t\varepsilon_t^T\Omega^{-1}\partial_\vartheta\varepsilon_t(\vartheta_0)\right) \\
&= \lim_{N\to\infty}\frac{1}{N}E\left(\partial_\vartheta\varepsilon_t(\vartheta_0)^T\Omega^{-1}E\left(\varepsilon_t\varepsilon_t^T \mid s_t, s_{t-1}\right)\Omega^{-1}\partial_\vartheta\varepsilon_t(\vartheta_0)\right).
\end{aligned}$$

The result now follows from Slutzky's theorem. □

A.2 Further empirical evidence

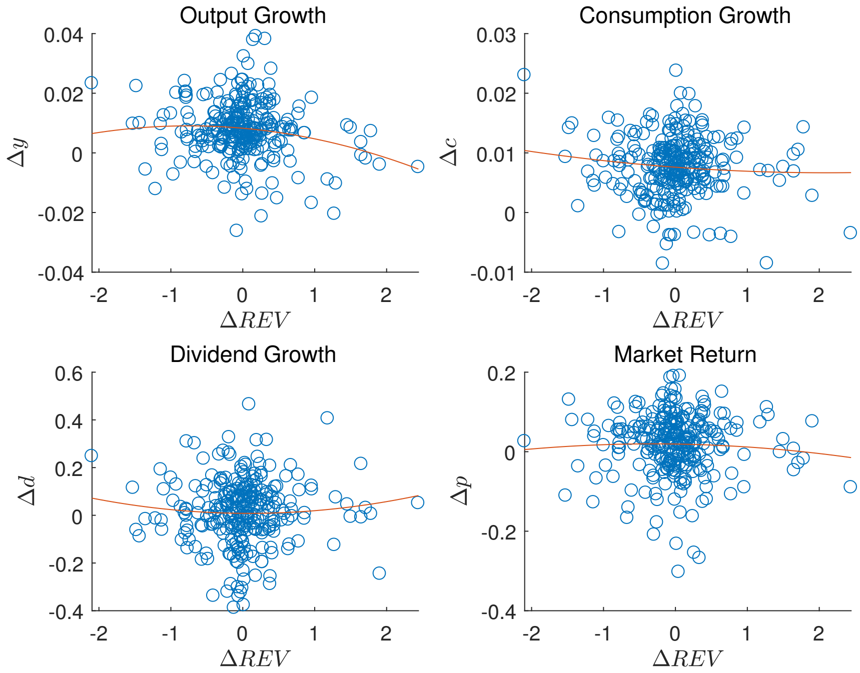


Figure 14: Quarterly changes in log Realized Economic Variance (REV) of Industrial Production growth versus annual log return on Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016.

