EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Davide Delle Monache, Ivan Petrella,
Fabrizio Venditti

Price dividend ratio and long-run stock returns: a score driven state space model

No 2369 / February 2020

Abstract

In this paper we develop a general framework to analyze state space models with time-varying system matrices where time variation is driven by the score of the conditional likelihood. We derive a new filter that allows for the simultaneous estimation of the state vector and of the time-varying parameters. We use this method to study the time-varying relationship between the price dividend ratio, expected stock returns and expected dividend growth in the US since 1880. We find a significant increase in the long-run equilibrium value of the price dividend ratio over time, associated with a fall in the long-run expected rate of return on stocks. The latter can be attributed mainly to a decrease in the natural rate of interest, as the long-run risk premium has only slightly fallen.

JEL codes: C22, C32, C51, C53, E31.

Keywords: State space models, time-varying parameters, score-driven models, equity premium, present-value models.

Non-technical summary

A decade after the Great Recession the global economy is mired in an environment of low real interest rates, low growth and high stock valuations. Whether this is a "new normal" or we can expect to move to a different steady state is a question that is at the centre of a new research agenda on macro-financial trends in a changing environment. The issue of structural breaks is back in the spotlight, as it is the use of time series models that allow for parameter instability. In this paper we contribute to this debate by developing a general method to analyse state space models where parameters change over time and by applying this method to study the evolving relationship between stock valuations, stock returns and dividend growth since 1880 in the US.

In the methodological part of the paper we introduce a general framework to analyse so called score driven state space models. In these models, time variation in the parameters is driven by the score of the conditional likelihood. The intuition is simple. At each point in time, the score of the likelihood, that is the derivative of the likelihood with respect to the parameters, provides information on how well the model is fitting the data. When the score is close to zero, the likelihood is close to its maximum and parameters can remain relatively stable. Yet, when the score is far from zero, a change in the parameters can help better fitting the data. Things become more complicated when the model also features unobserved components, which need to be filtered out from the data together with the parameters. We propose an algorithm that, together with the Kalman filter, solves this problem. Notably, any time series model that admits a state space representation (unobserved component models, ARMA models, Vector Autoregressions) fits into our general framework.

Our empirical investigation uses this methodology to shed light on the secular rise in the price dividend (PD) ratio of stocks in the US since 1880. We start from the well-known Campbell and Shiller decomposition (Campbell and Shiller, 1995), according to which the PD ratio can be split in two components: the expected discounted sum of dividend growth minus the expected discount sum of the return on stocks. When analysed over the long time period that we consider, an apparent contradiction emerges: stock valuations display a secular upward trend, steeper in the 1960s and in the 1990s, while both dividend growth and the return on stocks appear to be stationary. This is not just eyeball econometrics. Lettau and Van Nieuwerburgh (2006), for instance, document two structural breaks in the mean of the PD ratio but fail to provide an explanation of whether these are matched by significant changes in the long-run value of expected returns or of expected dividend growth. A plausible explanation is that these changes in the low frequency component of stock returns and dividend growth are overshadowed by the presence of very volatile transitory components. It is well known that, in such an

environment, structural break tests have low power (Cogley and Sargent, 2005 and Benati, 2007). We argue that the changing relationship between the PD ratio, returns and dividend growth could be better captured by a flexible unobserved component model that allows for gradual shifts in their long-run mean as well as in their volatility. We use our new method to estimate a generalized version of the Cambpell and Shiller decomposition, which allows for time-varying steady states as well as for time-varying volatilities.

Our findings provide a set of novel results in the literature. First, we show that expected returns and, to a lesser extent, expected dividend growth, exhibit slow moving steady states. Expected returns, in particular, have experienced a continuous decline in their long-run equilibrium value. This decline accelerated in the 1960s and in the 1990s, prior to the stock market crashes of the early 1970s and 2000s. This fall in the permanent component of expected returns is reflected in an upward trend of the price dividend ratio, which so far the literature failed to explain. According to our results, prices in the 60s and in the 90s were high not because of bright economic prospects (high expected dividend growth) but because of low discount rates (low expected returns). Second, at high frequencies, expected returns and expected cash flows contribute equally to the price dividend ratio. At business cycle frequencies, however, expected returns explain the bulk of the variance of the price dividend ratio in normal times, while sharp drops in expected cash flows play an important role in recessions. The stock market crashes of 1929 and 2008 were exceptional, in that they were determined by both a marked increase in discount rates as well as by a collapse of expected cash flows. Third, the conditional correlation between expected and actual returns is robustly negative over the whole sample and falls from -0.60 to -0.75, indicating increasing predictability in the last two decades. The correlation generally falls in recessions: bear markets predict higher future returns. Fourth, throughout the sample we find that the conditional variance of stock returns falls with the investment horizon. This confirms the notion that the stock market poses less risk for long-run than for the short-run investors (Siegel, 2008, Campbell and Viceira, 2005, Carvalho et al. 2018). Finally, we decompose long-run expected returns into a riskless component and a risk premium. The former is closely related to the real natural rate of interest (r-star), a concept that dates back to Knut Wicksell but that has been more recently popularized by Laubach and Williams (2003). Our findings indicate that the fall of r-star, particularly marked after the 1960s as documented also by Holston, Laubach and Williams (2017) and Del Negro et al. (2017) accounts for most of the decrease in expected stock returns. The risk premium, on the other hand, has only slightly fallen, from around 4 percent at the beginning of the 20th century, to reach a minimum of 3 percent in 2000 and to rebound thereafter to around 3.5 percent. In sum, our model reveals that the fall in r-star explains most of the fall in the long-run expected return on equity and that the latter accounts for the secular increase of the PD ratio in the US.

1 Introduction

A decade after the Great Recession the global economy is mired in an environment of low real interest rates, low growth and high stock valuations. Whether this is a "new normal" or we can expect to move to a different steady state is a question that is at the center of a new research agenda on macro-financial trends in a changing environment (Caballero et al., 2017). The issue of structural breaks is back in the spotlight, as it is the use of time series models that allow for parameter instability. The flexibility of these econometric tools, however, does not come without costs, either analytical or computational. Complications can be particularly burdensome for unobserved component models, where the latent state variables, together with the time-varying parameters (TVP), need to be inferred on the basis of the observed data. In this paper we contribute to this debate by developing a new method to analyze state space models where parameters change over time and by applying this method to analyze the evolving relationship between stock valuations, stock returns and dividend growth.

The econometric method that we propose posits a law of motion for the parameters that is a linear function of the score of the conditional likelihood, following Creal et al. (2008) and Harvey (2013). At each point in time, the score determines both the size and the sign of the adjustment of the model parameters. In a state space model with constant parameters new information (e.g. a data release) generates a prediction error that is the basis for updating and forecasting the unobserved states through the Kalman Filter (KF) recursions. Time variation in the parameters introduces an additional margin of adjustment to this process, calling for a simultaneous update of both the parameters and the latent states. We derive the analytical expressions for a new set of recursions that, running in parallel with the KF, update at each point in time both the vector of TVP and the latent states. Within this framework, the likelihood of any Gaussian state space model with TVP is available in closed form and the model can be estimated by maximum likelihood (ML). An important feature of our method is that it can easily handle parameter constraints, which might arise from economic theory or

from a deliberate choice of the econometrician. Constraints are taken into account directly in the estimation algorithm through a Jacobian function. Finally, missing observations, mixed frequencies and the shrinkage of the time-varying parameters towards desired values are easily dealt with. A Monte Carlo exercise shows that the method can replicate the salient features of various data generating processes. In particular, our method delivers constant coefficients when the data are simulated from a fixed coefficient model, and tracks time variation in parameters when this is present in the data.

We then use this methodology to revisit the relationship between the price dividend ratio, the return on stocks and dividend growth in present value models. State space models are particularly attractive to study present value relationships because they handle efficiently complex dynamics while avoiding over-parameterization (Binsbergen and Koijen, 2010). We estimate a generalization of the Campbell and Shiller decomposition (Campbell and Shiller, 1988), which allows for time variation in the steady state of expected dividend growth and of the expected return on stocks, as well as in the conditional variance of the shocks. Our findings provide a set of novel results in the literature. First, we show that expected returns and, to a lesser extent, expected dividend growth, exhibit slow moving steady states. Expected returns, in particular, have experienced a continuous decline in their long-run equilibrium value. This decline accelerated in the 1960s and in the 1990s, prior to the stock market crashes of the early 1970s and 2000s. This fall in the permanent component of expected returns is reflected in an upward trend of the price dividend ratio, which so far the literature failed to explain. According to our results, prices in the 60s and in the 90s were high not because of bright economic prospects (high expected dividend growth) but because of low discount rates (low expected returns). Second, at high frequencies expected returns and expected cash flows contribute equally to the price dividend ratio. At business cycle frequencies, however, expected returns explain the bulk of the variance of the price dividend ratio in normal times, while sharp drops in expected cash flows play an important role in recessions. The stock market crashes of 1929 and 2008 were

exceptional, in that they were determined by both a marked increase in discount rates as well as by a collapse of expected cash flows. Third, we decompose long-run expected returns into a riskless component and a risk premium. We find that the former has remained relatively stable until the beginning of the 1960s, to decrease rapidly thereafter. This corroborates the findings in the literature that analyzes the evolution of the natural rate of interest (r-star). The long-run equity risk premium has only slightly fallen, from 4 to 3 percent.

Relationship with the econometric literature. We are not the first ones to tackle the issue of parameter variation in state space models. Early work by Harvey et al. (1992) shows how to enrich unobserved components models with ARCH disturbances, Koopman et al. (2010) introduce GARCH dynamics in a specific factor model, Eickmeier et al. (2015) and Koop and Korobilis (2013, 2014) estimate factor models with changing loadings and volatilities using either Maximum Likelihood or forgetting factors. All these papers focus on specific models for which they develop ad-hoc estimation methods. The algorithm that we develop in this paper is instead general and can be used to analyze any model that can be cast in state space form, including factor models, Vector Autoregressions, and unobserved component models. Notably, all these models can be analyzed within the general framework that we lay out. A second strand of the literature includes papers in which the TVP of the model are driven by additional stochastic processes (also known as "parameter-driven" models). In this setting, Bayesian simulation techniques are needed (Stock and Watson, 2007; Del Negro and Otrok, 2008; Bianchi et al., 2009; Marcellino et al., 2016). The increase of computer power has made these methods relatively more attractive. Yet time-varying parameters in these models are typically modeled as a random walk plus noise processes, providing limited flexibility. Think, for instance, of identification restrictions on factor loadings, steady state restrictions that can be dictated by economic theory or stationarity restrictions that are sometimes imposed on autoregressive models. While some of these restrictions could be handled via rejection sampling, Bayesian methods tend to be cumbersome in the presence of non-linear restrictions in model

parameters. In our method, restrictions on the system matrices of the state space are embedded in the updating steps of the unobserved components and of the model parameters with no additional computational costs.[1]

Relationship with the macro-finance literature. Instability in the Campbell and Shiller decomposition has been studied mainly in the context of stock returns predictability. Time variation either in the conditional means or in the conditional variances of dividend growth and of the price dividend ratio has been documented, respectively, by Lettau and Nieuwerburgh (2008) and Piatti and Trojani (2017). Our first contribution to this literature consists of modeling time variation *both* in the conditional means *and* in the conditional variances. This is important because it is well known that a mis-specification of the conditional mean can lead to spurious evidence of time-varying volatility (Julliard and Bryzgalova, 2019) but also that neglecting changes in volatility leads to overstating time variation in the conditional mean (Sims, 2001). Second, we are the first to provide solid econometric evidence that the rise in valuations in the past century, and in particular after the 1950s, is due to a fall in the long-run mean of expected stock returns. The intuition for this result is already provided by Fama and French (2002) in their seminal paper on the equity risk premium. They use long historical data on dividend growth and dividend yields to obtain an estimate of expected stock returns and of the implied risk premium. Upon noticing that dividend growth is broadly stationary and unpredictable over this long sample but that the price dividend ratio has instead risen considerably, they conclude that the only logical explanation is a decline in expected returns. We provide an econometric method and a model specification that formalizes their intuition and that confirms its empirical validity. Noticeably, the estimate of the risk premium that we obtain (around 3 to 4 percent) is in line with their calculation based on simple unconditional means. We argue that these trends in valuations and discount rates are too small and gradual to be detected by conventional break tests like the ones used by Lettau

[1] The only requirement is that these constraints need to be be expressed through a function that is continuous and twice differentiable. A large class of commonly used restrictions can be indeed cast in this form.

and Nieuwerburgh (2008) but are large enough to be captured by our score driven state space model. Finally, we add to the recent literature on the relative role played by r-star and by the equilibrium risk-premium in explaining long-term trends in the return on stocks (Greenwald et al., 2019; Farhi and Gourio, 2018).

Structure of the paper. Section 2 constitutes the methodological body of the paper, where we present the general framework of score driven state space models. In Section 3 we describe the Monte Carlo exercise. In Section 4 we present the empirical application. Sections 5 and 6 discuss the main results and their implications for the literature on secular changes in the equity premium and the fall in the natural rate of interest. Section 7 concludes the paper.

2 Score driven state space models

Let us assume that a given time series model has the following state space representation:

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, H_t), \\ \alpha_t &= T_t\alpha_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t), \quad t = 1, ..., n, \end{aligned} \tag{1}$$

where y_t is the $N \times 1$ vector of observed variables, ϵ_t is the $N \times 1$ vector of measurement errors, α_t is the $m \times 1$ vector of state variables and η_t is the corresponding $m \times 1$ vector of disturbances. The two disturbances are assumed to be Gaussian distributed and uncorrelated for all time periods, that is $E(\epsilon_t\eta_s') = 0$ for $\forall t, s$.[2] The initial value of the state vector is also assumed to be Gaussian, $\alpha_0 \sim \mathcal{N}(a_0, P_0)$ and uncorrelated $\forall t$ with ε_t and η_t.

Following Harvey (1989, sec. 3.1) it is usually assumed that the *system matrices* Z_t, H_t, T_t and Q_t are non-stochastic. As a result the system (1) is *linear* with respect to the state vector. Conditional on the information set $Y_{t-1} = \{y_{t-1}, ..., y_1\}$ and on the vector of parameters θ, the state vector and the observations are both Gaussian distributed; i.e. $y_t|Y_{t-1}; \theta \sim \mathcal{N}(Z_t a_t, F_t)$

[2]This assumption can be relaxed at the cost of a complication in the filtering formulae (see Appendix C.3).

and $\alpha_t|Y_{t-1};\theta \sim \mathcal{N}(a_t, P_t)$, and the log-likelihood function at time t is:

$$\ell_t = \log p(y_t|Y_{t-1},\theta) \propto -\frac{1}{2}\left(\log|F_t| + v_t' F_t^{-1} v_t\right). \tag{2}$$

The prediction error v_t, its covariance matrix F_t, the conditional mean of the state vector a_t, and its mean square error (MSE) matrix P_t, are recursively estimated by means of the KF:

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ a_{t|t} &= a_t + P_t Z_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' F_t^{-1} Z_t P_t, \\ a_{t+1} &= T_{t+1} a_{t|t} & P_{t+1} &= T_{t+1} P_{t|t} T_{t+1}' + Q_{t+1}, \quad t = 1, ..., n. \end{aligned} \tag{3}$$

Specifically, we have that $a_t = \mathrm{E}(\alpha_t|Y_{t-1},\theta)$ is the so-called *predictive* filter with its MSE matrix being $P_t = \mathrm{E}[(a_t - \alpha_t)(a_t - \alpha_t)'|Y_{t-1},\theta]$, while $a_{t|t} = \mathrm{E}(\alpha_t|Y_t,\theta)$ is the so-called *real time* filter with MSE equal to $P_{t|t} = \mathrm{E}[(a_{t|t} - \alpha_t)(a_{t|t} - \alpha_t)'|Y_{t-1},\theta]$. Given the initial condition on α_0, the predictive filter is initialized as follows $a_1 = T_1 a_0$ and $P_1 = T_1 P_0 T_1' + Q_1$, and often it is the case to express the KF recursions (3) in terms of the predictive filter only; see details in the Appendix A.2. The state space model in (1) is the so-called *contemporaneous* form used in Harvey (1989). Durbin and Koopman (2012) use instead the so-called *forward* form. In this paper we prefer to use the former so that the system matrices have all the same time dependency with respect to the vector of time-varying parameters. Using the forward form, the time dependency of matrices T and Q needs to be adapted; for details see Appendix C.1.

The assumption that the system matrices are non-stochastic implies that the model is linear and the MSEs are independent from the observations, see Harvey (1989) sec. 3.2.3. Here instead, we assume that the changes in the system matrices over time are endogenous and depend on past observations. Thus, although stochastic the system matrices are *predetermined*, meaning that conditional to past observation they can be regarded ad being fixed. As a result, the model is still conditionally Gaussian like the one introduced by Harvey (1989, sec. 3.7.1).[3]

[3]The KF generates the conditional Gaussian distribution of the state vector where the mean is no longer linear in the observations and the MSE (conditional error covariance) depends upon the particular realization of the observations in the sample. The case of conditional Gaussian state space model in which the system matrices are random and adapted to the observation process has been also considered by Chen et al. (1989)

This setup has three attractive features. First, both the state vector and the observations are conditionally Gaussian. Second, the likelihood function ca be written in the form of the prediction error decomposition (2) and computed by means of the KF (3). Third, although the model is not linear in the observations, the KF delivers the minimum mean square estimates of the state vector (see Harvey, 1989, p. 342). The key analytic challenge here is represented by the joint updating of both the system matrices and of the state vector. To solve this problem, we propose a new set of recursions that run in parallel with the KF.[4]

2.1 Score driven system matrices

The time-varying elements of the system matrices in (1) are collected in the vector f_t. As in Creal et al. (2008) and Harvey (2013), we posit the following score driven law of motion for such vector:

$$f_{t+1} = \mathrm{c} + \mathrm{A}f_t + \mathrm{B}s_t, \quad s_t = \mathcal{S}_t \nabla_t, \quad t = 1, ..., n, \tag{4}$$

with

$$\nabla_t = \frac{\partial \ell_t}{\partial f_t}, \quad \mathcal{S}_t = -\mathrm{E}_t \left(\frac{\partial \ell_t^2}{\partial f_t \partial f_t'} \right)^{-1}, \tag{5}$$

where ℓ_t is the conditional log-likelihood function of the model (1), ∇_t is the score (gradient) with respect to f_t and the scaling matrix, $\mathcal{S}_t$, is the inverse of the information matrix $\mathcal{I}_t$. In this case s_t has zero conditional mean and conditional variance equal to inverse of information matrix.[5]

The system matrices may possibly contain both time-varying and constant elements and we collect those static parameters in the vector θ_m. Thus, at each point in time, the system

who state the necessary conditions in order for the KF to generate the conditional mean and covariance of the Gaussian distributed state vector.

[4] In the existing literature this challenge is typically solved by assuming that the vector and the time-varying system matrices can be somehow estimated in two *separate steps*. This is done in papers that use classical methods (Eickmeier et al., 2015; Koop and Korobilis, 2013) and implicitly also in studies that use Bayesian methods, as cycling through the Gibbs sampler implies conditioning on a given estimate of the whole state vector.

[5] One could choose $\mathcal{S}_t = \mathcal{I}_t^{-1/2}$, in which case the conditional variance of the score is equal to the identity matrix. Alternatively, one can set $s_t = \nabla_t$, in which case the score has conditional variance equal to the information matrix. In general, to avoid numerical instability in the scaling matrix we may replace $\mathcal{S}_t$ with its smoothed estimator $\tilde{\mathcal{S}}_t = (1 - \kappa_h)\mathcal{S}_t + \kappa_h \tilde{\mathcal{S}}_{t-1}$.

matrices depend upon f_t and θ_m, namely $Z_t = Z(f_t, \theta_m)$, $T_t = T(f_t, \theta_m)$, $H_t = H(f_t, \theta_m)$, and $Q_t = Q(f_t, \theta_m)$. This dependence can be linear or non-linear. The score vector s_t is computed conditional on the information up to time t and the vector f_t is entirely determined by past observations and by the vector of static parameters $\theta_f = \{\mathrm{c}, \mathrm{A}, \mathrm{B}\}$. Since the dynamic of the system matrices is observation-driven, i.e. entirely determined by past observations and by the vector $\theta = (\theta_f', \theta_m')'$, the model is conditional Gaussian and the log-likelihood (2) can be evaluated recursively through the KF (3).

Let us focus briefly on the updating rule (4). At each point in time, f_t is updated so as to maximize the local fit of the model; i.e. the magnitude of the update depends on the slope and on the curvature of the likelihood function, and the resulting score-driven filter can be rationalized as the stochastic analogue of the Gauss–Newton search direction.[6] The parameters in B determine the sensitivity of the time-varying parameters to the score of the conditional likelihood, and therefore to the information contained in the prediction error. The special case of constant system matrices is obtained by setting this matrix to 0.

Result 1 *Given the model (1)-(2), the score and the information matrix are:*

$$\begin{aligned} \nabla_t &= \tfrac{1}{2}\left[\dot{F}_t'(F_t \otimes F_t)^{-1} vec(v_t v_t' - F_t) - 2\dot{V}_t' F_t^{-1} v_t\right] \\ \mathcal{I}_t &= \tfrac{1}{2}\left[\dot{F}_t'(F_t \otimes F_t)^{-1}\dot{F}_t + 2\dot{V}_t' F_t^{-1}\dot{V}_t\right], \quad t = 1, ..., n, \end{aligned} \tag{6}$$

where $\dot{V}_t = \partial v_t / \partial f_t'$ and $\dot{F}_t = \partial vec(F_t)/\partial f_t'$ measure the sensitivity of the prediction error v_t and its variance F_t with respect to f_t. Proofs in Appendix A.1.

Notice that all the elements of the information matrix $\mathcal{I}_t$ are computed using information up to time $t-1$. On the other hand, the gradient ∇_t contains the current observation y_t via the prediction error v_t. The terms $\dot{V}_t$ and $\dot{F}_t$ play a key role in the gradient ∇_t. They measure the sensitivity of, respectively, the first and second moment of the state vector with respect to f_t. Together with the variance of the prediction error (F_t) and with the curvature of the conditional likelihood (proportional to $\mathcal{I}_t$), they determine the impact that new information,

[6] Blasques et al. (2015) show that updating the parameters using the score of the likelihood function is optimal in the sense that it locally reduces the Kullback-Leibler divergence between the true conditional density and the one implied by the model even when the model is mispecified.

summarized in the prediction error v_t, has on the time-varying parameters. Notice that v_t and F_t are recursively computed by means of the KF (3), while their Jacobian counterparts, $\dot{V}_t$ and $\dot{F}_t$, are recursively computed through the new filter presented below.

Result 2 *The Jacobian counterpart of the KF leads to the following set of expressions:*

$$\begin{array}{rcll} \dot{V}_t & = & -[(a_t' \otimes I_N)\dot{Z}_t + (a_{t-1|t-1}' \otimes Z_t)\dot{T}_t], & t = 1, ..., n, \\ \dot{F}_t & = & 2N_N(Z_t P_t \otimes I_N)\dot{Z}_t + 2(Z_t \otimes Z_t)N_m(T_t P_{t-1|t-1} \otimes I_m)\dot{T}_t + \dot{H}_t + (Z_t \otimes Z_t)\dot{Q}_t, & \end{array} \tag{7}$$

where $\dot{Z}_t = \partial vec(Z_t)/\partial f_t'$, $\dot{H}_t = \partial vec(H_t)/\partial f_t'$, $\dot{T}_t = \partial vec(T_t)/\partial f_t'$ *and* $\dot{Q}_t = \partial vec(Q_t)/\partial f_t'$ *are the Jacobians of the system matrices with respect to* f_t*, and* N_m *is a symmetrizer matrix (i.e., for any* $n \times n$ *matrix,* S*,* $N_n vec(S) = vec[\frac{1}{2}(S + S')]$*). Proofs in Appendix A.2.*

Putting together Results 1 and 2, we have all we need to compute the scaled score $s_t = \mathcal{S}_t \nabla_t$ and therefore to recursively estimate the vector f_t using the the score-driven filter (4). Such *auxiliary filter* runs in parallel with the standard KF (3) as exemplified in the Algorithm described below.

Algorithm for the score driven state space model
Initialize $a_{0\|0}$, a_1, $P_{0\|0}$, P_1, f_1.
For $t = 1, ..., n$:
i. evaluate Z_t, T_t, H_t, Q_t, $\dot{Z}_t$, $\dot{T}_t$, $\dot{H}_t$, $\dot{Q}_t$
ii. compute v_t, F_t, $\dot{V}_t$, $\dot{F}_t$
iii. compute $a_{t\|t}$, $P_{t\|t}$, ∇_t $\mathcal{I}_t$, s_t
iv. compute f_{t+1}
v. evaluate Z_{t+1}, T_{t+1}, H_{t+1}, Q_{t+1}
vi. compute a_{t+1}, P_{t+1}

The vector of parameters θ can be estimated by ML, that is $\widehat{\theta} = \arg\max \sum_{t=1}^{n} \ell_t(\theta)$. Given the above algorithm, the evaluation of the log-likelihood function is straightforward and the maximization can be obtained numerically.[7]

Remark: the expressions in (7) require the current values of the predictive filter (a_t and P_t) and past values of the real time filter ($a_{t-1|t-1}$ and $P_{t-1|t-1}$). It is possible to express the new filter with respect to the predictive filter only but the implied set of recursions are more cumbersome; see expressions (A.10)-(A.15) in the appendix A.2.

