Semiparametric Estimation of Latent Variable Asset Pricing Models

Jeroen Dalderop*

December 30, 2022

Abstract

This paper studies semiparametric identification and estimation of consumption-based asset pricing models with latent state variables. We model consumption, dividends, and prices via unknown functions of Markovian state variables describing aggregate output growth. Subsequently, we identify state-dependent components in stochastic discount factor models based on the Euler equation. We develop tractable algorithms for filtering, smoothing, and sieve maximum likelihood estimation, and establish its consistency. Empirically, we find sizable nonlinearities in the impacts of expected growth and volatility on the price-dividend ratio and the discount factor.

Keywords: Asset Prices, Volatility, Risk Aversion, Latent Variables, Semiparametric Estimation

JEL Codes: C14, C58, G12

---

*Department of Economics, University of Notre Dame, 3060 Jenkins-Nanovic Halls, Notre Dame, IN 46556, USA. Email address: jdalderop@nd.edu. This paper has been previously circulated as "Estimating Policy Functions Implicit in Asset Prices". I am grateful to Yacine Aït-Sahalia, Xiaohong Chen, Drew Creal, Yingyao Hu, Dennis Kristensen, Oliver Linton, Hamish Low, Ralph Koijen, Alexei Onatski, Richard Smith, and seminar participants at Cambridge, Erasmus University, Federal Reserve Board, LSE, Notre Dame, Penn State, Princeton, Queen Mary, University of Amsterdam, Wisconsin School of Business, and several conferences, for helpful comments. Any errors are my own.

1 Introduction

Standard consumption-based asset pricing models with moderately risk averse households have difficulty reconciling episodes of highly volatile asset prices with relatively smooth fluctuations in macroeconomic fundamentals. Model extensions in which fundamentals and preferences are driven by persistent yet unobserved state variables have made substantial progress in rationalizing the distribution of asset returns.[1] For the sake of tractability, such models commonly assume that variables such as consumption, dividends, and the stochastic discount factor depend log-linearly on the latent state variables. The resulting log-linear pricing formulas imply that the volatility of asset returns is proportional to that of the state variables. However, heightened economic uncertainty during the 1950s and the early 1980s did not trigger excessive stock market volatility, nor did large swings in stock prices around the 2001 dot-com bubble coincide with significant volatility in economic growth. Such episodes suggest an important role for nonlinear state-dependence in fundamentals, preferences, or both.

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

Subsequently, we use the consumption, dividend, and pricing functions to identify stochastic discount factor models with state-dependent components. In particular, we focus on models with components that can be identified as the positive eigenfunction of an operator derived from the Euler equation for consumption and investment. We demon-

[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).

strate this for a multiplicative state-dependent discount factor and for the continuation value under recursive preferences. We thereby extend nonparametric identification and estimation results for similar eigenfunction problems in Christensen (2017) and Escanciano et al. (2020), to accommodate unobserved state variables.

To avoid the curse of dimensionality of fully nonparametric models, we impose some parametric structure on the stochastic discount factor and the distribution of the state variables. In particular, the stochastic discount factor is specified semiparametrically as the product of the marginal rate of substitution of consumption under power utility and an unknown function of the stationary state variables. The latter component can be interpreted as a 'taste shifter' or derived from structural models, such as recursive preferences. Parametric models for the state variables allow increasing their dimension and analytic characterization of their dynamic properties. In particular, for affine state variables, the framework parsimoniously generalizes the class of affine equilibrium asset pricing models (Eraker and Shaliastovich, 2008) towards nonlinear consumption and dividend dynamics. Their nonlinear state-dependence endogenously generates variation in the mean and volatility of their growth rates, instead of modeling these with additional exogenous state variables.

The framework is highly tractable when the unknown functions are approximated by linear sieves as in Chen (2007). In particular, with polynomial bases, expected growth rates of consumption, dividends, and asset prices can be expressed in closed-form as polynomials of affine state variables. We establish the nonparametric identification of the measurement and pricing equations by limiting their feedback to output growth, and provide conditions for their consistent estimation by sieve maximum likelihood together with the parameters of the state variable dynamics. The estimates are computed using a sequential Monte Carlo variant of the EM-algorithm which analytically solves the maximization step for the approximating coefficients in terms of simulated smoothed moments of the state variables.

In a second stage method-of-moments step, we estimate the stochastic discount factor components by integrating out the latent variables of the Euler equation using their smoothed distribution. For both the state-dependent discount factor and the continuation value under recursive preferences, we implement this using a computationally attractive profiling estimator based on a feasible eigenvector problem.

The empirical application illustrates the methodology by analyzing the impact of long-run risk and stochastic volatility of aggregate output growth on equity valuation ratios. We estimate the model using quarterly data on post-war U.S. macroeconomic variables, aggregate stock prices and dividends, and short term Treasury Bill rates. We also consider high-frequency measures of return volatility in order to discipline the filtered economic volatility state similar to financial volatility in Andersen et al. (2015). We find periods of high growth volatility clustered around episodes such as the post-war years, the 1980s energy crisis, and to some extent the 2008 financial crisis. The frequency and duration of high volatility periods declines steadily over the sampling period, reaching its lows during the high growth 1990s. The consumption-output share mainly responds to expected growth, while dividend growth is U-shaped in volatility. High expected growth and low growth volatility lift the expected price-dividend ratio to at least one standard deviation above its mean, but only when combined. Meanwhile, return volatility peaks around median levels of growth volatility, suggesting a trade-off between the size and price-sensitivity of economic shocks. The state-dependence in the price-dividend ratio is only partially explained by that of consumption and dividends, as the state-dependent discount factor and the recursive preference continuation value both increase in expected volatility and decrease in expected growth. This suggests state-dependent preferences play an important role in relating asset prices to future economic growth and volatility.

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

Gallant and Tauchen (1989), Chen and Ludvigson (2009), and others, estimate the stochastic discount factor semi- or nonparametrically based on conditional moment restrictions in the form of Fredholm Type I integral equations, which are common in nonparametric instrumental variables studies. Meanwhile, Hansen and Scheinkman (2009), Chen et al. (2014), Christensen (2017), and Escanciano et al. (2020) formulate the Euler equation as a Fredholm Type II integral equation, establishing identification based on the Kreĭn-Rutman theorem. Similarly, Ross (2015) uses the finite-dimensional Perron-Frobenius theorem to recover discrete-state option-implied probabilities. While Hansen and Scheinkman (2009) and Christensen (2017) identify the eigenfunction of the long-term valuation operator of a given stochastic discount factor, Escanciano et al. (2020) nonparametrically identify the marginal utility function of consumption. Our paper extends the latter object to include unobserved state variables, covering a wide range of asset pricing models.

The identification of nonlinear dynamic latent variable models has been primarily studied for large cross-sections and panel data, such as Hu and Shum (2012) and Arellano et al. (2017), respectively. These papers focus on individual-specific state variables instead of common state variables. Gagliardini and Gourieroux (2014) and Andersen et al. (2019) extract nonlinear common factors from a large number of cross-sectional units. In our paper, the latent variable dynamics are identified through their relation to a low-dimensional growth series observed over many time periods. In finance, latent variables are often dealt with by inverting observations, such as in affine models for the term structure (Piazzesi, 2010), option prices (Pan, 2002; Ait-Sahalia and Kimmel, 2010), and price-dividend ratios (Constantinides and Ghosh, 2011; Jagannathan and Marakani, 2015). In nonlinear and multi-state models, the inverse mapping may not be unique. For volatility, direct proxies could be constructed using high-frequency measures (Andersen et al., 2003) or the option-implied VIX (Berger et al., 2020). However stock market volatility does not translate one-to-one into the volatility of economic fundamentals, and is subject to time-varying risk aversion. Realized variation of low frequency macroeconomic series suffers from non-vanishing measurement error, while cross-sectional dispersion measures based on firm level data (Bloom, 2009) require correctly specifying the conditional means and covariance structure (Jurado et al., 2015). Without reliable proxies, state variables may still be accurately filtered using forward-looking asset prices, which this paper focuses on.

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

Organization. The remainder of this paper is organized as follows. Section 2 introduces the model assumptions and the asset pricing Euler equations. Section 3 outlines the estimation procedure and its asymptotic properties. Section 4 discusses the empirical application. Section 5 concludes.

2 Setting

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

2.1 Aggregate growth

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

$$\mathcal{S}_{t+1} \mid \mathcal{F}_t^{\mathcal{S}} \sim \mathcal{S}_{t+1} \mid s_t. \tag{1}$$

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

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

$$\begin{aligned} \Delta y_{t+1} &= \mu + x_t + \sigma_t \eta_{y,t+1} \\ x_{t+1} &= \rho_x x_t + \phi_x \sigma_t \eta_{x,t+1}, \end{aligned} \tag{2}$$

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

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

This process, introduced by Gourieroux and Jasiak (2006), is the discrete-time analogue of the continuous-time Cox-Ingersoll-Ross process. This formulation ensures that the variance is positive, and that its conditional moments are available in closed form. In particular, its conditional mean and variance equal $E\left(\sigma_{t+1}^2 \mid \sigma_t^2\right) = \nu\sigma_t^2 + (1-\nu)\bar{\sigma}^2$ and $\mathrm{Var}\left(\sigma_{t+1}^2 \mid \sigma_t^2\right) = \frac{(1-\nu)\bar{\sigma}^2}{\phi_\sigma}(2\nu\sigma_t^2 + (1-\nu)\bar{\sigma}^2)$, respectively. As an alternative, we consider the conditionally log-Normal volatility specification used, among others, by Schorfheide et al. (2018):

$$\sigma_t = \bar{\sigma}\exp(h_{\sigma,t}), \qquad h_{\sigma,t+1} = \rho_h h_{\sigma,t} + \sigma_h \eta_{h,t+1}, \qquad \eta_{h,t+1} \sim \text{i.i.d.N}(0,1). \tag{3}$$

It implies the marginal distribution is known as $h_{\sigma,t} \sim N\left(0, \frac{\sigma_h^2}{1-\rho_h^2}\right)$ assuming $|\rho_h| < 1$.

2.2 Consumption and dividend policy

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

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

The parameter restriction $0 < \rho^c < 1$ guarantees the cointegration of consumption and output, as each period the consumption share partially adjusts towards a target level that depends on the stationary state variables.[2] The presence of Δy_t allows for the temporal smoothing of income shocks, or for measurement error in y_t that enters into $c_t - y_t$.

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

[2] Alternatively, Hansen et al. (2008) model consumption as cointegrated with corporate earnings, with a similar motivation.

the consumption share by the semiparametric additive specification

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

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

$$m_t = R m_{t-1} + \psi(s_t) + \delta^T \Delta y_t + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0, \Sigma_\varepsilon), \tag{4}$$

with $\varepsilon_t = (\varepsilon_t^c, \varepsilon_t^d)$ the combined error term, and Σ_ε its covariance matrix. The process reduces to a standard first-order vector autoregression with exogenous variables when $\psi(s_t)$ is linear. For example, Bansal et al. (2009) include the cointegration residuals of consumption and dividends in a linear vector autoregression with other stationary variables. Furthermore, setting R as the zero matrix yields hidden Markov models, which do not allow for direct lag dependence. For example, Schorfheide et al. (2018) model the consumption-output and dividend-consumption residuals using the linear function $\psi(x_t) = \mu + \psi x_t$, where x_t is a persistent growth component, while allowing for a serially correlated error term. Finally, the process (4) can be expressed in terms of growth rates as

$$\begin{aligned} \Delta c_{t+1} &= r_c^T m_t + (1 + \delta^c)\Delta y_{t+1} + \psi^c(s_{t+1}) + \varepsilon_{t+1}^c \\ \Delta d_{t+1} &= r_d^T m_t + (\lambda + \delta^d)\Delta y_{t+1} + \psi^d(s_{t+1}) + \varepsilon_{t+1}^d, \end{aligned} \tag{5}$$

where $r_c^T = (R - I_2)_{1\cdot}$ and $r_d^T = (R - I_2)_{2\cdot}$. When $R = I_2$, output, consumption, and dividends are not subject to cointegration relations. When Δy_{t+1} follows (2) and $\psi(\cdot)$ is linear, this yields the baseline long-run risk model from Bansal and Yaron (2004). Without cointegration, re-defining $m_t = (\Delta c_t, \Delta d_t)$ in model (4) allows for short-term lag dependence.[3]

An advantage of formulation (4) is that it does not introduce additional state variables for consumption and dividend growth. Instead, time-variation in their mean and volatility derives from those of output growth. This parsimony may appear restrictive, compared to for example Schorfheide et al. (2018) who allow for three separate stochastic volatility

[3]Our empirical application covers models with and without a cointegration relation between dividends and output.

processes. However, any nonlinear state dependence in $\psi(s_t)$ captures variation in conditional means that would otherwise show up as conditionally heteroskedastic, non-Gaussian errors.

The theoretical results in Section 3 allow for state-dependence beyond the mean, such as in the variance $\Sigma_\varepsilon(s_t)$ or the autoregression parameter $R(s_t)$. Our nonparametric identification argument requires (m_t, s_t) to be first-order Markov and the future augmented states $\mathcal{S}_{t+1} = (\Delta y_{t+1}, s_{t+1})$ to be independent of current and past m_t, given s_t. This 'no feedback' assumption allows future growth $(\Delta y_{t+h})_{h\geq 1}$ to serve as instruments for s_t. It implies that consumption and dividend shocks can only affect future output growth *through* the latent states s_t. For the process (4) this holds when $\mathcal{S}_t = (\Delta y_t, s_t)$ is strongly exogenous, in the sense that $\mathbb{E}(\varepsilon_t \mid (\mathcal{S}_{t+j})_{j=-\infty}^{\infty}) = 0$.

2.3 Stochastic discount factor

The absence of arbitrage implies the existence of a positive stochastic discount factor $M_{t,t+1}$ such that the cum-dividend return R^d_{t+1} on any traded equity price satisfies

$$1 = E\left(M_{t,t+1} R^d_{t+1} \mid \mathcal{F}_t\right).$$

In rational expectations equilibrium models, the stochastic discount factor typically depends on aggregate consumption growth and potentially other macroeconomic variables. Consider the general class of consumption-based models that can be represented as

$$M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} \tilde{M}(s_{t+1}, m_{t+1}, s_t, m_t), \tag{6}$$

for some positive multiplicative function $\tilde{M}(\cdot)$. For models within this class, the Markovian consumption and dividend dynamics imply that the joint distribution of $(M_{t,t+1}, \Delta d_{t+1})$ given $\mathcal{F}_t^{\mathcal{S},m}$ depends on (s_t, m_t) only. As a result, when the extended state vector $(s_t, m_t) \in \mathcal{F}_t$, the price-dividend ratio $\frac{P_t}{D_t}$ equals a function $\pi(s_t, m_t)$ satisfying the recursive relation

$$\pi(s_t, m_t) = E\left(M_{t,t+1} \frac{D_{t+1}}{D_t} \left(1 + \pi(s_{t+1}, m_{t+1})\right) \mid s_t, m_t\right). \tag{7}$$

The Euler equation of the asset return R^d_{t+1} can thus be represented as the conditional moment restriction

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

Without further restrictions, the function $\tilde{M}(\cdot)$ is not identified as it depends on two periods of the state vector (s_t, m_t), even if their transition density is known and the price-dividend function is known. The next subsections discuss two main specifications of consumption-based utility functions that are special cases of formulation (6). Both reduce the dimension of $\tilde{M}(\cdot)$ using possibly unknown functions of the Markovian state-vector (s_t, m_t). This gives rise to semiparametric models for which identification results can be obtained.

Of practical interest are partially log-linear SDF models that imply

$$\log \tilde{M}(s_{t+1}, m_{t+1}, s_t, m_t) = \log \breve{M}(s_{t+1}, s_t) + \alpha_0^T m_t + \alpha_1^T m_{t+1}, \tag{9}$$

for some positive function $\breve{M}(s_{t+1}, s_t)$. In particular, the price of a risk-free bond with one-period maturity under (9) and the partially linear growth dynamics (4) equals

$$\begin{aligned} P_t^f &= E\left(\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\breve{M}(s_{t+1}, s_t)e^{\alpha_0^T m_t + \alpha_1^T m_{t+1}} \mid s_t, m_t\right) \\ &\equiv \pi^f(s_t)\exp(\alpha_f^T m_t), \end{aligned}$$

where $\alpha_f = -\gamma r_c + \alpha_0 + R^T\alpha_1$. Hence the risk-free rate $r_t^f = \log P_t^f = \log \pi^f(s_t) + \alpha_f^T m_t$ inherits the partially linear form. Similarly, the price-dividend function $\pi(s_t, m_t) = \tilde{\pi}(s_t)\exp(\alpha^T m_t)$ solves (7) up to a log-linear approximation in the m_t argument.

2.3.1 State-dependent discount factor

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

$$U_t = E\left(\sum_{s=t}^{\infty}\beta^{s-t}v(C_t;\gamma)\phi(s_t) \mid \mathcal{F}_t\right), \tag{10}$$

where β is a fixed discount parameter, and the instantaneous utilities multiply the isoelastic utility function $v(\cdot)$

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

with a state-dependent discount factor $\phi(\cdot)$ that could be fully or partially unspecified. This specification extends the standard power utility of consumption model with the aim of better explaining equity risk premia. The component $\phi(s_t)$ can be directly interpreted as a taste shifter, describing how the marginal utility of consumption changes with the state of the economy. Since economic theory may not prescribe how variables such as expected growth and volatility affect such time preferences, it is desirable to not restrict the functional form $\phi(\cdot)$. For example, evidence from option markets suggests marginal utility may be U-shaped with respect to financial market volatility (Song and Xiu, 2016). Our specification allow for similar non-monotonic preferences over economic volatility.

Specification (10) also covers more structural models of the form $u(C_t, M_t) = \frac{C_t^{1-\gamma}}{1-\gamma}M_t$, where the multiplicative component M_t can be written in terms of the Markovian state variables s_t. For example, it covers the utility over wealth models in Bakshi and Chen (1996) with $M_t = W_t^{\lambda}$ for some parameter λ and absolute or relative wealth W_t, as long as W_t or the wealth-consumption ratio $\frac{W_t}{C_t}$ is a function of s_t. Models with reference utility imply $M_t = Q_t^{\gamma}$, where Q_t is the inverse consumption surplus ratio relative to a reference level, which may be determined by s_t. Habit models specify Q_t in terms of lagged consumption growth Δc_t, which could be added into $\phi(s_t, \Delta c_t)$ as an extension discussion below. Models with incomplete markets or private information imply $M_t = E\left(\left(C_t^i/C_t\right)^{-\gamma} \mid s_t\right)$ and $M_t = E\left(\left(C_t^i/C_t\right)^{\gamma} \mid s_t\right)^{-1}$, respectively, where C_t^i is consumption by each ex-ante identical consumer i, see e.g. Hansen and Renault (2010). For certain types of subjective beliefs $\phi(s_t)$ may represent probability overweighting of possible outcomes of the state variable. Finally, $\phi(s_t)$ could be used to detect statistical misspecification of the transitory component of the stochastic discount factor, and thereby aid the search for appropriate structural models.