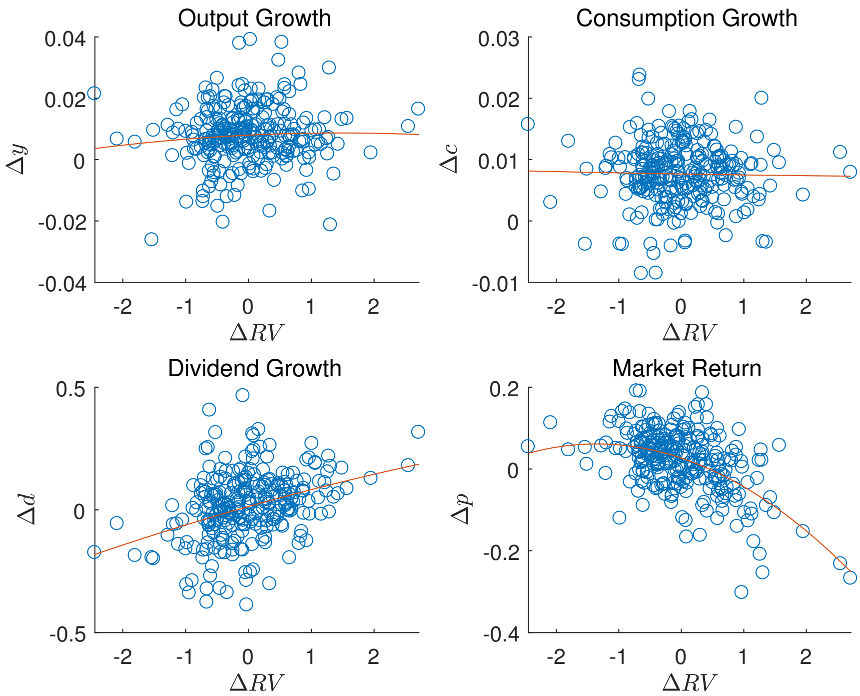
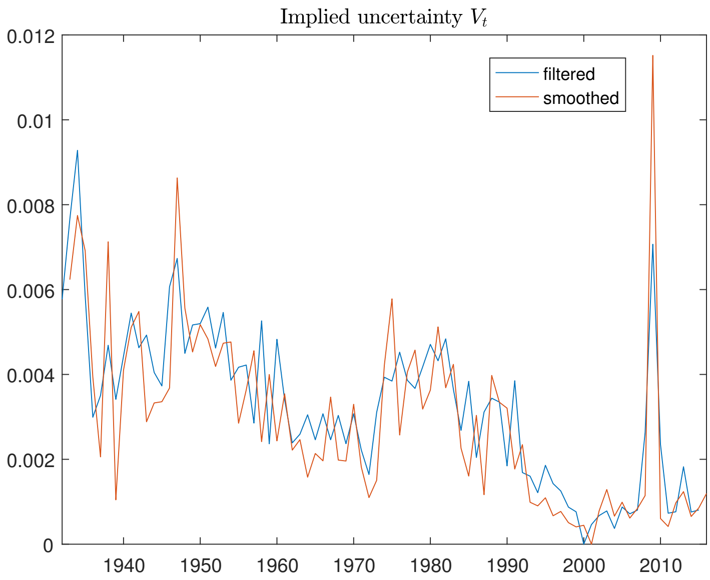
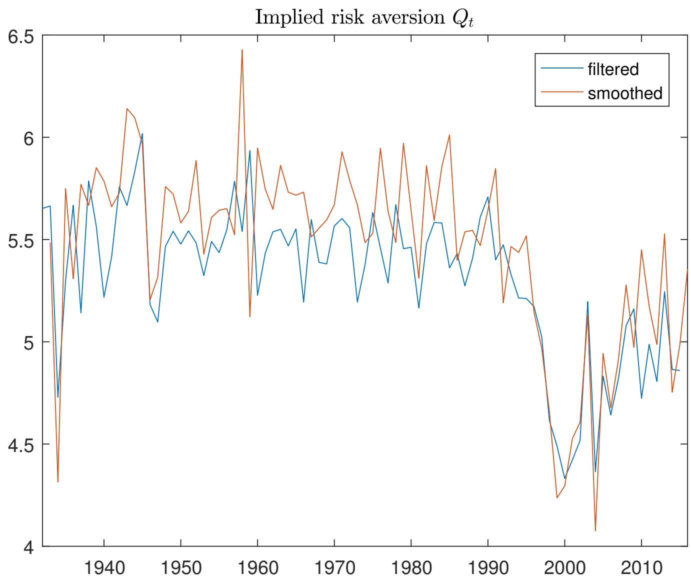


Figure 15: Quarterly changes in log Realized Variance (RV) versus Annual log return Output, Consumption, Dividends, and the S&P 500 Index from 1947-2016.



(a) Implied uncertainty.



(b) Implied risk aversion.

Figure 16: Comparison of the pairwise concentrated 'filtered' states $\hat{s}'_t$ and 'smoothed' states $\hat{s}_t$. Filtered time t states use observations from dates $(t, t+1)$; smoothed time t states use observations from dates $(t-1, t)$.