[7] As in Creal et al. (2013, sec. 2.3) one can conjecture that the usual ML results hold.

Examples. In Appendix B we provide some examples of popular models with time-varying parameters that have been used in the literature and that can be analyzed within our framework. In particular we discuss the local-level model with time-varying volatilities used by Stock and Watson (2007); the autoregressive models with Guassian innovations in Delle Monache and Petrella (2017) and Blasques et al. (2014) and the vector autoregression with time-varying parameters by Koop and Korobilis (2013).

2.2 Non-linearity in the system matrices

In many applications it is desirable to constrain the time-varying parameters at each point in time. For instance, steady state relationships often imply nonlinear relationships between parameters. Alternatively one might want to impose that the parameters stay in a certain region (it is the case, for instance, of positive semidefinite covariances or of stationary roots), or to be exactly equal to given values.

Such constraints must be accounted for when estimating the model. In this context, the score driven setup provides a clear advantage with respect to alternative methods, including Bayesian methods, because it provides a general framework to deal with parameter restrictions. The intuition of the mechanism at play is very simple. The score of the conditional likelihood is the derivative of the likelihood with respect to the vector of parameters f_t. The effect of these parameters on the score is mediated via the Jacobians $(\dot{Z}_t,\ \dot{T}_t,\ \dot{H}_t,\ \dot{Q}_t)$. When the system matrices $(Z_t,\ T_t,\ H_t,\ Q_t)$ are a non-linear function of f_t, these Jacobians will take this dependence into account via the chain-rule. In other words, in the score driven framework the presence of constraints does not add any particular complication, besides the derivation of the Jacobian matrices. Notably, the model can still be estimated by ML.

Although the exact expression of the Jacobians $\dot{Z}_t,\ \dot{T}_t,\ \dot{H}_t,\ \dot{Q}_t$ is model specific, we offer a flexible expression for dealing with them. Let M_t be a generic system matrix of dimension

$r \times c$, and decompose this matrix as follows:

$$\text{vec}\,(M_t) = S_0 + S_1\psi(S_2 f_t), \tag{8}$$

where S_0 is a $rc \times 1$ vector containing all the time invariant elements in each of the elements of the system matrix M_t, S_1 and S_2 are selection matrices that select respectively the time-varying elements of M_t and the sub-vector of f_t belonging to M_t. Finally, $\psi(\cdot)$ denotes the mapping function, often called *link function*, between f_t and the corresponding elements in M_t embedding parameter restrictions. Such function is assumed to be time-invariant, continuous, invertible, twice differentiable, and its Jacobian matrix is denoted by Ψ_t. Given the representation (8), the generic Jacobian matrix $\dot{M}_t$ can be computed as follows:

$$\dot{M}_t = \frac{\partial \text{vec}(M_t)}{\partial f_t'} = S_1 \Psi_t S_2. \tag{9}$$

While equations (8)-(9) can be directly used to deal with Z_t and T_t, when modelling a generic covariance, Ω_t, it is often useful to decompose this in terms of volatilities and correlations $\Omega_t = D_t R_t D_t$, where D_t is the diagonal matrix containing the standard deviations and R_t is the correlation matrix.[8] Now R_t and D_t can be expressed in the form of (8), $\dot{R}_t$ and $\dot{D}_t$ can be computed as in (9) and $\dot{\Omega}_t$ is computed using standard rules of matrix differentiation (see Appendix E.1).

2.3 Extensions

Our algorithm can be adapted to accommodate data irregularities (like an arbitrary pattern of missing data or data sampled at different frequencies) and parameters shrinkage. In this section, we sketch briefly how to treat these cases. In Appendix C we offer a detailed analysis and also discuss the case of correlated disturbances.

[8] The case where only the correlations or the volatilities are time-varying can also be dealt with using this decomposition.

2.3.1 Missing observations and data sampled at mixed frequencies

When some data are missing the observed vector is represented by $W_t y_t$, where W_t is an $N_t \times N$ selection matrix with $1 \leq N_t \leq N$, meaning that at least one observation is available at time t. Note that W_t is obtained by eliminating the $i-th$ row from I_N when the $i-th$ variable is missing. In this setting, at each time t the conditional likelihood ℓ_t is computed using N_t observations; i.e. $\ell_t = \log p(W_t y_t | Y_{t-1}, \theta)$. The score is then computed based on this modified likelihood that only takes into account available information, see Section C.2 for details. Mixed frequencies do not pose additional challenges, as they involve missing observations and temporal aggregation.[9]

2.3.2 Shrinking parameters with a L2 penalty

As the dimension of the system grows, it could be desirable to impose some shrinkage on the model parameters to avoid an increase in the estimation variance (Hastie et al., 2001). In a Bayesian framework this is achieved through the prior distribution. In a classical setting, like the one hereby adopted, shrinkage can be achieved by means of stochastic constraints that lead to a mixed estimator (Theil and Goldberger, 1961). In this framework actual data are mixed with artificial observations that are generated by the stochastic constraints.[10] Let us consider for instance the case in which we might want to shrink a linear combination of the parameters $R_t f_t$ in a given direction r_t. We augment the model with the following artificial observations:

$$r_t = R_t f_t + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma_t), \tag{10}$$

where r_t, R_t and Σ_t are known and the random vector u_t is Gaussian and uncorrelated with the other disturbances in the model. The matrix Σ_t determines the tightness of the constraints, i.e.

[9] Low frequency indicators can be modeled as a latent process that is observed at regular low frequency intervals and missing at higher frequency dates. The relation between the observed low frequency variable and the corresponding (latent) higher frequency indicator depends on whether the variable is a flow or a stock and on how the variable is transformed before entering the model. The variable can be rewritten as a weighted average of the unobserved high frequency indicator (Banbura et al., 2013).

[10] See Kapetanios et al. (2019) for an application to large Vector Autoregressions.

the degree of shrinkage. For $\Sigma_t \to \infty$ the constraints vanishes, while for $\Sigma_t \to 0$ the constraints holds exactly.

Recall that the state space model (1) with the score driven system matrices described by (4) is conditionally Gaussian with conditional likelihood equal to (2). Since u_t is Gaussian, the log-likelihood function for the vector $\tilde{y}_t = (y_t', r_t')'$ is equal to:

$$\ell_t^p = \log p(\tilde{y}_t|Y_{t-1}, \theta) \propto -\frac{1}{2}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) - \frac{1}{2}\left(\log|\Sigma_t| + u_t'\Sigma_t^{-1}u_t\right). \tag{11}$$

This expression can be interpreted as a penalized log-likelihood with a quadratic penalty function. The resulting score mixes the unrestricted score s_t in (5), which is based on actual data, with that computed on the basis of the stochastic constraint in (10):

$$s_t^p = (\mathcal{I}_t + R_t'\Sigma_t^{-1}R_t)^{-1}(\nabla_t + R_t'\Sigma_t^{-1}u_t) = (I - \Upsilon_t R_t)s_t + \Upsilon_t u_t, \tag{12}$$

where $\Upsilon_t = \mathcal{I}_t^{-1}R_t'(R_t\mathcal{I}_t^{-1}R_t' + \Sigma_t)^{-1}$. It is easy to see that when $\Sigma_t \to \infty$ the penalty vanishes and the score equals s_t. For more details on the filter and examples see Appendix C.4.

3 A Monte Carlo analysis

Before testing our method on actual data we assess its ability to replicate the features of a number of data generating processes (DGPs) through a Monte Carlo exercise. The details on the exercise, including model specifications, tables and graphs are contained in Appendix D. In what follows we briefly summarize the main results.

We experiment with simple DGPs that feature time variation either in the measurement or in the transition equation. Moreover, we assess time variation in the coefficients separately from time variation in the volatilities. The law of motion of the parameters can take six forms. We start with a baseline case in which we keep the parameters constant over time. We then move to four cases where the parameters change according to a deterministic process. In case 2 they follow a cyclical pattern determined by a sine function. In cases 3 and 4 they break

(either once or twice) at discrete points in time. In case 5 they increase for some time before returning abruptly to their starting levels. Finally, case 6 is the only one in which we let the parameters vary stochastically, following a persistent AR(1) model.[11]

We base our assessment on five different statistics, namely the Root Mean Squared Error (RMSEs), the Mean Absolute Error (MAE), the correlation between actual and estimated coefficients, the coverage (i.e. percentage of times that the estimated latent states fall in a given estimated confidence interval) and the number of cases in which a pile-up occurs.[12] There are four takeaways from the Monte Carlo exercise. First, for all the DGPs when the true parameters do not change over time the model correctly estimates them as being constant. As a result, RMSEs and MAEs are virtually nil, the coverage extremely precise and a pile-up occurs in about 75 percent of the cases for the models with time-varying volatility and more than half of the cases for the models with time-varying loadings and AR coefficients.[13] Our estimation method passes an essential test, i.e. it does not generate spurious time variation in the coefficients when this is not present in the DGP. Second, when the parameters actually change over time the pile-up problem, which plagues maximum likelihood estimators of models with time-varying coefficients (Stock and Watson, 1998), is not of primary concern with our method. The number of instances in which our method incorrectly concludes that there is no time variation is basically zero in most cases. Third, the model finds it more challenging to estimate parameters that are subject to sudden breaks. This is not surprising, since our model is, by construction, designed to detect smooth changes. Fourth, our new filter is rather conservative in the estimation of the time-varying variances, especially when these are driven by a near unit root process. However, when time variation *is* detected, the algorithm yields

[11] We consider two cases, one with a near unit root process (i.e. with an AR root of 0.99) and a low variance, one with lower persistence (AR root of 0.97) but substantially higher variance. We obtain very similar results in these two specifications.

[12] The last statistics consists of the number of simulations in which the static coefficients that pre-multiply the score end up being lower than 10^{-6}, which we take as sufficient evidence that the estimated parameters are effectively zero, i.e. that the model does not detect any time variation.

[13] For the latter two cases, in an additional 20% of the simulations the estimated parameters are virtually constant, despite not being classified as a pile-up according to our criterion.

relatively low RMSEs and MAEs and a satisfactory coverage. We take these results as evidence that, in the case of time-varying variances, the algorithm needs relatively more evidence of breaks in the parameters to move away from zero. A larger sample size basically eliminates the problem.

4 Price dividend ratio, expected returns and expected dividend growth

In this section, we revisit the relationship between the price dividend ratio, the return on stocks and dividend growth. This topic provides the ideal setting for score driven state space models. First, it involves present value relationships that can be conveniently analyzed via state space models (Binsbergen and Koijen, 2010). Second, there is significant evidence of instability in regressions of stock returns on the price dividend ratio (Paye and Timmermann, 2006; Lettau and Nieuwerburgh, 2008), suggesting that a state space model with time-varying parameters is the appropriate modelling choice. Third, the parameters that link these three objects are subject to a set of non-linear restrictions that pose a non-trivial challenge for other methods but can be easily dealt with in our framework. We start by recalling the steady state relationship between the return on stocks, the price dividend ratio and dividend growth. We then use break tests to document the presence of structural breaks in the mean of the price dividend ratio. We then move to specifying a score driven state space model that reconciles this motivating evidence with slow movements in the steady states of expected returns and of dividend growth.

4.1 Stock return and time-varying steady states

Let P_t and D_t denote stock prices and dividends. From the simple definition of gross return of an asset it follows that:

$$R_{t+1} \equiv \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{D_{t+1}}{D_t} \frac{P_{t+1}/D_{t+1} + 1}{P_t/D_t}. \tag{13}$$

Lettau and Nieuwerburgh (2008) show that this implies the following relationship in logs:

$$\overline{\mathtt{pd}}_t = \overline{g}_t - \log(\exp \overline{\mu}_t - \exp \overline{g}_t), \tag{14}$$

where $\overline{\mathtt{pd}}_t$, $\overline{\mu}_t$ and $\overline{g}_t$ denote the steady state level of the price dividend ratio, of the return on stocks and of dividend growth, respectively.[14] Equation (14) has two important implications. First, changes in the steady state of the price dividend ratio reflect either changes in the steady state of the return on stocks or in the steady state of dividend growth or in both. Second, small changes in long-run growth (reflected in the steady state of dividend growth) and/or in the steady state of the return on stocks have large effects on the steady state of the price dividend ratio.

4.2 Preliminary evidence on parameter instability

We start our empirical exercise by revisiting the evidence presented by Lettau and Nieuwerburgh (2008) that the long-run mean of the price dividend ratio is subject to two structural breaks. We extend their analysis by considering not only the price dividend ratio but also the return on stocks and dividend growth and by testing for instability in the variance of these variables. Our analysis is based on annual data between 1873 and 2018. Annual data for the Standard and Poor Composite Stock Price Index and associated dividends are sourced from Robert Shiller's website.[15] To account for changes in purchasing power, we deflate total returns and dividends using data on US CPI, also available from the same source.

We start by testing the null hypothesis of no breaks against the alternative hypotheses of one, two or three breaks with unknown dates using the Bai and Perron (2003) test.[16] The results of this first battery of tests, reported in the top panel of Table 1, convey two clear

[14] Lettau and Nieuwerburgh (2008) assume that, at the steady state, the level of the price dividend ratio is constant (i.e. $\overline{P/D}_{t+1} \approx \overline{P/D}_t$). Moreover, denoting with $\overline{DY}_t$ the steady state dividend yield, equation (14) implies that $\overline{DY}_t = \overline{R}_t - \overline{G}_t$, consistent with the Gordon model (see e.g. Campbell, 2018, p. 130).

[15] See http://www.econ.yale.edu/shiller/. Further details on the construction of the data are available in Shiller (1989)'s Chapter 26.

[16] We do not go beyond three breaks for two reasons. First, we want to capture large secular changes in the mean of these series. Second, as discussed below, evidence for more than three breaks is rejected by sequential break tests.

messages. First, the null hypothesis of no breaks can not be rejected for returns and dividend growth.[17] The second result is that there is strong evidence of structural breaks in the mean of the price dividend ratio. Two of the dates for which the Bai and Perron procedure detects a break (1954 and 1995) are consistent with the findings in Lettau and Nieuwerburgh (2008).[18] Evidence of a third break in 1913 is somewhat weaker. In fact, the null hypothesis of two breaks against the alternative of a third one cannot be rejected on the basis of sequential break tests (Table 1, central panel). These sequential tests also cannot reject the null that the mean of the return on stocks and of dividend growth has remained stable over time.

Last, we employ the Nyblom (1989) test, which confronts the null hypothesis of constant parameters (both mean and variance) with an alternative that the parameters follow a martingale.[19] The test detects significant shifts in the mean and volatility of the price dividend ratio as well as in the volatility of dividend growth. This result casts doubts on the hypothesis of constant variances maintained by Lettau and Nieuwerburgh (2008) and lends support to the model with heteroskedastic disturbances in Piatti and Trojani (2017). However, it also confirms that shifts in the steady state of the price dividend ratio, ignored by Piatti and Trojani (2017), are a robust feature of the data.[20]

This preliminary analysis leaves one question open, that is how to reconcile the evidence of breaks in the price dividend ratio with the apparent stability of returns and dividend growth, given that the former is a function of the latter two. Two related answers come to mind. First, this relationship is *non-linear*, so that even small changes in $\overline{\mu}_t$ and $\overline{g}_t$ can generate large swings

[17] Lack of breaks for dividend growth seems to contradict the results in Timmermann (2001), who finds the mean of dividend growth to be unstable. We attribute such difference to the fact that he allows for up to eight structural breaks and uses monthly data. With a high data frequency and a large number of candidate breaks, the data end up being cut in much narrower sub-samples, and changes in the mean can simply reflect short-term variations of dividend growth associated with standard business cycles.

[18] Remarkably, these two dates are identified by *all* the information criteria of the Bai and Perron test.

[19] This test is therefore designed to detect gradual changes in parameters rather than discrete breaks.

[20] After Lettau and Nieuwerburgh (2008) a number of papers have used the deviation of the price dividend ratio from a trend as a predictor of returns (see e.g., Koijen and Nieuwerburgh, 2011, and Herwartz et al., 2016). To the extent that also the equilibrium expected rate of return changes over time, such predictive regressions are misspecified, i.e. they lack a time-varying intercept. Moreover changes in the variances need to be taken into account.

in $\overline{\mathtt{pd}}_t$.[21] A second related explanation is that changes in the long-run mean of returns and dividend growth are overshadowed by the presence of a very volatile transitory component. In such an environment, when changes in the low frequency component are small and gradual, state of the art break tests have low power against the alternative of no breaks (Benati, 2007; Cogley and Sargent, 2005). The changing relationship between the price dividend ratio, returns and dividend growth could be better captured by a flexible model that allows for gradual shifts in their long-run mean as well as in their volatility. In the next section, we specify such a model.

4.3 A score driven present value model with drifting steady states

In this section, we describe the extension of the the Campbell-Shiller approximation introduced by Lettau and Nieuwerburgh (2008), which allows for time-varying steady states in expected returns and in expected dividend growth. Taking a first order approximation of (13) around a time-varying steady state yields:

$$\mathtt{pd}_t - \overline{\mathtt{pd}}_t \simeq \rho_{t+1}(\mathtt{pd}_{t+1} - \mathtt{pd}_{t+1}) + (\Delta d_{t+1} - \overline{g}_t) - (r_{t+1} - \overline{\mu}_t) + \Delta\overline{\mathtt{pd}}_{t+1} + \Delta\overline{\mu}_{t+1} - \Delta\overline{g}_{t+1}, \quad (15)$$

where $\rho_t = \exp\overline{\mathtt{pd}}_t/(1+\exp\overline{\mathtt{pd}}_t)$ and, through $\overline{\mathtt{pd}}_t$, depends on $\overline{\mu}_t$ and $\overline{g}_t$ according to (14). Assuming that $\overline{\mathtt{pd}}_t$, $\overline{\mu}_t$, $\overline{g}_t$ and thus ρ_t are martingales, one can do forward substitution and take expectations of (15) to obtain[22]

$$\mathtt{pd}_t - \overline{\mathtt{pd}}_t \simeq \sum_{j=1}^{\infty} \rho_t^j(\Delta d_{t+j} - \overline{g}_t) - \sum_{j=1}^{\infty} \rho_t^j(r_{t+j} - \overline{\mu}_t). \quad (16)$$

[21] This point is underscored by John Cochrane in "Stock Gyrations", February, 7th 2018, available at https://johnhcochrane.blogspot.com/2018/02/stock-gyrations.html, accessed on the 19th of September 2019.

[22] The martingale assumption on the steady state log returns and dividend growth (i.e. $\mathrm{E}_t(\overline{\mu}_{t+j}) = \overline{\mu}_t$ and $\mathrm{E}_t(\overline{g}_{t+j}) = \overline{g}_t$) is consistent with the specification of the model that we describe below. The additional assumption that the steady state log P/D ratio is a martingale (i.e. $\mathrm{E}_t(\overline{\mathtt{pd}}_{t+j}) = \overline{\mathtt{pd}}_t$ and $\mathrm{E}_t(\rho_{t+j}) = \rho_t$) is, in general, inconsistent with the martingale assumption on $\overline{\mu}_t$ and $\overline{g}_t$, since $\overline{\mathtt{pd}}_t$ and ρ_t are a non-linear function of the previous two. However, Lettau and Nieuwerburgh (2008) show that the martingale assumptions are satisfied to a very good approximation for reasonable break processes. Moreover, in line with Lettau and Nieuwerburgh (2008) we also assume that that deviations from the mean price-dividend ratio are uncorrelated with ρ_t, i.e. $\mathrm{E}_t[\rho_{t+j}(\mathtt{pd}_{t+j} - \overline{\mathtt{pd}}_t)] = 0$.

Equation (16) shows that stock prices fall relative to dividends when there is bad news about future cash flows, or when discount rates (i.e. expected returns) rise. It also implies that the price dividend ratio should forecast returns and/or dividend growth, especially at long horizons (Cochrane, 2008b). Most importantly, the presence of shifts in long-run expected returns and expected dividend growth implies that the sensitivity of $\mathtt{pd}_t - \overline{\mathtt{pd}}_t$ to news about cash flows and discount rates also changes over time (i.e. the higher the equilibrium level of the price dividend ratio, the higher ρ_t).

To take the model to the data one needs to make some additional assumption on expected returns and expected dividend growth. Binsbergen and Koijen (2010) assume that these expected values can be modeled as AR(1) processes. Here we depart from their specification and assume that expected returns and expected dividend growth are the sum of a transitory and a persistent component. The latter, which constitutes a shift in long-run expected returns and expected dividend growth, is the key novelty of our model and we discuss it more in detail in the next sub-section. Specifically, we assume that:

$$\mathrm{E}_t(\Delta d_{t+1}) = \overline{g}_{t+1|t} + \tilde{g}_{t+1|t} \tag{17}$$

$$\mathrm{E}_t(r_{t+1}) = \overline{\mu}_{t+1|t} + \tilde{\mu}_{t+1|t}, \tag{18}$$

where the equilibrium levels of expected returns and expected dividend growth are defined as $\lim_{h\to\infty} \mathrm{E}_t(r_{t+h}) = \overline{\mu}_{t+1|t}$ and $\lim_{h\to\infty} \mathrm{E}_t(\Delta d_{t+h}) = \overline{g}_{t+1|t}$. Using the notation in Binsbergen and Koijen (2010) we denote the transitory component of the expectations ($\tilde{g}_{t+1|t}$ and $\tilde{\mu}_{t+1|t}$) with $\tilde{g}_t$ and $\tilde{\mu}_t$ and assume that they can be characterized by simple AR(1) models:

$$\tilde{g}_{t+1} = \phi_g \tilde{g}_t + \varepsilon_{g,t+1} \tag{19}$$

$$\tilde{\mu}_{t+1} = \phi_\mu \tilde{\mu}_t + \varepsilon_{\mu,t+1}. \tag{20}$$

The present value relationship (16) implies that the transitory component of the price dividend

ratio is related to $\tilde{\mu}_t$ and $\tilde{g}_t$:

$$\mathtt{pd}_t - \overline{\mathtt{pd}}_{t|t-1} \simeq -b_{1,t|t-1}\tilde{\mu}_t + b_{2,t|t-1}\tilde{g}_t \tag{21}$$

with the following constraints on the parameters:

$$b_{1,t|t-1} = \frac{1}{1-\rho_{t|t-1}\phi_\mu}, \qquad b_{2,t|t-1} = \frac{1}{1-\rho_{t|t-1}\phi_g}. \tag{22}$$

Hence the loadings of the transitory component of the price dividend ratio are not only time-varying, but also a non-linear function of (i) the persistence of the transitory components of expected returns and of expected dividend growth and (ii) of the steady state level of the price dividend ratio. These restrictions must be imposed exactly when estimating the model, a challenge that our score driven modeling approach can easily overcome.

The decomposition of dividend growth into the expected dividend growth plus an unexpected shock, $\varepsilon_{d,t+1}$, provides the first measurement equation:

$$\Delta d_{t+1} - \mathtt{E}_t(\Delta d_{t+1}) = \varepsilon_{d,t+1}. \tag{23}$$

The second measurement equation is (21), which we augment with a measurement error $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$ to take into account the approximation error associated with the solution of the present value model. In the model there are three sources of stochastic variation, namely $\varepsilon_{d,t}$, $\varepsilon_{g,t}$ and $\varepsilon_{\mu,t}$, respectively the shock to dividend growth, the shock to expected dividend growth and the shock to expected returns. It is easy to see that they also map into unexpected changes in returns. In fact, combining (15) with (21) the unexpected component of returns is obtained:

$$r_{t+1} - \mathrm{E}_t(r_{t+1}) = -\rho_{t+1|t}b_{1,t+1|t}\varepsilon_{\mu,t+1} + \rho_{t+1|t}b_{2,t+1|t}\varepsilon_{g,t+1} + \varepsilon_{d,t+1}. \tag{24}$$

4.3.1 Slow-moving trends and time-varying risk

We assume that the long-run expected dividend growth and long-run expected returns are martingales. They are driven by the score of the conditional likelihood:

$$\overline{\mu}_{t+1|t} = \overline{\mu}_{t|t-1} + b_{\mu} s_{\mu,t} \tag{25}$$

$$\overline{g}_{t+1|t} = \overline{g}_{t|t-1} + b_{g} s_{g,t}, \tag{26}$$

where $s_{\mu,t}$ $s_{g,t}$ are the appropriate elements of the score vector. The steady states, $\overline{\mu}_{t|t-1}$ and $\overline{g}_{t|t-1}$, are therefore updated (through the scaled score) using information on dividend growth and on the price dividend ratio, as highlighted by equation (14), as well as the present value restrictions embodied in equation (21) which imply that long-run returns and dividend growth can also change, through $\rho_{t|t-1}$, the sensitivity of the transitory components of price dividend ratio, $\mathtt{pd}_t - \overline{\mathtt{pd}}_{t|t-1}$, to expected returns and dividend growth.