Under this semiparametric specification, the stochastic discount factor between times t

and $t+\tau$ is a multiplicatively separable special case of (6), given by

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

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

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

The Euler equation under state-dependent discounting reduces to, provided $s_t \in \mathcal{F}_t$,

$$\frac{1}{\beta}\phi(s_t) = E\left(\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \phi(s_{t+1}) R^d_{t+1} \mid s_t \right), \tag{11}$$

which can be recognized as a Type-II Fredholm integral equation. Using infinite-dimensional versions of the Perron-Frobenious theorem, Christensen (2017) and Escanciano et al. (2020) provide conditions for the existence and uniqueness of a positive eigenvalue-eigenfunction pair (β, ϕ) that solves this type of equation.

Computing this solution requires knowing the expectation $E\left(\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R^d_{t+1} \mid s_{t+1}, s_t \right)$ given both the current and next period state variables. With $m_t = (c_t - y_t, d_t - \lambda y_t)$, the logarithm of the consumption-discounted cash flow $\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \frac{D_{t+1}}{D_t}$ can be expressed as:

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

The integral equation (11) can therefore be stated in terms of the price-dividend function $\pi(s_t, m_t)$ and the Markovian density $f(\mathcal{S}_{t+1}, m_{t+1} \mid s_t, m_t)$. The latter decomposes into the product of the parametric state density $f(\mathcal{S}_{t+1} \mid s_t)$ and the semiparametric measurement density $f(m_{t+1} \mid \mathcal{S}_{t+1}, m_t) = f_\varepsilon(m_{t+1} - \psi(\mathcal{S}_{t+1}) - (R-I)m_t)$, both of which can be identified without using asset prices.

2.3.2 Recursive preferences

Under the recursive preference framework by Epstein and Zin (1989), the representative agent's continuation value is defined via the recursion

$$V_t = \left[(1-\beta) C_t^{1-\theta} + \beta E(V_{t+1}^{1-\gamma} \mid \mathcal{F}_t)^{\frac{1-\theta}{1-\gamma}} \right]^{\frac{1}{1-\theta}},$$

which introduces the elasticity of intertemporal substitution (EIS) parameter θ. Hansen and Scheinkman (2012) provide terminal conditions such that the scaled continuation value $\frac{V_t}{C_t}$ is a function of the Markovian state vector that solves the fixed-point equation

$$V(s_t, m_t) = \left\{ (1-\beta) + \beta E \left[\left(V(s_{t+1}, m_{t+1}) \frac{C_{t+1}}{C_t} \right)^{1-\gamma} \mid s_t, m_t \right]^{\frac{1-\theta}{1-\gamma}} \right\}^{\frac{1}{1-\theta}}.$$

The stochastic discount factor has the representation

$$M_{t,t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\frac{V(s_{t+1}, m_{t+1})}{\mathcal{R}_t \left(V(s_{t+1}, m_{t+1}) C_{t+1}/C_t \right)} \right)^{\theta-\gamma} \tag{12}$$

where $\mathcal{R}_t(V_{t+1}) = E\left(V_{t+1}^{1-\gamma}\right)^{\frac{1}{1-\gamma}}$. With Markovian consumption growth, the SDF specifies the function $\tilde{M}(\cdot)$ in (6) in terms of the continuation value and transition density of (s_t, m_t).

With unit EIS (i.e. when $\theta = 1$), the continuation value satisfies the fixed-point equation

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

Re-defining this equation in terms of $h(s_t, m_t) = V(s_t, m_t)^{\frac{1-\gamma}{\beta}}$ yields

$$h(s_t, m_t) = E \left(\left(\frac{C_{t+1}}{C_t} \right)^{1-\gamma} |h(s_{t+1}, m_{t+1})|^{\beta} \mid s_t, m_t \right). \tag{14}$$

Taking the absolute value inside the conditional expectation does not change the positive solution we seek. Following Christensen (2017), we recognize (14) as a nonlinear Perron-Frobenius problem, with positive eigenfunction $\chi(s_t, m_t) = \frac{h(s_t, m_t)}{\|h\|}$ and associated eigenvalue $\lambda = \|h\|^{1-\beta}$. The norm of $\chi(\cdot)$ is fixed to unity to prevent its eigenvalue from changing with scale due to the nonlinearity induced by $|\chi(\cdot)|^\beta$.

Under the partially linear growth dynamics (4), the fixed-point equation (13) has a separable solution of the form $V(s_t, m_t) = \tilde{V}(s_t)e^{\alpha_v^T m_t}$, where

$$\alpha_v^T = \beta r_c^T (I_2 - \beta R)^{-1} \tag{15}$$

captures the discounted impulse response to consumption following a unit shock to m_t (Hansen et al., 2008). This solution allows focusing on lower-dimensional eigenfunctions of (14) of the form $h(s_t, m_t) = \tilde{h}(s_t)e^{\alpha_h^T m_t}$ with $\alpha_h = \frac{1-\gamma}{\beta}\alpha_v$.

2.4 Generalizing affine models

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

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

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

where l is a $D \times 1$ multi-index of degree $|l| = \sum_{d=1}^D l_d$, $s_l = \prod_{d=1}^D s_d^{l_d}$, and $\bar{s}^L$ is a column vector that stacks monomials up to degree L in lexicographic order. Orthogonal polynomials such as the Hermite or Chebyshev polynomials are spanned by elementary polynomials and can be represented in this way.

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

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

for some coefficient functions $a(\cdot)$ and $b(\cdot)$ that satisfy $a(0) = b(0) = 0$. The conditional moments of monomials of order $0 \leq |l| \leq L$ under such models can be computed as

$$E\left(s_{t+1}^{l} \mid s_t\right) = \frac{\partial^{|l|}}{\partial s_1^{l_1} \cdots \partial s_D^{l_D}} e^{a(u)+b(u)^T s_t}\bigg|_{u=0} = Q_{L,l\cdot}^{T} \bar{s}_t^{\,L}, \tag{16}$$

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

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

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

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

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

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

3 Estimation

This section first discusses the identification and estimation of the policy and pricing functions ψ and π that form the measurement equation of a nonlinear state space model, and of the parameters θ^s describing the growth and state variable dynamics. The functional parameters are combined into $h = (\psi, \pi)$, the finite-dimensional parameters into $\theta = (\theta^s, \delta, R, \Sigma)$, and both are collected in $\vartheta = (\theta, h)$. Afterwards, we discuss estimating the components of the stochastic discount factor, for both the state-dependent discount factor and the recursive preference specifications.

3.1 State space formulation

The measurements m_t and normalized prices p_t contain aggregate quantities whose conditional distribution depends on the latent state variables s_t. Let $\mathcal{S}_t = (\Delta y_t, s_t)$ be the partially observed Markovian state vector introduced in (1). The following assumptions describe the interaction between the observations and states:

Assumption 1.

a) $(m_t, p_t, \mathcal{S}_t)_{t\in\mathcal{Z}}$ *is a jointly stationary ergodic process*

b) The joint process is first-order Markov:

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

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

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

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

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

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

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

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

With measurements $m_t = (c_t - y_t, d_t - \lambda y_t)^T$ and normalized prices $p_t = \log \frac{P_t}{D_t}$, stationarity of (m_t, p_t) implies the cointegration of the logarithms of output, consumption, dividends, and prices. The resulting mean-reverting behavior of m_t is a well-known source of return predictability (Lettau and Ludvigson, 2001; Bansal et al., 2009). The joint first-order Markov assumption 1*b)* rules out any dependence on past states or errors. Multi-period dependence can be allowed for by including further lags in the state vector. The no feedback assumption 1*c)* implies that the partially observed $\mathcal{S}_t$ forms a hidden Markov process, and is not caused in the sense of Granger (1969) by the observations (m_t, p_t). This allows for an interpretation of exogenous variation in the state variables generating endogenous responses in the observations. The hidden Markov assumption does not require that observations are themselves Markovian, as it allows for their dependence at all leads and lags. The contemporaneous state-dependence of measurements and prices in 1*d)* and 1*e)* rules out their direct dependence on past states, which is a timing assumption also made by Hu and Shum (2012) and describes rational forward-looking behavior. Finally, prices p_t are distinguished from measurements m_t by depending on their own lag only through a Markovian pricing error. The distinction is motivated by the presence of habits or frictions in consumption and cash flow choices, while the pricing error is attributed to market sentiments unrelated to fundamentals. However, if the other type of lag dependence is deemed more appropriate, a series of prices could be included in m_t, or a series of measurements in p_t.

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

$$\begin{aligned} m_t &= R_m m_{t-1} + \psi(s_t) + \delta^T \Delta y_t + \varepsilon_t, && \varepsilon_t \sim \text{i.i.d.} N(0, \Sigma_\varepsilon) \\ p_t &= \tilde{\pi}(s_t) + \alpha^T m_t + \eta_t, && \\ \eta_t &= R_p \eta_{t-1} + \omega_t, && \omega_t \sim \text{i.i.d.} N(0, \Sigma_\omega) \\ \mathcal{S}_{t+1} &\sim f(\mathcal{S}_{t+1} \mid s_t), && \end{aligned} \tag{17}$$

where ε_t and ω_t are uncorrelated, and independent of s_t. The serially correlated pricing error η_t allows for the presence of persistent deviations from the fundamental price given the state variables and cointegration residuals. This allows for an autoregressive stochastic discount factor component unrelated to fundamentals, as in Albuquerque et al. (2016) and Schorfheide et al. (2018).

3.2 Identification

The identification of the functional parameters follows a sequential argument. First, we study the identification of the policy functions ψ and pricing functions π under the hidden Markov assumptions. Given these, section 3.7 studies the identification of stochastic discount factor components using the Euler equation.

3.2.1 Nonparametric identification of the policy and pricing functions

Under Assumption 1, our semiparametric formulation is a special case of the nonparametric dynamic latent variable models considered in Hu and Shum (2012). Their main result yields high-level invertibility conditions under which the four-period joint density of $(m_t, \Delta y_t)$ identifies the first-order Markovian distribution of $(m_t, \Delta y_t, s_t)$. Intuitively, they exploit the conditional independence of past and future observations given the current partially observed state variable. A related argument is used in Arellano et al. (2017) to identify the consumption rule in terms of a persistent earnings component using only *future* observed earnings. Our no feedback assumption makes this strategy possible as future growth realizations are independent of the current measurement given the current latent state. However, while Hu and Shum (2012) and Arellano et al. (2017) use a large number of cross-sectional units to estimate the multi-period densities, we instead observe a stationary time series $(m_t, \Delta y_t)$ over a large number of periods T.

Let $\Delta y_{t-k:t}$ denote the growth realizations $(\Delta y_{t-k}, \ldots, \Delta y_t)$ for any non-negative integer k, Θ^s denote the finite-dimensional parameter space for θ^s, which contains the true parameter θ_0^s, and $\mathcal{L}^2(s_t)$ be the space of squared-integrable functions with respect to the marginal density of the random variable s_t. Consider the following set of assumptions:

Assumption 2.

a) For any $\theta^s \in \Theta$ and $s \in \mathcal{S}$, the Markov probability kernel has a density $f(\cdot \mid s; \theta^s)$ w.r.t. the Lebesgue measure on $\mathcal{Y} \times \mathcal{S} \subseteq \mathbb{R}^{D+1}$.

b) For some non-negative integer K_1 and any $\theta^s, \tilde{\theta}^s \in \Theta^s$ with $\theta^s \neq \tilde{\theta}^s$, there exists some $\Delta y_{-K_1:1} \in \mathcal{Y}^{K_1+1}$ such that $f(\Delta y_1 \mid \Delta y_{-K_1:0}; \theta^s) \neq f(\Delta y_1 \mid \Delta y_{-K_1:0}; \tilde{\theta}^s)$.

c) The linear operator $L^y_{s_0 \mid \bar{y}_0, \Delta y_{1:K_2}} : \mathcal{L}^2(s_0) \mapsto \mathcal{L}^2(\Delta y_{1:K_2})$ defined by

$$\left(L^y_{s_0 \mid \bar{y}_0, \Delta y_{1:K_2}} g\right)(\bar{y}_{1:K_2}) = \int_S g(s_0, \bar{y}_0) f(s_0 \mid \Delta y_{0:K_2} = \bar{y}_{0:K_2}; \theta_0^s) ds_0$$

is one-to-one for some non-negative integer K_2 and all $\bar{y}_0 \in \mathcal{Y}$. Moreover, the measurement density $f(m_t \mid m_{t-1}, s_t, \Delta y_t)$ w.r.t the Lebesgue measure exists, and

$f(m' \mid m, \cdot, \Delta y) \in \mathcal{L}^2(s_0)$ for every $(m', m, \Delta y)$.

d) The linear operator $L^m_{s_0|\bar{m}_0,\Delta y_{1:K_3}} : \mathcal{L}^2(s_0) \mapsto \mathcal{L}^2(\Delta y_{1:K_3})$ defined by

$$\left(L^m_{s_0|\bar{m}_0,\Delta y_{1:K_3}} g\right)(\bar{y}_{1:K_3}) = \int_S g(s_0, \bar{m}_0) f(s_0 \mid m_0 = \bar{m}_0, \Delta y_{1:K_3} = \bar{y}_{1:K_3}) ds_0$$

is one-to-one for some non-negative integer K_3 and all $\bar{m}_0 \in \mathbb{R}$. Moreover, $E(p_t \mid m_t, s_t) < \infty$ almost surely, and $E(p \mid m, \cdot) \in \mathcal{L}^2(s_0)$ for every (p, m).

The identification argument proceeds sequentially. First, Assumption 2b) implies that the data-generating parameters θ^s_0 of the state transition density $f(\mathcal{S}_{t+1} \mid s_t; \theta^s)$ are identified from the density of multi-period observed growth $\Delta y_{t-K_1:t+1}$. For affine models this condition can be verified from the Laplace transform of the state vector (Gagliardini and Gouriéroux, 2019). Second, consider the current and future growth realizations $\Delta y_{t:t+K_2}$ for a finite number of leads $K_2 > 0$. The no feedback condition implies the conditional independence $(m_t, m_{t-1}) \mid s_t, \Delta y_{t:t+K} \sim (m_t, m_{t-1}) \mid s_t, \Delta y_t$. Therefore the conditional density of the pairwise measurements equals an integral over one period of the state vector:

$$f(m_t, m_{t-1} \mid \Delta y_{t:t+K_2}) = \int f(m_t, m_{t-1} \mid s_t, \Delta y_t) f(s_t \mid \Delta y_{t:t+K_2}; \theta^s_0) ds_t. \tag{18}$$

Assumption 2c) formalizes that the backward filtering density $f(s_t \mid \Delta y_{t:t+K_2}, \theta^s_0)$ under the true parameter is complete for every Δy_t. Given knowledge of the conditional density of observations $f(m_t, m_{t-1} \mid \Delta y_{t:t+K_2})$, the integral equation (18) therefore identifies a unique pairwise measurement density $f(m_t, m_{t-1} \mid s_t, \Delta y_t)$, which in turn identifies the lag-dependent density $f(m_t \mid m_{t-1}, s_t, \Delta y_t)$. Furthermore, identification of the conditional density $f(m_t \mid m_{t-1}, s_t, \Delta y_t)$ implies that all its existing conditional moments, in particular the conditional mean $\psi(m_{t-1}, s_t, \Delta y_t)$, are identified as well.

Since the measurement density characterizes the transition density of the augmented state vector $(m_t, \Delta y_t, s_t)$, the updating density $f(s_t \mid m_t, \Delta y_{t+1:t+K_3}) = \frac{f(m_t, s_t \mid \Delta y_{t+1:t+K_3})}{f(m_t \mid \Delta y_{t+1:t+K_3})}$ can be computed for any K_3. Moreover, by Assumption 2d) it is complete for every m_t. The pricing function $\pi(s_t, m_t) = E\left(p_t \mid \mathcal{F}^{m,s}_{-\infty:\infty}\right)$ is therefore nonparametrically identified from the integral equation

$$E\left(p_t \mid m_t, \Delta y_{t+1:t+K_3}\right) = \int \pi(s_t, m_t) f(s_t \mid m_t, \Delta y_{t+1:t+K_3}) ds_t, \tag{19}$$

using the hidden Markov and no feedback Assumptions 1a-c). Since Assumptions 1d) and 1e) imply the pricing errors are independent of future and past measurements, respectively, leads and lags of the latter could be added as conditioning variables. Finally, the identification strategy assumes knowledge of the joint densities $f(\Delta y_{-K_1:1})$, $f(m_1, m_0, \Delta y_{1:K_2})$, and $f(p_1 \mid m_0, \Delta y_{1:K_3})$. For all it suffices to know the density $f(p_{0:K^*}, m_{0:K^*}, \Delta y_{0:K^*})$ for $K^* = \max(K_1, K_2, K_3)$, which can be consistently estimated when the sample period grows using the stationarity of $(p_t, m_t, \Delta y_t)$.

The completeness Assumptions 2cd) are interpreted as requiring the updating of the latent state distribution by the growth observations to be relevant for a sufficient number of leads. For example, in the Gaussian homoskedastic long-run risk model, $x_t \mid y_{t:t+1} \sim N(\beta_0 y_t + \beta_1 y_{t+1}, \sigma^2)$ with $\beta_1 \neq 0$ if $\rho_x \neq 0$. Given y_t, completeness in y_{t+1} therefore derives from the persistence of the latent growth component. More general sufficient conditions are provided in Newey and Powell (2003) and Hu and Shiu (2018). The next subsection provides an intuitive analogue with parametric models.

3.2.2 Parameter identification in polynomial models

For models with a finite-order polynomial measurement equation, identification of its coefficients can be expressed in terms of rank conditions on the conditional moments of the state variables given future and/or past growth realizations. For example, let $s^l_{t|T} = E(s^l_t \mid \mathcal{F}^y_{1:T})$ be the smoothed conditional l-th mixed moment of the state s_t given the full sample of growth realizations, and let the vector $\bar{s}^L_{t|T}$ stack the smoothed moments up to order L. Consider the univariate specification $m_t = c_L' \bar{s}^L_t + \varepsilon_t$, where ε_t is independent of $(\Delta y_t)_{t\in\mathbb{Z}}$. The latter implies $E\left(\varepsilon_t s^l_{t|T}\right) = 0$ for any $l = 0, \ldots, L$, which yields the linear system of $L+1$ equations

$$E\left(\bar{s}^{L\,\prime}_{t|T} m_t\right) = c_L' E\left(\bar{s}^L_t \bar{s}^{L\,\prime}_{t|T}\right). \tag{20}$$

The prediction error $\bar{e}^L_t = \bar{s}^L_{t|T} - \bar{s}^L_t$ by construction satisfies $E\left(e^k_t s^l_{t|T}\right) = 0$ for each $(k, l) \in \{0, \ldots, L\}$, so that $E\left(\bar{s}^L_t \bar{s}^{L\,\prime}_{t|T}\right) = E\left(\bar{s}^L_{t|T} \bar{s}^{L\,\prime}_{t|T}\right)$. The coefficient vector c_L could therefore be identified from the regression of m_t on $\bar{s}^L_{t|T}$, provided the outer product matrix $E\left(\bar{s}^L_{t|T} \bar{s}^{L\,\prime}_{t|T}\right)$ is invertible. Similar rank conditions on filtered moments can be used to identify the parameters of a finite-order polynomial pricing function for the log price-dividend ratio $p_t = \log(\frac{P_t}{D_t}) = b_L' \bar{s}_t^L + \alpha^T m_t + \eta_t$, provided η_t is independent of leads and/or lags of Δy_t.

This result can be seen as a two-stage version of the identification strategy of polynomial models with measurement error by Hausman et al. (1991), who solve a linear system involving moments of the mismeasured and instrumental variables. In our case, the first stage uses the path of growth realizations to instrument the moments of the unobserved regressors.