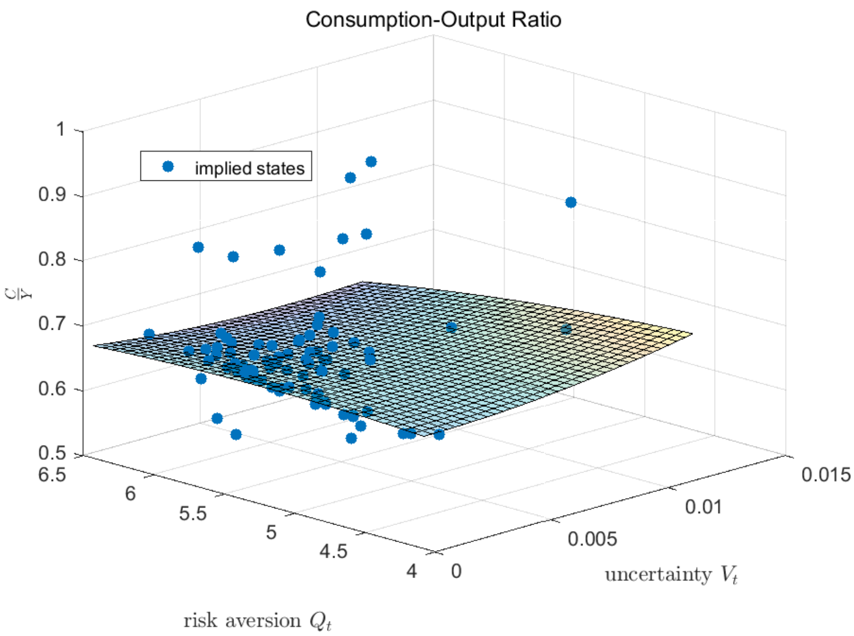
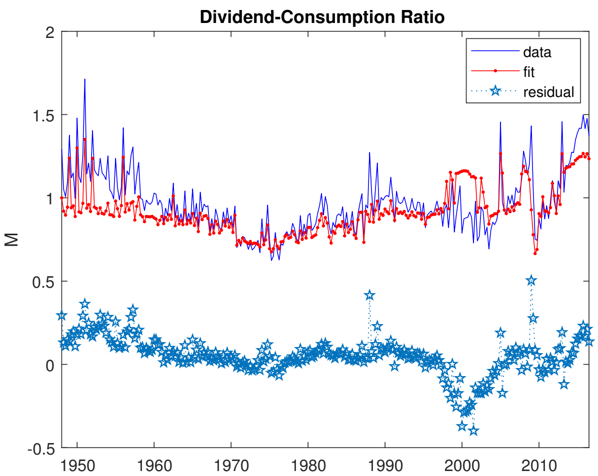
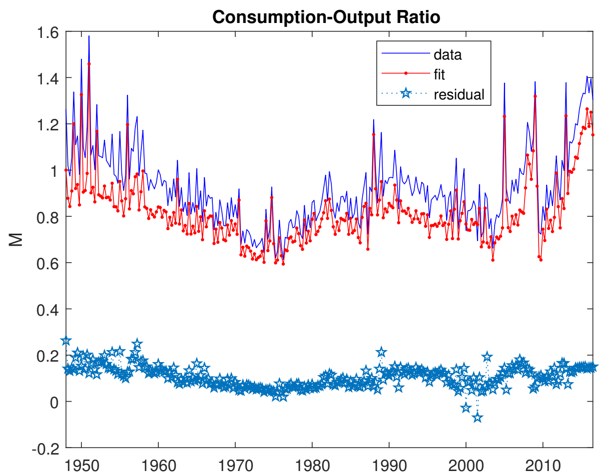


Figure 17: Fitted consumption-output ratio as a function of the estimated states using a second-order bivariate expansion. Time t states are estimated using observations from dates $t-1$ and t.



(a) Dividend-Consumption ratio.



(b) Consumption-Output ratio.

Figure 18: Time-series fitted values of the quarterly macroeconomic ratios using the estimated states and a second-order bivariate expansion.

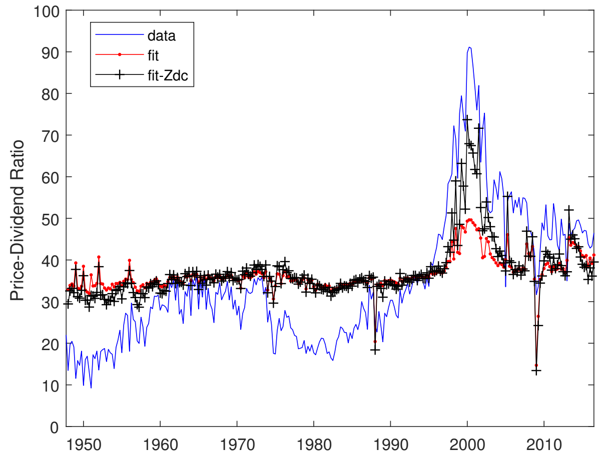


Figure 19: Time-series fitted values of the quarterly price-dividend ratio as a function of the estimated states using a second-order bivariate expansion, with and without correcting for the dividend-consumption residual $\hat{Z}_t^d$.

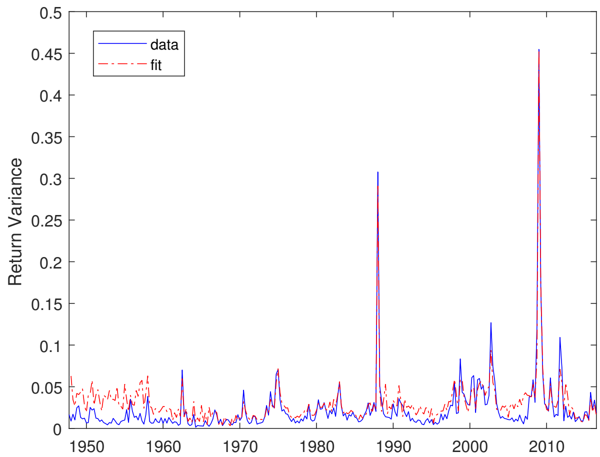


Figure 20: Time-series fitted values of the quarterly realized variance as a function of the estimated states using a second-order bivariate expansion.