Last, we assume that the innovations of the model are normally distributed with time-varying covariances that are themselves driven by the score. Formally, $(\varepsilon_{d,t}, \varepsilon_{g,t}, \varepsilon_{\mu,t})' \sim \mathcal{N}\left(0, \Omega_{t|t-1}\right)$. We decompose $\Omega_{t|t-1} = D_{t|t-1} R_{t|t-1} D_{t|t-1}$ where $R_{t|t-1}$ denotes the time varying correlations among the disturbances in the model and $D_{t|t-1} = \text{diag}\left[\sigma_{d,t|t-1}, \sigma_{g,t|t-1}, \sigma_{\mu,t|t-1}\right]$. Whereas the three shocks $(\varepsilon_{d,t}, \varepsilon_{\mu,t}, \varepsilon_{g,t})$ could be all correlated, as highlighted by Cochrane (2008a) not all elements of the covariance matrix can be separately identified. In Appendix E.2 we show that, in order to identify the model, one restriction needs to be imposed on the correlations among the innovations of the model. In the application, we follow Binsbergen and Koijen (2010) and Rytchkov (2012) and assume that the measurement error in dividend growth $(\varepsilon_{d,t})$ and the stochastic disturbance in expected dividend growth $(\varepsilon_{g,t})$ are uncorrelated (i.e. $\text{Corr}(\varepsilon_{g,t}, \varepsilon_{d,t}) = 0$). In practice, we take advantage of the one-to-one mapping between the correlations and the partial correlations (see, e.g., Joe, 2006).[23] This allows us to model each of the partial correlations separately (with the only constraint that each of them has to

[23] See Appendix C.3 for additional details.

lay in the unit circle) and yet guarantees a well-defined correlation matrix in every period. Modeling the partial correlation has the additional advantage that we can easily impose that one of the correlation coefficients is always zero. In fact, it is easy to show that, ordering the innovations so that the restriction is placed on the first column of the correlation matrix, one has that $\varrho_{dg,t} = \pi_{dg,t}$ (where ϱ_{ij} and π_{ij} are, respectively, the generic ij element of the correlation and partial correlation matrix) and the identification restriction simply requires that $\pi_{dg,t} = 0, \forall t$. Therefore, through the Jacobian, the algorithm translates the score of the likelihood with respect to the correlation matrix into the appropriate updating of the unrestricted partial correlations.

4.3.2 State Space Representation and estimation

Let $y_t = (\Delta d_t, \mathtt{pd}_t)'$ be the vector of observed variables and $\alpha_t = (1, \tilde{g}_t, \tilde{\mu}_t, \tilde{g}_{t-1}, \varepsilon_{d,t}, \varepsilon_{g,t}, \varepsilon_{\mu,t})'$ be the state vector. The measurement equations of the model are (21)-(23). The law of motion of the transitory components (19)-(20) as well as the definition of the innovations of the model constitute the transition equations.[24] The state space representation of the model is:

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, H), \\ \alpha_t &= T\alpha_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t), \end{aligned}$$

where $\epsilon_t = (0, \nu_t)'$, $\eta_t = S\varepsilon_t$, S is a selection matrix, and $Q_t = S\Omega_{t|t-1}S'$. All the details on the state space representation can be found in Appendix E.1. The time-varying elements of the system matrices Z_t and Q_t are collected in the vector:

$$f_{t+1} = (\overline{\mu}_{t+1|t}, \overline{g}_{t+1|t}, \log\sigma_{d,t+1|t}, \log\sigma_{g,t+1|t}, \log\sigma_{\mu,t+1|t}, \text{atanh}\pi^{d\mu}_{t+1|t}, \text{atanh}\pi^{g\mu}_{t+1|t})'$$

where atanh$(\cdot)$ denotes the inverse hyperbolic tangent so that the partial correlations $\pi_{ij,t} \in (-1, 1)$, $\forall t$. The law of motion of f_{t+1} is a restricted version of (4). Specifically, we assume A

[24] Notice that the measurement equation for the price dividend ratio and for dividend growth imply a measurement equation for returns (24), which is therefore redundant for the estimation of the model (see also, Binsbergen and Koijen, 2010).

and B to be diagonal and we further restrict some of the elements of c and A in accordance with the random walk specification for the slow moving trends discussed in section 4.3.1. The model likelihood is computed using the Kalman filter, coupled with the updating algorithm for the score driven parameters as discussed in section 2 and parameters are easily estimated by maximising this likelihood. Confidence intervals can be obtained following Blasques et al. (2016). Appendix E.1 details the non-linear mapping between the score driven time-varying parameters and the system matrices and the associated Jacobians that are required in the score driven filter.

5 Results

Parameter estimates. The estimation results are shown in Table 2, which is organized in three columns. The one on the left presents the results for the autoregressive roots of the transitory components of expected returns and dividend growth (ϕ_μ and ϕ_g), the average volatility and correlations of the shocks in the model over the entire sample as well as the variance of the measurement error in the price dividend equation. Expected returns have a root of 0.829, which implies a half life of 4.7 years. Expected dividend growth is less persistent, with an autoregressive root of 0.345, implying a half life of 1.7. These numbers are, as expected, lower than those estimated by Binsbergen and Koijen (2010) and Piatti and Trojani (2017), as part of the persistence is captured in our case by the shifts in the steady states. The volatility and correlation for the innovations of the model map into an average volatility of the shocks to returns of roughly 0.18 and correlation between the innovations to the return on stocks and expected returns in a range between -0.6 and -0.7, in line with the estimates in Carvalho et al. (2018). The variance of the measurement error of the price dividend ratio is negligible. The second column reports the autoregressive roots of the five time-varying parameters that are modelled as stationary processes, namely the three volatilities and the partial correlations. All of them are very persistent, but far from having a unit root, justifying our modeling choice.

The third column presents the seven loadings on the score and the smoothing coefficient of the Hessian term. The most interesting result regards the loadings of the time-varying steady states. The long-run mean of expected returns has a loading of 0.151 on the likelihood score, three times as large as that of dividend growth (0.052). This is a signal that the low frequency component of expected returns exhibits much more time variation than that of dividend growth. Volatilities and partial correlations, instead, have similar properties. The main takeaway of these estimation results is that expected returns are more predictable than expected dividend growth. They are characterized by a slowly changing steady state, as well as by a persistent transitory component.

Filtered steady states. Figure 1 shows returns and dividend growth together with the expected components. For both series, movements in the long-run steady states are completely dominated by very volatile transitory shocks, and seem relatively flat. As we shall see below, this is in fact not the case. Consistently with the parameter estimates in Table 2, the transitory component of expected dividend growth ($\tilde{g}_t$) is more volatile and falls considerably in recessions. It is worth stressing that its fall in the 2008 recession is not exceptional by historical standards and is in fact significantly milder than those observed in 1929 and during WWII.

Figure 2 shows the estimated trend components. These are the central results of our empirical analysis. In the left panel we plot $\overline{\mu}_{t|t-1}$ and $\overline{g}_{t|t-1}$, which are estimated in our model as score driven martingales. It is clear that what seemed to be a flat line, i.e. the permanent component of expected returns, is in fact a downward sloping trend. It starts off at around 9 percent at the end of the 19^{th} century and fluctuates between 7 and 8 percent until the 50s. Thereafter it experiences two sharp falls: one between the 50s and the 70s, and one in the 90s, the former being interrupted by the stock market crash in 1973 and the latter by the burst of the dotcom bubble in 2000. At the end of the sample, the long-run expected return on stocks stands at around 4 percent, less than half of its initial value. The steady state of expected dividend growth has also fallen slightly over the sample, from around 2 percent to 1.3 percent

at the turn of the century, to then rebound to 1.5 percent. This minor fall in long-run dividend growth is quantitatively consistent with the rise of alternative forms of payout (such as shares repurchases and issuance) which have become quantitatively more important since the 90s as documented by Boudoukh et al. (2007). Overall, however, it has remained more stable, as it could be expected, given the lower loading on the score displayed in Table 2. The right hand side panel shows the implications of these results for the steady state of the price dividend ratio. Our estimate of $\overline{\mathtt{pd}}_t$ has an upward trend over the whole sample, with three large changes, one early in the sample, one after WWII and one in the Nineties. These are indeed the dates for which the Bai and Perron test detects significant structural breaks. However, structural break tests are unable to offer a structural explanation for such an upward trend in valuations, nor can they provide a narrative behind these episodes. Our method reveals that these long waves of strong valuations have largely built on falling expected returns. The rally of stock prices in the 90s, in particular, was not associated with improving cash flows prospects, but with a sharp fall in discount rates. These results formalize and provide evidence for the intuition put forward by Fama and French. In their seminal paper on the equity premium (Fama and French, 2002) they notice that, between 1950 and 2000, realized average stock returns in excess of the riskless interest rate have been substantially higher than the equity risk premium. Given the relative stationarity of dividend growth, they argue that realized capital gains must have come from a fall in discount rates. Our empirical analysis quantifies such shifts in long-run discount rates and identifies their timing within a formal model.

Dynamic price dividend ratio decompositions. The left panel of Figure 3 decomposes the cyclical component of the price dividend ratio in a linear combination of the transitory components of expected returns and of expected dividend growth, based on equation (21). The result is quite striking. First, most of the cyclical variation in valuations is due to changes in expected returns. In particular, high stock prices in the 60s and in the 90s were due to a slow fall (of transitory nature) in the discount rate, which eventually reverted back towards its

mean in the bear markets of the early 70s and early 2000s. This finding is consistent with the relatively higher persistence of $\tilde{\mu}_t$ compared to $\tilde{g}_t$ (AR root of 0.829 versus 0.345) as well as with the fact that changes in expected dividends affect both prices as well as actual dividends, with relative little impact on valuations (Cochrane, 2011; Campbell and Ammer, 1993). Second, the role of cash flows is episodic but non-negligible. A fall in cash flows expectations, for instance, contributed as much as discount rates to the stock market crash in 1929 and in WWII. It had an even larger role in explaining the 2008 crash.

The right hand side panel decomposes the (one step ahead) conditional variance of the the price dividend ratio.[25] At short horizons, dividend and returns shocks contribute in almost equal parts to shifts in valuations. The reason for the stark difference between these two decompositions lies in the frequency on which they focus. At high frequencies, shocks to cash flows have the power to affect valuations (which explains their role in the right hand side graph), but their effect dies out quickly. In the medium run valuations are dominated by discount rate shocks (which is the message of the left hand side chart).

Term Structure of expected returns and dividend growth. Iterating forward equations (19)-(20) we can obtain an expression for expected returns and expected dividend growth at a given horizon n:

$$\mu_t^{(n)} = \frac{1}{n}\mathbb{E}_t\left[\sum_{j=1}^{n} r_{t+j}\right] = \overline{\mu}_{t+1|t} + \frac{1}{n}\frac{1-\phi_\mu^n}{1-\phi_\mu}\tilde{\mu}_t \tag{27}$$

$$g_t^{(n)} = \frac{1}{n}\mathbb{E}_t\left[\sum_{j=1}^{n} \Delta d_{t+j}\right] = \overline{g}_{t+1|t} + \frac{1}{n}\frac{1-\phi_g^n}{1-\phi_g}\tilde{g}_t. \tag{28}$$

In Figure 4, left hand side panel, we plot expected returns at the ten years horizon (blue dashed line) as well as the slope of the term structure of expected returns (red solid line), that is the difference between expected returns ten and two years out. At low frequencies, expected returns inherit the properties of the shifting steady state $\bar{\mu}_{t|t-1}$: they fall from an average of around 7 percent in the first part of the sample to around 2 percent at the end of the sample. The slope,

[25] Details on this decomposition are reported in appendix E.4.

despite some short run movements, fluctuates around zero, suggesting that over this period the whole term structure of discount rates has shifted downwards. The level and the slope are clearly negatively correlated. This negative correlation is driven by a number of a recessionary episodes in which long term discount rates rise but the slope dips, i.e. discount rates rise more sharply at short than at long maturities. It is also directly connected to the predictability of returns: high expected returns today predict low expected returns tomorrow.[26] The right hand side panel shows the level and the slope of the term structure of expected dividend growth. Long-term expectations are rather smooth (much more than those of expected returns) and have remained relatively stable at 2 percent,[27] consistently with the relative stability of $\overline{g}_{t|t-1}$. The slope, on the other hand, is countercyclical and quite volatile fluctuating between minus 10 to 10 percent, a range that is almost five times as wide as that in which the slope of expected returns moves. Indeed, as shown in table 2, the mean volatility of shocks to $\tilde{g}_t$ is almost four times as high as that of $\tilde{\mu}_t$ (0.083 versus 0.024) and this is reflected in the volatility of expected annual cash flows. This slope has some notable peaks, corresponding to sharp recessions, such as the post WWI recession, the Great Depression, the 1973 recession and more recently the Great Financial Crisis. Therefore, the term structure of dividend growth highlights that short-term cash flows expectations fall substantially during recessions, but are anticipated to mean revert relatively quickly. The countercyclicality of the slope of the term structure of dividend growth expectations is in line with the finding of van Binsbergen et al. (2013) who document similar properties from dividend derivatives for the post 2003 sample. In Appendix E we look more in detail at some specific historical stock market corrections.[28] We find that discount

[26] These findings stand partially in contrast with those in Piatti and Trojani (2017) who keep the steady state of expected returns constant. Keeping the steady state of expected returns constant has two consequences. First, it constraints long-term expectations to remain relatively stable, the more the longer the horizon that one considers. This is hard to reconcile with the secular upward trend in the price dividend ratio. Second, in a model with constant steady states the terminal point of the term structure is fixed. This implies that a model with constant steady states ends up overstating the predictability of returns, especially at very long horizons.

[27] Few rare exceptions when pessimistic expectations on cash flows persisted for more than 10 years include the height of the Great Depression, the oil price shock of 1973 and the Great Financial Crisis.

[28] In Figure E.5 for these three historical episodes we plot the whole term structure of expected returns and expected dividend growth for the year before the recession, the peak of the recession and the year after the recession.

rates shocks, especially at the short end of the curve, contributed greatly to the severity of the recessions in 1929 and 2008, while they played a relatively minor role in the 2001 recession episode. These results are consistent with the ones of Campbell et al. (2013).

Time-varying risk. Our model also allows us to analyze how the conditional second moments have changed over time. We refer to Appendix E.5 for the detailed analysis but summarize here two interesting findings. First, the conditional correlation between expected and actual returns is robustly negative over the whole sample. This speaks directly to the notion that actual returns are predictable due to mean reversion: surprisingly high returns today correspond to low future returns. This correlation falls in our sample, from -0.60 to -0.75, indicating increasing predictability in the last two decades, and generally falls in recessions: bear markets predict higher future returns. Second, throughout the sample we find that the conditional variance of stock returns falls with the investment horizon. This confirms the results in Campbell and Viceira (2005) and Carvalho et al. (2018) that the stock market poses less risk for the long-run investor than for the short-run investor (see, e.g., Siegel, 2008).

6 Expected excess returns and the equilibrium real rate

The central result of our empirical analysis is that the long-run expected return on equity has fallen over time. In this section, we elaborate further on this finding by splitting long-run expected returns in a riskless component and a risk premium. The riskless component is closely related to the so called natural rate of interest (r-star), a concept that dates back to Knut Wicksell but that has been more recently popularized by Laubach and Williams (2003). A wealth of research has shown that the natural rate has fallen dramatically over time not only in the US but also in a set of advanced countries (Holston et al., 2017, and Del Negro et al., 2019).[29]

[29]Three main competing explanations have been put forward to rationalize the fall in equilibrium real rate: a permanent decline in the rate of growth of the economy (secular stagnation), an increase in desired savings due to aging population (saving glut) and a rise in economic risk (or a fall in its tolerance) that has raised the liquidity and safety premium of safe assets like US Treasuries, see Del Negro et al. (2017) for an extensive

We follow Cochrane (2008b) and rewrite the present value decomposition of the price dividend ratio in equation (16) in terms of excess returns, by simply subtracting a measure of the safe real rate of interest (let us call it r^f) from both expected discounted returns and expected discounted dividend growth.[30] Our baseline model can then be re-estimated using as observable variables for the measurement equations the price dividend ratio and the difference between dividend growth and r_t^f.[31] Besides this simple modification, the model is essentially unchanged, apart from the fact that, following the suggestion in Campbell and Thompson (2008), we impose the restriction that the long run equity premium needs to be always positive.[32] This alternative model provides us with three time series, $\overline{\mathtt{pd}}_{t|t-1}$, $\overline{g}^{ex}_{t|t-1}$, $\overline{\mu}^{ex}_{t|t-1}$ that measure, respectively, the time-varying equilibrium price to dividend ratio for this alternative model specification, the long-run expected *excess* dividend growth (i.e. the long-run expected dividend growth minus the equilibrium riskless real rate) and the steady state equity risk premium or *excess* return. The first (comforting) result of this exercise is that the estimate of the equilibrium price to dividend ratio that we retrieve using this alternative specification is very close to the one obtained in the baseline model in Section 4.[33] Subtracting the riskless real rate turned out to be indeed an innocuous econometric twist. Yet the novelty is that we now have a measure of the equilibrium excess returns $\overline{\mu}^{ex}_{t|t-1}$. Subtracting this from $\overline{\mu}_{t|t-1}$ we can then obtain a measure of the riskless long-run real rate, $\overline{r}_{t|t-1} = \overline{\mu}_{t|t-1} - \overline{\mu}^{ex}_{t|t-1}$.[34]

The left panel of figure 5 shows that the long-run expected excess return, $\overline{\mu}^{ex}_{t|t-1}$, has only slightly fallen, from around 4 percent at the beginning of the 20th century, to reach a minimum of 3 percent in 2000 and to rebound thereafter to around 3.5 percent. These numbers are in

discussion.

[30] Returns and dividend growth enter the decomposition with opposite signs, so that r_t^f cancels out leaving unaffected the price dividend ratio.

[31] A long annual series of the risk free rate is taken from Amit Goyal's website http://www.hec.unil.ch/agoyal/ (see Welch and Goyal, 2008, for additional details).

[32] This constraint does not seem to play an important role as the estimated values for the long-run excess return are always far from zero.

[33] See Figure E.8 in Appendix E.6 for a comparison.

[34] Alternatively one could compute the implied riskless real rate as $\overline{r}_{t|t-1} = \overline{g}_{t|t-1} - \overline{g}^{ex}_{t|t-1}$. The two offer a similar picture, as we show in the Figure E.8 in Appendix E.6. Figure 5 reports the average between these two alternative measures.

the ballpark of the estimates provided by Avdis and Wachter (2017) who report an annual equity premium of 3.86 per cent in the post WWII period and of around 4.5 in a longer sample (1927-2011) as well as with the estimate of the equity risk premium (between 2.6 percent and 4.3 percent) for the period 1950-2000 in Fama and French (2002). These results are also in line with Greenwald et al. (2019) who find that the fall in the equity premium played a limited role in the overall fall in the equilibrium rate of return on stocks after the 80s. We add to their analysis by showing that the trend in valuations is almost entirely driven by the fall in the long-run riskless rate. Moreover, the uptick in the equity premium in the 90s, albeit modest in size, is qualitatively in line with the narrative in Farhi and Gourio (2018).

The implied long-run natural rate of interest $\overline{r}_t$, our measure of r-star, has remained stable (between 3 and 4 percent) for about a century until 1960. It has then fallen by 1 full percentage point in the 60s, and by 2 percentage points in the 90s. Its terminal estimate is only slightly above zero. The sudden fall in the equilibrium rate of interest after the 60s is also documented, with a completely different method, by Del Negro et al. (2019). The behaviour of our measure of r-star is qualitatively in line with that of Laubach and Williams (2003) and Holston et al. (2017), also plotted in the chart.[35]

Concluding, our results confirm that the natural rate of interest has fallen substantially over the last three decades, but add to previous studies by drawing a clear link between this secular trend and the increase in stock valuations (Caballero et al., 2017). These results bear important implications for the conduct of monetary policy. In fact, as highlighted by Williams (2019), in a world of low r-star "central banks will be grappling with the challenges of life near the Zero Lower Bound (ZLB), which is why its so critical to consider how the ZLB alters strategies related to monetary policy."

[35] Quantitatively, our measures are lower in the 60s and 70s. In defence of our results we point out that, given our estimate of the expected return on equity, the level of r-star implied by the Laubach and Williams (2003) model prior to the 90s would require a level of the equity risk premium substantially lower than what is typically found in the literature.

7 Conclusions

State space models with time-varying parameters can help us better understanding the co-movement in macro financial aggregates, in a world in which returns, long-run growth and asset valuations appear to have undergone long-lasting shifts. These models present computational as well as analytical challenges. In this paper we propose a method for analyzing state space models with time-varying system matrices where the parameters are driven by the score of the conditional likelihood. We derive a new set of recursions that, running in parallel with the KF, update at each point in time both the vector of TVP and the latent states. The method can easily incorporate in the estimation a broad class of parameter constraints. These are taken into account directly in the estimation algorithm through a Jacobian function, without the need for rejection sampling or complicated filtering techniques. A Monte Carlo analysis provides support for the proposed method. Recent empirical analyses that use our method also testify its usefulness (Delle Monache et al., 2016; Buccheri et al., 2018).

We have then used this framework to fill a gap in the literature that studies the relationship between the price dividend ratio, the expected return on stocks and expected dividend growth in present value models. Our estimates reveal that the secular upward trend in the price dividend ratio, so far unexplained in the literature, is associated with a persistent decline (from 9 to 4 percent) of the long-run expected return on stocks. A decomposition into a riskless component and a risk premium further reveals that most of this decline (four percentage points) is accounted for by the riskless component (i.e. the natural rate of interest, r-star), that is virtually zero at the end of our sample. The long-term equity risk premium has instead remained relatively stable over the past 150 years.

Table 1: Structural Break Tests

	r_t	Δd_t	pd_t
$SupF_T(k)$			
k=1	0.954	0.458	48.836***
	[1921]	[1996]	[1991]
k=2	1.078	0.365	49.804***
	[1921; 1962]	[1918; 1944]	[1954; 1995]
k=3	1.619	0.488	35.862***
	[1921; 1960; 1982]	[1896; 1918; 1944]	[1913; 1954; 1995]
$SupF_T(k+1\|k)$			
k=2	1.676	0.233	14.896***
k=3	2.229	0.846	2.227
$U_d$max	1.619	0.488	49.804***
Nyblom Test			
μ	0.023	0.057	7.401***
σ^2	0.319	0.693**	2.608***
Joint L_c	0.348	0.750	7.873***

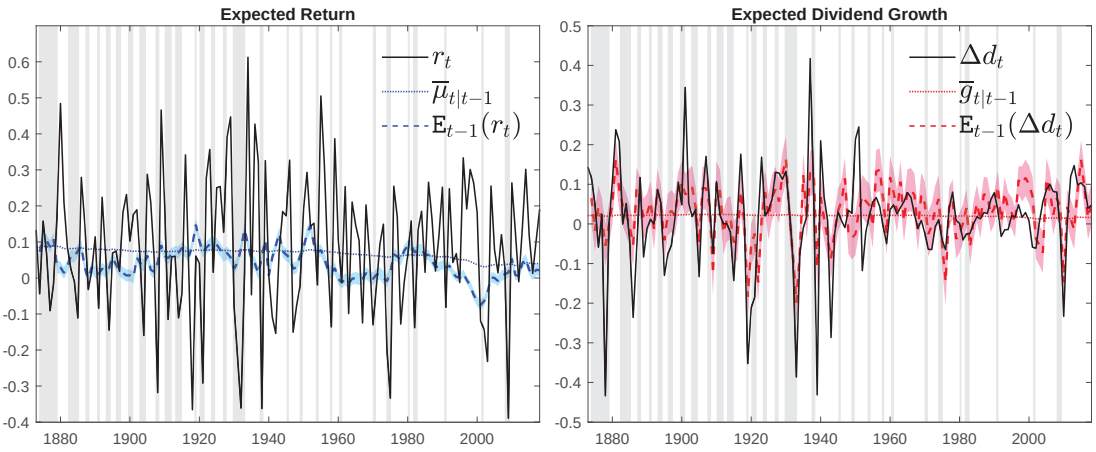
Note. Table 1 reports the results of a number of structural breaks tests applied to the return on stocks, to dividend growth and to the price dividend ratio for the full sample $1873-2018$. The upper panel ($SupF_T(k)$) reports the results of the Bai and Perron (2003) test where null hypothesis of no breaks is tested against the alternatives of $k = 1, 2,$ or 3 breaks. Dates in square brackets are the most likely break date(s) for each of the specifications. The central panel ($SupF_T(k+1|k)$) shows the results of the Bai and Perron (2003) test of k breaks against the alternative of $k+1$ breaks. The U_d max statistics is the result of testing the null hypothesis of absence of breaks against the alternative of an unknown number of breaks. The bottom panel reports the Nyblom (1989) test, as described in Hansen (1992). For each variable y_t we specify the following model: $y_t = \mu + \sigma\epsilon_t$, where ϵ_t is a white noise process, and test the null hypothesis that either μ or σ are constant against the alternative that they evolve as random walks. In the last row we report the results of a joint test that both μ and σ are constant rather than martingales. The symbols $^*/^{**}/^{***}$ indicate significance at the 10/5/1% level.