Given the correct specification of the joint Markovian transition density of (p_t, m_t, s_t), the policy and pricing functions can be efficiently estimated using full-information likelihood methods. Still, the moment-based estimator (20) only requires correct specification of the density $f(\mathcal{S}_t \mid s_t; \theta^s)$. This robustness motivates using the moment-based estimators as initial values for the Expectation-Maximization or similar algorithm.

3.3 Likelihood formulation

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

$$\begin{aligned}\ell_T(\vartheta) &= \log f(p_{1:T}, m_{1:T}, \Delta y_{1:T}; \vartheta) \\ &= \log f(p_{1:T} \mid \mathcal{F}_{1:T}^{m,y}; \pi, \psi, \theta^s) + \log f(m_{1:T} \mid \mathcal{F}_{1:T}^{y}; \psi, \theta^s) + \log f(\Delta y_{1:T}; \theta^s).\end{aligned}$$

The parameter structure allows for both joint and sequential estimation procedures. In particular, θ^s can be consistently estimated from the series Δy_t alone, ψ from $(m_t, \Delta y_t)$ given $\hat{\theta}^s$, and π from the full observation vector $(p_t, m_t, \Delta y_t)$ given $(\hat{\theta}^s, \hat{\psi})$.

The time $t+1$ contribution to the joint log-likelihood $\ell_T(\vartheta) = \sum_{t=0}^{T-1} l_{t+1}(\vartheta)$ is given by

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

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

$$f(\Delta y_{t+1} \mid \mathcal{F}_{1:t}^{p,m,y}; \vartheta) = \int f\left(\Delta y_{t+1} \mid s_t; \theta^s\right) f\left(s_t \mid \mathcal{F}_{1:t}^{p,m,y}; \vartheta\right) ds_t,$$

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

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

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

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

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

3.4 Sequential Monte Carlo filtering and smoothing

In nonlinear dynamic models it is generally not possible to integrate out the latent variables analytically from the likelihood components, unlike in linear models with Gaussian errors where the updating density $f\left(s_t \mid \mathcal{F}_{1:t}^{y}; \theta^s\right)$ can be computed recursively by the Kalman filter. In line with Taylor expansion methods of solving equilibrium models (e.g. Schmitt-Grohé and Uribe, 2004), a second order approximation to the measurement equation could be performed to identify parameters corresponding to volatility shocks (Fernández-Villaverde and Rubio-Ramírez, 2007). However, this may cause parameters related to higher order moments to become unidentified. Instead, particle filtering or sequential Monte Carlo simulation can be used to recursively approximate expectations of any nonlinear function of the state vector, see Doucet and Johansen (2009) for an overview.

The filtering density of the latent states satisfies the recursion

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

This motivates the following recursive algorithm. Let $(s_{i,t})_{i=1}^{N^s}$ be a set of N^s particles drawn from $\mathcal{F}_{1:t}^{p,m,y}$ with weights $(w_{i,t})_{i=1}^{N^s}$. For $t = 0$, the particles can be drawn from the marginal state density $f_{\theta^s}(s_0)$, if available, or from an approximate distribution χ_0 that is easy to simulate from.[4] First, draw next period's states $(s_{i,t+1})_{i=1}^{N^s}$ from the transition density $f(s_{t+1} \mid s_t; \theta^s)$. Second, compute the updated sampling weights of $(s_{i,t+1})_{i=1}^{N^s}$ given

[4]Proposition 1 below shows the choice of χ_0 becomes irrelevant when $T \to \infty$ under some conditions.

$(p_{t+1}, m_{t+1}, \Delta y_{t+1})$ as

$$w_{i,t+1} \propto f_\eta\left(\eta_{t+1}(s_{i,t+1}) \mid \eta_t(s_{i,t}); \vartheta\right) f\left(m_{t+1} \mid s_{i,t+1}, m_t; \vartheta\right) f(\Delta y_{t+1} \mid s_{i,t+1}, s_{i,t}; \theta^s), \tag{21}$$

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

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

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

Alternatively, one could draw from an auxiliary transition density $\omega(s_{t+1} \mid s_t)$ that is easy to simulate from, and multiply the weights (21) by the importance sampling factors $f(s_{i,t+1} \mid s_{i,t}; \theta^s)/\omega(s_{t+1} \mid s_t)$ before normalizing. The auxiliary densities can be chosen to improve the efficiency of the simulations. In particular, the variance of the importance sampling factors is minimized by choosing the updated density $\omega_{t+1}(s_{t+1} \mid s_t) = f(s_{t+1} \mid s_t, p_{t+1}, m_{t+1}, \Delta y_{t+1})$ given next period's observations (e.g. Doucet and Johansen, 2009). For nonlinear models the latter density is typically not available in closed form, but can be approximated as Gaussian using the Unscented Kalman Filter (e.g. Fulop et al., 2021).

The smoothing distribution can be approximated similarly based on the backward recursive relation

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

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

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

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

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

The particle filter approximates this as

$$w^{\dagger}_{i,t+1} = w_{a(i),t} \frac{w^{*}_{i,t+1} f(s_{i,t+1} \mid s_{a(i),t})}{\sum_{k=1}^{N^s} f(s_{i,t+1} \mid s_{k,t}) w_{k,t}}, \tag{22}$$

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

3.5 Expectation-Maximization algorithm

Global maximization of the approximated likelihood function is computationally unattractive when the parameter space is large-dimensional, as is the case when approximating functional parameters. However, when the measurement equations are approximated by linear series, their coefficients can be estimated by the method-of-moments based on the conditional moments of the states, in line with the identification argument in Section 3.2. This motivates the following variant of the Expectation-Maximization (EM) algorithm, illustrated for approximating polynomials, in which the M-step of optimizing the expected log-likelihood given the observations reduces to a linear regression. Related iterative algorithms in which the M-step is performed analytically are available for linear Gaussian models (Watson and Engle, 1983), finite mixture models (Arcidiacono and Jones, 2003), and conditional quantile models (Arellano et al., 2017).

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

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

where the price, measurement, and growth components equal

$$
\begin{aligned}
Q_p(\vartheta_p,\tilde{\vartheta}) &= E_{\tilde{\vartheta}}\left(\sum_{t=2}^{T}\log f_\omega\left(\eta_t(s_t)-R_p\eta_{t-1}(s_{t-1})\right)\;\Big|\;\mathcal{F}_T^{p,m,y}\right)\\
Q_m(\vartheta_m,\tilde{\vartheta}) &= E_{\tilde{\vartheta}}\left(\sum_{t=2}^{T}\log f_\varepsilon\left(m_t-\psi_L(s_t)-R_m m_{t-1}\right)\;\Big|\;\mathcal{F}_{1:T}^{p,m,y}\right)\\
Q_y(\theta^s,\tilde{\vartheta}) &= E_{\tilde{\vartheta}}\left(\log f(s_1;\theta^s)+\sum_{t=2}^{T}\log f(\mathcal{S}_t\mid s_{t-1};\theta^s)\;\Big|\;\mathcal{F}_{1:T}^{p,m,y}\right),
\end{aligned}
$$

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

M-step. The three-way decomposition indicates that $Q(\vartheta,\tilde{\vartheta})$ can be maximized component-wise using corresponding subsets of parameters. Let $\tilde{\bar{s}}_{t|1:T}^L=E_{\tilde{\vartheta}}(\bar{s}_t^L\mid\mathcal{F}_{1:T}^{p,m,y})$ and $\tilde{V}_{t|1:T}^L=\mathrm{Var}_{\tilde{\vartheta}}(\bar{s}_t^L\mid\mathcal{F}_{1:T}^{p,m,y})$ denote the smoothed means and variances, respectively, of the polynomials $\bar{s}_t^L$ given the initial parameter values. For Gaussian measurement errors ε_t, $Q_m(\cdot)$ is maximized over c_L and R_m as

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

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

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

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

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

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

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

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

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

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

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

3.6 Consistency

We are interested in the limiting properties of the maximum likelihood estimator

$$(\hat{\theta}, \hat{h}) = \arg\max_{\theta \in \Theta, h \in \mathcal{H}} \frac{1}{T} \ell_T(\theta, h), \tag{23}$$

where Θ is a finite-dimensional parameter space, and $\mathcal{H} = \prod_{k=1}^{K} \mathcal{H}_k$ is a Cartesian product of $K = K_m + K_p$ infinite-dimensional parameter spaces for the policy functions $(\psi_k)_{k=1}^{K_p}$ and the pricing function $(\pi_p)_{k=1}^{K_m}$, which are assumed to contain the true parameter (θ_0, h_0). Let each space $\mathcal{H}_k$ be equipped with a weighted Sobolev L^2 norm, which sums integrals of the squared partial derivatives of a function. In particular, define the $D \times 1$ vector l of non-negative integers with $|l| = \sum_{d=1}^{D} l_d$, the partial derivative operator $D^l = \frac{\partial^{|l|}}{\partial y_1^{l_1} \cdots \partial y_D^{l_D}}$, and the normalized state vector $\tilde{s} = \Sigma_s^{-1}(s - \mu_s)$, assuming its mean μ_s and covariance matrix Σ_s exist. Then, for some integers $r > 0$, $r_0 > \frac{D}{2}$ and $\delta_0 > \frac{D}{2}$, define this norm by

$$\|g\|_{r+r_0,2} = \left\{ \sum_{|l| \le r+r_0} \int \left(D^l g(\tilde{s})\right)^2 (1 + \tilde{s}'\tilde{s})^{\delta_0} ds \right\}^{1/2}.$$

For vector-valued functions define $\|g\|_{r,2} = \sum_{m=1}^{K} \|g_m\|_{r,2}$. Instead of maximizing $\ell_T(\vartheta)$ over the infinite-dimensional functional space $\mathcal{H}$, the method of sieves (Chen, 2007) controls the complexity of the model in relation to the sample size by minimizing over approximating finite-dimensional spaces $\mathcal{H}_L \subseteq \mathcal{H}_{L+1} \subseteq ... \subseteq \mathcal{H}$ which are dense in $\mathcal{H}$. For some positive vectors B_0 and B_1, and a vector of known functions $g_0(s)$, we follow Newey and Powell (2003) and define the functional spaces $\mathcal{H}_k$ for $k = 1, \ldots, K$ as

$$\mathcal{H}_k = \left\{ g(s) = c^T g_0(s) + g_1(s) : \mathbb{R}^D \mapsto \mathbb{R} : c^T c \le B_{0k}, \|g_1\|^2_{r+r_0,2} \le B_{1k} \right\}.$$

All unknown functions $g_1(s)$ that meet the second requirement in $\mathcal{H}_k$ have at least $r + r_0$ partial derivatives with bounded squared weighted integrals. In particular, this implies the nonparametric component flattens out when $|\tilde{s}|$ becomes large, at a rate faster than $|\tilde{s}|^{-\delta_0}$. The parametric component $c^T g_0(s)$ therefore serves primarily to describe the tails of the functions in $\mathcal{H}$, though it can also accommodate some prior information about the functional form. Linear combinations in this space can be conveniently characterized in terms of their coefficients. Let $p^L(\cdot) = (p_1(\cdot), ..., p_L(\cdot))$ be a set of basis functions, and consider the finite-dimensional series approximator $g_L(s) = \sum_{l=1}^{L} \gamma_l p_l(s) = \gamma^T p^L(s)$. Define

$$\Lambda_L = \sum_{|l| \le r} \int D^l p^L(\tilde{s}) D^l p^L(\tilde{s})^T (1 + \tilde{s}'\tilde{s})^{\delta_0} ds,$$

which implies that $\|g_1\|^2_{r+r_0,2} = \gamma^T \Lambda_L \gamma$ takes a quadratic form (Newey and Powell, 2003). Therefore the optimization in (23) is redefined over the compact finite-dimensional subspace $\mathcal{H}_{L_T} = \prod_{k=1}^{K} \mathcal{H}_{L_T,k}$:

$$(\hat{\theta}, \hat{h}_L) = \arg \max_{\theta \in \Theta, h \in \mathcal{H}_{L_T}} \frac{1}{T} \ell_T(\theta, h), \tag{24}$$

where

$$\mathcal{H}_{L,k} = \left\{ g(s) = c^T g_0(s) + \gamma^T p^L(s) : \; c^T c \leq B_{0k}, \gamma^T \Lambda_L \gamma \leq B_{1k} \right\}. \tag{25}$$

Also, for some positive constant $\frac{D}{2} < \delta_c < \delta_0$, define the Sobolev sup-norm

$$\|g\|_{r,\infty} = \max_{|l| \leq r} \sup_{s} \left| D^l g(\tilde{s}) \right| (1 + \tilde{s}'\tilde{s})^{\delta_c}.$$

Then the closure $\bar{\mathcal{H}}$ of $\mathcal{H}$ with respect to the norm $\|g\|_{r,\infty}$ is compact (Gallant and Nychka, 1987). This facilitates obtaining consistency under the norm $\|\vartheta\|_c = (\theta'\theta)^{1/2} + \sum_{k=1}^{K} \|h_k\|_{r,\infty}$. See Freyberger and Masten (2019) for other combinations of norms that allow such compact embedding.

Due to filtering the latent variables, the conditional log-likelihoods $l_{t+1}(\vartheta)$ depend on a growing number of past observations. We can ensure that the normalized log-likelihood has a well-defined stationary limit using the 'forgetting' properties of ergodic Markov processes. In the following, let $z_t \subseteq (p_t, m_t, \Delta y_t)$ be (a subset of) the observations taking values in $\mathcal{Z} \subseteq \mathbb{R}^K$, which together with the latent state vector s_t in $\mathcal{S} \subseteq \mathbb{R}^D$ (which may include the pricing error η_t) follows a hidden Markov model (HMM).[6] Consider the following assumptions, adapted from Douc and Moulines (2012, Assumption NL):

Assumption 3.

a) *The parameter product space $\varTheta = \Theta \times \bar{\mathcal{H}}$ is compact.*

b) *The state transition density $f_{\theta^s}(s' \mid s)$ is positive and continuous in (θ^s, s', s) on $\Theta^s \times \mathcal{S} \times \mathcal{S}$, and satisfies $\sup_{\theta^s \in \Theta} \sup_{(s',s) \in \mathcal{S} \times \mathcal{S}} f_{\theta^s}(s' \mid s) < \infty$.*

c) *The measurement density $f_\vartheta(z \mid s)$ is positive and continuous in (z, s) on $\mathcal{Z} \times \mathcal{S}$ for any $\vartheta \in \varTheta$, and satisfies $E\left(\log^+ \sup_{\vartheta \in \varTheta} \sup_{s \in \mathcal{S}} f_\vartheta(Z_1 \mid s)\right) < \infty$.*

d) *For any compact subset $\mathcal{K} \subset \mathcal{Z}$, and $\vartheta \in \varTheta$, $\lim_{|s| \to \infty} \sup_{z \in \mathcal{K}} \frac{f_\vartheta(z|s)}{\sup_{s' \in \mathcal{S}} f_\vartheta(z|s')} \to 0$.*

e) *There exists a compact subset $\mathcal{D} \subset \mathcal{S}$ such that $E\left(\log^- \inf_{\vartheta \in \varTheta} \inf_{s \in \mathcal{D}} f_\vartheta(Z_1 \mid s)\right) < \infty$.*

Most of these assumptions can be directly verified. For example, the log-volatility model (3) satisfies 3b) when $\bar{\sigma}$ and σ_h are bounded away from zero, and $|\rho_h|$ is bounded below unity, so that the transition density $f_{\theta^s}(\sigma_{t+1} \mid \sigma_t)$ is bounded from above. For the

[6] In addition to the pairwise Markov process defined by Assumption 1, (z_t, s_t) being an HMM requires $z_t \perp\!\!\!\perp z_{-\infty:t-1} \mid s_t$. In example (17), this requires R_m to equal zero or be known.

Gaussian location-scale model $Z_1 \mid s_1 \sim \mathcal{N}(\mu_\vartheta(s_1), \Sigma_\vartheta^{1/2}(s_1))$, Assumption 3c) is satisfied when $\inf_{s\in\mathcal{S}} \lambda_{\min}(\Sigma_\vartheta(s)) > 0$ for all $\vartheta \in \Theta$, where $\lambda_{\min}$ is the minimum eigenvalue. For this model, Assumption 3e) is satisfied by $E(\|Z_1\|_2^2) < \infty$ and continuity of $\mu_\vartheta(s)$ and $\Sigma_\vartheta(s)$.

Proposition 1. *Under Assumption 3, for any $\vartheta \in \Theta$ and initial state probability distribution χ_0 with $\chi_0(D) > 0$, a measurable function $\bar{f}_\vartheta : \mathcal{Z}^{1+\mathbb{Z}^-} \mapsto \mathbb{R}^+$ exists such that*

(i) $\sup_{\vartheta\in\Theta} |\log f_{\vartheta,\chi_0}(Z_1 \mid Z_{-k:0}) - \log \bar{f}_\vartheta(Z_1 \mid Z_{-\infty:0})| \leq C\kappa^k$ *a.s. for some $\kappa \in (0,1)$ and random variable C with $\mathbb{P}(C < \infty) = 1$.*

(ii) $E\left(|\log \bar{f}_\vartheta(Z_1 \mid Z_{-\infty:0})|\right) < \infty$.

(iii) $\lim_{T\to\infty} \ell_{T,\chi_0}(\vartheta) = E\left(\log \bar{f}_\vartheta(Z_1 \mid Z_{-\infty:0})\right) \equiv \bar{\ell}(\vartheta)$ *a.s.*

Proposition 1(i) shows that $\log f_{\vartheta,\chi_0}(Z_1 \mid Z_{-k:0})$ converges to $\bar{f}_\vartheta(Z_1 \mid Z_{-\infty:0})$ almost surely and uniformly over Θ, and the influence of past observations decays at a geometric rate. Moreover, the limiting conditional density does not depend on the initial state distribution χ_0. Part (ii) shows that its logarithm is integrable, which allows using the ergodic theorem to establish the pointwise convergence in part (iii).

For Markov processes with autoregressive dependence in the observables, Douc et al. (2004) obtain similar results while assuming the state space is compact, which excludes our application to long-run risk with stochastic volatility models. Still, we conjecture that the results of Proposition 1 extend to this case by suitably adapting Assumptions 3c-d) to the measurement density $f_\vartheta(z' \mid s, z)$.