Table 2: Model Estimation Results

ϕ_μ	0.829			γ_μ	0.151
	[0.010]				[0.010]
ϕ_g	0.345			γ_g	0.052
	[0.010]				[0.010]
$\bar{\sigma}_d$	0.075	a_{σ_d}	0.881	b_{σ_d}	0.015
	[0.074; 0.083]		[0.026]		[0.002]
$\bar{\sigma}_g$	0.083	a_{σ_g}	0.899	b_{σ_g}	0.012
	[0.082; 0.127]		[0.050]		[0.004]
$\bar{\sigma}_\mu$	0.024	a_{σ_μ}	0.902	b_{σ_μ}	0.014
	[0.023; 0.040]		[0.040]		[0.004]
$\bar{\rho}_{d,\mu}$	0.339	$a_{\pi_{d,\mu}}$	0.820	$b_{\pi_{d,\mu}}$	0.013
	[0.280; 0.371]		[0.068]		[0.003]
$\bar{\rho}_{g,\mu}$	-0.232	$a_{\pi_{g,\mu}}$	0.844	$b_{\pi_{g,\mu}}$	0.017
	[-0.253; -0.130]		[0.040]		[0.003]
σ^2_ν	0.001			κ_h	0.020
	[0.0005]				[0.001]
Log Lik.	311.567				

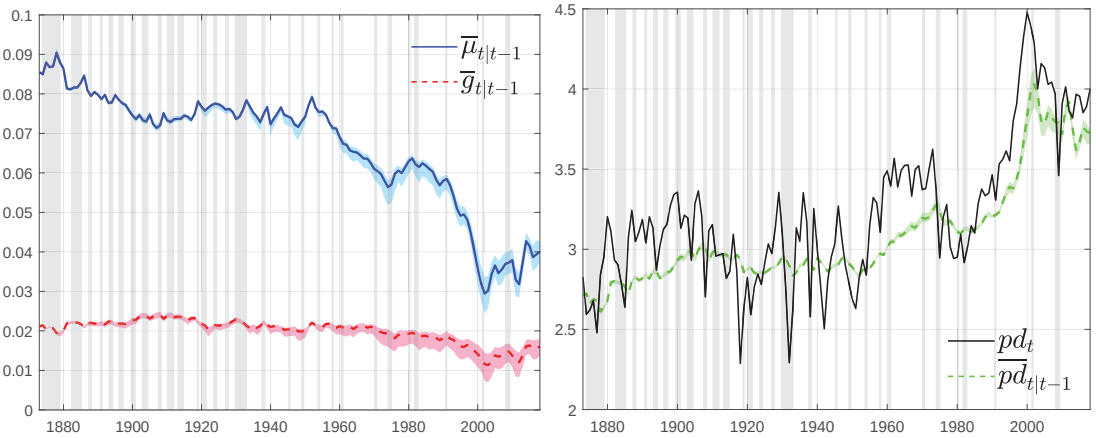
Note. Table 2 reports parameter estimates for our baseline model. In the first two entries of the first column we report the estimates of the autoregressive coefficients of expected returns and expected dividend growth (ϕ_μ and ϕ_g). Then we show the average (over the whole sample) estimates of the volatilities ($\bar{\sigma}_d$, $\bar{\sigma}_g$ and $\bar{\sigma}_\mu$) and correlations ($\bar{\rho}_{d,\mu}$ and $\bar{\rho}_{g,\mu}$) that form the matrix Q_t in the state space model. Finally, we report the estimated volatility of the measurement error for the price dividend ratio. The second and third columns show the estimates of the coefficients that enter the law of motion of the score driven time-varying processes: $f_{t+1} = Af_t + Bs_t$ where A and B are diagonal matrices. Recall that the first two elements of f_t are martingales. This implies that the first two elements of the diagonal of A are zeros. Estimates for the remaining five entries are reported in the second column of this table. In the third column we report the seven elements that form the diagonal of B, that is the loadings on the likelihood score, and the smoothing coefficient applied to the Hessian term (κ_h). For each coefficient we report in square brackets the associated standard error. For the average volatilities and correlations in the first column we report the 68% confidence interval from 1000 simulations of the model (calculated as in Blasques et al., 2016).

Figure 1: Expected Returns and Expected Dividend Growth



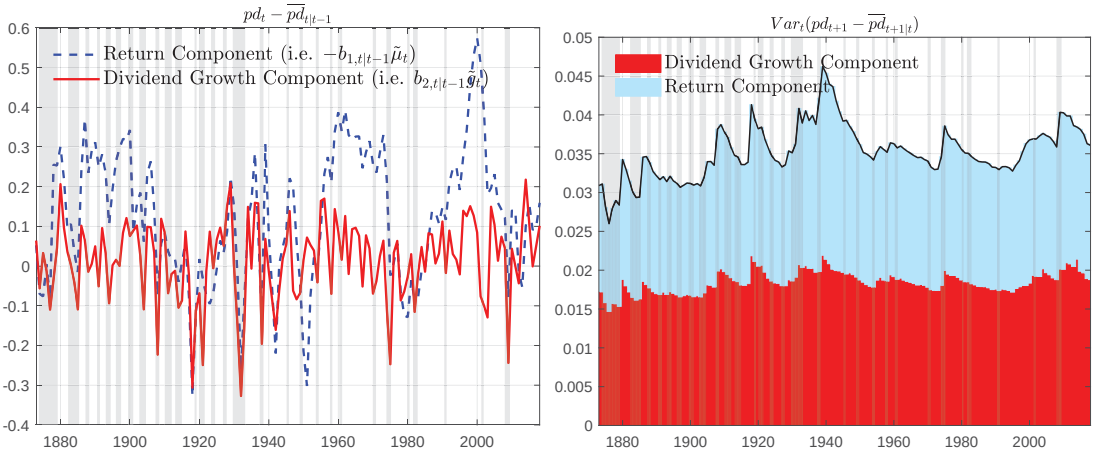
Note. The panel on the left of Figure 1 shows the return on stocks (r_t), expected returns ($\mathrm{E}_{t-1}(r_t)$), and the long-run component of expected returns ($\bar{\mu}_{t|t-1}$). The panel on the right reports dividend growth (Δd_t), together with expected dividend growth ($\mathrm{E}_{t-1}(\Delta d_t)$) and the long-run component of expected dividend growth ($\bar{g}_{t|t-1}$). In both panels the colored bands denote the 68% confidence interval. Vertical shadows indicate recessions as identified by the National Bureau of Economic Research (NBER).

Figure 2: Long-run Expected Returns, Expected Dividend Growth and Price-Dividend Ratio



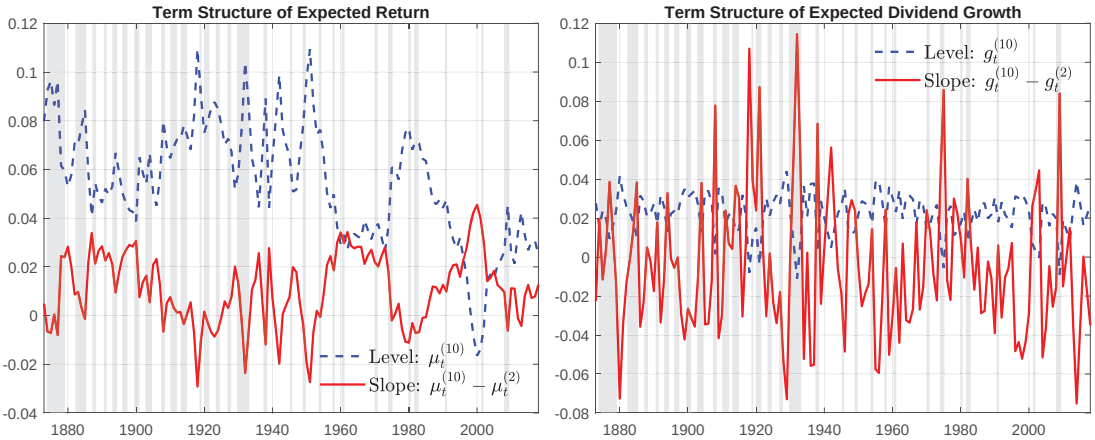
Note. The panel on the left of Figure 2 shows long-run expected returns ($\overline{\mu}_{t|t-1}$, continuous blue line) and long-run expected dividend growth ($\overline{g}_{t|t-1}$, broken red line). The implied long-run price dividend ratio, obtained from $\overline{\mu}_{t|t-1}$ and $\overline{g}_{t|t-1}$ on the basis of equation (14), is shown in the panel on the right (green broken line) together with the actual level of the (log) price dividend ratio (black solid line). Bands around the estimates correspond to the 68% confidence interval. They are obtained through simulation, as discussed in Blasques et al. (2016). Vertical shadows indicate recessions as identified by the National Bureau of Economic Research (NBER).

Figure 3: Price Dividend Ratio Decompositions



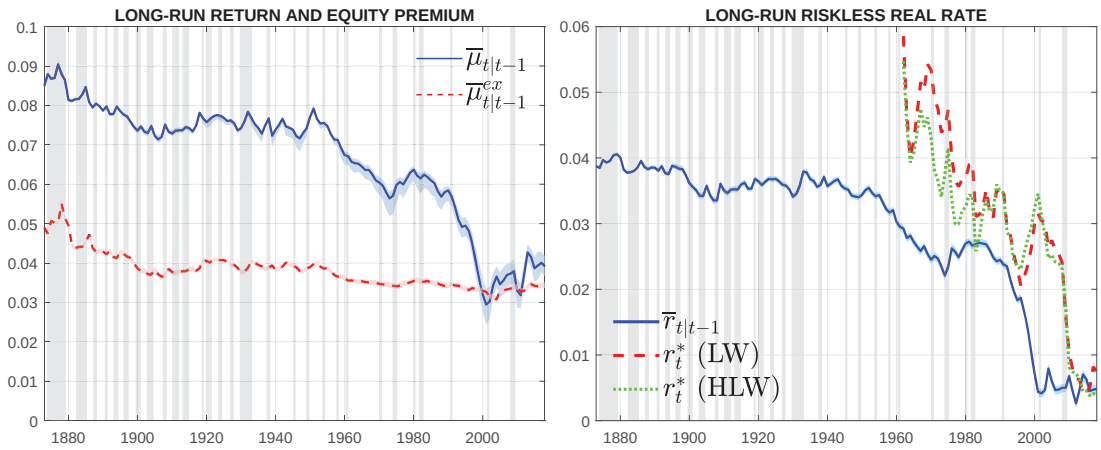
Note. The panel on the left of Figure 3 reports the decomposition of the transitory component of the price dividend ratio into the contribution of the transitory components of, respectively, expected returns and expected dividend growth according to equation (21). The panel on the right shows the decomposition of the conditional variance of the transitory component of the price dividend ratio into the contribution of the shocks to expected returns and of the shocks to expected dividend growth. To obtain this latter decomposition notice that equation (21) implies that $Var_t(pd_{t+1} - \overline{pd}_{t+1|t}) = b^2_{1,t+1|t}\mathtt{Var}_t(\varepsilon^{\mu}_{t+1}) + b^2_{2,t+1|t}\mathtt{Var}_t\left(\varepsilon^{g}_{t+1}\right) - 2b_{1,t+1|t}b_{2,t+1|t}\mathtt{Cov}_t(\varepsilon^{\mu}_{t+1}, \varepsilon^{g}_{t+1})$. The covariance term is split equally between the return and the dividend growth components, as in Piatti and Trojani (2017). Vertical shadows indicate recessions as identified by the National Bureau of Economic Research (NBER).

Figure 4: Term Structure of Expected Returns and Expected Dividend Growth



Note. Figure 4 reports with a broken blue line the level (i.e. 10 years) and with a continuous red line the slope (i.e 10 year minus 2 years) of conditional expected returns (panel on the left) and of conditional expected dividend growth (panel on the right). Vertical shadows indicate recessions as identified by the National Bureau of Economic Research (NBER).

Figure 5: Long-Run Excess Return on Stocks and Long-Run Riskless Rate



Note. The panel on the left of Figure 5 reports in blue the long-run expected return on stocks ($\overline{\mu}_{t|t-1}$) estimated using the baseline model specification described in Section 4 and in red the long run equity premium ($\overline{\mu}^{ex}_{t|t-1}$) estimated as described in Section 6. The panel on the right shows (blue line) the long-run real riskless rate constructed as $\overline{r}_{t|t-1} = 0.5(\overline{\mu}_{t|t-1} - \overline{\mu}^{ex}_{t|t-1}) + 0.5(\overline{g}_{t|t-1} - \overline{g}^{ex}_{t|t-1})$. The panel on the right also shows two alternative measures of r-star for the US, as estimated in Laubach and Williams (2003) denoted with "LW" and in Holston et al. (2017) denoted with "HLW". Vertical shadows indicate recessions as identified by the National Bureau of Economic Research (NBER).

References

Avdis, E. and Wachter, J. A. (2017). Maximum likelihood estimation of the equity premium. *Journal of Financial Economics*, 125(3):589–609.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). *Now-Casting and the Real-Time Data Flow*, volume 2 of *Handbook of Economic Forecasting*, chapter 0, pages 195–237. Elsevier.

Benati, L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics and Control*, 31(8):2847–2877.

Bianchi, F., Mumtaz, H., and Surico, P. (2009). The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics*, 56(6):856–871.

Binsbergen, J. H. V., Hueskes, W., Koijen, R., and Vrugt, E. (2013). Equity yields. *Journal of Financial Economics*, 110(3):503–519.

Binsbergen, J. H. V. and Koijen, R. S. J. (2010). Predictive Regressions: A Present Value Approach. *Journal of Finance*, 65(4):1439–1471.

Blasques, F., Koopman, S. J., Lasak, K., and Lucas, A. (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting*, 32(3):875–887.

Blasques, F., Koopman, S. J., and Lucas, A. (2014). Optimal Formulations for Nonlinear Autoregressive Processes. Tinbergen Institute Discussion Papers 14-103/III, Tinbergen Institute.

Blasques, F., Koopman, S. J., and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika*, 102(2):325–343.

Boudoukh, J., Michaely, R., Richardson, M., and Roberts, M. R. (2007). On the Importance of Measuring Payout Yield: Implications for Empirical Asset Pricing. *Journal of Finance*, 62(2):877–915.

Buccheri, G., Bormetti, G., Corsi, F., and Lillo, F. (2018). A Score-Driven Conditional Correlation Model for Noisy and Asynchronous Data: an Application to High-Frequency Covariance Dynamics. Papers 1803.04894, arXiv.org.

Caballero, R. J., Farhi, E., and Gourinchas, P.-O. (2017). Rents, Technical Change, and Risk Premia Accounting for Secular Trends in Interest Rates, Returns on Capital, Earning Yields, and Factor Shares. *American Economic Review*, 107(5):614–620.

Campbell, J. Y. (2018). *Financial Decisions and Markets*. Princeton University Press, Princeton, NJ.

Campbell, J. Y. and Ammer, J. (1993). What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns. *Journal of Finance*, 48(1):3–37.

Campbell, J. Y., Giglio, S., and Polk, C. (2013). Hard Times. *Review of Asset Pricing Studies*, 3(1):95–132.

Campbell, J. Y. and Shiller, R. J. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Viceira, L. (2005). The Term Structure of the Risk-Return Tradeoff. NBER Working Papers 11119, National Bureau of Economic Research, Inc.

Carvalho, C. M., Lopes, H. F., and McCulloch, R. E. (2018). On the long-run volatility of stocks. *Journal of the American Statistical Association*, 113(523):1050–1069.

Chen, H., Kumar, P., and van Schuppen, J. (1989). On Kalman filtering for conditionally Gaussian systems with random matrices. *Systems & Control Letters*, 13(5):397–404.

Cochrane, J. H. (2008a). State-Space vs. VAR Models for Stock Returns. Unpublished manuscript.

Cochrane, J. H. (2008b). The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies*, 21(4):1533–1575.

Cochrane, J. H. (2011). Discount Rates. *Journal of Finance*, 66(4):1047–1108.

Cogley, T. and Sargent, T. J. (2005). Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics*, 8(2):262–302.

Creal, D., Koopman, S. J., and Lucas, A. (2008). A General Framework for Observation Driven Time-Varying Parameter Models. Tinbergen Institute Discussion Papers 08-108/4, Tinbergen Institute.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized Autoregressive Score Models With Applications. *Journal of Applied Econometrics*, 28(5):777–795.

Del Negro, M., Giannone, D., Giannoni, M. P., and Tambalotti, A. (2017). Safety, Liquidity, and the Natural Rate of Interest. *Brookings Papers on Economic Activity*, 48(1 (Spring):235–316.

Del Negro, M., Giannone, D., Giannoni, M. P., and Tambalotti, A. (2019). Global trends in interest rates. *Journal of International Economics*, 118(C):248–262.

Del Negro, M. and Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Reports 326, Federal Reserve Bank of New York.

Delle Monache, D. and Petrella, I. (2017). Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting*, 33(2):482–501.

Delle Monache, D., Petrella, I., and Venditti, F. (2016). Common Faith or Parting Ways? A Time Varying Parameters Factor Analysis of Euro-Area Inflation. In *Dynamic Factor Models*, volume 35 of *Advances in Econometrics*, pages 539–565. Emerald Publishing Ltd.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press.

Eickmeier, S., Lemke, W., and Marcellino, M. (2015). Classical time varying factor-augmented vector auto-regressive models. *Journal of the Royal Statistical Society Series A*, 178:493–533.

Fama, E. F. and French, K. R. (2002). The Equity Premium. *Journal of Finance*, 57:637–659.

Farhi, E. and Gourio, F. (2018). Accounting for Macro-Finance Trends: Market Power, Intangibles, and Risk Premia. *Brookings Papers on Economic Activity*, (Forthcoming).

Greenwald, D. L., Lettau, M., and Ludvigson, S. C. (2019). How the Wealth Was Won: Factors Shares as Market Fundamentals. NBER Working Papers 25769, National Bureau of Economic Research, Inc.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14(4):517–533.

Harvey, A., Ruiz, E., and Sentana, E. (1992). Unobserved component time series models with Arch disturbances. *Journal of Econometrics*, 52(1-2):129–157.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Herwartz, H., Rengel, M., and Xu, F. (2016). Local Trends in PricetoDividend Ratios: Assessment, Predictive Value, and Determinants. *Journal of Money, Credit and Banking*, 48(8):1655–1690.

Holston, K., Laubach, T., and Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108(S1):59–75.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10):2177 – 2189.

Julliard, C. and Bryzgalova, S. (2019). Consumption in Asset Returns. Technical report, mimeo.

Kapetanios, G., Marcellino, M., and Venditti, F. (2019). Large time-varying parameter VARs: a non-parametric approach. *Journal of Applied Econometrics, forthcoming*.

Koijen, R. S. and Nieuwerburgh, S. V. (2011). Predictability of Returns and Cash Flows. *Annual Review of Financial Economics*, 3(1):467–491.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198.

Koop, G. and Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, 71(C):101–116.

Koopman, S. J., Mallee, M. I. P., and Van der Wel, M. (2010). Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson-Siegel Model With Time-Varying Parameters. *Journal of Business & Economic Statistics*, 28(3):329–343.

Laubach, T. and Williams, J. C. (2003). Measuring the Natural Rate of Interest. *The Review of Economics and Statistics*, 85(4):1063–1070.

Lettau, M. and Nieuwerburgh, S. V. (2008). Reconciling the Return Predictability Evidence. *Review of Financial Studies*, 21(4):1607–1652.

Marcellino, M., Porqueddu, M., and Venditti, F. (2016). Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility. *Journal of Business & Economic Statistics*, 34(1):118–127.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Piatti, I. and Trojani, F. (2017). Predictable Risks and Predictive Regression in Present-Value Models. Working paper, Said Business School.

Rytchkov, O. (2012). Filtering Out Expected Dividends and Expected Returns. *Quarterly Journal of Finance*, 2(03):1–56.

Shiller, R. J. (1989). *Market Volatility*. MIT Press, Cambridge, MA.

Siegel, J. J. (2008). *Stocks for the Long Run*. McGraw Hill, New York, NY.

Sims, C. (2001). Comment on Sargent and Cogley's "Evolving US Postwar Inflation Dynamics". Unpublished.

Stock, J. H. and Watson, M. W. (1998). Asymptotically Median Unbiased Estimation of Coefficient Variance in a Time Varying Parameter Model. *Journal of the American Statistical Association*, 83(441):349–358.

Stock, J. H. and Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39(s1):3–33.

Theil, H. and Goldberger, A. S. (1961). On pure and mixed statistical estimation in economics. *International Economic Review*, 2(1):65–78.

Timmermann, A. (2001). Structural Breaks, Incomplete Information, and Stock Prices. *Journal of Business & Economic Statistics*, 19(3):299–314.

Welch, I. and Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, 21(4):1455–1508.

Williams, J. C. (2019). Living Life Near the ZLB. Remarks at 2019 Annual Meeting of the Central Bank Research Association, New York City. Speech 327, Federal Reserve Bank of New York.

A Proofs

We follow the notation and the results in Abadir and Magnus (2005, ch. 13). Given a $N \times m$ matrix X, $\text{vec}(X)$ is the vector obtained by stacking the columns of X one underneath the other. The $Nm \times Nm$ commutation matrix $C_{N,m}$ is such that $C_{N,m}\text{vec}(X) = \text{vec}(X')$. For $N = m$ the $m^2 \times m^2$ commutation matrix is denoted by C_m. The identity matrix of order k is denoted by I_k, and '$\otimes$' is the Kronecker product. Given a square matrix U, the symmetrizer matrix is $N_n = \frac{1}{2}(I_{n^2} + C_n)$ and $N_n\text{vec}(U) = \text{vec}\left[\frac{1}{2}(U + U')\right]$.