Assumption 4.

a) The population log-likelihood $\bar{\ell}(\vartheta)$ has a unique maximum at $\vartheta_0 = (\theta_0, h_0) \in \Theta \times \mathcal{H}$.

b) For every $h \in \bar{\mathcal{H}}$, there exists γ_L with $\gamma_L^T \Lambda_L \gamma_L \leq B_1$ such that $\|h - \gamma_L^T p^L\|_c \to 0$ when $L \to \infty$. For any $c \neq 0$, $\|c^T g_0\|_{r+r_0,2}^2 > B_1$.

c) The measurement density $f_\vartheta(z \mid s)$ is continuous in ϑ under the metric $d(\vartheta, \tilde{\vartheta}) = \|\vartheta - \tilde{\vartheta}\|_c$ on $\Theta \times \mathcal{Z} \times \mathcal{S}$.

d) For every $\vartheta \in \Theta$, there is a neighborhood $\mathcal{N}_\vartheta$ such that

$$E\left(\sup_{\tilde{\vartheta}\in\mathcal{N}_\vartheta} \log^- \bar{f}_{\tilde{\vartheta}}(Z_1 \mid Z_{-\infty:0})\right) < \infty.$$

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

Theorem 1. *Suppose that Assumptions 3 and 4 hold. When $T \to \infty$ and $L_T \to \infty$, the estimator $(\hat{\theta}, \hat{h}_L)$ in (24) over the approximating spaces $\mathcal{H}_L$ given by (25) satisfies*

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

Assumption 4a) follows from identification of ϑ_0 based on a finite-period joint density under Assumption 2, which extends to the ergodic conditional density (Douc et al., 2004, Lemma 6). Assumption 4b) defines the set of allowed basis functions, and separates the parametric and nonparametric components. Due to the polynomial weighting function, setting $p_l(s) = s^l \exp(-\delta_h s^T s)$ for some $\delta_h > 0$ ensures that the approximating subspaces $\mathcal{H}_L$ are dense (Gallant and Nychka, 1987, Theorem 2).[7] Assumptions 4cd) are used to establish the uniform convergence of $\ell_T(\vartheta)$ to $\bar{\ell}(\vartheta)$. Assumption 4d) is a local version of the dominance condition that ensures the relative entropy is well-defined, similar to White (1982, Assumption 3A(a)) for i.i.d. observations. For compact state spaces, 4d) can be established recursively from a lower bound on the transition density (Douc and Moulines, 2012). For infinite state spaces, a sufficient condition is the stochastic equicontinuity of $\bar{f}_\vartheta(Z_1|Z_{-\infty:0})$, or that $\mathbb{P}_\vartheta(s_1 \in D|Z_{-\infty:0}) > 0$ uniformly and almost surely on a subset D that satisfies 3e).

3.7 Method of moments estimation of the SDF components

The recursive pricing equation (7) pins down the dependence of the expected price-dividend ratio $\pi(s_t; \beta, \gamma, \theta, h)$ on unknown structural parameter functions in h, such as the state-dependent discount factor ϕ or the continuation value function χ. When $\pi(s_t; \cdot)$ can be quickly and accurately computed, these could be efficiently estimated by maximizing the restricted likelihood function. However, in general this requires an additional numerical approximation of the pricing functions, which may no longer be linear in parameters. Instead, we consider method-of-moments estimators for the SDF parameters based on the

[7] See Rodriguez (2003) for approximation results for general weighted Sobolev spaces.

smoothing distribution of s_t given the observations, which is obtained during the M-step of the EM-algorithm. Moreover, for both the state-dependent discount factor and the continuation value under recursive preferences, we construct a computationally attractive profiling estimator as the eigenfunction of a certain operator.

3.7.1 Identification of the state-dependent discount factor

Given the transition density and the policy and pricing functions, the identification of the stochastic discount function proceeds essentially as if the state variables are observable. In particular, the stochastic discount function is identified from the price-dividend function as long as there is unique eigenvalue-eigenfunction pair $(\phi, \frac{1}{\beta})$ that solves (11). Let $\mathbb{M} : \mathcal{L}^2(s) \mapsto \mathcal{L}^2(s)$ be the linear operator defined by

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

where

$$\mathcal{K}(s_t, s_{t+1}) = E\left(\left(C_{t+1}/C_t\right)^{-\gamma} R^d_{t+1} \mid s_t, s_{t+1}\right).$$

Under the following assumption, Christensen (2017) and Escanciano et al. (2020) show that $\mathbb{M}$ has a unique (up to scale) positive eigenfunction $\phi \in \mathcal{L}^2(s)$ with positive eigenvalue $\frac{1}{\beta}$:

Assumption 5.

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

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

The positivity assumption facilitates the use of an infinite-dimensional extension of the Perron-Frobenious theorem for positive valued matrices. This theorem also underpins the pricing kernel recovery theorem for finite Markov chains in Ross (2015). In our setting, a sufficient condition for the positivity of $\mathcal{K}$ is that the price-dividend function $\pi(s_t, m_t)$ is positive almost everywhere. Some mild sufficient conditions on $\mathcal{K}$ for $\mathbb{M}$ to be bounded and compact are given in Christensen (2017) and Escanciano et al. (2020). In particular, compactness follows from

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

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

The operator $\mathbb{M}$ can be conditioned on further observed variables m_t to identify additional arguments in $\phi(s_t, m_t)$. Christensen (2015) shows how Assumption 5 can be suitably relaxed to accommodate main examples such as external habit formation, in which case lagged and current state variables overlap. Chen and Ludvigson (2009) identify this model under the completeness of an expected return-weighted density of the state variables. Chen et al. (2014, Thms. 9 and 10) provide conditions for the local identification of the joint parameter $(\beta, \gamma, \phi(\cdot))$.

3.7.2 Estimation of the state-dependent discount factor

Let $b^L(\cdot) = (b_1(\cdot), ..., b_L(\cdot))$ be a set of basis functions, possibly different from p^L. Approximating the state-dependent discount factor by the series $\phi_L(s) = e_L' b^L(s)$ yields the finite-dimensional counterpart of the eigenfunction problem:

$$G_L^{-1} M_L e_L = \frac{1}{\beta} e_L, \tag{26}$$

where

$$\begin{aligned} G_L &= E\left(b^L(s_t) b^L(s_t)'\right) \\ M_L &= E\left(b^L(s_t) b^L(s_{t+1})' \left(C_{t+1}/C_t\right)^{-\gamma} R_{t+1}^d\right). \end{aligned}$$

Let $\Pi_L : \mathcal{L}^2(s) \mapsto B_L$ be the orthogonal projection on the subspace $B_L \subset \mathcal{L}^2(s)$ spanned by the basis functions b^L. Let $\hat{G}_L$ and $\hat{M}_L$ be estimators of G_L and M_L, and for any matrix A_L let the superscript 'o' define its orthogonalized version $A_L^o = G_L^{-1/2} A_L G_L^{-1/2}$. Let $\|\cdot\|$ denote the $\mathcal{L}^2(s)$-norm, Euclidian norm, or operator norm when applied to functions, vectors, and matrices, respectively. Let $(\hat{\beta}^{-1}, \hat{e}_L)$ be the largest eigenvalue and corresponding eigenvector of $\hat{G}_L^{-1} \hat{M}_L$, and $\hat{\phi}_L(s) = \hat{e}_L' b^L(s)$ be the estimated eigenfunction. The following result, adapted from Christensen (2017, Thm. 3.1), establishes their consistency:

Assumption 6.

a) $\|\Pi_L \mathbb{M} - \mathbb{M}\| = o(1)$.

b) $\|\hat{G}_L^o - I_L\| = o_p(1)$ *and* $\|\hat{M}_L^o - M_L^o\| = o_p(1)$.

Proposition 2. *Let Assumptions 5 and 6 hold. Then* $\hat{\beta} \xrightarrow{p} \beta$ *and* $\|\hat{\phi}_L - \phi\| \xrightarrow{p} 0$ *when* $T \to \infty$.

Assumption 6a) requires the basis functions to span the range of $\mathbb{M}$ when $L \to \infty$. Assumption 6b) requires the matrix $G_L^{-1} M_L$ to be consistently estimated, while implicitly controlling the rate at which L grows with T. Lemma A1 in the Appendix provides sufficient conditions for 6b) for the plug-in estimators

$$\begin{aligned}
\hat{G}_L &= E_{\hat{\theta}^s}\left(b^L(s_t) b^L(s_t)'\right) \\
\hat{M}_L &= E_{\hat{\theta}^s}\left(b^L(s_t) b^L(s_{t+1})' \mathcal{K}(s_t, s_{t+1}; \hat{\vartheta})\right).
\end{aligned}$$

where $\mathcal{K}(s_t, s_{t+1}; \vartheta) = E_\vartheta\left((C_{t+1}/C_t)^{-\gamma} R^d_{t+1} \mid s_t, s_{t+1}\right)$ can be computed analytically in special cases such as the conditionally Gaussian HMM, or otherwise approximated by simulation.

This eigenvector problem requires knowledge of the risk aversion parameter γ. We consider its estimation based on moment restrictions derived from the Euler equation (11):

$$E\left[\left(\frac{1}{\beta} e_L' b^L(s_t) - \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} e_L' b^L(s_{t+1}) R^d_{t+1}\right) z_t\right] = 0, \tag{27}$$

for any variables $z_t \in \mathcal{F}_t$ in the investor information set. Setting $z_t = b^L(s_t)$ reproduces the eigenvector problem. Therefore we add moments based on variables observed at time t. In particular, $z_t = (m_t, p_t)$ is motivated by the Markovian assumption under which $(\Delta c_{t+1}, R^d_{t+1}, s_{t+1})$ only depends on $\mathcal{F}_t^{m,y,p,s}$ through (m_t, p_t, s_t).

The moment conditions (27) depend linearly on functions of the latent variables, whose conditional expectation given observed variables can be computed during the EM algorithm. For example, when $b^L(s_t) = \bar{s}_t^L$ are mixed polynomials, the Law of iterated expectations yields the feasible moment restrictions

$$g(\beta, \gamma, e_L) = E\left[\left(\frac{1}{\beta} e_L' \hat{\bar{s}}^L_{t|T} - \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} e_L' \hat{\bar{s}}^L_{t+1|T} R^d_{t+1}\right) z_t\right] = 0. \tag{28}$$

in terms of the smoothed moments $\hat{\bar{s}}^L_{t|T} = E_{\hat{\vartheta}}(\bar{s}_t^L \mid \mathcal{F}^{m,y,p}_{1:T})$. For estimation we replace the unconditional moments by their empirical averages

$$g_T(\beta,\gamma,e_L)=\frac{1}{T}\sum_{t=0}^{T-1}e_L'\left(\frac{1}{\beta}\hat{\hat{s}}_{t|T}^L-\hat{\hat{s}}_{t+1|T}^L\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}R_{t+1}^d\right)z_t.$$

Instead of joint GMM estimation of (β,γ,e_L), we profile the eigenvalue-eigenvector pair $(\hat{\beta}(\gamma),\hat{e}_L(\gamma))$ over the risk aversion parameter γ, which is then estimated as

$$\hat{\gamma}=\arg\min_{\gamma}\|g_T(\hat{\beta}(\gamma),\gamma,\hat{e}_L(\gamma))\|^2.$$

3.7.3 Identification and estimation under recursive preferences

For recursive preferences, we propose a similar profiling procedure to estimate the components of the SDF. First, for given values of (β,γ), we solve the nonlinear Perron-Frobenius problem (14) to find the scaled continuation value function $\chi(s;\beta,\gamma)$.[8] Profiling this function, we estimate (β,γ) based on a set of unconditional moments obtained by instrumenting the Euler equation for equity returns. For simplicity, we focus on the partially linear model with solutions of the form $h(s_t,m_t)=\tilde{h}(s_t)e^{\alpha_h^T m_t}$, with coefficients α_h determined by (15).

Define the operator $\mathbb{T}:\mathcal{L}^2(s)\mapsto\mathcal{L}^2(s)$ by $\mathbb{T}h=\mathbb{G}h^\beta$, where $\mathbb{G}$ is the linear operator

$$\mathbb{G}\phi(s)=E\left(\phi(s_{t+1})(C_{t+1}/C_t)^{1-\gamma}e^{\alpha_h'(\beta m_{t+1}-m_t)}\mid s_t=s\right).$$

Hansen and Scheinkman (2012) study the existence and uniqueness of a fixed point of $\mathbb{T}$ based on Perron-Frobenius theory applied to $\mathbb{G}$. Furthermore, Christensen (2017, Corollary 4.1) shows that a positive fixed point of $\mathbb{T}$ is locally identified when $\mathbb{G}$ is positive and bounded, and $\mathbb{T}$ has a Fréchet derivative with spectral radius less than one.

For estimation, we approximate the scaled continuation value by $h_L(s)=v_L'b^L(s)$ for some vector v_L. Define the operator $\mathbb{T}_L:\mathbb{R}^L\to\mathbb{R}^L$ as

$$\mathbb{T}_Lv_L=E\left(b^L(s_t)\left(\frac{C_{t+1}}{C_t}\right)^{1-\gamma}e^{\alpha_h'(\beta m_{t+1}-m_t)}|v_L'b^L(s_{t+1})|^\beta\right),$$

The plug-in estimator of $\mathbb{T}_L$ based on $E_{\hat{\vartheta}}$ is analytically intractable due to the nonlinearity

[8] Chen et al. (2013) develop an alternative profiling procedure for a recursive specification of the continuation value.

in β. We therefore consider its simulation-based estimator

$$\hat{\mathbb{T}}_L v_L = \frac{1}{TN^s} \sum_{t=0}^{T-1} \sum_{i=0}^{N^s} w_{i,t+1}^{\dagger} b^L(s_{a(i)t}) \left(\frac{C_{t+1}}{C_t} \right)^{1-\gamma} e^{\hat{\alpha}_h'(\beta m_{t+1} - m_t)} |v_L' b^L(s_{it+1})|^{\beta},$$

using the smoothed pairwise particle distribution $(w_{i,t+1}^{\dagger}, s_{a(i)t}, s_{it+1})$ from (22), and the plug-in estimate $\hat{\alpha}_h'(\beta,\gamma) = (1-\gamma)\hat{r}_c^T(I_2 - \beta\hat{R})^{-1}$. Define the estimated fixed point $\hat{v}_L$ as

$$\hat{G}_L^{-1} \hat{\mathbb{T}}_L \hat{v}_L = \hat{v}_L. \tag{29}$$

While $\hat{\mathbb{T}}_L$ is non-linear in the coefficients v_L, $\hat{v}_L$ can still be computed quickly using the iterative algorithm in Christensen (2017). Furthermore, Christensen (2017, Theorem 4.1) establishes the consistency of the estimators of h, χ and λ, given by

$$\hat{h}_L(s) = \hat{v}_L' b^L(s), \qquad \hat{\chi}_L(s) = \frac{\hat{v}_L' b^L(s)}{(\hat{v}_L' \hat{G}_L \hat{v}_L)^{\frac{1}{2}}}, \qquad \hat{\lambda} = (\hat{v}_L' \hat{G}_L \hat{v}_L)^{\frac{1-\beta}{2}}.$$

The parameters (β,γ) are estimated using the moment conditions

$$g(\beta,\gamma,\alpha_h,\lambda,\chi_L) = E\left[\left(\frac{\lambda}{\beta}\chi_L(s_t) - \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} e^{\alpha_h'(\beta m_{t+1} - m_t)} |\chi_L(s_{t+1})|^{\beta} R_{t+1}^d \right) z_t \right] = 0, \tag{30}$$

By setting $z_t = (b^L(s_t), m_t, p_t)$, we exploit the conditional nature of the Euler equation based on the SDF (12). Unlike the stochastic discount function $\phi(s_t)$, including $b^L(s_t)$ is not redundant as the continuation value under recursive preferences is an eigenfunction of an operator depending on consumption growth, rather than equity returns. For implementation, we compute the sample equivalent $g_T(\beta,\gamma,\lambda,\chi_L)$ of the moments in (30) using the smoothed particle distribution, and minimize the profiled criterion

$$(\hat{\beta},\hat{\gamma}) = \arg\min_{(\beta,\gamma)} \|g_T\left(\beta,\gamma,\hat{\alpha}_h(\beta,\gamma),\hat{\lambda}(\beta,\gamma),\hat{\chi}_L(\beta,\gamma) \right)\|^2.$$

4 Empirical Results

4.1 Data

Aggregate output and consumption data are obtained from the NIPA Tables published by the U.S. Bureau of Economic Analysis. We consider quarterly data from 1947:Q1 until 2019:Q4. Output is measured by U.S. real gross domestic product in 1992 chained dollars (Table 1.1.6). Consumption is measured as the real expenditure on nondurables and services (Table 2.3.5, lines 8 plus 13). Output and consumption are then converted to per capita basis using population data (Table 2.1, line 40).

Stock market prices and dividends are based on the CRSP value-weighted portfolio of all stocks traded on the NYSE, AMEX, and NASDAQ. Dividends per share are computed from the difference in value-weighted returns with (R^d_{t+1}) and without (R^x_{t+1}) dividends:

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

Quarterly dividends are then aggregated at the annual frequency to diminish seasonal effects. Prices are also constructed on a per share basis as the cumulative product of daily ex-dividend returns up to the end of each quarter. Prices and dividends are then expressed in real per capita terms using the Price Index for Personal Consumption Expenditures (Table 2.3.4, line 1) and the population data. We proxy the ex ante real risk-free rate using the fitted value from regressing the ex post real rate on the 90-day U.S. Treasury Bill on its nominal rate and CPI inflation over the previous year, following Schorfheide et al. (2018). Finally, a Realized stock market Variance (RV) proxy is constructed from daily log returns $R_{t,d}$ after centering with the quarterly mean $\overline{R}_t$ as[9]

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

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

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

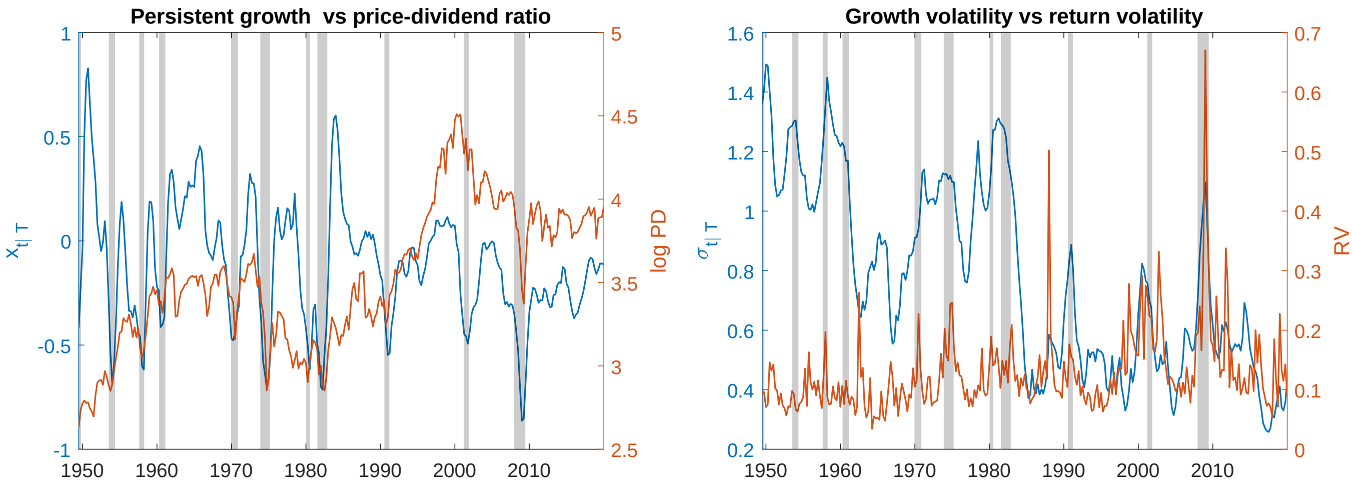
4.2 Economic Growth Risk and Stock Market Volatility

First, we compare the fit of several long-run risk stochastic volatility models estimated to output growth. Table 1 shows model selection criteria for variants of model (2) estimated using the particle filter EM algorithm with using $N^s = 5,000$ particles and re-sampling threshold $ESS = 0.75N^s$ (these settings apply throughout). The log-likelihood and information criteria greatly improve when allowing for stochastic volatility, regardless of whether an autoregressive Gamma or log-Normal volatility process is specified. Meanwhile, extending the default specification (2) to allow for variance-in-mean effects, or for Student t-distributed instead of Gaussian errors, does not substantially improve model fit. Next, we heuristically compare the estimated mean and volatility of economic growth with

Table 1: Simulated maximum log-likelihoods with Akaike and Bayesian Information Criteria of various long-run risk models for output growth Δy_{t+1} from 1947:Q1 until 2019:Q4. LRR, LRR-ARG and LRR-LOGV refer to homoskedastic, autoregressive Gamma, and log-Normal stochastic volatility models, respectively. 'σ_t^2 in mean' adds a linear variance term to $E_t(\Delta y_{t+1})$, while 't shocks' means $\eta_{y,t+1}$ in (2) is t-distributed instead of Gaussian. Estimation based on the EM algorithm with $N^s = 5,000$ particles and $ESS = 0.75N^s$.