A.1 Gradient and information matrix

The gradient vector is

$$\begin{aligned}
\nabla_t = \left(\frac{\partial \ell_t}{\partial f_t'}\right)' &= -\frac{1}{2}\left[\frac{\partial \log|F_t|}{\partial f_t'} + \frac{\partial v_t' F_t^{-1} v_t}{\partial f_t'}\right]' \\
&= -\frac{1}{2}\left[\frac{1}{|F_t|}\frac{\partial |F_t|}{\partial \text{vec}(F_t)'}\frac{\partial \text{vec}(F_t)}{\partial f_t'} + \frac{\partial v_t' F_t^{-1} v_t}{\partial v_t}\frac{\partial v_t}{\partial f_t'} + \frac{\partial v_t' F_t^{-1} v_t}{\partial \text{vec}(F_t^{-1})'}\frac{\partial \text{vec}(F_t^{-1})}{\partial \text{vec}(F_t)'}\frac{\partial \text{vec}(F_t)}{\partial f_t'}\right]' \\
&= -\frac{1}{2}\left[\text{vec}(F_t^{-1})'\dot{F}_t + 2v_t' F_t^{-1}\dot{V}_t - (v_t' \otimes v_t')(F_t^{-1} \otimes F_t^{-1})\dot{F}_t\right]' \\
&= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})(v_t \otimes v_t) - \dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\text{vec}(F_t) - 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\text{vec}(v_t v_t' - F_t) - 2\dot{V}_t' F_t^{-1} v_t\right]. \qquad \text{(A.1)}
\end{aligned}$$

We now compute the information matrix as the expected value of the Hessian matrix

$$\mathcal{I}_t = -\text{E}_t\left[\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'}\right]. \qquad \text{(A.2)}$$

From the third line of (A.1) we can write that

$$\begin{aligned}
\nabla_t &= -\frac{1}{2}\left[\dot{F}_t'[\text{vec}(F_t^{-1}) - \text{vec}(F_t^{-1} v_t v_t' F_t^{-1})] + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t'\text{vec}(F_t^{-1} - F_t^{-1} v_t v_t' F_t^{-1}) + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t'\text{vec}[F_t^{-1}(I_N - v_t v_t' F_t^{-1})] + 2\dot{V}_t' F_t^{-1} v_t\right] \\
&= -\frac{1}{2}\left[\dot{F}_t'(I_N \otimes F_t^{-1})\text{vec}(I_N - v_t v_t' F_t^{-1}) + 2\dot{V}_t' F_t^{-1} v_t\right]. \qquad \text{(A.3)}
\end{aligned}$$

The negative Hessian is equal to:

$$
\begin{aligned}
-\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'} &= \frac{1}{2}\frac{\partial \Phi_t}{\partial f_t'} + \frac{\partial \text{vec}(\dot{V}_t' F_t^{-1} v_t)}{\partial \text{vec}(\dot{V}_t')'}\frac{\partial \text{vec}(\dot{V}_t')}{\partial f_t'} \qquad (A.4)\\
&\quad + \frac{\partial \text{vec}(\dot{V}_t' F_t^{-1} v_t)}{\partial \text{vec}(F_t^{-1})}\frac{\partial \text{vec}(F_t^{-1})}{\partial \text{vec}(F_t)'}\frac{\partial \text{vec}(F_t)}{\partial f_t'} + \frac{\partial \text{vec}(\dot{V}_t' F_t^{-1} v_t)}{\partial v_t}\frac{\partial v_t}{\partial f_t'}\\
&= \frac{1}{2}\frac{\partial \Phi_t}{\partial f_t'} + (v_t' F_t^{-1} \otimes I_N)\ddot{V}_t - (v_t \otimes \dot{V}_t)'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + \dot{V}_t' F_t^{-1}\dot{V}_t, \qquad (A.5)
\end{aligned}
$$

where $\Phi_t = \dot{F}_t'(I_N \otimes F_t^{-1})\text{vec}(I_N - v_t v_t' F_t^{-1})$. Let us now compute the following Jacobian:

$$
\begin{aligned}
\frac{\partial \Phi_t}{\partial f_t'} &= \frac{\partial \Phi_t}{\partial \text{vec}(\dot{F}_t')'}\frac{\partial \text{vec}(\dot{F}_t')}{\partial f_t'} + \frac{\partial \Phi_t}{\partial \text{vec}(I_N \otimes F_t^{-1})'}\frac{\partial \text{vec}(I_N \otimes F_t^{-1})}{\partial \text{vec}(F_t^{-1})'}\frac{\partial \text{vec}(F_t^{-1})}{\partial \text{vec}(F_t)'}\frac{\partial \text{vec}(F_t)}{\partial f_t'}\\
&\quad - \frac{\partial \Phi_t}{\partial \text{vec}(I_N - v_t v_t' F_t^{-1})'}\frac{\partial \text{vec}(v_t v_t' F_t^{-1})'}{\partial f_t'}\\
&= \left[\text{vec}(I_N - v_t v_t' F_t^{-1})'(I \otimes F_t^{-1}) \otimes I\right]\ddot{F}_t - \left[\text{vec}(I_N - v_t v_t' F_t^{-1})' \otimes \dot{F}_t'\right](F_t^{-1} \otimes F_t^{-1})\dot{F}_t\\
&\quad - \dot{F}_t'(I \otimes F_t^{-1})\left[\frac{\partial \text{vec}(v_t v_t' F_t^{-1})}{\partial \text{vec}(v_t v_t')'}\frac{\partial \text{vec}(v_t v_t')}{\partial v_t'}\frac{\partial v_t}{\partial f_t'} + \frac{\partial \text{vec}(v_t v_t' F_t^{-1})}{\partial \text{vec}(F_t^{-1})'}\frac{\partial \text{vec}(F_t^{-1})}{\partial \text{vec}(F_t)'}\frac{\partial \text{vec}(F_t)}{\partial f_t'}\right]\\
&= [\text{vec}(I_N - v_t v_t' F_t^{-1})'(I_N \otimes F_t^{-1}) \otimes I_N]\ddot{F}_t - \left[\text{vec}(I_N - v_t v_t' F_t^{-1})' \otimes \dot{F}_t'\right](F_t^{-1} \otimes F_t^{-1})\dot{F}_t\\
&\quad - \dot{F}_t'(F_t^{-1} \otimes F_t^{-1})(v_t \otimes I_N + I_N \otimes v_t)\dot{V}_t + \dot{F}_t'(F_t^{-1} \otimes F_t^{-1} v_t v_t' F_t^{-1})\dot{F}_t. \qquad (A.6)
\end{aligned}
$$

where $\ddot{F}_t = \frac{\partial \text{vec}(\dot{F}_t')}{\partial f_t'}$. Putting together (A.5) and (A.6) we obtain the following expression:

$$
\begin{aligned}
-\frac{\partial^2 \ell_t}{\partial f_t f_t'} &= \frac{1}{2}[\text{vec}(I_N - v_t v_t' F_t^{-1})'(I_N \otimes F_t^{-1}) \otimes I_N]\ddot{F}_t - \frac{1}{2}\left[\text{vec}(I_N - v_t v_t' F_t^{-1})' \otimes \dot{F}_t'\right](F_t^{-1} \otimes F_t^{-1})\dot{F}_t\\
&\quad - \frac{1}{2}\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})(v_t \otimes I_N + I_N \otimes v_t)\dot{V}_t + \frac{1}{2}\dot{F}_t'\left(F_t^{-1} \otimes F_t^{-1} v_t v_t' F_t^{-1}\right)\dot{F}_t\\
&\quad + (v_t' F_t^{-1} \otimes I_N)\ddot{V}_t - (v_t \otimes \dot{V}_t)'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + \dot{V}_t' F_t^{-1}\dot{V}_t, \qquad (A.7)
\end{aligned}
$$

where $\ddot{V}_t = \frac{\partial \text{vec}(\dot{V}_t')}{\partial f_t'}$. Following Harvey (1989, p.141), taking the conditional expectation of (A.7) the fourth and the seventh term in (A.7) are the only nonzero elements and the information matrix is equal to

$$
\mathcal{I}_t = \frac{1}{2}\dot{F}_t'\left(F_t^{-1} \otimes F_t^{-1}\right)\dot{F}_t + \dot{V}_t' F_t^{-1}\dot{V}_t. \qquad (A.8)
$$

A.2 Jacobians of the Kalman filter

Let us write the KF recursions (3) in terms of the predictive filter:

$$
\begin{aligned}
&v_t = y_t - Z_t a_t, && F_t = Z_t P_t Z_t' + H_t,\\
&K_t = T_{t+1} P_t Z_t' F_t^{-1}, && L_t = T_{t+1} - K_t Z_t,\\
&a_{t+1} = T_{t+1} a_t + K_t v_t && P_{t+1} = T_{t+1} P_t L_t' + Q_{t+1}, \quad t = 1, ..., n.
\end{aligned} \qquad (A.9)
$$

Given the model specific Jacobian matrices:

$$\dot{Z}_t = \frac{\partial \text{vec}(Z_t)}{\partial f_t'}, \quad \dot{H}_t = \frac{\partial \text{vec}(H_t)}{\partial f_t'}, \quad \dot{T}_t = \frac{\partial \text{vec}(T_t)}{\partial f_t'}, \quad \dot{Q}_t = \frac{\partial \text{vec}(Q_t)}{\partial f_t'},$$

we compute the following Jacobian matrices:

$$\dot{V}_t = \frac{\partial v_t}{\partial f_t'} = \left[\frac{\partial v_t}{\partial \text{vec}(Z_t)'}\frac{\partial \text{vec}(Z_t)}{\partial f_t'} + \frac{\partial v_t}{\partial a_t'}\frac{\partial a_t}{\partial f_t'}\right] = -[(a_t' \otimes I_N)\dot{Z}_t + Z_t\dot{A}_t]. \tag{A.10}$$

$$\begin{aligned}\dot{F}_t = \frac{\partial \text{vec}(F_t)}{\partial f_t'} &= \frac{\partial \text{vec}(F_t)}{\partial \text{vec}(Z_t)'}\frac{\partial \text{vec}(Z_t)}{\partial f_t'} + \frac{\partial \text{vec}(F_t)}{\partial \text{vec}(P_t)'}\frac{\partial \text{vec}(P_t)}{\partial f_t'} + \dot{H}_t \\ &= 2N_N(Z_tP_t \otimes I_N)\dot{Z}_t + (Z_t \otimes Z_t)\dot{P}_t + \dot{H}_t. \end{aligned} \tag{A.11}$$

$$\dot{K}_t = \frac{\partial \text{vec}(K_t)}{\partial f_{t+1}'} = \frac{\partial \text{vec}(K_t)}{\partial \text{vec}(T_{t+1})'}\frac{\partial \text{vec}(T_{t+1})}{\partial f_{t+1}'} = (F_t^{-1}Z_tP_t \otimes I_m)\dot{T}_{t+1}. \tag{A.12}$$

$$\begin{aligned}\dot{L}_t = \frac{\partial \text{vec}(L_t)}{\partial f_{t+1}'} &= \frac{\partial \text{vec}(L_t)}{\partial \text{vec}(T_{t+1})'}\frac{\partial \text{vec}(T_{t+1})}{\partial f_{t+1}'} + \frac{\partial \text{vec}(L_t)}{\partial \text{vec}(K_t)'}\frac{\partial \text{vec}(K_t)}{\partial f_{t+1}'} \\ &= \dot{T}_{t+1} - (Z_t' \otimes I_m)\dot{K}_t. \end{aligned} \tag{A.13}$$

$$\begin{aligned}\dot{A}_{t+1} = \frac{\partial a_{t+1}}{\partial f_{t+1}'} &= \frac{\partial T_{t+1}a_t}{\partial \text{vec}(T_{t+1})'}\frac{\partial \text{vec}(T_{t+1})}{\partial f_{t+1}'} + \frac{\partial K_tv_t}{\partial \text{vec}(K_t)'}\frac{\partial \text{vec}(K_t)}{\partial f_{t+1}'} \\ &= (a_t' \otimes I_m)\dot{T}_{t+1} + (v_t' \otimes I_m)\dot{K}_t. \end{aligned} \tag{A.14}$$

$$\begin{aligned}\dot{P}_{t+1} = \frac{\partial \text{vec}(P_{t+1})}{\partial f_{t+1}'} &= \frac{\partial \text{vec}(T_{t+1}P_tL_t')}{\partial \text{vec}(T_{t+1})'}\frac{\partial \text{vec}(T_{t+1})}{\partial f_{t+1}'} + \frac{\partial \text{vec}(T_{t+1}P_tL_t')}{\partial \text{vec}(L_t')'}\frac{\partial \text{vec}(L_t')}{\partial \text{vec}(L_t)'}\frac{\partial \text{vec}(L_t)}{\partial f_{t+1}'} + \dot{Q}_{t+1} \\ &= (L_tP_t \otimes I_m)\dot{T}_{t+1} + (I_m \otimes T_{t+1}P_t)C_m\dot{L}_t + \dot{Q}_{t+1}. \end{aligned} \tag{A.15}$$

Plugging (A.12) in (A.14) we obtain

$$\begin{aligned}\dot{A}_{t+1} &= [(a_t' \otimes I_m) + (v_t'F_t^{-1}Z_tP_t \otimes I_m)]\dot{T}_{t+1} \\ &= [(a_t' + v_t'F_t^{-1}Z_tP_t) \otimes I_m]\dot{T}_{t+1} \\ &= (a_{t|t}' \otimes I_m)\dot{T}_{t+1}. \end{aligned} \tag{A.16}$$

Plugging (A.12) and (A.13) in (A.15) we obtain

$$\begin{aligned}
\dot{P}_{t+1} &= (L_t P_t \otimes I_m)\dot{T}_{t+1} + (I_m \otimes T_{t+1}P_t)C_m[I_{m^2} - (Z_t' F_t^{-1} Z_t P_t \otimes I_m)]\dot{T}_{t+1} + \dot{Q}_{t+1} \\
&= (L_t P_t \otimes I_m)\dot{T}_{t+1} + C_m[(T_{t+1}P_t \otimes I_m) - (T_{t+1}P_t Z_t' F_t^{-1} Z_t P_t \otimes I_m)]\dot{T}_{t+1} + \dot{Q}_{t+1} \\
&= (L_t P_t \otimes I_m)\dot{T}_{t+1} + C_m[(T_{t+1}P_t - T_{t+1}P_t Z_t' F_t^{-1} Z_t P_t) \otimes I_m]\dot{T}_{t+1} + \dot{Q}_{t+1} \\
&= (T_{t+1}P_{t|t} \otimes I_m)\dot{T}_{t+1} + C_m(T_{t+1}P_{t|t} \otimes I_m)\dot{T}_{t+1} + \dot{Q}_{t+1} \\
&= 2N_m(T_{t+1}P_{t|t} \otimes I_m)\dot{T}_{t+1} + \dot{Q}_{t+1}. \qquad \text{(A.17)}
\end{aligned}$$

Note that expressions (A.10), (A.11), (A.16) and (A.17) can be be also obtained by differentiating the recursions in (3), therefore avoiding the computation of (A.12)-(A.15). Shifting one period backward (A.16) and substituting into (A.10) we obtain:

$$\dot{V}_t = -[(a_t' \otimes I_N)\dot{Z}_t + (a_{t-1|t-1}' \otimes Z_t)\dot{T}_t]. \qquad \text{(A.18)}$$

Shifting one period backward (A.17) and substituting into (A.11) we obtain

$$\dot{F}_t = 2N_N(Z_t P_t \otimes I_N)\dot{Z}_t + 2(Z_t \otimes Z_t)N_m(T_t P_{t-1|t-1} \otimes I_m)\dot{T}_t + \dot{H}_t + (Z_t \otimes Z_t)\dot{Q}_t. \qquad \text{(A.19)}$$

B Examples

In this section, we look at some examples of time-varying state space models that have been considered in the literature and show how they can be analyzed with our score driven algorithm. In particular, in section B.1 we consider the local level model with time-varying volatility, a model that has been popularized by Stock and Watson (2007) to study inflation dynamics. In section, B.2 we consider autoregressive processes with time-varying parameters.

B.1 Local level model

Let us consider the local level model with time-varying volatilities:

$$\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \qquad & \varepsilon_t &\sim \mathcal{N}(0, \sigma^2_{\varepsilon,t}), \\
\mu_t &= \mu_{t-1} + \eta_t, \qquad & \eta_t &\sim \mathcal{N}(0, \sigma^2_{\eta,t}).
\end{aligned} \qquad \text{(B.1)}$$

This model has been proposed by Stock and Watson (2007) to model US inflation. The estimation of (B.1) using the score-driven approach was initially proposed by Creal et al. (2008, sec. 4.4). Their algorithm, however, contains some inconsistencies and below we show how it should be modified. First, the state vector and the system matrices are equal to $\alpha_t = \mu_t$, $Z_t = T_t = 1$, $H_t = \sigma^2_{\varepsilon,t}$, $Q_t = \sigma^2_{\eta,t}$. Thus, the application of the Kalman Filter leads to the

following recursions:

$$\begin{aligned} v_t &= y_t - a_t, & a_{t+1} &= a_t + k_t v_t, \\ d_t &= p_t + \sigma^2_{\varepsilon,t}, & p_{t+1} &= (1-k_t)p_t + \sigma^2_{\eta,t+1}, \\ k_t &= p_t/d_t, & t &= 1, ..., n. \end{aligned} \tag{B.2}$$

Second, the vector of time-varying parameters, which is recursively estimated using the score-driven filter (4), is equal to $f_t = (\log \sigma_{\varepsilon,t}, \log \sigma_{\eta,t})'$. Third, the corresponding Jacobian matrices are $\dot{H}_t = \left(2\sigma^2_{\varepsilon,t}, 0\right)$, $\dot{Q}_t = \left(0, 2\sigma^2_{\eta,t}\right)$, $\dot{Z}_t = \dot{T}_t = (0,0)'$. Finally, the conditional log-likelihood is equal to $\ell_t \propto -\frac{1}{2}(\log d_t + v_t^2/d_t)$, and the gradient vector and information matrix in (6) are:[1]

$$\nabla_t = \frac{1}{2d_t^2}\dot{d}_t'(v_t^2 - d_t), \quad \mathcal{I}_t = \frac{1}{2d_t^2}\left[\dot{d}_t'\dot{d}_t\right], \tag{B.3}$$

where $\dot{d}_t = \left(2\sigma^2_{\varepsilon,t}, 2\sigma^2_{\eta,t}\right)$. The score's driving mechanism is represented by $(v_t^2 - d_t)$, that is the deviation of the current estimation of the prediction error variance (by means of v_t) from its past estimation (d_t). Such term is weighted by $\dot{d}_t$, which determines the different impact on the two time-varying volatilities.[2] A multivariate extension of the score driven model considered in this section has been used by Buccheri et al. (2018) to model high-frequency multivariate financial time-series.

B.2 Autoregressive models

Here, we consider models that are perfectly observable. In this case, our algorithm collapses to the score-driven filter proposed in the literature by Blasques et al. (2014) and Delle Monache and Petrella (2017).

Let us consider the following autoregressive model with time-varying parameters:

$$y_t = \phi_t y_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma_t^2). \tag{B.4}$$

The model can be easily cast in state space form by setting $\alpha_t = y_t$, $Z_t = 1$, $\epsilon_t = 0$, $H_t = 0$, $T_t = \phi_t$, $\eta_t = \xi_t$ and $Q_t = \sigma_t^2$. The vector of interest is $f_t = (\phi_t, \sigma_t^2)'$, and the corresponding Jacobians are $\dot{T}_t = (1,0)$, $\dot{Q}_t = (0,1)$, $\dot{Z}_t = \dot{H}_t = 0$. For simplicity we do not impose any restrictions on f_t.[3] Combining the KF (3) with the new filter (6)-(7) leads to the following

[1] Note that the information matrix is singular. Therefore, it needs to be smoothed to be used as scaling matrix.

[2] Note that the resulting algorithm is different from the one derived in Creal et al. (2008). In fact, the two volatilities are only updated using information in the second moments of the data and the level of the prediction error, v_t, does not enter directly the filter.

[3] Delle Monache and Petrella (2017) show how to impose stable roots.

expression for the scaled-score vector:

$$s_t = \mathcal{I}_t^{-1}\nabla_t = \begin{bmatrix} \frac{1}{y_{t-1}^2}(y_{t-1}\xi_t) \\ (\xi_t^2 - \sigma_t^2) \end{bmatrix}. \tag{B.5}$$

The driving process for the coefficient ϕ_t is the prediction error scaled by the regressor, while the volatility σ_t^2 is driven by the squared prediction error. These match exactly those in Blasques et al. (2014) and Delle Monache and Petrella (2017).

We can easily extend to the case of vector autoregressive model of order p:

$$y_t = \Phi_{1,t}y_{t-1} + \cdots + \Phi_{p,t}y_{t-p} + c_t + \xi_t, \quad \xi_t \sim \mathcal{N}(0, \Gamma_t). \tag{B.6}$$

The SSF representation can be obtain by setting:

$$\alpha_t = (y_t', \ldots, y_{t-p}', 1)', \quad Z_t = I, \quad T_t = \begin{bmatrix} \Phi_{1,t} & \ldots & \Phi_{p,t} & c_t \\ I & & & \\ & \ddots & & \\ & & I & \\ 0 & & & 0 \end{bmatrix}, \quad Q_t = \begin{bmatrix} \Gamma_t & & & \\ & 0 & & \\ & & \ddots & f \\ & & & 0 \end{bmatrix},$$

where $f_t = (\text{vec}(\Phi_t)', \text{vec}(\Gamma_t)')'$, with $\Phi_t = [\Phi_{1,t}, \ldots, \Phi_{p,t}, c_t]$. We therefore have that $\dot{Z}_t = \dot{H}_t = 0$, while $\dot{T}_t$ and $\dot{Q}_t$, are selection matrices. After some algebra, the scaled-score can be specialized in two sub-vectors driving the coefficients and the volatilities:

$$s_t = \mathcal{I}_t^{-1}\nabla_t = \begin{bmatrix} (X_t'\Gamma_t^{-1}X_t)^{-1}X_t'\Gamma_t^{-1}\xi_t \\ \text{vec}(\xi_t\xi_t') - \text{vec}(\Gamma_t) \end{bmatrix}, \tag{B.7}$$

where $X_t = (\alpha_{t-1}' \otimes I)$. Interestingly, the algorithm proposed by Koop and Korobilis (2013) can be obtained as a special case of ours by imposing some restrictions on the model. First, set the law of motion (4) c = 0, A = I and let B depend on two scalar parameters (one driving the coefficients and one the volatility). Second, replace the information matrix for the time-varying coefficients by its smoothed estimator: $\tilde{\mathcal{I}}_{c,t} = (1-\kappa)\tilde{\mathcal{I}}_{c,t-1} + \kappa(X_t'\Gamma_t^{-1}X_t)$.

It is well known that VAR models tend to suffer from the 'curse of dimensionality' and to overfit the data (see, e.g., Litterman, 1979; Doan et al., 1986). In the context of fixed coefficients, Bayesian techniques are used to shrink the parameters, therefore reducing estimation variance. Our algorithm can easily accommodate shrinkage, as discussed in section 2.3.2 and as detailed in Appendix C.4. A regularized version of the model proposed by Koop and Korobilis (2013) can then be easily obtained, where the parameters can be shrunk towards any type of prior that can be reformulated in the form of stochastic constraint (10). These include the most popular priors typically considered in the Bayesian VAR literature, including the Min-

nesota prior, the sum of coefficients prior and the long-run prior (see, e.g., Del Negro, 2012; Kapetanios et al., 2019).

C Additional results on the score driven algorithm

C.1 State space model in forward form

Let us consider the state space model in the so-called *forward form*:

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, H_t), \\ \alpha_{t+1} &= T_t\alpha_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t), \quad \alpha_1 \sim \mathcal{N}(a_1, P_1), \quad t = 1, ..., n. \end{aligned} \tag{C.1}$$

The log-likelihood function is the same as in (2), thus ∇_t and $\mathcal{I}_t$ are the same as in (6). The recursions in (3) are replaced by the following ones:

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ K_t &= T_t P_t Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t, \\ a_{t+1} &= T_t a_t + K_t v_t & P_{t+1} &= T_t P_t L_t' + Q_t, \quad t = 1, ..., n. \end{aligned} \tag{C.2}$$

Here we assume the following time dependency in the system matrices: $Z_t = Z(f_t, \theta_m)$, $H_t = H(f_t, \theta_m)$, but $T_t = T(f_{t+1}, \theta_m)$ and $Q_t = Q(f_{t+1}, \theta_m)$. The formulae (A.10)-(A.19) remain unchanged by simply replacing T_{t+1}, $\dot{T}_{t+1}$, Q_{t+1}, and $\dot{Q}_{t+1}$ with T_t, $\dot{T}_t$, Q_t, and $\dot{Q}_t$.

C.2 Mixed frequencies and missing observations

Assume to have a data set containing missing observations. The observed vector is represented by $W_t y_t$, where W_t is an $N_t \times N$ selection matrix with $1 \leq N_t \leq N$, meaning that at least one observation is available at time t. Note that W_t is obtained by eliminating the $i-th$ row from I_N when the $i-th$ variable is missing. In this setting, at each time t the likelihood ℓ_t is computed using N_t observations; i.e. $\ell_t = \log p(W_t y_t | Y_{t-1}, \theta)$, that is the *marginal* likelihood. In practice, the score of the marginal likelihood is computed and the updating of f_t is based on the available information.[4] Given this reparameterization, the measurement equation in (1) is modified as follows:

$$W_t y_t = W_t Z_t \alpha_t + W_t \varepsilon_t, \quad W_t \epsilon_t \sim \mathcal{N}(0, W_t H_t W_t'), \tag{C.3}$$

and the first four expressions of the KF (3) are modified as follows:

$$\begin{aligned} v_t &= W_t(y_t - Z_t a_t), & F_t &= W_t(Z_t P_t Z_t' + H_t)W_t', \\ a_{t|t} &= a_t + P_t Z_t' W_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' W_t' F_t^{-1} W_t Z_t P_t. \end{aligned} \tag{C.4}$$

[4] A formal discussion of dealing with missing values in score-driven models can be found in Lucas et al. (2016).

The expressions in (7) become

$$\begin{aligned}\dot{V}_t &= -[(a_t' \otimes W_t)\dot{Z}_t + (a_{t-1|t-1}' \otimes W_t Z_t)\dot{T}_t], \\ \dot{F}_t &= 2N_{N_t}(W_t Z_t P_t \otimes W_t)\dot{Z}_t + (W_t Z_t \otimes W_t Z_t)2N_m(T_t P_{t-1|t-1} \otimes I_m)\dot{T}_t \\ &\quad +(W_t \otimes W_t)\dot{H}_t + (W_t Z_t \otimes W_t Z_t)\dot{Q}_t.\end{aligned} \tag{C.5}$$

Mixed frequencies typically involve missing observations and temporal aggregation. The case of mixed frequencies is of particular interest for a number of applications, like for instance forecasting low frequency variables using higher frequency predictors (nowcasting). Indeed low frequency indicators can be modeled as a latent process that is observed at regular low frequency intervals and missing at higher frequency dates. The relation between the observed low frequency variable and the corresponding (latent) higher frequency indicator depends on whether the variable is a flow or a stock and on how the variable is transformed before entering the model. In all cases, the variable can be rewritten as a weighted average of the unobserved high frequency indicator (see e.g., Banbura et al., 2013).

C.3 Correlated disturbances

Let consider the case in which the disturbances in (1) are correlated, that is $\mathrm{E}(\eta_t \epsilon_t') = G_t$. In this case the following expressions in (3) need to be modififed:

$$\begin{aligned}F_t &= Z_t P_t Z_t' + Z_t G_t + G_t' Z_t' + H_t, \\ a_{t|t} &= a_t + (P_t Z_t' + G_t)F_t^{-1} v_t, \\ P_{t|t} &= P_t - (P_t Z_t' + G_t)F_t^{-1}(P_t Z_t' + G_t)'.\end{aligned} \tag{C.6}$$

Therefore, the expression for $\dot{F}_t$ in (7) need to be modified as follows:

$$\begin{aligned}\dot{F}_t &= [2N_N(Z_t P_t \otimes I_N) + (G_t' \otimes I_N) + (I_N \otimes G_t')C_{Nm}]\dot{Z}_t \\ &\quad +(Z_t \otimes Z_t)2N_m(T_t P_{t-1|t-1} \otimes I_m)\dot{T}_t + \dot{H}_t + (Z_t \otimes Z_t)\dot{Q}_t \\ &\quad +[(I_N \otimes Z_t) + (Z_t \otimes I_N)C_{mN}]\dot{G}_t.\end{aligned} \tag{C.7}$$

C.4 Shrinking the vector of parameters by the L2 penalty

The state space model (1), with the score driven system matrices described by (4) is conditionally Gaussian with conditional likelihood equal to (2). We augment the model using the following Gaussian constraints:

$$r_t = R_t f_t + u_t \quad u_t \sim \mathcal{N}(0, \Sigma_t), \tag{C.8}$$

where r_t, Σ_t and R_t are known and the random disturbance u_t is Gaussian and uncorrelated with the other two disturbances ε_t and η_t. Since u_t is Gaussian, the resulting penalty will be quadratic with the matrix Σ_t regulating the degree of shrinkage; e.g for $\Sigma_t \to \infty$ the constraints

vanish, while for $\Sigma_t \to 0$ the constraints hold exactly. The vector r_t can be considered a vector of artificial observations. The likelihood function for the vector of 'augmented data', $\tilde{y}_t = (y_t', r_t')'$ is equal to:

$$\ell_t^p = \log p(\tilde{y}_t|Y_{t-1}, \theta) \propto -\frac{1}{2}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) - \frac{1}{2}\left(\log|\Sigma_t| + u_t'\Sigma_t^{-1}u_t\right), \tag{C.9}$$

which can be interpreted as a 'penalized likelihood' with a quadratic penalty function. The resulting 'penalized' score is:

$$s_t^p = (\mathcal{I}_t^p)^{-1}\nabla_t^p = (\mathcal{I}_t + R_t'\Sigma_t^{-1}R_t)^{-1}(\nabla_t + R_t'\Sigma_t^{-1}u_t). \tag{C.10}$$

Using the Woodbury identity we obtain[5]:

$$(\mathcal{I}_t^p)^{-1} = (\mathcal{I}_t + R_t'\Sigma_t^{-1}R_t)^{-1} = (\mathcal{I}_t^{-1} - \mathcal{I}_t^{-1}R_t'(R_t\mathcal{I}_t^{-1}R_t' + \Sigma_t)^{-1}R_t\mathcal{I}_t^{-1}) = (I - \Upsilon_t R_t)\mathcal{I}_t^{-1}, \tag{C.11}$$

where $\Upsilon_t = \mathcal{I}_t^{-1}R_t'(R_t\mathcal{I}_t^{-1}R_t' + \Sigma_t)^{-1}$. Finally, we can express the penalized (regularized) score as follows:

$$s_t^p = (I - \Upsilon_t R_t)s_t + \Upsilon_t u_t, \tag{C.12}$$

From last expression it is straightforward to obtain the two polar cases for $\Sigma_t \to 0$ and $\Sigma_t \to \infty$. To illustrate with a simple example how such 'penalized score' works let us $R_t = I$, $r_t = \bar{f}$, and $\Sigma_t = \frac{1}{\lambda}I$ so that the constraint (10) reduces to $f_t \sim \mathcal{N}(\bar{f}, \frac{1}{\lambda}I)$. This corresponds to a Ridge-type penalty with λ regulating the degree of penalization. The penalized score will be $s_t^p = (I - \Lambda_t)s_t + \Lambda_t(\bar{f} - f_t)$, where $\Lambda_t = \lambda(\mathcal{I}_t + \lambda I)^{-1}$. If we assume a simple random-walk law of motion for f_t and information matrix equal to identity matrix, the resulting regularized score-driven filter is:[6]

$$f_{t+1} = \frac{\mathrm{B}\lambda}{1+\lambda}\bar{f} + \left(I - \frac{\mathrm{B}\lambda}{1+\lambda}\right)f_t + \frac{\mathrm{B}}{1+\lambda}s_t. \tag{C.13}$$

The law of motion is now mean reverting towards the desired value $\bar{f}$. Everything else equal, the larger is λ, the lower is the weight assigned to actual data and the more binding is the constraint. Notice that the strategy of stochastic constraints is very flexible, and generalizes a number of shrinkage methods, including Ridge regressions, Minnesota priors, sum of coefficients priors as well as the long-run prior in Giannone et al. (2019). Kapetanios et al. (2019) discuss in details these additional cases.

[5] In case $\mathcal{I}_t$ is not invertible we use its smoothed estimator $\tilde{\mathcal{I}}_t = (1-\kappa)\tilde{\mathcal{I}}_{t-1} + \kappa\mathcal{I}_t$ which is invertible.

[6] The same regularized filter can be obtained by setting $f_t \sim \mathcal{N}(\bar{f}, \frac{1}{\lambda}\mathcal{I}_t)$, which is the Litterman-type of prior.

D Monte Carlo exercise

We start from a battery of simple bivariate models that share a common component and study the performance of our method in tracking time variation of different subsets of parameters. We simulate two time series ($y_{1,t}$ and $y_{2,t}$) that load (with parameters $\lambda_{1,t}$ and $\lambda_{2,t}$) on a common factor μ_t and are each affected by idiosyncratic noise. The common factor evolves as an AR(1), with coefficient ρ_t. In this model, we look at the time variation of a subset of parameters one at the time. Specifically, DGP1 lets the loading on the common factor vary over time and keeps all the remaining parameters fixed. In DGP2 we keep both factor loadings constant ($\lambda_{1,t} = 1$ and $\lambda_{2,t} = 1$) while allowing for time variation in the AR coefficient of the common factor, ρ_t. In DGP3 and DGP4 we experiment with time-varying volatility, either in the measurement or in the transition equations, keeping everything else fixed. In these latter cases, the simulated model is a univariate signal plus noise model. In all cases, we consider two different sample sizes, $n = 250$ and $n = 500$. As for the laws of motions for the TVPs entering the various DGPs, we experiment with 6 different possibilities:

Case 1: CONSTANT $f_t = a_1, \forall t;$
Case 2: SINE $f_t = a_2 + b_2 \sin\left(\frac{2\pi t}{T/2}\right);$
Case 3: SINGLE STEP $f_t = a_3 + b_3 \left(t \geq \tau\right);$
Case 4: DOUBLE STEP $f_t = a_4 + b_4 I\left(t \geq \tau_1\right) + c_4 I\left(t \geq \tau_2\right);$
Case 5: RAMP $f_t = a_5 + \left(\frac{b_5}{T/c_5}\right) \mod (t);$
Case 6: AR(1) MODEL $f_t = a_6(1 - b_6) + b_6 f_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{N}\left(0, c_6\right);$

where $f_t = \lambda_t$ in DGP1, $f_t = \rho_t$ in DGP2, $f_t = \sigma^2_{\varepsilon,t}$ in DGP3, and $f_t = \sigma^2_{u,t}$ in DGP4.[7] We start with a baseline case in which we keep the parameters constant over time. We then move to four cases where the parameters change according to a deterministic process. In case 2 the parameters follow a cyclical pattern determined by a sine function. In cases 3 and 4 we let the parameters break at discrete points in time, allowing for either one or two breaks. We set the location of the discrete breaks at given point in the sample. In the case of a single break, this occurs in the middle of the sample. When we consider two breaks, they are located at 1/3 and 2/3 of the sample. Case 5 (RAMP) is a rather challenging case, whereby the parameters increase for some time before returning abruptly to their starting levels. Finally, case 6 is the only one in which we let the parameters vary stochastically, following a persistent AR(1) model. We consider two cases, one with a near unit root process (i.e. with an AR root of 0.99) and a low variance, one with lower persistence (AR root of 0.97) but substantially higher variance. We obtain very similar results in these two specifications. The DGPs that we design

[7] Moreover, f_t in the AR(1) model is transformed to be within the unit circle for DGP2, and to be positive for DGP3 and DGP4.

are simple, in that time variation is introduced in all the channels in which it can manifest itself, but only one at the time. By focusing on a single channel at the time, we can better identify the situations in which our model either succeeds or fails at tracking parameter variation. Further details on the calibration of the parameters are provided in the next subsections.

The results of the Monte Carlo exercise are shown in Table D.1. The table is organized in four panels, corresponding to each of the four DGPs. On the left hand side of the table we show the results for a sample size of $n = 250$, on the right hand side those obtained when setting $n = 500$. For each DGP the analysis is based on 300 simulations. In each panel we report the results obtained for the six alternative laws of motion described above. We base our assessment on five different statistics, namely the Root Mean Squared Error (RMSEs), the Mean Absolute Error (MAE), the correlation between actual and estimated coefficients, the Coverage (i.e. percentage of times that the actual parameters fall in a given estimated confidence interval) and the number of cases in which a pile-up occurs (#Pile-up). The last statistics consists of the number of simulations in which the static coefficients that pre-multiply the score end up being lower than 10^{-6}, which we take as sufficient evidence that the estimated parameters are effectively zero, i.e. that the model does not detect any time variation.

We highlight five results. First, for all the DGPs in which the true parameters are constant the model performs extremely well. This means that the adaptive filter correctly collapses the parameters to a constant. As a result, RMSEs and MAEs are virtually nil, the actual coverage extremely precise and a pile-up at zero occurs in about 75 percent of the cases for the volatility models and more than half of the cases for the loadings and AR coefficients.[8] This result implies that our estimation method does not generate spurious time variation in the coefficients when this is not present in the data generating process. Second, the pile-up problem is not of primary concern for our estimator. The number of instances in which our method (incorrectly) concludes that there is no time variation is basically zero in most cases. Third, for all the DGPs and across all the specifications for the parameters we obtain extremely good coverage. Coverage is almost perfect when time variation involves the autoregressive coefficients, somewhat lower when it affects the volatility of the measurement and of the transition equation, in particular when parameters break at discrete points in the sample. Fourth, across all DGPs the RAMP specification is the one that the model finds more challenging to estimate. This specification generally leads to low correlation between actual and estimated parameters, higher RMSEs and MAEs and lower coverage. This is not surprising, since our model is, by construction, designed to detect smooth changes, whereas in the RAMP model periods of small, continuous changes are suddenly interrupted by large breaks. Fifth, the adaptive filter is very effective in estimating time-varying loadings and autoregressive coefficients, but it is rather conservative

[8] For the latter two cases, in an additional 20% of the simulations the estimated parameters are virtually constant, despite not being classified as a pile-up according to the criterion we have set above.

in the estimation of the time-varying variances, especially when these are driven by a near unit root process. For this DGP, in one third of the cases the filter ends in a pile-up when the sample is relatively small (T=250). However, when time variation *is* detected, the algorithm yields relatively low RMSEs and MAEs and a satisfactory coverage. We take these results as evidence that, in the case of time-varying variances, the algorithm needs relatively more evidence of breaks in the parameters to move away from zero. A larger sample size (of the type encountered in financial applications that use high frequency data) basically eliminates the problem.

D.1 Specification of the DGPs

DGP1 - Time-Varying loadings

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_t \end{bmatrix} \mu_t + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim \mathcal{N}(0, I),$$
$$\mu_t = 0.8\mu_{t-1} + u_t \qquad u_t, \sim \mathcal{N}(0, 1).$$

DGP2 - Time-Varying AR coefficient

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu_t + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \sim \mathcal{N}(0, I),$$
$$\mu_t = \rho_t \mu_{t-1} + u_t, \qquad u_t \sim \mathcal{N}(0, 1).$$

DGP3 - Time-Varying Volatility in the measurement equation

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2_{\varepsilon,t}), \\ \mu_{t+1} &= 0.8\mu_t + u_t, \quad u_t \sim \mathcal{N}(0, 1). \end{aligned}$$

DGP4 - Time-Varying Volatility in the transition equation

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \\ \mu_{t+1} &= 0.8\mu_t + u_t, \quad u_t \sim \mathcal{N}(0, \sigma^2_{\eta,t}). \end{aligned}$$

D.2 Calibration

DGP1:Time-varying loadings

CONSTANT: $a_1 = 1$;

SINE: $a_2 = 2$, $b_2 = 1.5$;

SINGLE STEP: $a_3 = 1$, $b_3 = 2$, $\tau = (2/5)n$;

DOUBLE STEP: $a_4 = 1$, $b_4 = c_4 = 1.5$, $\tau_1 = (1/5)n$, $\tau_2 = (3/5)n$;

RAMP: $a_5 = 0.5$, $b_5 = 4$, $c_5 = 2$;

AR(1) $[b_6 = 0.99]$: $a_6 = 1$, $b_6 = 0.99$, $c_6 = 0.08^2$.

AR(1) $[b_6 = 0.97]$: $a_6 = 1$, $b_6 = 0.97$, $c_6 = 30.24^2$.

DGP2: Time-varying autoregressive coefficient

CONSTANT: $a_1 = 0.7$;

SINE: $a_2 = 0$, $b_2 = 0.7$;

SINGLE STEP $a_3 = 0.8$, $b_3 = -0.6$, $\tau = (2/5)n$;

DOUBLE STEP:$a_4 = 0.8$, $b_4 = c_4 = -0.5$, $\tau_1 = (1/5)n$, $\tau_2 = (3/5)n$;

RAMP: $a_5 = 0.3$, $b_5 = -0.9$. $c_5 = 2$;

AR(1) $[b_6 = 0.99]$: $a_6 = 0.2$, $b_6 = 0.99$, $c_6 = 0.08^2$;

AR(1) $[b_6 = 0.97]$: $a_6 = 0.2$, $b_6 = 0.97$, $c_6 = 0.24^2$;

and in the latter two cases we also impose the restriction that $|\rho_t| < 1$.

DGP3 and DGP4: Time-varying volatilities

CONSTANT: $a_1 = 1$;

SINE: $a_2 = 1$, $b_2 = 0.9$;

SINGLE STEP: $a_3 = 1$, $b_3 = 4$, $\tau = (2/5)n$;

DOUBLE STEP:$a_4 = 1$, $b_4 = c_4 = 3$, $\tau_1 = (1/5)n$, $\tau_2 = (3/5)n$;

RAMP: $a_5 = 0.5$, $b_5 = 8$, $c_5 = 2$;

AR(1) $[b_6 = 0.99]$: $a_6 = 0$, $b_6 = 0.99$, $c_6 = 0.08^2$;

AR(1) $[b_6 = 0.97]$: $a_6 = 0$, $b_6 = 0.97$, $c_6 = 0.24^2$;

In DGP3 and DGP4, after having simulated the dynamic of the volatility the time-varying volatilities are rescaled so as to have a fixed ratio between the measurement and transition error variances equal to 1.

For each DGP we target 300 simulations. However, the actual number of samples changes depending on the specifications. In the case of constant coefficients, where we would like to see our estimator to end up in a pile-up situation as often as possible, we perform 300 simulations and compute all the statistics on these samples. For the remaining specifications, on the other hand, we keep on drawing artificial samples until we obtain 300 simulations in which the estimated parameters are different from zero and compute RMSEs, MAEs, correlations and coverage ratios on these 300 artificial samples. At the same time, we also keep track of the number of times in which the pile-up problem arises. To better understand how we proceed, let us take a concrete example, that is the bottom-left panel of Table 1 (DGP4, i.e. the model with time-varying volatility in the transition equation, $n = 250$). In the first row we report the results for the constant coefficient case. As explained, for this case we simulate 300 artificial samples and estimate the model using our algorithm. It turns out that in 236 out of 300 simulations our estimation method ends up in a pile-up. The RMSEs, MAEs, Correlations and Coverages, are estimated on all the 300 simulations. Now let us take in the same panel the

last line, referring to one of the AR(1) specifications. In this case we need to draw up to 314 samples to obtain 300 simulations in which the estimation algorithm does not end being stuck in a region of the likelihood where the model loading is zero. Now, in this case all the remaining statistics are computed on the 300 'good' samples. We proceed in this way because we want to appraise two different issues. The former is the percentage of cases in which the algorithm ends up in the pile-up even if the true DGP implies time variation. The second is how well it estimates the parameters *conditional on the model correctly detecting time variation*. The two points are of independent interest. If we were to find that the model often ends up in the pile-up but it is very precise when it does not, one could decide to force the algorithm to stay away from zero, for instance by using a grid-based estimation method. This is the choice made, for instance, by Koop and Korobilis (2013). Similarly, in their Monte Carlo Markov Chain estimation, Stock and Watson (2007) reject draws in which the variances are very close to zero.

In figures D.1-D.8 we report the simulated true process for the time-varying parameters (red line), and the fan chart associated to the $5^{th}, 10^{th}, 20^{th}, 30^{th}, 40^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th}$ and 95^{th} quantile of the filtered parameters. In the case of the AR(1) specification we focus on the more persistent AR(1) DGP and report the difference between actual and estimated parameters. The figures are based on 300 replications.

Table D.1: Monte Carlo Exercise

	T = 250						T = 500					
DGP 1: time-varying LOADINGS COEFFICIENT												
	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up
CONSTANT	0.003	0.003		0.678	0.900	165	0.000	0.000		0.680	0.900	180
SINE	0.473	0.380	0.909	0.636	0.852	0	0.386	0.305	0.940	0.648	0.868	0
SINGLE STEP	0.406	0.280	0.927	0.656	0.876	0	0.335	0.229	0.951	0.660	0.882	0
DOUBLE STEP	0.462	0.339	0.936	0.640	0.860	0	0.390	0.277	0.953	0.652	0.874	0
RAMP	0.695	0.461	0.723	0.648	0.856	0	0.575	0.367	0.817	0.658	0.870	0
AR(1) $[b_6 = 0.99]$	0.265	0.213	0.727	0.676	0.892	0	0.274	0.217	0.807	0.676	0.892	0
AR(1) $[b_6 = 0.97]$	0.523	0.415	0.803	0.660	0.872	0	0.527	0.413	0.828	0.662	0.872	0
DGP 2: time-varying AUTOREGRESSIVE COEFFICIENT												
	T = 250						T = 500					
	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up
CONSTANT	0.006	0.006		0.676	0.900	147	0.005	0.004		0.680	0.900	156
SINE	0.330	0.267	0.780	0.684	0.900	0	0.268	0.212	0.866	0.682	0.900	0
SINGLE STEP	0.228	0.166	0.769	0.684	0.900	0	0.203	0.140	0.811	0.686	0.902	0
DOUBLE STEP	0.240	0.185	0.872	0.684	0.900	0	0.209	0.160	0.892	0.686	0.900	0
RAMP	0.341	0.261	0.392	0.682	0.900	0	0.299	0.221	0.548	0.683	0.900	0
AR(1) $[b_6 = 0.99]$	0.297	0.237	0.608	0.684	0.900	4	0.301	0.241	0.695	0.686	0.902	1
AR(1) $[b_6 = 0.97]$	0.477	0.377	0.575	0.686	0.900	0	0.478	0.375	0.593	0.685	0.900	0
DGP 3: TIME-VARYING VOLATILITY - MEASUREMENT EQUATION ERROR												
	T = 250						T = 500					
	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up
CONSTANT	0.000	0.000		0.676	0.896	231	0.000	0.000		0.682	0.899	232
SINE	0.981	0.768	0.747	0.672	0.876	1	0.829	0.637	0.813	0.678	0.882	0
SINGLE STEP	0.808	0.605	0.843	0.618	0.848	0	0.659	0.477	0.883	0.652	0.870	0
DOUBLE STEP	0.702	0.551	0.856	0.628	0.848	0	0.595	0.460	0.889	0.648	0.870	0
RAMP	0.960	0.764	0.498	0.640	0.860	20	0.803	0.599	0.646	0.656	0.874	1
AR(1) $[b_6 = 0.99]$	0.717	0.571	0.568	0.664	0.880	93	0.748	0.578	0.608	0.668	0.886	24
AR(1) $[b_6 = 0.97]$	1.446	0.998	0.600	0.664	0.868	22	1.489	1.005	0.626	0.674	0.878	3
DGP 4: TIME-VARYING VOLATILITY - TRANSITION EQUATION ERROR												
	T = 250						T = 500					
	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up
CONSTANT	0.000	0.000		0.676	0.896	236	0.000	0.000		0.680	0.898	241
SINE	1.052	0.832	0.714	0.672	0.880	1	0.871	0.673	0.794	0.672	0.886	0
SINGLE STEP	0.829	0.614	0.834	0.644	0.864	0	0.680	0.485	0.874	0.656	0.878	0
DOUBLE STEP	0.754	0.592	0.849	0.644	0.868	0	0.620	0.481	0.885	0.656	0.876	0
RAMP	1.015	0.822	0.468	0.644	0.868	1	0.829	0.640	0.615	0.659	0.881	0
AR(1) $[b_6 = 0.99]$	0.768	0.623	0.622	0.668	0.888	95	0.776	0.607	0.613	0.668	0.887	34
AR(1) $[b_6 = 0.97]$	1.533	1.069	0.590	0.664	0.876	14	1.523	1.042	0.619	0.664	0.880	5

Note. The results shown in the first and in the second panel (DGP1 and DGP2) refer to a bivariate factor model in which two variables are driven by a single common factor that evolves as an autoregressive process of order 1. In the first case (DGP1) the loading of the second variable on the common factor varies over time and all the other parameters are kept constant. In the second case (DGP2) the autoregressive component of the common factor varies over time and all the other parameters are kept constant. The results shown in the third and in the fourth panel (DGP3 and DGP4) refer to ARMA(1,1) models that are cast in state space and feature time-varying variances of the random disturbance in, respectively, the measurement and the transition equation. The abbreviations *Corr.* and *Cov.* stand, respectively for Correlation and Coverage, while # Pile-up denotes the number of simulations in which the algorithm delivers constant parameters. The different laws of motion of the parameters in the first column (Constant, Sine, Single Step, Double Step, Ramp and AR(1) are described in Section 4).