	LRR	LRR-ARG			LRR-LOGV		
	default	default	σ_t^2 in mean	t shocks	default	σ_t^2 in mean	t shocks
l^Y	-1.291	-1.166	-1.163	-1.165	-1.175	-1.175	-1.177
AIC	741.4	674.2	674.9	675.4	679.2	681.3	682.3
BIC	756.0	696.2	700.4	701.0	701.1	706.8	707.8

stock market valuations and return volatility. Figure 1a) plots the smoothed persistent output growth component $x_{t|T}$ and the aggregate log price-dividend ratio over time. While the growth component appears to trend downward, price-dividend ratios generally moved upward. At business cycle frequencies, economic booms do not reliably coincide with stock market rises, while economic recessions such as the mid 1970s stagflation and the 2008-09 Great Recession are associated with large stock market drops. These asymmetries do not support a simple log-linear, positive relation between the price-dividend ratio and expected growth. Figure 1b) plots the estimated volatilities of output growth and aggregate stock returns. Most episodes of high economic volatility occurred before or during the early 1980s, while most financial market turbulence occurred thereafter. In particular, the 1950s saw substantial economic uncertainty but historically calm financial markets. Meanwhile, output growth remained relatively stable during the stock market volatility spikes of the 1987 Black Monday crash and following the 1998 LTCM collapse. With the notable exception of the 2008-09 Great Recession, economic and financial volatility measures are only weakly correlated, against the predictions of standard log-linear valuation models.



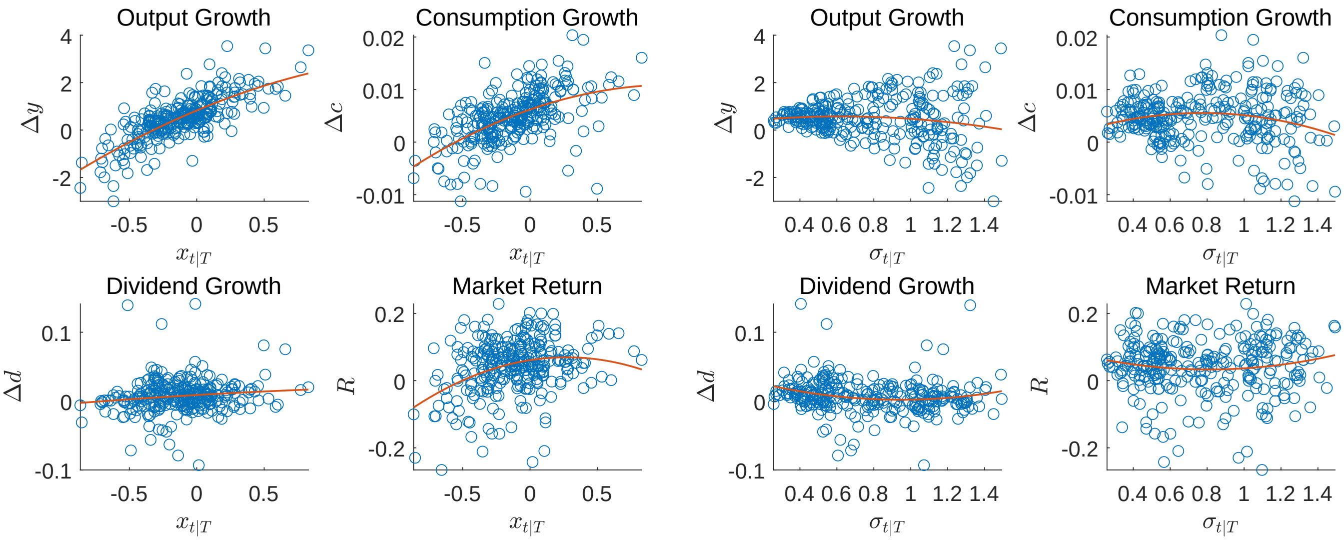
(a) Economic growth and stock market valuations.

(b) Economic versus stock return volatility.

Figure 1: Smoothed latent states of the estimated LRR-ARG model for U.S. real GDP growth (left: means $x_{t|T}$, right: square roots of $\sigma^2_{t|T}$) against log price-dividend ratio (left) and realized return volatility (right) of the CRSP value-weighted portfolio.

Figure 2 plots the quarterly growth rates of output, consumption, dividends, and the CRSP value-weighted portfolio against the smoothed means and volatilities of output growth estimated by the LRR-ARG model. The top-left panels confirm that the state variables impact output growth as described by the model. Figure 2a) shows that consumption and stock returns respond positively to increases in the persistent growth component, with the quadratic fit suggesting this relation strengthens during downturns. Figure 2b) suggests output and consumption growth decrease when volatility becomes sufficiently high, in line with the findings in Bloom (2009) and Nakamura et al. (2017). The market return shows no clear response to output growth volatility, including no heteroskedasticity. Similarly, annual dividends do not show a clear relation to either state variable, although this might be due to the presence of outliers.

While the evidence in this section suggests the observed variables may depend non-linearly on the latent states, a formal analysis should consider the measurement error in estimating the persistent growth and volatility paths. The next sections therefore discuss maximum likelihood estimates of the measurement equations under various specifications.



(a) Impact of mean growth component.

(b) Impact of growth volatility.

Figure 2: Quarterly growth rates of output, consumption, annual dividends, and the CRSP value-weighted portfolio (cum-dividend), against the smoothed mean and volatility of output growth estimated by the LRR-ARG model for 1947:Q1-2019:Q4. Growth rates other than output are lagged one quarter. Fitted lines depict the quadratic least squares fit.

4.3 Estimated consumption and dividend dynamics

First, we review whether consumption and dividends are cointegrated with output, as is a common feature of macroeconomic equilibrium models with balanced growth restrictions. For consumption and output, both the augmented Dickey-Fuller (ADF) test with imposed cointegration parameter of unity and the Engle-Granger (EG) test with estimated parameter 0.998, find p-values against the null of no cointegration below 0.01. Meanwhile, for dividends and output, neither the ADF test with imposed cointegration parameter $\lambda = 1$ or the EG test with estimated $\hat{\lambda} = 0.94$, allows rejecting the null of no cointegration, allowing for five lags based on information criteria.[10] Using a longer sample of annual dividend and output data starting from 1930, the Engle-Granger test with estimated parameter $\hat{\lambda} = 0.75$ and one lag yields a p-value of 0.034. Bansal et al. (2007) similarly find dividends and consumption are cointegrated using annual data. This suggests the post-war

[10]We reach the same conclusion after controlling dividends for repurchases, in line with Bansal et al. (2009), who find dividend and consumption are cointegrated only after removing deterministic trends.

sample may simply be too short to detect mean reversion in the error correction term. Therefore, we report separate maximum likelihood estimates for the measurement vectors $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$ with cointegration and $m_t = (c_t - y_t, \Delta d_t)$ without. In the former case, we estimate $\hat{\lambda}$ using OLS and treat it as known afterwards.

While our focus is on measuring nonlinear dependence on the latent state variables, we can also identify nonlinear dependence on the observables $(m_{t-1}, \Delta y_t)$ or interactions between observed and latent variables. Such nonlinearities often arise from economic frictions or amplification mechanisms, and their inclusion increases robustness against misspecification. Therefore, we consider four variants of the partially linear measurement equation (4):

$$\textit{Model 1:} \quad m_t = R_m m_{t-1} + \psi(s_t) + \delta^T \Delta y_t + \varepsilon_t \qquad \text{(y linear)}$$

$$\textit{Model 2:} \quad m_t = R_m m_{t-1} + \psi(s_t) + \delta(s_t)^T \Delta y_t + \varepsilon_t \qquad \text{(y state-dep)}$$

$$\textit{Model 3:} \quad m_t = R_m m_{t-1} + \psi(s_t, \Delta y_t) + \varepsilon_t \qquad \text{(ys non-linear)}$$

$$\textit{Model 4:} \quad m_t = R_m(s_t) m_{t-1} + \delta_m^T \Delta y_t m_{t-1} + \psi(s_t) + \delta^T \Delta y_t + \varepsilon_t \qquad \text{(m state-dep)}$$

The first model is the benchmark with linear dependence on observables. In the second model, the impact of output growth is a function $\delta(s_t)$ of the latent states, which for example allows centering and scaling output growth by its current mean and volatility. The third model generalizes this towards a nonlinear impact of output growth, as is for example the case when consumers respond asymmetrically to positive and negative income shocks. Finally, the fourth model covers state-dependence in the autoregressive coefficients, allowing the speed of mean-reversion to depend on the state and growth rate of the economy. The Markovian assumptions allow for several further generalizations, such as state-dependent heteroskedasticity, or nonlinear lag dependence. However, the plots in Figure 2 show no clear evidence of the former, while quadratic autoregressions of the observables yield no significant evidence of the latter.

We estimate all four model variants by approximating the unknown functions $\psi(\cdot)$, and any functional coefficients $(R_m(\cdot), \delta(\cdot))$, by polynomials of order L and $L-1$, respectively.[11] Specifically, we approximate each of these functions by $g_L(s) = \sum_{j:|l_j| \leq L} \gamma_j p_j(\tilde{s})$, where

[11] Using one lesser order for the functional coefficients assures that interaction terms are of the same order, and that $L = 1$ reproduces the linear model.

$\tilde{s}_d = \frac{s_d - \hat{\mu}_d}{\hat{\sigma}_d}$ are the state variables normalized by their model-estimated mean and standard deviation, and the basis functions are the dampened polynomials

$$p_j(s) = \begin{cases} s^{l_j} & \text{if } |l_j| \leq L_h \\ s^{l_j} \exp\left(-\delta_h s^T s\right) & \text{if } |l_j| > L_h. \end{cases}$$

The basis functions for $|l_j| > L_h$ are also known as generalized Hermite polynomials. The results are obtained by setting $L_h = 2$ and $\delta_h = \frac{1}{9}$. The former implies that the tails of the function eventually take a quadratic shape. The latter implies that the exponential dampening factor at three standard deviations equals that at one standard deviation with $\delta_h = 1$ as considered in Newey and Powell (2003). This higher value therefore reduces the impact of restricting the tails parametrically. Finally, we initialize the EM-algorithm for each $L > 1$ using the estimates for $L - 1$ with added zero coefficients for the higher-order basis functions, while for $L = 1$ we use estimates from a linear VAR in observables.

Table 2 reports likelihood-based model selection criteria for the four model variants when the measurements m_t are the cointegration residuals and Δy_t follows the LRR-ARG model. The likelihood of output growth for all twenty models reported is higher than without including the measurements (-1.16, see Table 1). Meanwhile, for all models the estimated autoregressive coefficient of the dividend-output residual lies in the range $(0.98, 1.01)$, while the predictive coefficient of the dividend-output residual by the consumption-output ratio falls in the range $(-0.03, 0)$. Therefore allowing for nonlinear state-dependence yields no new evidence that dividends and output are cointegrated, corroborating the findings in Schorfheide et al. (2018) based on Bayesian likelihood methods.

Table 3 reports the same model selection criteria when dividend growth replaces the dividend-output cointegration residual. For virtually all models, including dividend growth further improves the likelihood of output growth. Unsurprisingly, the measurement likelihoods ℓ^M monotonically increase with the approximation order L of the measurement equation. The output growth likelihoods ℓ^Y also initially increase with L, but peak at $L = 4$ for two of the four model variants. The AIC favors quadratic models after considering the number of parameters, while the BIC, expressing a stronger preference for parsimony, only favors a nonlinear model ($L = 2$) under the 'y linear' model variant. Since none of the variants with nonlinearity in the observables consistently improves the infor-

Table 2: Simulated maximum log-likelihoods and information criteria of augmented LRR-ARG models for $(\Delta y_{t+1}, m_t)$ with $m_t = (c_t - y_t, d_t - \hat{\lambda} y_t)$ from 1947:Q1-2019:Q4 by varying type and order L of measurement equation. Criteria are normalized by sample size T. Bold numbers indicate optimal L by type.

	ℓ^Y					ℓ^M				
L	1	2	3	4	5	1	2	3	4	5
y linear	-1.06	-1.06	-1.06	-1.05	**-1.04**	0.97	1.07	1.10	1.17	**1.21**
y state-dep	-1.07	-1.05	**-1.05**	-1.05	-1.07	0.98	1.08	1.13	1.22	**1.28**
ys non-linear	-1.06	-1.06	-1.05	-1.04	**-1.03**	0.99	1.08	1.16	1.25	**1.32**
m state-dep	-1.06	-1.04	-1.05	**-1.04**	-1.05	0.99	1.10	1.19	1.30	**1.44**
	AIC					*BIC*				
L	1	2	3	4	5	1	2	3	4	5
y linear	0.34	0.18	0.17	0.09	**0.07**	0.62	**0.53**	0.62	0.67	0.80
y state-dep	0.32	0.18	0.15	**0.12**	0.18	0.59	**0.58**	0.73	0.93	1.27
ys non-linear	0.31	0.19	**0.15**	0.18	0.33	**0.58**	0.62	0.83	1.25	1.94
m state-dep	0.33	0.16	0.14	0.10	**0.07**	**0.65**	0.66	0.90	1.19	1.57

mation criteria, we report the estimates for the 'y linear' model with $L = 4$.[12] The count of coefficients in $\hat{\vartheta}_L$ is then 45, which corresponds to 18.8 observations per coefficient.[13]

Figure 3 shows the estimated response functions $\hat{\psi}^c_L(s_t)$ and $\hat{\psi}^d_L(s_t)$ of the log consumption-output ratio and dividend-growth responses under the 'y linear' model. The consumption-output ratio monotonically increases in expected growth, but does not appear strongly related to growth volatility, in line with the scatterplots in Figure 2. Conversely, dividend growth shows a U-shaped response to volatility, but no clear relation to expected growth.

To relate the response functions to specific episodes, Figure 4 plots the growth rates of the measurements against their means under the smoothing distribution of the state variables, whose own smoothed means are shown in Figure 5 for different observation sets. Comparing the first two rows of panels of Figure 5 shows that the added measurements narrow the confidence intervals for the latent expected growth and variance components, while their point estimates remain on similar trajectories. The consumption share of output in Figure 4a tends to peak during recessions, as consumption does not immediately respond

[12] Table 8 in the Appendix reports similar findings based on the LRR-LOGV models. Table 9 shows that extending the sample period until 2021:Q4 increases support for higher order L and for the 'y-state dep' variant, which helps explain opposing movements in consumption and income during the pandemic.

[13] For comparison, Gallant and Tauchen (1989, p. 1101) state that for their SNP densities a saturation ratio of about 10 observations per coefficient yields stable polynomial approximations.

Table 3: Simulated maximum log-likelihoods and information criteria of augmented LRR-ARG models for $(\Delta y_{t+1}, m_t)$ with $m_t = (c_t - y_t, \Delta d_t)$ from 1947:Q1-2019:Q4 by varying type and order L of measurement equation. Criteria are normalized by sample size T. Bold numbers indicate optimal L by type.

	ℓ^Y					ℓ^M				
L	1	2	3	4	5	1	2	3	4	5
y linear	-1.08	-1.07	-1.06	**-1.06**	-1.06	-1.09	-1.02	-1.01	-0.98	**-0.93**
y state-dep	-1.08	-1.07	-1.06	-1.05	**-1.05**	-1.09	-1.02	-0.99	-0.94	**-0.88**
ys non-linear	-1.08	-1.07	-1.06	**-1.04**	-1.04	-1.09	-1.01	-0.98	-0.89	**-0.82**
m state-dep	-1.08	-1.07	-1.06	-1.05	**-1.04**	-1.08	-0.98	-0.91	-0.85	**-0.76**
	AIC					BIC				
L	1	2	3	4	5	1	2	3	4	5
y linear	4.49	**4.37**	4.38	4.39	4.39	4.76	**4.72**	4.84	4.97	5.13
y state-dep	4.49	**4.39**	4.42	4.43	4.45	**4.76**	4.79	5.00	5.25	5.55
ys non-linear	4.49	**4.39**	4.44	4.45	4.62	**4.76**	4.82	5.12	5.53	6.23
m state-dep	4.49	**4.36**	4.36	4.40	4.44	**4.81**	4.86	5.12	5.50	5.95

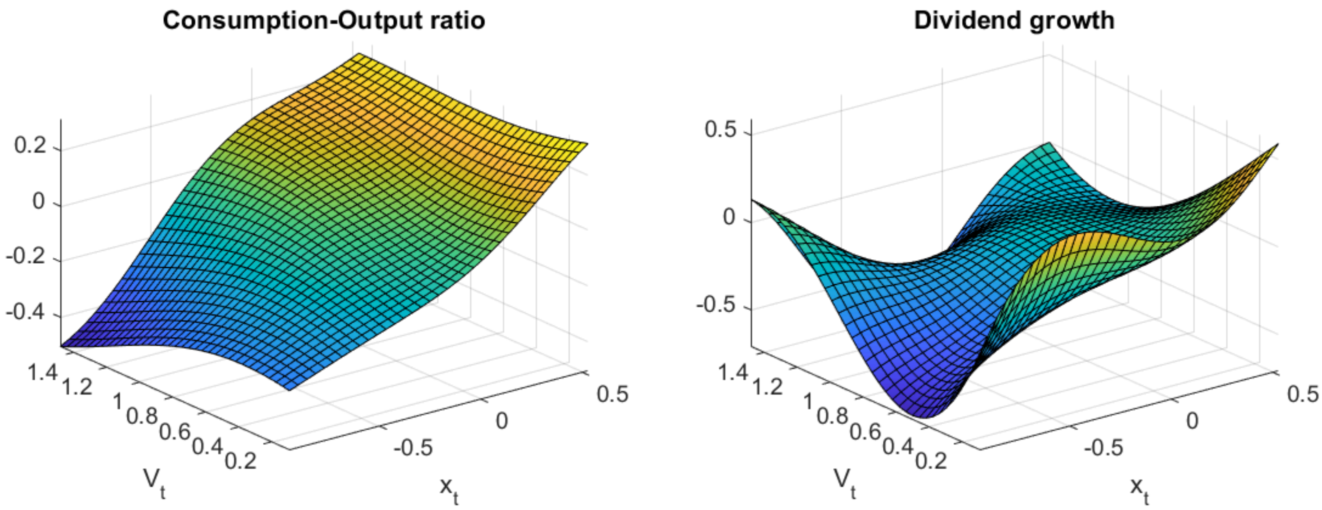
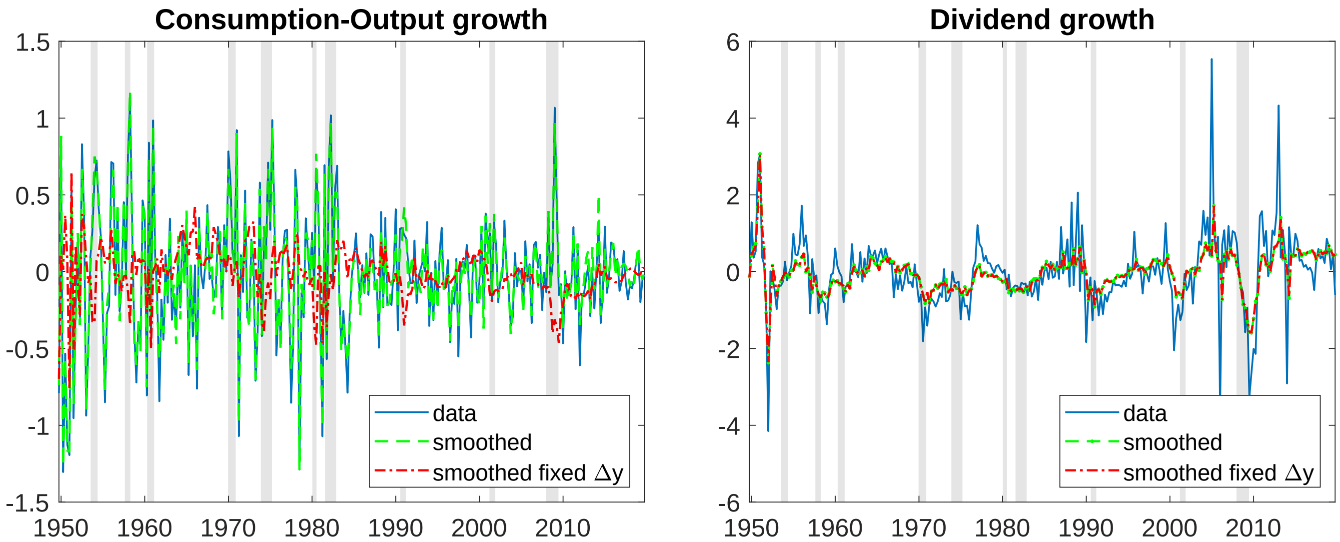