Figure D.1: TIME-VARYING LOADINGS, N=250

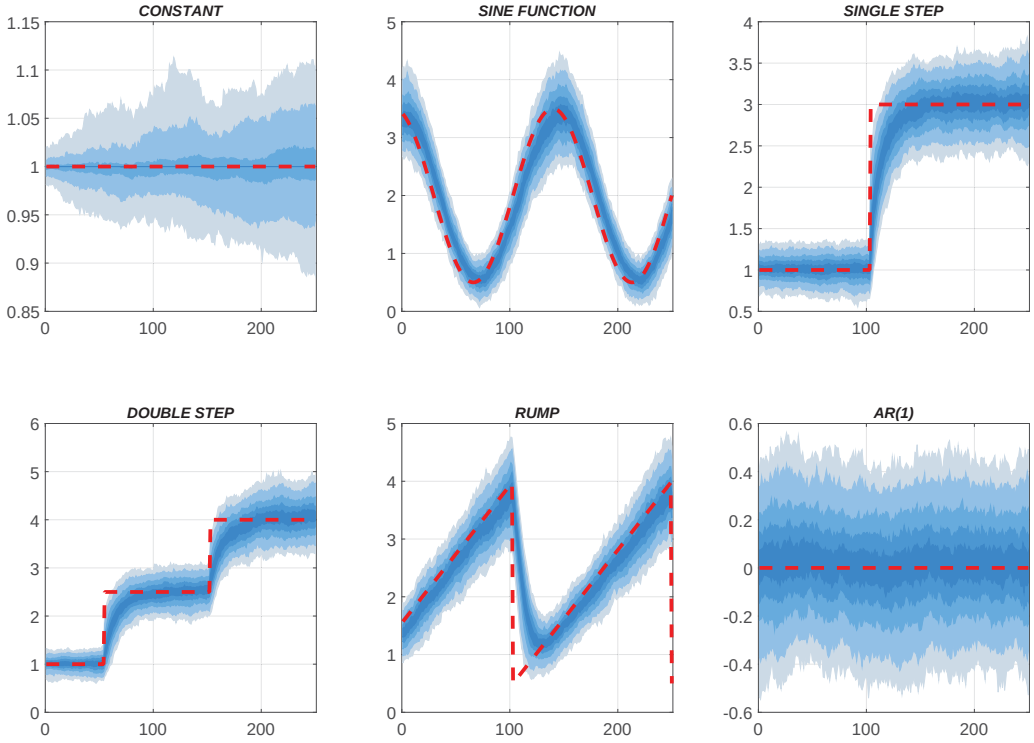


Figure D.2: TIME-VARYING LOADINGS, N=500

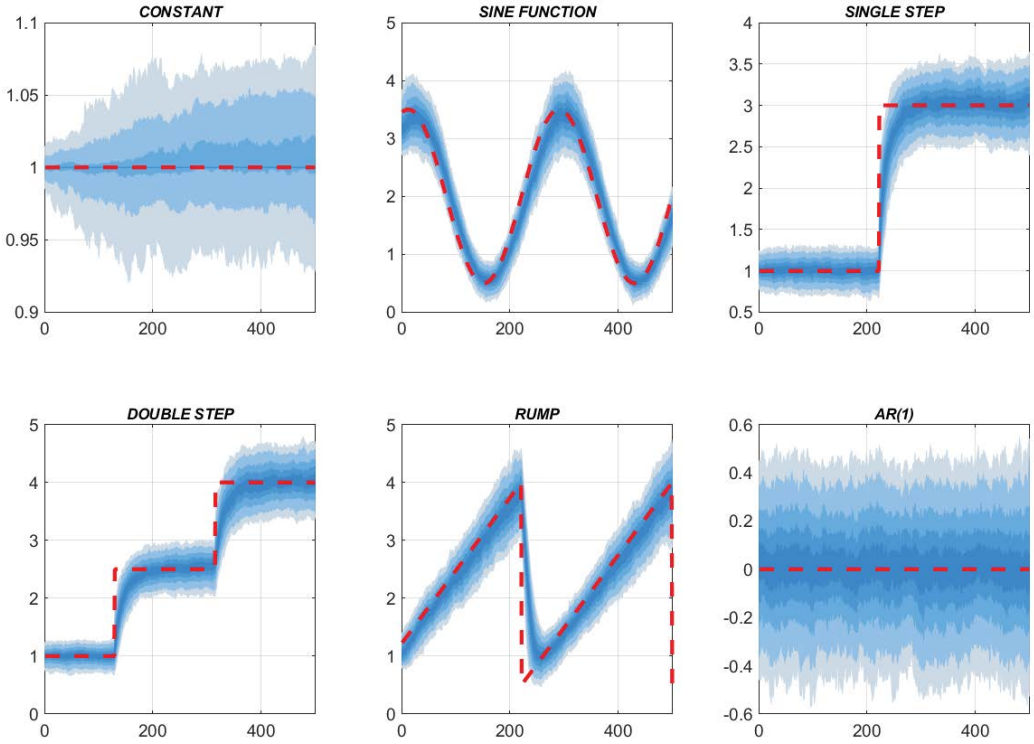


Figure D.3: TIME-VARYING AUTOREGRESSIVE COEFFICIENTS, n=250

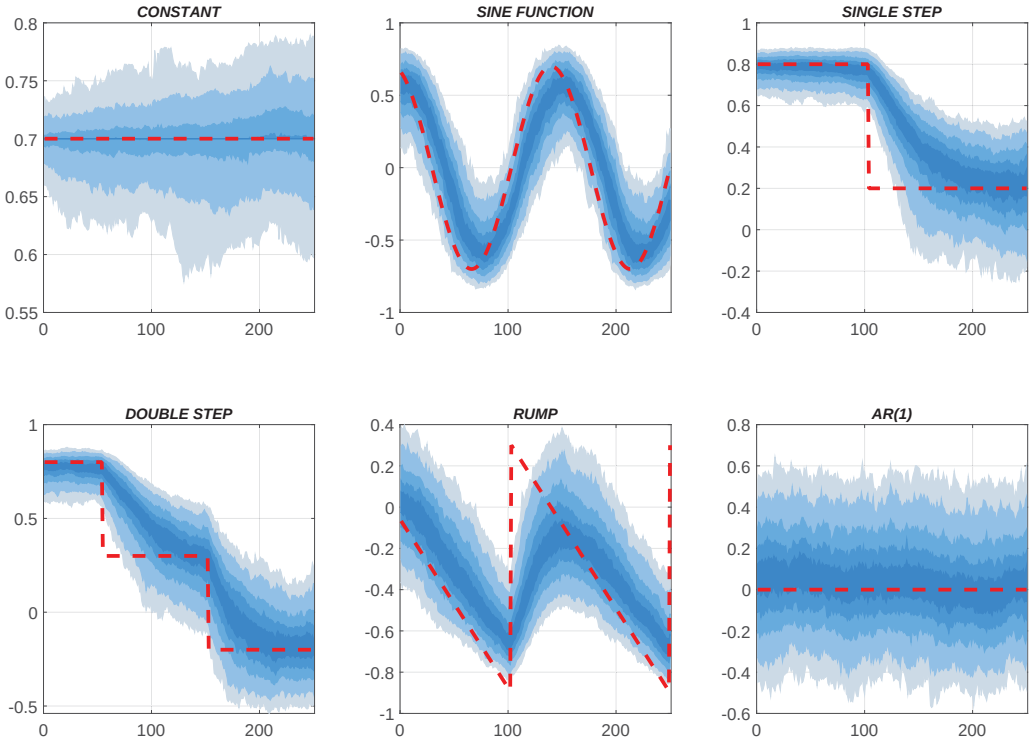


Figure D.4: TIME-VARYING AUTOREGRESSIVE COEFFICIENTS, n=500

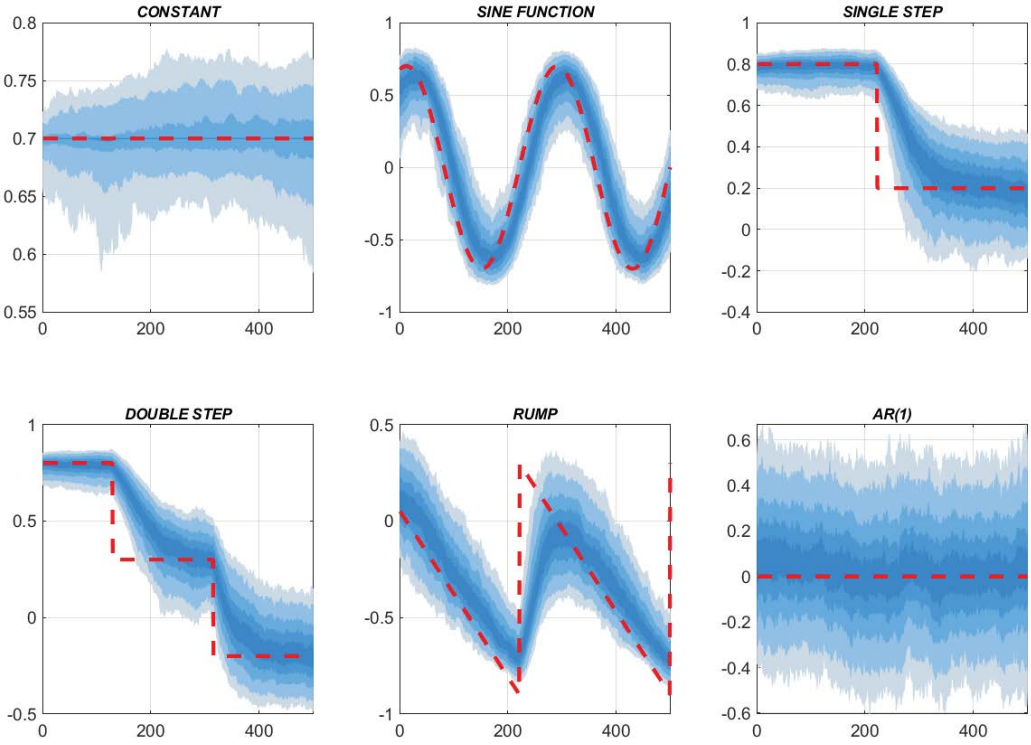


Figure D.5: TIME-VARYING MEASUREMENT EQUATION ERROR VARIANCE, n=250

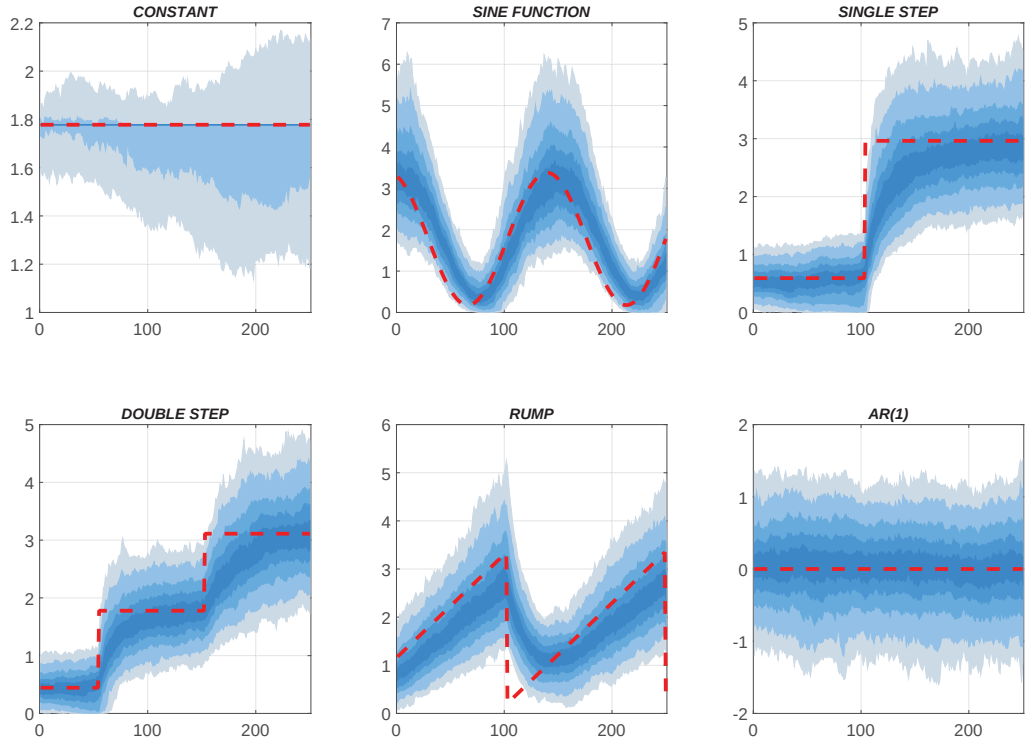


Figure D.6: TIME-VARYING MEASUREMENT EQUATION ERROR VARIANCE, n=500

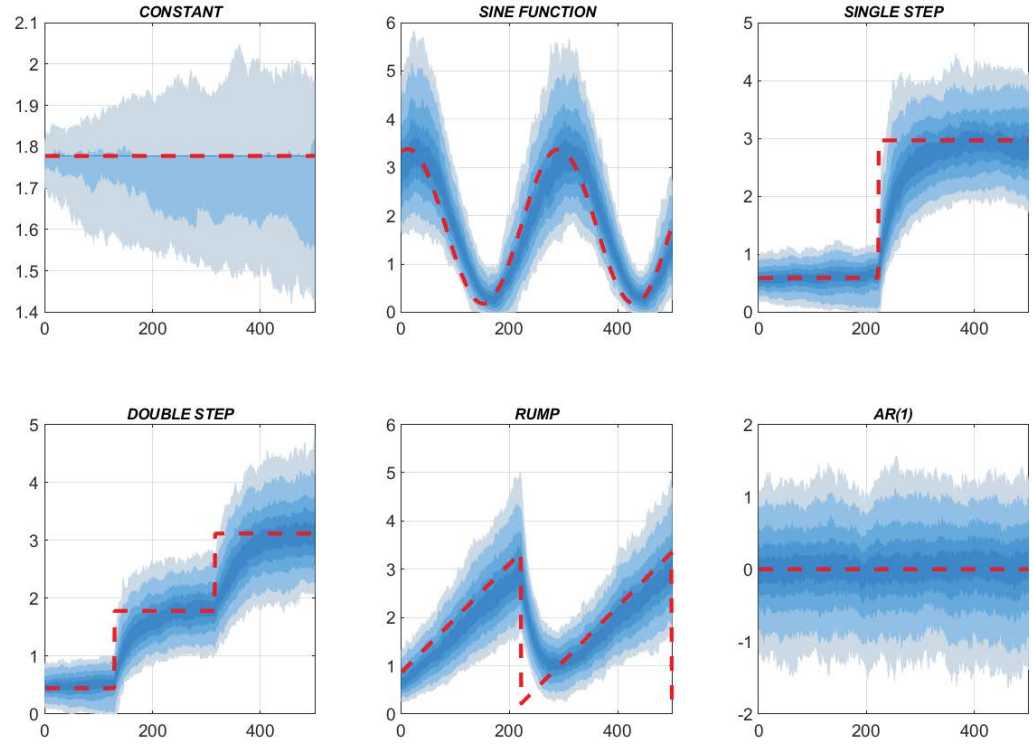


Figure D.7: TIME-VARYING TRANSITION EQUATION ERROR VARIANCE, n=250

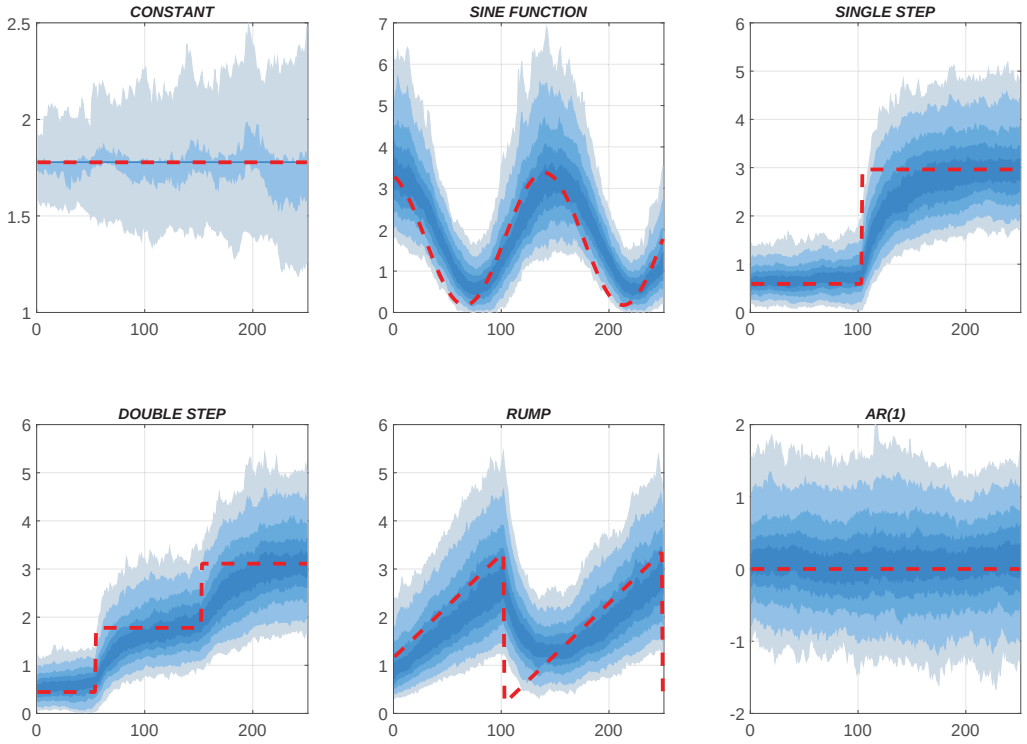
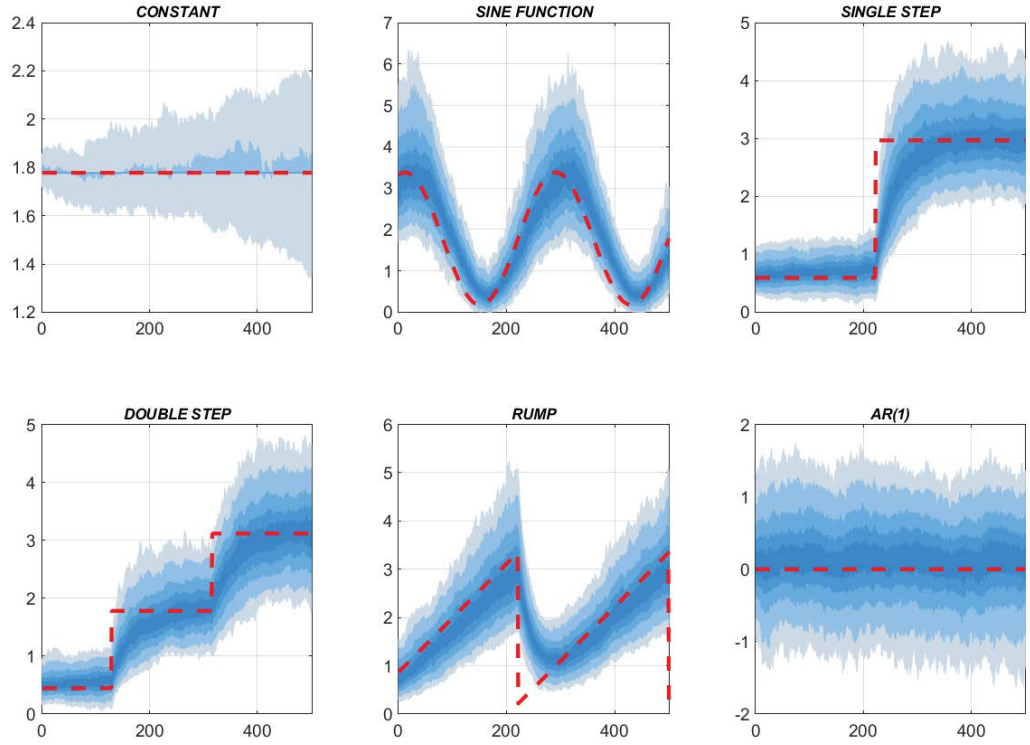


Figure D.8: TIME-VARYING TRANSITION EQUATION ERROR VARIANCE, n=500



E Additional details on the Empirical Application

To simplify the notation, in this section any generic score driven time-varying parameter, $x_{t|t-1}$, is simply denoted with x_t.

E.1 State state representation and Jacobians of the system matrices

State Space representation. The state space representation of the model is:

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, H), \\ \alpha_t &= T\alpha_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t), \end{aligned}$$

where

$$y_t = \begin{bmatrix} \Delta d_t \\ \mathtt{pd}_t \end{bmatrix}, \quad Z_t = \begin{bmatrix} \bar{g}_t & 0 & 0 & 1 & 1 & 0 & 0 \\ \overline{\mathtt{pd}}_t & \frac{1}{1-\rho_t\phi_g} & -\frac{1}{1-\rho_t\phi_\mu} & 0 & 0 & 0 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\nu^2 \end{bmatrix},$$

$$\alpha_t = \begin{bmatrix} 1 \\ \tilde{g}_t, \\ \tilde{\mu}_t, \\ \tilde{g}_{t-1} \\ \varepsilon_{d,t} \\ \varepsilon_{g,t} \\ \varepsilon_{\mu,t} \end{bmatrix}, \quad T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_g & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_\mu & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \eta_t = S_\eta \begin{bmatrix} \varepsilon_{d,t} \\ \varepsilon_{g,t} \\ \varepsilon_{\mu,t} \end{bmatrix}, \quad S_\eta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus, we have that $Q_t = S_\eta \Omega_t S_\eta'$, $\Omega_t = D_t R_t D_t$, where $D_t = \text{diag}[\sigma_{d,t}, \sigma_{g,t}, \sigma_{\mu,t}]$ contains the standard deviations, and R_t denotes the correlation matrix. Not all the parameters of the correlation matrix can be identified (see Section 4.3.1 and Appendix E.2 for additional details). The zero correlation between the measurement error in dividend growth ($\varepsilon_{d,t}$) and the stochastic disturbance in expected dividend growth ($\varepsilon_{g,t}$), which is required for the identification of the model, is appropriately imposed and the resulting correlations matrix is:

$$R_t = \begin{bmatrix} 1 & 0 & \varrho_{d\mu,t} \\ 0 & 1 & \varrho_{g\mu,t} \\ \varrho_{d\mu,t} & \varrho_{g\mu,t} & 1 \end{bmatrix}.$$

Recall that to have positive variances we model the log standard deviations σ_i, and for a well defined correlation matrix we model a transformation of the partial correlations π_{ij}. For more details on the function mapping the correlations into the partial correlations see section E.3. Therefore, the system matrices Z_t and Q_t are function of the vector of time-varying parameters $f_t = (\varphi_t', \delta_t', \gamma_t')'$, where $\varphi_t = (\overline{\mu}_t, \overline{g}_t)'$, $\delta_t = (\log\sigma_{d,t}, \log\sigma_{g,t}, \log\sigma_{\mu,t})'$, $\gamma_t = (\pi_{d\mu,t}, \pi_{g\mu,t})'$. The

vector f_t follows the score driven model discussed in section 2, with the following specification of the static parameters: $\mathrm{c} = (0, 0, \mathrm{c}_{\delta_1}, \mathrm{c}_{\delta_2}, \mathrm{c}_{\delta_3}, \mathrm{c}_{\gamma_1}, \mathrm{c}_{\gamma_2})'$, $\mathrm{A} = \mathrm{diag}[1, 1, a_{\delta_1}, a_{\delta_2}, a_{\delta_3}, a_{\gamma_1}, a_{\gamma_2}]$, $\mathrm{B} = \mathrm{diag}[\kappa_\mu, \kappa_g, \kappa_{\delta_1}, \kappa_{\delta_2}, \kappa_{\delta_3}, \kappa_{\gamma_1}, \kappa_{\gamma_2}]$. In the remainder of this section we discuss the parametrization of the time-varying system matrices and the associated Jacobians that are required by the score driven algorithm introduced in section 2.