Figure 3: Estimated response functions $\hat{\psi}_L$ of log consumption-output ratio (left panel) and dividend growth (right panel) to the conditional mean x_t and variance V_t of output growth under the LRR-ARG model, using quarterly observations (y_t, m_t) and the 'y linear' variant with $L = 4$ order expansion. Horizontal axes capture 95% of marginal distributions. Vertical axis measures standard deviations $\sqrt{\mathrm{Var}(m_t)}$ from the mean.

to output declines. However, controlling for this contemporaneous 'denominator' effect, the state-dependent component $\hat{\psi}^c_L(\cdot)$ actually falls during the low-growth early 1960s, early 1980s, and Great Recession, in order to rise again during their recoveries, reflecting forward-looking consumption choices. Dividend growth in Figure 4b held up relatively well throughout the high economic volatility of the late 1950s and early 1980s, explaining why $\hat{\psi}^d_L(\cdot)$ is positive for large σ^2_t. Meanwhile, its U-shape in σ^2_t partially reconciles the sudden increases in dividends around 2005 and 2015 when economic volatility dropped, with the persistently low dividend growth during the economic uncertainty of the Great Recession.



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

(b) Δd_t against smoothed values.

Figure 4: Growth rates in quarterly consumption-output ratio (left) and annual dividends (right) against smoothed values $(\hat{R}_m - I)m_{t-1} + \hat{c}'_L \hat{\hat{s}}^L_{t|T} + \hat{\delta}_y \Delta y_t$ in the 'y linear' model, using $L = 4$ order functions of LRR-ARG state variables (x_t, σ^2_t) given full sample of $(\Delta y_t, c_t - y_t, \Delta d_t)$. Vertical axes show standard deviations from the mean.

Table 4 shows the estimated parameters $(\hat{R}_m, \hat{\delta}_y, \hat{\Sigma}_\varepsilon)$ of the measurement equation. Comparing panels (a) and (b) shows that latent state dependence reduces the unexplained variance in consumption growth by 76% and in dividend growth by 26%. Both series maintain large autoregressive components, but are not strongly mutually correlated. The estimates confirm that consumption depends on both observed and unobserved components of output growth. In particular, consumption growth is expected to increase by around 0.56 per additional percentage point of contemporaneous output growth, ceteris paribus, while dividends would only grow by 0.04 percentage points.

Table 4: Estimated autoregression matrix $\hat{R}_m$, contemporaneous growth impact $\hat{\delta}_y$, and error covariance matrix $\hat{\Sigma}_\varepsilon$ of 'y linear' model for quarterly measurements $m_t = (c_t - y_t, \Delta d_t)$ based on (a) least squares regression of m_t on $(m_{t-1}, \Delta y_t)$, and (b-c) LRR-ARG SML estimates with $L = 4$ for two sets of observations. Measurements are centered and standardized to unit variance.

(a) Initial estimates.

$\hat{R}_m$	0.968	-0.004
	-0.081	0.394
$\hat{\delta}_y$	-0.371	0.111
$\hat{\Sigma}_\varepsilon$	0.045	0.002
	0.002	0.732

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

$\hat{R}_m$	0.931	-0.024
	-0.143	0.274
$\hat{a}_y$	-0.437	0.037
$\hat{\Sigma}_\varepsilon$	0.011	-0.011
	-0.011	0.544

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

$\hat{R}_m$	0.917	-0.013
	-0.125	0.293
$\hat{\delta}_y$	-0.424	0.044
$\hat{\Sigma}_\varepsilon$	0.018	-0.001
	-0.001	0.616

4.4 Full-information estimates

Table 5 reports likelihood-based model selection criteria for the four model variants when the prices $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ are added to the measurements $m_t = (c_t - y_t, \Delta d_t)$ and growth observations Δy_t. For each model variant, the pricing functions allow for the same type of nonlinear state-dependence as the measurements.[14] The likelihood of output growth for all reported models is higher than when using the growth observations alone (-1.16, see Table 1). While the likelihoods of prices ℓ^P monotonically increase with the approximation order L of the pricing equation, the output growth likelihoods ℓ^Y appear to plateau around $L = 3$ for most variants. After considering the number of parameters, AIC favors quadratic or cubic models, while the BIC favors the quadratic model ($L = 2$) for three of four model variants. We therefore find somewhat stronger evidence of nonlinearity in the pricing than in the measurement equation.

Table 6 reports the simulated maximum likelihood estimates of the transition density parameters θ_s for the long-run risk model (2) with autoregressive Gamma stochastic variance. The estimates are for three datasets created by consecutively adding the measurements $m_t = (c_t - y_t, \Delta d_t)$ and prices $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ to the growth observations Δy_t, using the partially linear specification (17) with fourth order dampened polynomials (ψ_L, π_L). Differences in parameter estimates across datasets are generally less than two standard errors, suggesting no major misspecification of the measurement and pricing equations. Moreover, standard errors fall when adding measurements and prices, confirming that they help to

[14]For example, in Model 4 prices depend on the interaction term $R_p(s_t)m_t$.

Table 5: Simulated maximum log-likelihoods and information criteria of augmented LRR-ARG models for $(\Delta y_{t+1}, m_t, p_t)$ with $m_t = (c_t - y_t, \Delta d_t)$ and prices $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ from 1947:Q1-2019:Q4 by varying type and order L of measurement equation. Criteria are normalized by sample size T. Bold numbers indicate optimal L by type.

	ℓ^Y					ℓ^P				
L	1	2	3	4	5	1	2	3	4	5
y linear	-1.12	-1.07	**-1.06**	-1.10	-1.13	-0.83	-0.71	-0.71	-0.65	**-0.54**
y state-dep	-1.12	-1.07	-1.06	-1.06	**-1.05**	-0.84	-0.67	-0.66	-0.67	**-0.61**
ys non-linear	-1.12	-1.06	**-1.05**	-1.07	-1.10	-0.82	-0.71	-0.74	-0.67	**-0.46**
m state-dep	-1.12	-1.10	-1.09	**-1.09**	-1.12	-0.80	-0.64	-0.61	-0.63	**-0.51**
	AIC					*BIC*				
L	1	2	3	4	5	1	2	3	4	5
y linear	6.35	**5.98**	6.05	6.03	6.03	6.84	**6.63**	6.90	7.14	7.45
y state-dep	6.35	5.95	**5.92**	6.06	6.12	6.84	**6.65**	6.90	7.40	7.90
ys non-linear	6.31	6.04	**6.04**	6.14	6.18	6.80	**6.77**	7.13	7.74	8.47
m state-dep	6.29	**5.95**	6.09	6.44	6.56	**6.84**	6.85	7.51	8.53	9.48

identify the state variables. The estimated mean reversion parameters (ρ_x, ν) increase after adding prices, and correspond to half-lives of the expected growth and variance components of 1 and 7 years, respectively, suggesting the latter is particularly persistent.

The bottom row of Figure 5 confirms the persistence of the variance component of output growth, which has steadily declined since the post-war years despite some elevated periods during the 1980s energy crisis and to a lesser extent the 2008 financial crisis. For all three sets of observations, smoothed economic growth volatility reaches its lows during the late 1980s and the 1990s, in between which valuation ratios steadily increased.

Figure 6 shows the estimated price-dividend ratio and risk-free rate functions in $\hat{\pi}_L$, after controlling for the measurements $m_t = (c_t - y_t, \Delta d_t)$. The price-dividend function appears to increase monotonically in expected growth, and to decrease monotonically in the variance. However, when expected growth is low, the price-dividend ratio responds less to volatility falling below its median. Intuitively, during recessions investors may actually prefer some volatility for the growth rate to revert to its mean faster. Meanwhile, the risk-free rate function increases fairly linearly with expected growth for low levels of uncertainty. However, it also peaks when low expected growth combines with high uncertainty, which may be explained by contractionary monetary policy aiming to break inflationary spirals.

Table 6: Simulated maximum likelihood estimates of the transition density parameters θ_s for the LRR-ARG model, for the sampling period 1947:Q1-2019:Q4 and subsets of Δy_t, $m_t = (c_t - y_t, \Delta d_t)$ and $p_t = (\log \frac{P_t}{D_t}, r_t^f)$. Estimates and standard errors (in brackets) based on the EM algorithm, using $N^s = 5,000$ simulated particles and $L = 4$.

	Based on Δy_t		Based on $(\Delta y_t, m_t)$		Based on $(\Delta y_t, m_t, p_t)$	
θ^s	Est	Std	Est	Std	Est	Std
μ	0.659	(0.035)	0.620	(0.029)	0.557	(0.026)
ρ_x	0.809	(0.120)	0.737	(0.066)	0.830	(0.050)
ϕ_x	0.325	(0.109)	0.464	(0.047)	0.257	(0.023)
ϕ_v	0.902	(0.340)	0.949	(0.261)	1.180	(0.416)
$\bar{\sigma}$	0.847	(0.315)	0.883	(0.329)	0.851	(0.387)
ν	0.865	(0.131)	0.935	(0.062)	0.975	(0.028)

Figure 7 shows the time series fit of the prices for both the linear and quartic pricing equations. Both models can explain drops in the price-dividend ratio during highly uncertain recessions such as the early 1980s and the Great Recession. However, the quartic model is better able to explain the sustained rise in valuation leading up to the dot-com bubble, due to the amplification of high expected growth and low volatility. Still, neither model can fully explain the peak valuations, confirming the need for the serially correlated pricing error η_t unrelated to fundamentals. The quartic model also better matches the high risk-free rates during the early 1980s contractionary monetary policy and the low rates during the 2010s. The pricing error parameter estimates in Table 7 confirm the negative relation between the unexplained components of the price-dividend ratio and risk-free rates, and their reduced scale and/or persistence after allowing for state-dependence.

Finally, we consider the use of realized variance proxies $\log RV_t$ in addition to the prices p_t using the measurement equation $\log RV_{t+1} = \psi_L^v(s_t) + \eta_{t+1}^v$, where the errors η_t^v may be possibly serially correlated. Figure 8a plots the realized variance proxies against their full-information smoothed values. The relatively low financial market volatility during the economically uncertain 1950s can be explained by the low sensitivity of the price-dividend ratio when growth volatility is high. In contrast, valuations were more sensitive to shocks during the moderate levels of economic uncertainty around the Great Recession. Meanwhile, short-lived bumps in return volatility outside recessions such as the 1987 Black Monday crash are treated as pure variance shocks. The estimated realized variance function $\hat{\psi}_L^v$ in Figure 8b reveals its nonlinear relation to the variance of economic growth. It appears to trade-off the effect of more volatile economic shocks against higher sensitivity to shocks when the price-dividend ratio is high due to low uncertainty. High valuation ratios can explain the highly volatile returns during 'good' times when growth is high and uncertainty is low. Overall, return volatility peaks during recessions with moderate rather than high levels of uncertainty, when there are still downside risks to valuations.

Table 7: Estimated regression coefficients $\hat{\alpha}$ of the prices on measurements, and autocovariance matrices $\hat{\Sigma}_\eta(j)$ of order $j = 0, 1$ of the pricing error η_t, based on (a) least squares regression of $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ on $m_t = (c_t - y_t, \Delta d_t)$, and (b) SML estimates based on $(\Delta y_t, m_t, p_t)$. Prices and measurements are centered and standardized to unit variance.

	(a) Initial estimates.			(b) SML estimates.	
$\hat{\alpha}$	-0.117	0.060	$\hat{\alpha}$	-0.101	-0.120
	0.233	-0.062		0.053	0.004
$\hat{R}_p$	0.972	-0.006	$\hat{R}_p$	0.888	-0.005
	0.010	0.887		-0.090	0.881
$\hat{\Sigma}_\omega$	0.054	–	$\hat{\Sigma}_\omega$	0.055	–
	-0.014	0.202		0.021	0.093

4.5 Estimated SDF components

Figure 9 shows the estimated state-dependent discount factor $\hat{\phi}_{L^*}$ and the profiled GMM criterion for various approximation orders L of the measurement and pricing equations. They are computed using on the eigenproblem (26) with the simulation-based estimates of G_L and M_L given by

$$\hat{G}_{L*} = \frac{1}{T} \sum_{t=0}^{T-1} \left(\hat{\bar{s}}_{t|T}^{L*} \hat{\bar{s}}_{t|T}^{L*\prime} + \hat{V}_{t|T}^{L*} \right)$$

$$\hat{M}_{L*} = \frac{1}{T} \sum_{t=0}^{T-1} \left(\hat{\bar{s}}_{t|T}^{L*} \hat{\bar{s}}_{t+1|T}^{L*}{}' + \hat{C}_{t,t+1|T}^{L*} \right) \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{t+1}^d,$$

where $\hat{V}_{t|T}^L$ and $\hat{C}_{t,t+1|T}^L = \mathrm{Cov}_{\hat{\vartheta}}(\bar{s}_t^L, \bar{s}_{t+1}^L \mid \mathcal{F}_{1:T}^{p,m,y})$ are the smoothed variance and first-order autocovariance matrices. The estimator of the positive eigenfunction is less sensitive to its approximation order L^* than to L, so that we choose a higher order for the former. The state-dependent discount factor tends to move inversely to the price-dividend ratio. In particular, it reaches its highest levels when expected growth is low and volatility is high, and vice versa. Moreover, it remains elevated at low volatility levels for low expected growth. These dependencies indicate that consumption and dividend dynamics cannot fully explain the state-dependence of the price-dividend ratio. Instead, the estimates rationalize the latter through the discount factor, giving higher marginal utility for payoffs during low growth and/or uncertain times, thus inducing a negative risk premium for output growth volatility. The profiled GMM-criterion for $L = 3$ is minimized by the risk aversion parameter $\hat{\gamma} = 11.25$ and yields a quarterly discount parameter of $\hat{\beta} = 0.992$. These moderate values suggest that power utility over consumption may not be unreasonable with some additional state-dependent discounting. However, the risk aversion parameter is relatively weakly identified, as evidenced by its sensitivity to L. Nonetheless, values in the range $5 < \gamma < 20$ yield qualitatively similar state-dependent discount factors.

Figure 10 shows the estimated scaled continuation value $\hat{\chi}_{L^*}$ under recursive preferences, and the profiled GMM criterion as a function of the preference parameters (β, γ). They are computed based on the fixed point of the approximated recursion (29). The scaled continuation value is comparable to the state-dependent discount factor, as it almost monotonically decreases in expected growth and increases in the variance component, although the latter relation is weaker. This make sense given that both specifications yield the same SDF whenever the time-discount factor $\beta = 1$, which is close to the estimate $\hat{\beta} = 0.994$. The risk aversion parameter estimate is similarly moderate at $\hat{\gamma} = 9.0$, although the GMM criterion shows that it is relatively weakly identified on the upside. This confirms that recursive preferences can rationalize nonlinear state-dependence in valuation ratios even with mostly linear state-dependence in consumption growth.

5 Conclusion

This paper develops a class of nonlinear Markovian asset pricing models in which consumption and dividends are described via general functions of latent state variables describing persistent components in the aggregate output growth distribution. We establish the nonparametric identification of the measurement and pricing equations under a limited feedback condition, and establish the consistency of a sieve maximum likelihood estimator under general conditions. We implement the estimator using a tractable simulation-based algorithm for filtering and smoothing. Subsequently, we study the identification and consistent estimation of semiparametric stochastic discount factor models by formulating the Euler equation as an eigenfunction problem. The estimated price-dividend ratio is increasing in expected growth, decreasing in growth volatility, with both effects interacting, and shows stronger state-dependence than can be rationalized under power utility of consumption. Instead, the evidence supports models in which investors have moderate relative risk aversion but higher marginal utility for payoffs in times of low expected growth and high volatility. Finally, the steeply declining price-dividend ratios for low levels of growth volatility help explain bursts of stock market volatility during periods of moderate economic uncertainty.

References

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

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

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

Andersen, T. G., Fusari, N., and Todorov, V. (2015). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

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

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

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

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

Bansal, R., Dittmar, R., and Kiku, D. (2009). Cointegration and consumption risks in asset returns. *The Review of Financial Studies*, 22(3):1343–1375.

Bansal, R., Gallant, A. R., and Tauchen, G. (2007). Rational pessimism, rational exuberance, and asset pricing models. *The Review of Economic Studies*, 74(4):1005–1033.

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

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

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

Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251.

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

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

Chen, X., Favilukis, J., and Ludvigson, S. C. (2013). An estimation of economic models with recursive preferences. *Quantitative Economics*, 4(1):39–83.

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

Christensen, T. M. (2015). Nonparametric identification of positive eigenfunctions. *Econometric Theory*, 31(6):1310–1330.

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

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

Douc, R. and Moulines, E. (2012). Asymptotic properties of the maximum likelihood estimation in misspecified hidden markov models. *The Annals of Statistics*, 40(5):2697–2732.

Douc, R., Moulines, E., and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime. *The Annals of statistics*, 32(5):2254–2304.

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.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

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

Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., and Srisuma, S. (2020). Nonparametric euler equation identification and estimation. *Econometric Theory*, pages 1–41.

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

Freyberger, J. and Masten, M. A. (2019). A practical guide to compact infinite dimensional parameter spaces. *Econometric Reviews*, 38(9):979–1006.

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

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

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

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

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

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

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

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

Hansen, L. P., Heaton, J. C., and Li, N. (2008). Consumption strikes back? measuring long-run risk. *Journal of Political economy*, 116(2):260–302.

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.

Hansen, L. P. and Scheinkman, J. A. (2012). Recursive utility in a markov environment with stochastic growth. *Proceedings of the National Academy of Sciences*, 109(30):11967–11972.

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

Hu, Y. and Shiu, J.-L. (2018). Nonparametric identification using instrumental variables: sufficient conditions for completeness. *Econometric Theory*, 34(3):659–693.

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

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

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

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

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

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

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

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

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

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

Rodriguez, J. M. (2003). Approximation by polynomials and smooth functions in sobolev spaces with respect to measures. *Journal of Approximation Theory*, 120(2):185–216.

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

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

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

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

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

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

A Appendix

A.1 Proofs

Proof of Proposition 1. The results are obtained in Douc and Moulines (2012, 'DM') under their Assumption (A1) and (A2) for HMM models with general state spaces, which in turn are shown to hold under their Assumptions (NL1)-(NL4) for models with infinite state spaces and existing densities with respect to the Lebesgue measure. Our assumptions are almost identical to (NL1)-(NL4), except do not require the measurement density to be continuous with respect to ϑ, which is only required for their continuity condition (A3). To check that (A1) still holds, the uniformly bounded and continuous transition density Assumption 3a) implies that any compact set $D \subset \mathcal{S}$ is 1-small, so that (A1)(ii) is satisfied with $r = 1$, and there exists a constant $\delta > 0$ and probability measure ν with $\nu(D) > 0$ such that $\inf_{\theta_s \in \Theta} \inf_{s \in \mathcal{D}} \mathbb{P}_{\theta_s}(s' \in D \mid s) \geq \delta\nu(D)$. Therefore (A1)(iii) follows from

$$E\left(\log^- \inf_{\vartheta\in\Theta} \inf_{s\in\mathcal{D}} \int_D f_\vartheta(Z_1 \mid s) f_{\theta_s}(s' \mid s) ds'\right) \leq E\left(\log^- \inf_{\vartheta\in\Theta} \inf_{s\in\mathcal{D}} f_\vartheta(Z_1 \mid s)\right) + \log^- \delta\nu(D) < \infty,$$

where the last part follows directly from Assumption 3d). (A2) is part of our Assumption 3c). Parts (i)-(iii) of Proposition 1 are stated, respectively, in the proof of DM Proposition 10(ii), and as Propositions 10(i) and 10(iii). DM Proposition 3 implies the results hold for any χ_0 with $\chi_0(D) > 0$. $\square$

Proof of Theorem 1. The proof is based on Lemma A1 in Newey and Powell (2003). This requires that (i) there is unique ϑ_0 that maximizes $\bar{\ell}(\vartheta)$ on $\mathit{\Theta}$, (ii) $\mathit{\Theta}_T$ are compact subsets of $\mathit{\Theta}$ such that for any $\vartheta \in \mathit{\Theta}$ there exists a $\tilde{\vartheta}_T \in \mathit{\Theta}_T$ such that $\tilde{\vartheta}_T \xrightarrow{p} \vartheta$, and (iii) $\ell_T(\vartheta)$ and $\bar{\ell}(\vartheta)$ are continuous, $\mathit{\Theta}$ is compact, and $\max_{\vartheta\in\mathit{\Theta}} |\ell_T(\vartheta) - \bar{\ell}(\vartheta)| \xrightarrow{p} 0$. We verify these conditions for $\mathit{\Theta} = \Theta \times \bar{\mathcal{H}}$, subspaces $\mathit{\Theta}_T = \Theta \times \mathcal{H}_{L(T)}$, and norm $\|\cdot\|_c$.

The identification condition (i) follows directly from Assumption 4a). The compact subset condition in (ii) holds for $\mathit{\Theta}_T$ by construction of the finite-dimensional subspaces $\mathcal{H}_T$, while $\mathit{\Theta}$ is compact under the chosen norm by Gallant and Nychka (1987, Theorem 1) and the Tychonoff Theorem. Assumption 4b) directly implies that for any $\vartheta \in \mathit{\Theta}$ we can find a sequence $\vartheta_T \in \mathit{\Theta}_T$ that satisfies $\|\vartheta_T - \vartheta\|_c \to 0$, as the constructed subspaces $\mathcal{H}_T$ are dense in $\bar{\mathcal{H}}$.

For (iii), continuity of $l_{1,k}(\vartheta) = \log f_{\vartheta,\chi_0}(Z_1 \mid Z_{-k:0})$ and thereby of $\ell_T(\vartheta)$ in ϑ carries over from the continuity of the measurement and transition densities and the fact that $\sup_{\vartheta\in\mathit{\Theta}} \prod_{t=1}^{k} \sup_{s\in\mathcal{S}} f_\vartheta(Z_t \mid s) < \infty$ a.s. for every $k \in \mathcal{Z}^+$. Since $l_{1,k}(\vartheta)$ converges uniformly in ϑ to $\log \bar{f}_\vartheta(Z_1 \mid Z_{-\infty:0}) \equiv l_{1,\infty}(\vartheta)$ when $k \to \infty$, the latter is itself continuous. Furthermore, Assumption 4d) implies that $E\left(\sup_{\tilde{\vartheta}\in\mathcal{N}_\vartheta} |l_{1,\infty}(\vartheta)|\right) < \infty$ for any $\vartheta \in \mathit{\Theta}$, since

$$E\left(\sup_{\tilde{\vartheta}\in\mathcal{N}_\vartheta} l^+_{1,\infty}(\vartheta)\right) \le E\left(\log^+ \sup_{\vartheta\in\mathit{\Theta}} \sup_{s\in\mathcal{S}} f_\vartheta(Z_1|s)\right) < \infty$$

by Assumption 3c). Using monotone convergence, for any $\vartheta \in \mathit{\Theta}$

$$\lim_{\delta\to 0} E\left(\sup_{\|\tilde{\vartheta}-\vartheta\|_c\le\delta} |l_{1,\infty}(\tilde{\vartheta}) - l_{1,\infty}(\vartheta)|\right) = E\left(\lim_{\delta\to 0} \sup_{\|\tilde{\vartheta}-\vartheta\|_c\le\delta} |l_{1,\infty}(\tilde{\vartheta}) - l_{1,\infty}(\vartheta)|\right) = 0. \quad (32)$$

As a result, $\bar{\ell}(\vartheta)$ is continuous, since for any $\vartheta \in \Theta$

$$\begin{aligned}\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |\bar{\ell}(\tilde{\vartheta}) - \bar{\ell}(\vartheta)| &= \sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |E\left(l_{1,\infty}(\tilde{\vartheta}) - l_{1,\infty}(\vartheta)\right)| \\ &\leq E\left(\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |l_{1,\infty}(\tilde{\vartheta}) - l_{1,\infty}(\vartheta)|\right) \to 0, \quad \delta \to 0.\end{aligned}$$

We establish uniform convergence by bounding the RHS terms in the triangle equality

$$|\ell_T(\vartheta) - \bar{\ell}(\vartheta)| \leq |\ell_T(\vartheta) - \ell_T^{\infty}(\vartheta)| + |\ell_T^{\infty}(\vartheta) - \bar{\ell}(\vartheta)|, \tag{33}$$

where $\ell_T^{\infty}(\vartheta) = \frac{1}{T}\sum_{t=1}^{T} l_t^{\infty}(\vartheta)$. The first term in (33) vanishes uniformly when $T \to \infty$, as

$$\sup_{\vartheta\in\Theta} |\ell_T(\vartheta) - \ell_T^{\infty}(\vartheta)| \leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\vartheta\in\Theta}|l_t(\vartheta) - l_t^{\infty}(\vartheta)| \leq \frac{1}{T}\sum_{t=1}^{T} C_t\kappa^t \xrightarrow{\text{a.s.}} 0,$$

using Proposition 1(i), since $\kappa \in (0,1)$ and $\mathbb{P}(C_t < \infty) = 1$ for $t = 1, \ldots, T$. Since Θ is compact, the second term in (33) vanishes uniformly if for any $\vartheta \in \Theta$,

$$\limsup_{\delta\to 0}\limsup_{T\to\infty}\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |\ell_T^{\infty}(\tilde{\vartheta}) - \bar{\ell}(\vartheta)| = 0 \quad \text{a.s.}$$

Following the proof of Douc et al. (2004, Prop. 2), we obtain that

$$\begin{aligned}\limsup_{\delta\to 0}\limsup_{T\to\infty}\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |\ell_T^{\infty}(\tilde{\vartheta}) - \bar{\ell}(\vartheta)| &= \limsup_{\delta\to 0}\limsup_{T\to\infty}\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |\ell_T^{\infty}(\tilde{\vartheta}) - \ell_T^{\infty}(\vartheta)| \\ &\leq \limsup_{\delta\to 0}\limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |l_t^{\infty}(\tilde{\vartheta}) - l_t^{\infty}(\vartheta)| \\ &= \limsup_{\delta\to 0} E\left(\sup_{\|\tilde{\vartheta}-\vartheta\|_c \leq \delta} |l_1^{\infty}(\tilde{\vartheta}) - l_1^{\infty}(\vartheta)|\right) = 0,\end{aligned}$$

using the almost sure pointwise convergence in the first step, and the Birkhoff ergodic theorem and (32) in the last two steps. The conclusion follows by applying Lemma A1 in Newey and Powell (2003). $\square$

Sufficient conditions for Assumption 6b). Let $\tilde{b}^L(s) = G_L^{-1/2} b^L(s)$ be the orthogonalized basis functions. Define the sequence $\xi_L = \sup_s \|\tilde{b}^L(s)\|$ to control convergence rates. For noncompact support, unbounded basis functions should be trimmed or its tails down-weighted so that ξ_L is finite. Define the operator $\mathbb{M}^{(\vartheta)}\phi(s) = E_{\theta^s}\left(\mathcal{K}(s_t, s_{t+1}; \vartheta)\phi(s_{t+1}) \mid s_t = s\right)$, and let $\ell^*(\vartheta) = \|\mathbb{M}^{(\vartheta)} - \mathbb{M}\|$. Note that $\ell^*(\vartheta_0) = 0$ since $\mathbb{M}^{(\vartheta_0)} = \mathbb{M}$. The following conditions assure the consistency of the plug-in estimators $\hat{G}_L$ and $\hat{M}_L$ in the operator norm. Related conditions for weakly dependent sample averages are in Christensen (2017, Appendix C.1) for various first-stage estimators and Escanciano et al. (2020, Appendix B) for kernel estimators.

Lemma A1. *Suppose the following conditions hold:*

a) $\|\hat{\theta}^s - \theta_0^s\| = O_p\left(T^{-1/2}\right)$.

b) $\xi_L^2/T = o(1)$.

c) The marginal state density $f(s;\theta^s)$ is twice continuously differentiable in θ^s in a neighborhood $\mathcal{N}$ of θ_0^s for all $s \in \mathcal{S}$, and

(i) $\sup_{\theta^s\in\mathcal{N}} E_{\theta^s}\left(\left\|\frac{\partial \log f(s_t;\theta^s)}{\partial\theta^s}\right\|^2\right) < \infty$

(ii) $\sup_{\theta^s\in\mathcal{N}} E_{\theta^s}\left(\left\|\frac{\partial^2 \log f(s_t;\theta^s)}{\partial\theta^s\partial'\theta^s}\right\|\right) < \infty$.

d) $\ell^*(\vartheta)$ is pathwise differentiable at ϑ_0 with $|\ell^*(\vartheta) - \ell^*(\vartheta_0) - \dot{\ell}^*(\vartheta_0)[\vartheta - \vartheta_0]| = O\left(\|\vartheta - \vartheta_0\|_c^2\right)$, $\|\vartheta - \vartheta_0\|_c = o_p(T^{-1/4})$, and $\dot{\ell}^*(\vartheta_0)[\vartheta - \vartheta_0] = O_p(T^{-1/2})$.

Then Assumption 6b) holds.

Proof: Since $\hat{G}_L^o = E_{\hat{\theta}^s}\left(\tilde{b}^L(s)\tilde{b}^L(s)'\right)$ and $I_L = E_{\theta_0^s}\left(\tilde{b}^L(s_t)\tilde{b}^L(s_t)'\right)$, we can write

$$\hat{G}_L^o - I_L = \int \tilde{b}^L(s)\tilde{b}^L(s)'\left(f(s;\hat{\theta}^s) - f(s;\theta_0^s)\right)ds.$$

Let $U^L = \{u \in \mathbb{R}^L : \|u\| = 1\}$ be the unit sphere. By a Taylor expansion and the mean value theorem, there exists a $\tilde{\theta}^s$ between $\hat{\theta}^s$ and θ_0^s such that $f(s;\hat{\theta}^s) - f(s;\theta_0^s) = \frac{\partial f(s;\theta_0^s)}{\partial'\theta^s}(\hat{\theta}^s - \theta_0^s) + \frac{1}{2}(\hat{\theta}^s - \theta_0^s)'\frac{\partial^2 f(s;\tilde{\theta}^s)}{\partial\theta^s\partial'\theta^s}(\hat{\theta}^s - \theta_0^s)$. Therefore

$$\begin{aligned}
\|\hat{G}_L^o - I_L\| &= \sup_{u,v\in U^L}\left|\int (u'\tilde{b}^L(s))(v'\tilde{b}^L(s))\left(f(s;\hat{\theta}^s) - f(s;\theta_0^s)\right)ds\right| \\
&\leq \xi_L \sup_{u\in U^L}\int \left|(u'\tilde{b}^L(s))\right| \times \left|\frac{\partial \log f(s;\theta_0^s)}{\partial'\theta^s}(\hat{\theta}^s - \theta_0^s)\right| f(s;\theta_0^s)ds \\
&\quad + \frac{1}{2}\xi_L^2\int\left|(\hat{\theta}^s - \theta_0^s)'\left(\frac{\partial \log f(s;\tilde{\theta}^s)}{\partial\theta^s}\frac{\partial \log f(s;\tilde{\theta}^s)}{\partial'\theta^s} + \frac{\partial^2 \log f(s;\tilde{\theta}^s)}{\partial\theta^s\partial'\theta^s}\right)(\hat{\theta}^s - \theta_0^s)\right| f(s;\theta_0^s)ds \\
&:= g_1 + g_2,
\end{aligned}$$

where the first line uses $\|A\| = \sup_{u,v \in U^L} |u'Av|$ for any matrix A, and the second line uses the triangle inequality and the identities $\partial f = f \partial \log f$ and $\partial^2 f = f(\partial \log f)^2 + f\partial^2 \log f$.

For the first term, the Cauchy-Schwartz and Hölder inequalities yield

$$
\begin{aligned}
g_1 &\leq \xi_L \sup_{u \in U^L} \int \left|(u'\tilde{b}^L(s))\right| \times \left\| \frac{\partial \log f(s;\theta_0^s)}{\partial \theta^s} \right\| f(s;\theta_0^s) ds \times \|\hat{\theta}^s - \theta_0^s\| \\
&\leq \xi_L \left(\sup_{u \in U^L} u' G_L^o u \right)^{1/2} \times E_{\theta_0^s} \left(\left\| \frac{\partial \log f(s_t;\theta_0^s)}{\partial \theta^s} \right\|^2 \right)^{1/2} \times \|\hat{\theta}^s - \theta_0^s\|,
\end{aligned}
$$

where $\sup_{u \in U^L} u'G_L^o u = 1$ since $G_L^o = I_L$. For the second term, the Cauchy-Schwartz inequality and the definition of the operator norm yield

$$
g_2 \leq \frac{1}{2}\xi_L^2 \left[E_{\tilde{\theta}^s} \left(\left\| \frac{\partial \log f(s_t;\tilde{\theta}^s)}{\partial \theta^s} \right\|^2 \right) + E_{\tilde{\theta}^s} \left(\left\| \frac{\partial^2 \log f(s_t;\tilde{\theta}^s)}{\partial \theta^s \partial' \theta^s} \right\| \right) \right] \times \|\hat{\theta}^s - \theta_0^s\|^2.
$$

Therefore conditions a) and b) imply $\|\hat{G}_L^o - I_L\| = O_p(\xi_L T^{-1/2}) = o_p(1)$.

Finally, condition d) implies that

$$
\|\hat{M}_L^o - M_L^o\| = \|\Pi_L \left(\mathbb{M}^{(\hat{\vartheta})} - \mathbb{M} \right) \| \leq \ell^*(\hat{\vartheta}) = \dot{\ell}^*(\vartheta_0)[\hat{\vartheta} - \vartheta_0] + O\left(\|\hat{\vartheta} - \vartheta_0\|_c^2 \right) = O_p(T^{-1/2}),
$$

which is of smaller order. $\square$

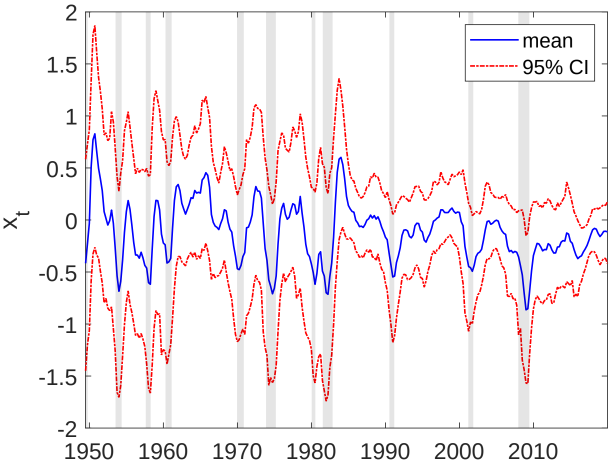
A.2 Further empirical results

Table 8: Simulated maximum log-likelihoods and information criteria of augmented LRR-LOGV models for $(\Delta y_{t+1}, m_t)$ with $m_t = (c_t - y_t, \Delta d_t)$ from 1947:Q1-2019:Q4 by varying type and order L of measurement equation. Criteria are normalized by sample size T. Bold numbers indicate optimal L by type.

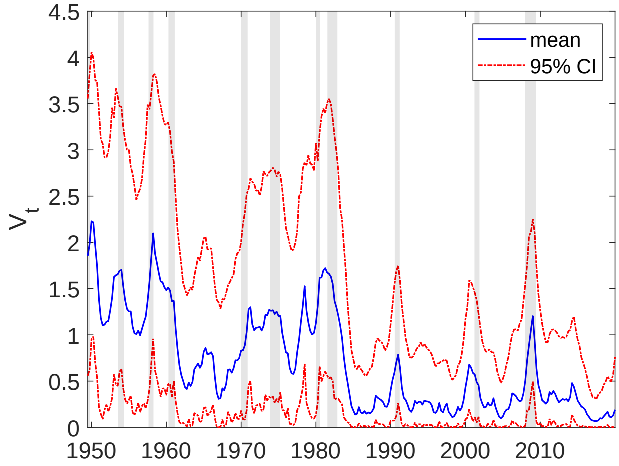
	ℓ^Y					ℓ^M				
L	1	2	3	4	5	1	2	3	4	5
y linear	-1.12	-1.10	-1.07	**-1.07**	-1.07	-1.10	-1.04	-1.03	-0.96	**-0.88**
y state-dep	-1.12	-1.09	-1.07	**-1.04**	-1.05	-1.10	-1.03	-1.00	-0.95	**-0.82**
ys non-linear	-1.12	-1.09	-1.07	**-1.04**	-1.05	-1.09	-1.02	-0.97	-0.92	**-0.76**
m state-dep	-1.12	-1.09	-1.07	-1.05	**-1.04**	-1.08	-0.99	-0.93	-0.82	**-0.72**
	AIC					*BIC*				
L	1	2	3	4	5	1	2	3	4	5
y linear	4.60	4.48	4.47	4.41	**4.34**	4.88	**4.83**	4.92	4.99	5.07
y state-dep	4.60	4.48	4.47	4.47	**4.37**	**4.88**	4.88	5.06	5.28	5.47
ys non-linear	4.58	4.47	**4.46**	4.52	4.51	**4.85**	4.89	5.15	5.59	6.13
m state-dep	4.60	4.47	4.44	**4.37**	4.38	**4.93**	4.97	5.20	5.47	5.89

Table 9: Simulated maximum log-likelihoods and information criteria of augmented LRR-ARG models for $(\Delta y_{t+1}, m_t)$ with $m_t = (c_t - y_t, \Delta d_t)$ from 1947:Q1-2021:Q4 by varying type and order L of measurement equation. Criteria are normalized by sample size T. Bold numbers indicate optimal L by type.

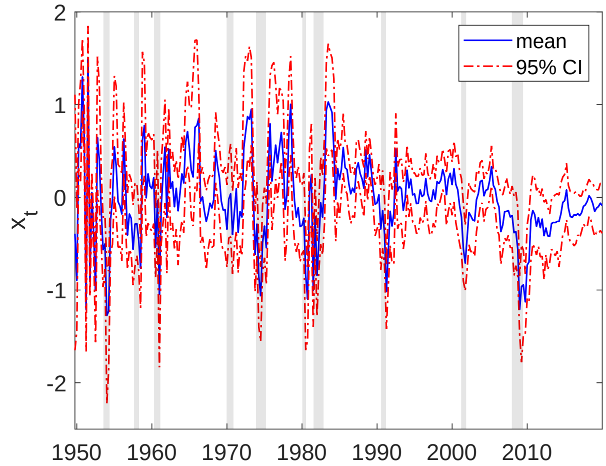
	ℓ^Y					ℓ^M				
L	1	2	3	4	5	1	2	3	4	5
y linear	-1.29	-1.26	-1.23	-1.21	**-1.20**	-1.53	-1.24	-1.16	-1.12	**-1.11**
y state-dep	-1.29	-1.21	-1.17	**-1.16**	-1.17	-1.53	-1.15	-1.10	-1.02	**-1.00**
ys non-linear	-1.30	-1.22	-1.17	-1.17	**-1.15**	-1.52	-1.14	-1.08	-0.98	**-0.91**
m state-dep	-1.25	**-1.21**	-1.23	-1.24	-1.27	-1.36	-1.15	-1.06	-1.04	**-1.01**
	AIC					*BIC*				
L	1	2	3	4	5	1	2	3	4	5
y linear	5.77	5.20	5.03	**4.97**	5.02	6.04	5.55	**5.48**	5.54	5.75
y state-dep	5.80	4.92	4.82	**4.74**	4.86	6.06	**5.26**	5.31	5.44	5.80
ys non-linear	5.80	4.96	**4.88**	4.88	4.9	6.07	**5.38**	5.55	5.93	6.568
m state-dep	5.39	5.00	**4.99**	5.16	5.36	5.71	**5.49**	5.74	6.24	6.84



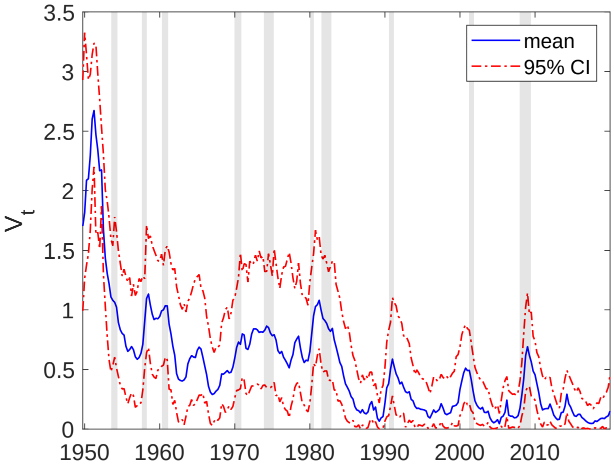
(a) Smoothed expected growth using (Δy_t).



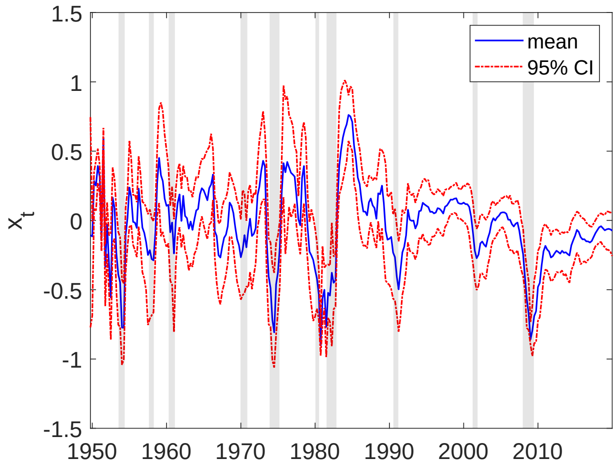
(b) Smoothed variances using (Δy_t).



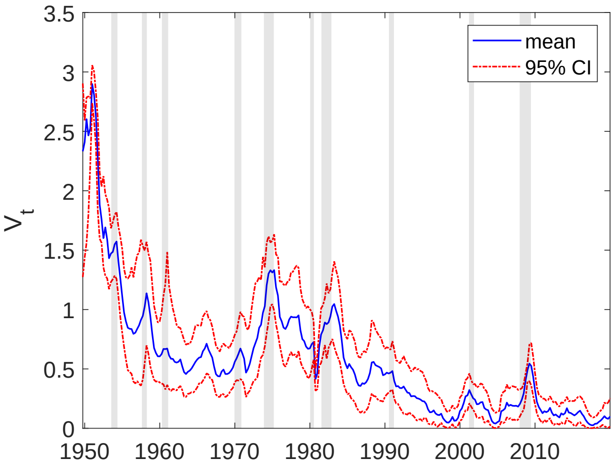
(c) Smoothed expected growth using $(\Delta y_t, m_t)$.



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

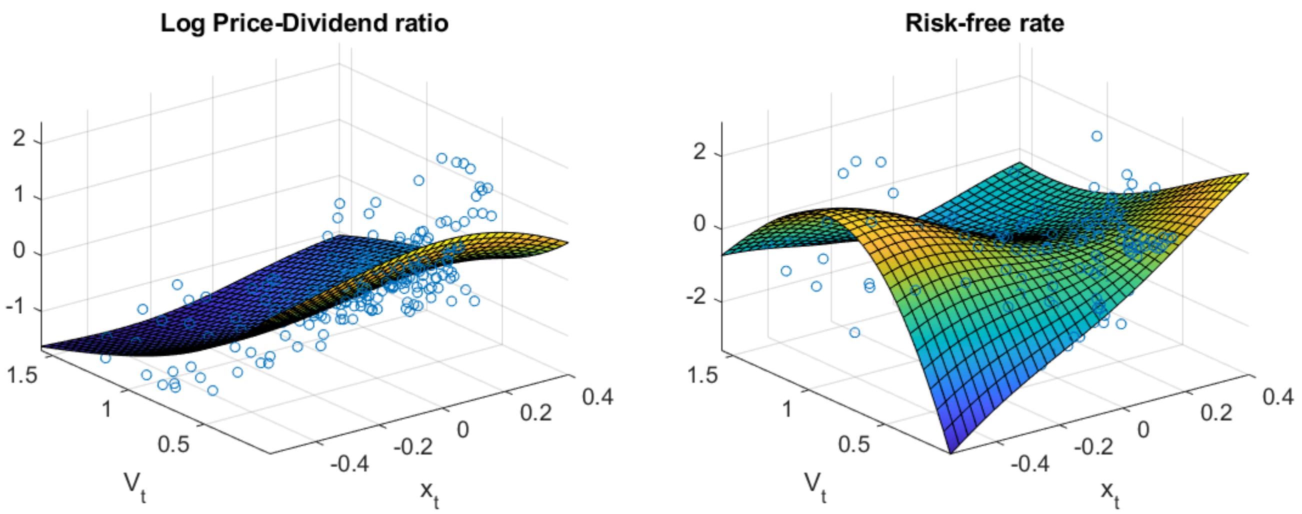


(e) Smoothed expected growth using $(\Delta y_t, m_t, p_t)$.



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

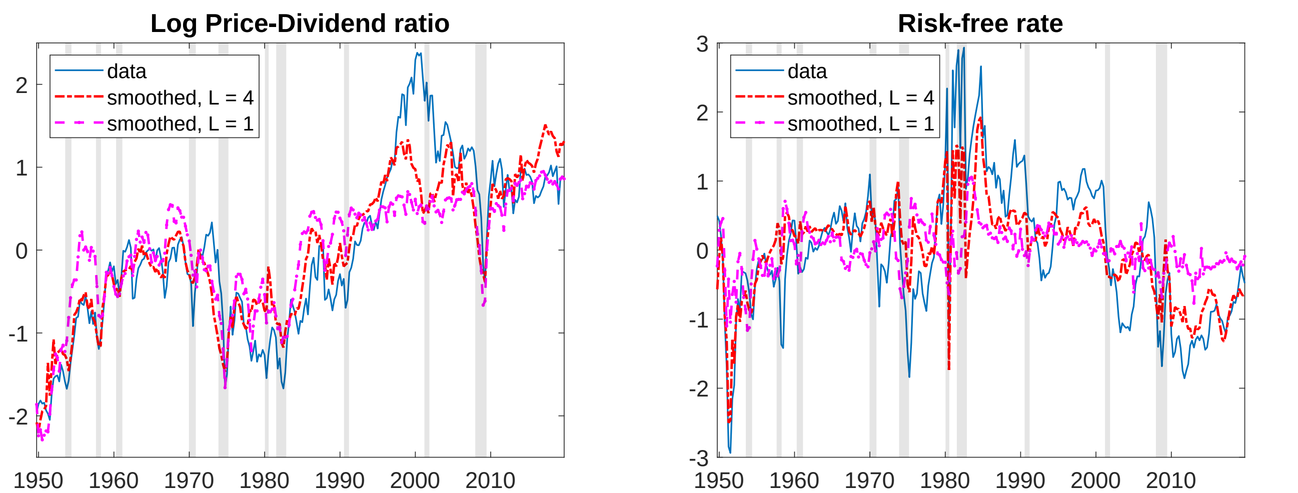
Figure 5: Smoothed conditional means x_t (left panels, in annualized percentages) and variances V_t (right panels) of the LRR-ARG model, based on observations of Δy_t, $m_t = (c_t - y_t, \Delta d_t)$ and $p_t = (\log \frac{P_t}{D_t}, r_t^f)$ from 1947:Q1-2019:Q4, for the 'y linear' variant with $L = 4$. Dashed lines represent 95% confidence intervals based on simulated particles. Grey shades indicate NBER recession dates.



(a) Log price-dividend ratio function.

(b) Risk-free rate function.

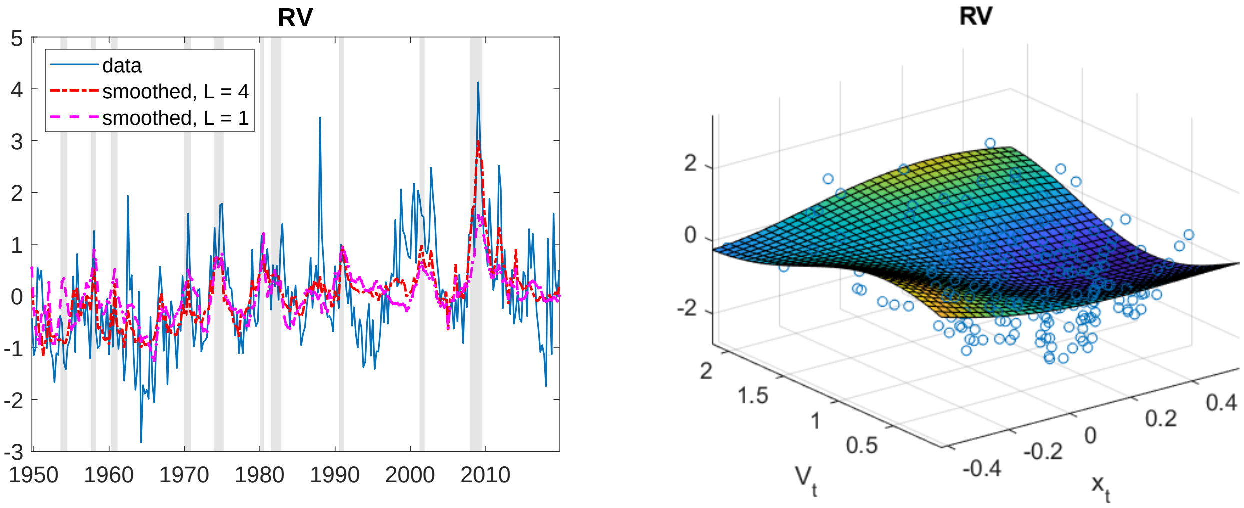
Figure 6: Estimated pricing functions in terms of the conditional mean x_t and variance V_t of output growth under the LRR-ARG model, using quarterly observations (y_t, m_t, p_t) and the 'y linear' variant with $L = 4$ order expansion. Blue circles plot data against the smoothed means of state variables. Horizontal axes capture 95% of marginal distributions. Vertical axis measures standard deviations $\sqrt{\mathrm{Var}(p_{it})}$ from the mean



(a) Log price-dividend ratio.

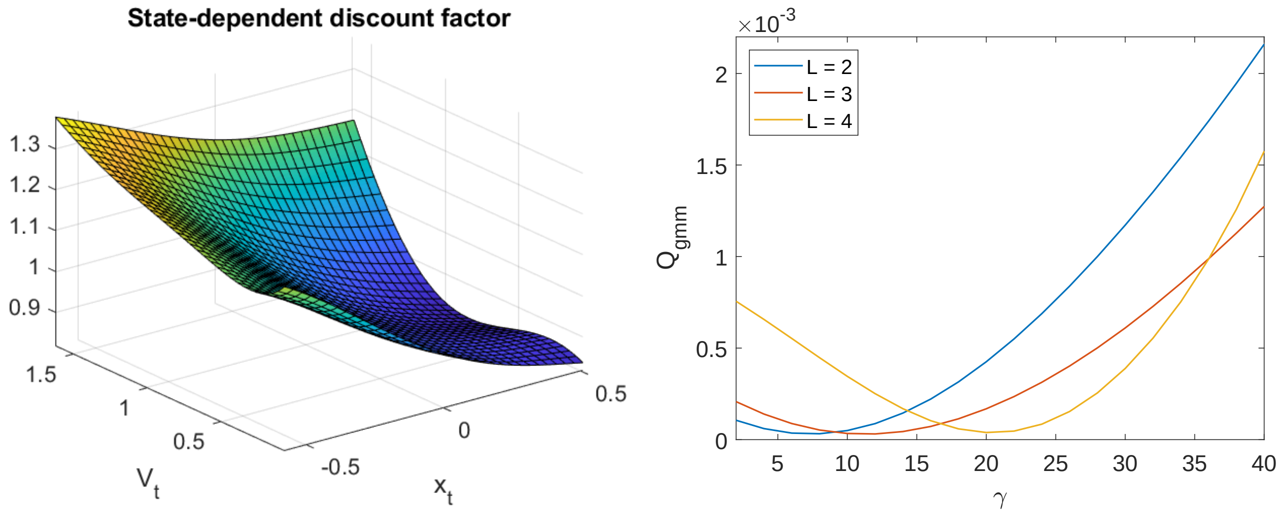
(b) Risk-free rate.

Figure 7: Time series of prices and realized variance proxies against smoothed values $\hat{b}_L' \hat{\hat{s}}^L_{t|T} + \hat{\alpha}' m_t$ using $L = 4$ order joint conditional moments of state variables (x_t, σ_t^2) given quarterly observations of $(\Delta y_t, m_t, p_t)$. Plots with prices include the least squares fitted values $\tilde{p}_t = \tilde{\alpha}' m_t$ based on measurements only. Vertical axis shows standard deviations from the mean.



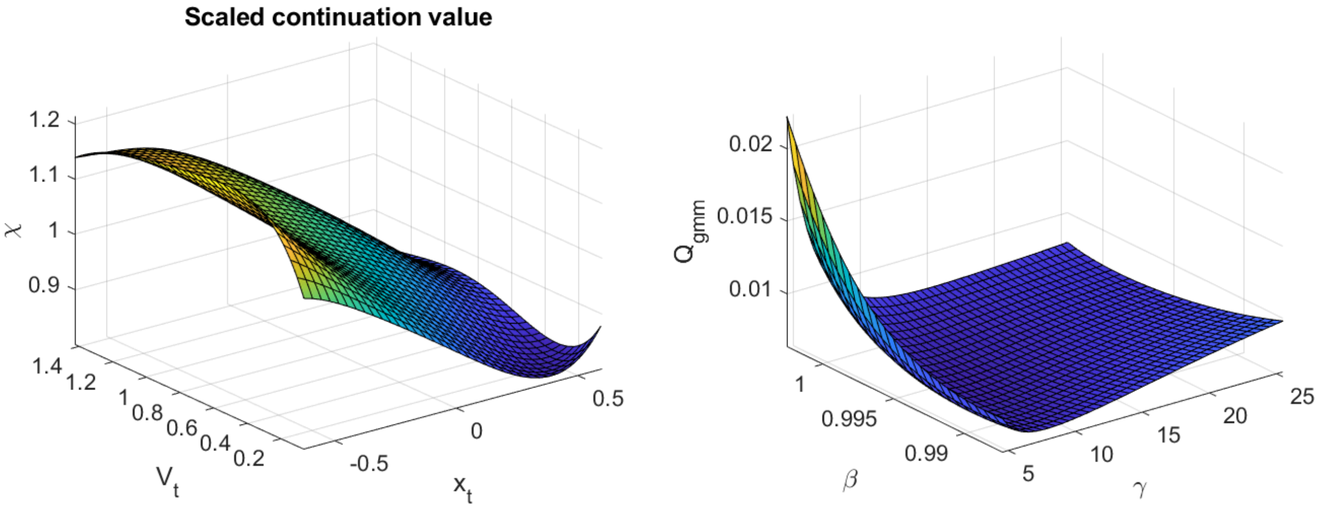
(a) Log realized return variance. (b) Fitted log realized variance function.

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



(a) State-dependent discount factor. (b) Profiled GMM criterion.

Figure 9: State-dependent discount factor $\hat{\phi}_{L^*}$ estimated as $L^* = 6$ order approximation to the positive eigenfunction for the augmented LRR-ARG model with $L = 3$ and $(\hat{\beta}, \hat{\gamma}) = (0.992, 11.25)$, and the profiled GMM-criterion as a function of the risk aversion parameter γ for various orders L of the measurement and pricing equations.



(a) Scaled continuation value.

(b) Profiled GMM criterion.

Figure 10: Scaled continuation value $\hat{\chi}_{L^*}$ estimated as $L^* = 6$ order approximation to the fixed point defined in (29) based on the augmented LRR-ARG model with $L = 3$ order measurement and pricing equations, and the profiled GMM-criterion as a function of the preference parameters (β, γ), which is minimized at $(\hat{\beta}, \hat{\gamma}) = (0.994, 9.0)$.