Time variation in the Z matrix. Using the notation in section 2.2 we have that

$$\mathrm{vec}(Z_{t|t-1}) = S_{0,z} + S_{1,z}\psi_z\left(S_{2,z}f_{t|t-1}\right)$$

where

$$S_{0,z} = \begin{bmatrix} 0_{6\times1} \\ 1 \\ 0 \\ 1 \\ 0_{5\times1} \end{bmatrix}, \quad S_{1,z} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ & 0_{8\times4} & & \end{bmatrix}, \quad S'_{2,z} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

$$\psi_z(\varphi_t) = \begin{bmatrix} \overline{g}_t \\ \overline{\mathtt{pd}}_t \\ \frac{1}{1-\rho_t\phi_g} \\ -\frac{1}{1-\rho_t\phi_\mu} \end{bmatrix}, \quad \overline{\mathtt{pd}}_t = \overline{g}_t - \log(\exp\overline{\mu}_t - \exp\overline{g}_t), \quad \rho_t = \frac{\exp\overline{\mathtt{pd}}_t}{1+\exp\overline{\mathtt{pd}}_t}.$$

Finally, we have the following Jacobian:

$$\dot{Z}_t = S_{1,z}\Psi_{z,t}S_{2,z}, \quad \Psi_{z,t} = \begin{bmatrix} 0 & 1 \\ \frac{\partial\overline{\mathtt{pd}}_t}{\partial\overline{\mu}_t} & \frac{\partial\overline{\mathtt{pd}}_t}{\partial\overline{g}_t} \\ \frac{\phi_g}{(1-\phi_g\rho_t)^2}\frac{\partial\rho_t}{\partial\overline{\mu}_t} & \frac{\phi_g}{(1-\phi_g\rho_t)^2}\frac{\partial\rho_t}{\partial\overline{g}_t} \\ -\frac{\phi_\mu}{(1-\phi_\mu\rho_t)^2}\frac{\partial\rho_t}{\partial\overline{\mu}_t} & -\frac{\phi_\mu}{(1-\phi_\mu\rho_t)^2}\frac{\partial\rho_t}{\partial\overline{g}_t} \end{bmatrix}, \quad \begin{array}{l} \frac{\partial\overline{\mathtt{pd}}_t}{\partial\overline{\mu}_t} = -\frac{\exp\overline{\mu}_t}{\exp\overline{\mu}_t-\exp\overline{g}_t}, \\ \frac{\partial\overline{\mathtt{pd}}_t}{\partial\overline{g}_t} = -\frac{\partial\overline{\mathtt{pd}}_t}{\partial\overline{\mu}_t}, \\ \frac{\partial\rho_t}{\partial\overline{\mu}_t} = -\frac{\rho_t(1-\rho_t)\exp\overline{\mu}_t}{\exp\overline{\mu}_t-\exp\overline{g}_t}, \\ \frac{\partial\rho_t}{\partial\overline{g}_t} = -\frac{\partial\rho_t}{\partial\overline{\mu}_t}. \end{array}$$

Time variation in the volatility and correlations. Recall that the covariance matrix of the transition equation is $Q_t = S_\eta\Omega_t S'_\eta$ where $\Omega_t = D_tR_tD_t$. Using the notation in Section 2.2, and the standard rules of matrix differentiation, we have that:

$$\begin{aligned} \dot{Q}_t &= (S_\eta \otimes S_\eta)\dot{\Omega}_t \\ &= (S_\eta \otimes S_\eta)\left[\frac{\partial\mathrm{vec}(\Omega_t)}{\partial\mathrm{vec}(D_t)'}\dot{D}_t + \frac{\partial\mathrm{vec}(\Omega_t)}{\partial\mathrm{vec}(R_t)'}\dot{R}_t\right] \\ &= (S_\eta \otimes S_\eta)\left[(D_tR_t \otimes I + I \otimes D_tR_t)\dot{D}_t + (D_t \otimes D_t)\dot{R}_t\right]. \end{aligned}$$

We now express the matrices of volatilities and correlations as follows:

$$\text{vec}(D_t) = S_{1,d}\psi_d\left(S_{2,d}f_t\right), \quad \text{vec}(R_t) = S_{0,r} + S_{1,r}\psi_r\left(S_{2,r}f_t\right),$$

where $S_{1,d}$ $S_{2,d}$, $S_{1,r}$, $S_{2,r}$ are selection matrices

$$S_{1,d} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad S'_{2,d} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad S_{0,r} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad S_{1,r} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad S'_{2,r} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The functions $\psi_d(\delta_t)$ and $\psi_r(\gamma_t)$ and their Jacobians are described in section E.3. Specifically, we have that:

$$\dot{D}_t = S_{1,d}\Psi_{d,t}S_{2,d}, \quad \dot{R}_t = S_{1,r}\Psi_{r,t}S_{2,r},$$

where

$$\Psi_{d,t} = D_t, \quad \Psi_{r,t} = \sqrt{1-\pi^2_{d\mu,t}}\begin{bmatrix} \sqrt{1-\pi^2_{d\mu,t}} & 0 \\ -\pi_{d\mu,t}\pi_{g\mu,t} & 1-\pi^2_{g\mu,t} \end{bmatrix}.$$

E.2 Model Identification

In this section, we consider the identification of the model. Let us start by looking at the constant parameter version of the model:

$$\begin{aligned} \texttt{pd}_{t+1} &= \overline{\texttt{pd}} - b_1\tilde{\mu}_{t+1} + b_2\tilde{g}_{t+1} + \nu_{t+1}, \qquad \nu_{t+1} \sim \mathcal{N}(0,\sigma^2_\nu), \\ \Delta d_{t+1} &= \overline{g} + \tilde{g}_t + \varepsilon_{d,t+1}, \\ \tilde{\mu}_{t+1} &= \phi_\mu\tilde{\mu}_t + \varepsilon_{\mu,t+1}, \\ \tilde{g}_{t+1} &= \phi_g\tilde{g}_t + \varepsilon_{g,t+1}. \end{aligned} \tag{E.1}$$

The measurement error ν_t is uncorrelated with the innovations of the model for which we assume a general a covariance structure, $\varepsilon_t = (\varepsilon_{d,t}, \varepsilon_{\mu,t}, \varepsilon_{g,t})' \sim \mathcal{N}(0,\Omega)$. Below we discuss the restrictions needed for this model to be identified.

Model (E.1) is equivalent to one estimated by Binsbergen and Koijen (2010), whose identification issue is discussed at lenght in Cochrane (2008a), with the key difference being that we have added the measurement error in the $\texttt{pd}_t$ equation. To be more precise, the specification for Δd_{t+1} implies an ARMA(1,1) process, while the model for $\texttt{pd}_t$ without the measurement

error ν_t is an ARMA(2,1). The resulting bivariate model is a restricted VARMA(2,1) with five parameters[9] to identify the covariance Ω. Specifically, we set one zero correlation as in Binsbergen and Koijen (2010) and Rytchkov (2012), i.e. $\text{Corr}(\varepsilon_{d,t}, \varepsilon_{g,t}) = 0$. Adding the measurement error ν_t in the $\texttt{pd}_t$ equation, the additional parameter σ^2_ν is identified by the additional moving average coefficients.[10] By introducing time-varying long-run mean $\overline{\texttt{pd}}_t$, and $\overline{g}_t$, the implied reduced form models for $\texttt{pd}_t$ and Δd_t become ARIMA(2,1,3) and ARIMA(1,1,2), respectively. Thus, the two additional moving average coefficients are used to identify the two parameters, b_μ and b_g, scaling the score-driven filters for $\overline{\mu}_t$ and $\overline{g}_t$ in (25)-(26). Since our model features a time-varying Ω_t, at each point in time for a given covariance matrix Ω the model is identified; i.e. the model is locally identified.

E.3 Modelling the correlation matrix via partial correlations

Here we show how to model a time-varying correlation matrix by imposing bounds on the partial correlations. In order to save in notation we drop the subscript t. Let consider the following covariance matrix $\Omega = DRD'$, where $D = \text{diag}(\sigma_1, \sigma_2, \sigma_3)$ and R is the correlation matrix:

$$R = \begin{bmatrix} 1 & \varrho_{12} & \varrho_{13} \\ \varrho_{12} & 1 & \varrho_{23} \\ \varrho_{13} & \varrho_{23} & 1 \end{bmatrix}.$$

To ensure positive standard deviations we model $\delta_i = \log \sigma_i$ so that $\sigma_i = \exp \delta_i$. For the correlations we model $\gamma = (\gamma_{12}, \gamma_{13}, \gamma_{23})'$, where $\gamma_{ij} = h(\varrho_{ij})$ and $h(\cdot)$ is the inverse function of the transformation $\varrho_{ij} = \psi_r(\gamma_{ij})$ that we describe below.

A well defined correlation matrix R must be positive semidefinite with ones on the main diagonal. This poses a non-trivial problem; see e.g. Budden et al. (2008).

On the other hand, the one-to-one mapping between the correlations and the partial correlations allows us to impose simple constraints to the partial correlations. Inspired by Joe (2006), Daniels and Pourahmadi (2009) and Lewandowski et al. (2009), we re-parametrize the correlation matrix with respect to the partial correlation matrix. Specifically, R is positive semidefinite if the corresponding partial correlation matrix

$$\Pi = \begin{bmatrix} 1 & \pi_{12} & \pi_{13} \\ \pi_{12} & 1 & \pi_{23} \\ \pi_{13} & \pi_{23} & 1 \end{bmatrix}$$

has all the elements $\pi_{ij} \in (-1, 1)$, where π_{ij} are the partial correlations between variables i

[9] The tree autoregressive coefficients and the two constants are identified by construction. The two moving average coefficients and three parameters of the covariance matrix are used to identify the six elements of the matrix Ω.

[10] Adding the measurement error ν_t, the reduced form model for $\texttt{pd}_t$ becomes an ARMA(2,2).

and j. To satisfy those bounds on π_{ij} we use the Fisher transformation, i.e. $\pi_{ij} = \tanh(\gamma_{ij})$. The function mapping the elements of R into the elements of Π is:[11]

$$\varrho_{12} = \pi_{12}, \quad \varrho_{13} = \pi_{13}, \quad \varrho_{23} = \pi_{23}\sqrt{(1-\pi_{12}^2)(1-\pi_{13}^2)} + \pi_{12}\pi_{13}. \tag{E.2}$$

In practice, we perform two transformations:

$$\varrho_{ij} = \psi_r(\gamma_{ij}) = \psi_{r,2}(\psi_{r,1}(\gamma_{ij})), \tag{E.3}$$

where $\psi_{r,1}(\cdot) = \tanh(\cdot)$, $\psi_{r,2}(\cdot)$ is defined in (E.2), and $\gamma_{ij} = \text{atanh}\pi_{ij}$. The resulting Jacobian is:

$$\frac{\partial \varrho}{\partial \gamma'} = \frac{\partial \varrho}{\partial \pi'}\frac{\partial \pi}{\partial \gamma'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \varkappa_{12} & \varkappa_{13} & \varkappa_{23} \end{bmatrix} \begin{bmatrix} 1-\pi_{12}^2 & 0 & 0 \\ 0 & 1-\pi_{13}^2 & 0 \\ 0 & 0 & 1-\pi_{23}^2 \end{bmatrix},$$

where

$$\varkappa_{12} = \pi_{13} - \pi_{12}\pi_{23}\sqrt{\frac{1-\pi_{13}^2}{1-\pi_{12}^2}}, \quad \varkappa_{13} = \pi_{12} - \pi_{13}\pi_{23}\sqrt{\frac{1-\pi_{12}^2}{1-\pi_{13}^2}}, \quad \varkappa_{23} = \sqrt{(1-\pi_{12}^2)(1-\pi_{13}^2)}.$$

Remark: It can be shown that if the partial correlations π_{ij} are bounded using the cosine function, i.e. $\psi_{r,1}(\cdot) = \cos\gamma_{ij}$, the transformation (E.3) turns out to be the same as the *hyperspherical coordinates* used in, e.g., Creal et al. (2011) and Buccheri et al. (2018). This means that by using hyperspherical coordinates they are implicitly modelling $\text{acos}\pi_{ij}$. The proof for a correlation matrix of general dimension is beyond the scope of this paper.

In our application, the identification of the model requires to set to zero one of the correlations. Without loss of generality we set to zero the correlation between the first and second innovation. Eploiting the mapping between the correlations and partial correlations we have that $\pi_{12} = 0$ implies $\varrho_{12} = 0$. Therefore, we model the following vectors $\varrho = (\varrho_{13}, \varrho_{23})'$, $\pi = (\pi_{13}, \pi_{23})'$, $\gamma = (\gamma_{13}, \gamma_{23})'$ The mapping between correlations and partial correlation is $\varrho_{13} = \pi_{13}$, $\varrho_{23} = \pi_{23}\sqrt{1-\pi_{13}^2}$, and the Jacobian is

$$\frac{\partial \varrho}{\partial \gamma'} = \sqrt{1-\pi_{13}^2}\begin{bmatrix} \sqrt{1-\pi_{13}^2} & 0 \\ -\pi_{13}\pi_{23} & 1-\pi_{23}^2 \end{bmatrix}.$$

[11] See also Yule and Kendall (1965, ch. 12) and Anderson (1984, p. 41).

E.4 Additional derivations of model objects

E.4.1 On the Conditional Variance and correlation of returns

In this section, we give additional details on the conditional variance and correlation of returns implied by the baseline model. Recall that:

$$r_{t+1} - \mathtt{E}_t(r_{t+1}) = \varepsilon^r_{t+1}$$

$$\varepsilon^r_{t+1} = -\rho_{t+1} b_{1,t+1} \varepsilon^{\mu}_{t+1} + \rho_{t+1} b_{2,t+1} \varepsilon^{g}_{t+1} + \varepsilon^{d}_{t+1}.$$

Therefore, one can calculate the conditional variance of returns:

$$\begin{aligned}
\mathtt{Var}_t\left(\varepsilon^r_{t+1}\right) &= \left(\rho_{t+1} b_{1,t+1}\right)^2 \mathtt{Var}_t\left(\varepsilon^{\mu}_{t+1}\right) + \left(\rho_{t+1} b_{2,t+1}\right)^2 \mathtt{Var}_t\left(\varepsilon^{g}_{t+1}\right) + \mathtt{Var}_t\left(\varepsilon^{d}_{t+1}\right) + \\
&\quad -2\left(\rho_{t+1}\right)^2 b_{1,t+1} b_{2,t+1} \mathtt{Cov}_t\left(\varepsilon^{g}_{t+1}, \varepsilon^{\mu}_{t+1}\right) - 2\rho_{t+1} b_{1,t+1} \mathtt{Cov}_t\left(\varepsilon^{\mu}_{t+1}, \varepsilon^{d}_{t+1}\right) \\
&= \left(\rho_{t+1} b_{1,t+1}\right)^2 \left(\sigma^{\mu}_{t+1}\right)^2 + \left(\rho_{t+1} b_{2,t+1}\right)^2 \left(\sigma^{g}_{t+1}\right)^2 + \left(\sigma^{d}_{t+1}\right)^2 \\
&\quad -2\left(\rho_{t+1}\right)^2 b_{1,t+1} b_{2,t+1} \left(\varrho^{g\mu}_{t+1} \sigma^{\mu}_{t+1} \sigma^{g}_{t+1}\right) - 2\rho_{t+1} b_{1,t+1} \left(\varrho^{d\mu}_{t+1} \sigma^{\mu}_{t+1} \sigma^{d}_{t+1}\right).
\end{aligned}$$

The conditional covariance can also be computed as follows:

$$\begin{aligned}
\mathtt{Cov}_t\left(\varepsilon^{\mu}_{t+1}, \varepsilon^r_{t+1}\right) &= -\rho_{t+1} b_{1,t+1} \mathtt{Var}_t\left(\varepsilon^{\mu}_{t+1}\right) + \rho_{t+1} b_{2,t+1} \mathtt{Cov}_t\left(\varepsilon^{g}_{t+1}, \varepsilon^{\mu}_{t+1}\right) + \mathtt{Cov}_t\left(\varepsilon^{\mu}_{t+1}, \varepsilon^{d}_{t+1}\right) \\
&= -\rho_{t+1} b_{1,t+1} \left(\sigma^{\mu}_{t+1}\right)^2 + \rho_{t+1} b_{2,t+1} \left(\varrho^{g\mu}_{t+1} \sigma^{\mu}_{t+1} \sigma^{g}_{t+1}\right) + \left(\varrho^{d\mu}_{t+1} \sigma^{\mu}_{t+1} \sigma^{d}_{t+1}\right)
\end{aligned}$$

E.4.2 On the Conditional Variance of price dividend

Note that

$$\begin{aligned}
pd_{t+1} - \overline{pd}_{t+1} &\simeq -b_{1,t+1} \widetilde{\mu}_{t+1} + b_{2,t+1} \widetilde{g}_{t+1} \\
&= -b_{1,t+1} \phi_1 \widetilde{\mu}_t + b_{2,t+1} \gamma_1 \widetilde{g}_t - b_{1,t+1} \varepsilon^{\mu}_{t+1} + b_{2,t+1} \varepsilon^{g}_{t+1},
\end{aligned}$$

and that

$$\begin{aligned}
\Delta d_{t+1} - \mathtt{E}_t(\Delta d_{t+1}) &= \varepsilon^{d}_{t+1} \\
r_{t+1} - E_t(r_{t+1}) &= \varepsilon^{r}_{t+1}
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon^{pd}_{t+1} &= -b_{1,t+1} \varepsilon^{\mu}_{t+1} + b_{2,t+1} \varepsilon^{g}_{t+1} \\
\varepsilon^{r}_{t+1} &= -\rho_{t+1} b_{1,t+1} \varepsilon^{\mu}_{t+1} + \rho_{t+1} b_{2,t+1} \varepsilon^{g}_{t+1} + \varepsilon^{d}_{t+1}.
\end{aligned}$$

It follows that

$$\mathtt{Var}_t\left(pd_{t+1} - \overline{\mathtt{pd}}_{t+1}\right) = b^2_{1,t+1} \mathtt{Var}_t(\varepsilon^{\mu}_{t+1}) + b^2_{2,t+1} \mathtt{Var}_t\left(\varepsilon^{g}_{t+1}\right) - 2 b_{1,t+1} b_{2,t+1} \mathtt{Cov}_t(\varepsilon^{\mu}_{t+1}, \varepsilon^{g}_{t+1}).$$

E.4.3 Term structure of Expectations and Risk

In this subsection we discuss the derivation of the term structure of expectations and risk for both dividend growth and returns. We define here long-horizon returns and dividend growth as the simple sum of annual log returns and dividend growth. By applying recursively equations (18)-(20) in the main text and taking conditional expectations, we obtain explicit expressions for the model-implied n-year return and dividend growth

$$\mu_t^{(n)} = \frac{1}{n}\mathtt{E}_t\left[\sum_{j=1}^{n} r_{t+j}\right] = \overline{\mu}_t + \frac{1}{n}\frac{1-\phi_\mu^n}{1-\phi_\mu}\tilde{\mu}_t \tag{E.4}$$

$$g_t^{(n)} = \frac{1}{n}\mathtt{E}_t\left[\sum_{j=1}^{n} \Delta d_{t+j}\right] = \overline{g}_t + \frac{1}{n}\frac{1-\phi_g^n}{1-\phi_g}\tilde{g}_t. \tag{E.5}$$

The model can be used also to obtain the annualized conditional variance of a n-years return. Ideally, when computing the uncertainty associated to multi-step forecasts, one should also account for the fact that the steady state parameters, $\bar{\mu}_t$ and $\bar{g}_t$, drift going forward from date t. But this is computationally challenging because it requires integrating a high-dimensional predictive density across all possible paths of future parameters. Consistent with a long-standing tradition in the learning literature (referred to as 'anticipated-utility' by Kreps, 1998), we instead assume that the steady state return and dividend growth are assumed to be constant going forward in time.[12] Therefore we can compute the term structure of risk in returns as

$$\begin{aligned}
\frac{1}{n}\mathtt{Var}_t\left[\sum_{j=1}^{n} r_{t+j}\middle|\overline{\mu}_{t+j}=\overline{\mu}_t\right] &= \frac{1}{n}\mathtt{Var}_t\left[\sum_{j=1}^{n-1}\frac{1-\phi_\mu^{n-j}}{1-\phi_\mu}\varepsilon_{t+j}^{\mu}+\sum_{j=1}^{n}\varepsilon_{t+j}^{r}\right] \\
&= \frac{1}{n}\sum_{j=1}^{n-1}\left(\frac{1-\phi_\mu^{n-j}}{1-\phi_\mu}\right)^2\mathtt{Var}_t\left(\varepsilon_{t+j}^{\mu}\right)+\frac{1}{n}\sum_{j=1}^{n}\mathtt{Var}_t\left(\varepsilon_{t+j}^{r}\right) \\
&\quad +\frac{2}{n}\sum_{j=1}^{n-1}\left(\frac{1-\phi_\mu^{n-j}}{1-\phi_\mu}\right)\mathtt{Cov}_t\left(\varepsilon_{t+j}^{\mu},\varepsilon_{t+j}^{r}\right)
\end{aligned}$$

where the conditional variance and covariance of returns are defined in E.4.1. Similarly, for dividend growth

$$\frac{1}{n}\mathtt{Var}_t\left[\sum_{j=1}^{n}\Delta d_{t+j}\middle|\overline{g}_{t+j}=\overline{g}_t\right] = \frac{1}{n}\sum_{j=1}^{n-1}\left(\frac{1-\phi_g^{n-j}}{1-\phi_g}\right)^2\mathtt{Var}_t\left(\varepsilon_{t+j}^{g}\right)+\frac{1}{n}\sum_{j=1}^{n}\mathtt{Var}_t\left(\varepsilon_{t+j}^{d}\right)$$

given that $\mathtt{Cov}_t\left(\varepsilon_{t+j}^{g},\varepsilon_{t+j}^{d}\right)=0$.

To compute the term structure of risk, we need to compute multiple step-ahead conditional

[12] This is consistent with the assumption that the long-run components move slowly overtime, therefore any variation of them is likely to be dominated by the variability of the other components of the model.

variances and covariances. We follow Creal et al. (2011) and for given parameter estimates at time t we simulate 1000 sample paths for the variables of the model, and use those to update the variances and correlations of the model. The h-step ahead conditional variances and covariances are therefore computed as the average over those simulations.

E.5 Time-varying risk: detailed results

The left hand side panel of Figure E.1 shows the evolution the conditional volatility of returns. This is very persistent and slightly trending upwards over the whole sample. Both the time profile and the average on the whole sample (at around 18 percent) are remarkably close to the estimates obtained by Carvalho et al. (2018). The right hand side panel shows the conditional correlation between expected and actual returns, that is the correlation between $r_{t+1} - E_t(r_{t+1})$ and $\tilde{\mu}_t$.[13]. The first point to notice is that it is robustly negative over the whole sample. This speaks directly to the notion that actual returns are predictable due to mean reversion: surprisingly high returns today correspond to low future returns. This correlation falls in our sample, from -.60 to -0.75, indicating increasing predictability in the last two decades, and generally falls in recessions, in line with the procyclical behaviour of the slope of the term structure of returns.

Mean reversion in returns is an important element in shaping investment strategies of long-run investors. The conventional view holds that the longer the investment horizon, the higher should the share of stocks in investment portfolios be. The reason is that the variance of stock returns at long horizons on a per year basis is lower than that at short-horizons, a result that would hold both unconditionally (Siegel, 2008), as well as conditionally on a given information set (Campbell and Viceira, 2005). Mean reversion plays a large role in rationalizing this result: if bear and bull markets roughly cancel out over long time spans, investors with a longer time horizon hold relatively less risk.[14]

Given time-varying covariances, our model produces time-varying profiles for the variance

[13] See Appendix E.4.3 for details on how this is computed

[14] Pastor and Stambaugh (2012) challenge this view and argue that if one takes into account others sources of uncertainty (in particular estimation risk and uncertainty about future expected returns) long-run variance is actually thirty percent higher than short-term variance. Recent support for the conventional wisdom comes from the work of Carvalho et al. (2018): when investors hold reasonable priors on the model parameters, the estimated conditional variance of stock returns falls with the investment horizon.

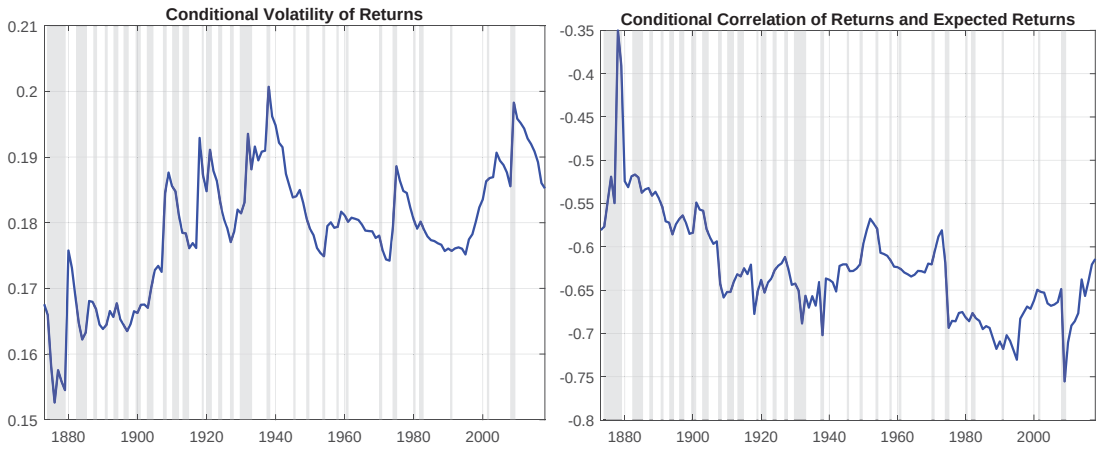
of stock returns at different horizons:[15]

$$\frac{1}{n}\mathtt{Var}_t\left[\sum_{j=1}^{n} r_{t+j}\middle|\overline{\mu}_{t+j|t}=\overline{\mu}_{t+1|t}\right] = \underbrace{\frac{1}{n}\sum_{j=1}^{n}\mathtt{Var}_t\left(\varepsilon^r_{t+j}\right)}_{iid\text{ uncertainty}} + \underbrace{\frac{1}{n}\sum_{j=1}^{n-1}\left(\frac{1-\phi_\mu^{n-j}}{1-\phi_\mu}\right)^2\mathtt{Var}_t\left(\varepsilon^\mu_{t+j}\right)}_{\text{future expected return uncertainty}} + \underbrace{\frac{2}{n}\sum_{j=1}^{n-1}\left(\frac{1-\phi_\mu^{n-j}}{1-\phi_\mu}\right)\mathtt{Cov}_t\left(\varepsilon^\mu_{t+j},\varepsilon^r_{t+j}\right)}_{\text{mean reversion}}$$

The left panel of Figure E.2 shows how the conditional variance for investment horizons of 5, 10 and 15 years ahead has evolved relative to the variance at the one year horizon. Three main points emerge. First, throughout the sample the conditional variance of the long run investor falls with the investment horizon. In fact, 5 to 15 years out the (per year) variance of stock returns is between 80 to 60 percent that of one year ahead, confirming the results in Campbell and Viceira (2002, 2005) and Carvalho et al. (2018). Secondly, reading this in conjunction with the right hand side panel of Figure E.1 shows that the evolution over time of the term structure mirrors closely that of return predictability: rising mean reversion is important in periods of high volatility (e.g. the depressions up to WW2, as well as the last 2 decades) and pushes down long horizon variance. Finally, the right hand side panel of Figure E.2 decomposes the (sample average) term structure of risk in its three components: the variance of innovations to expected returns (red solid line), the variance of innovations to actual returns (blue dotted line) and the mean reversion component (greed dashed line). The increase in predictability over long horizons dominates the other terms and gives rise to a downward sloping term structure of returns risk.

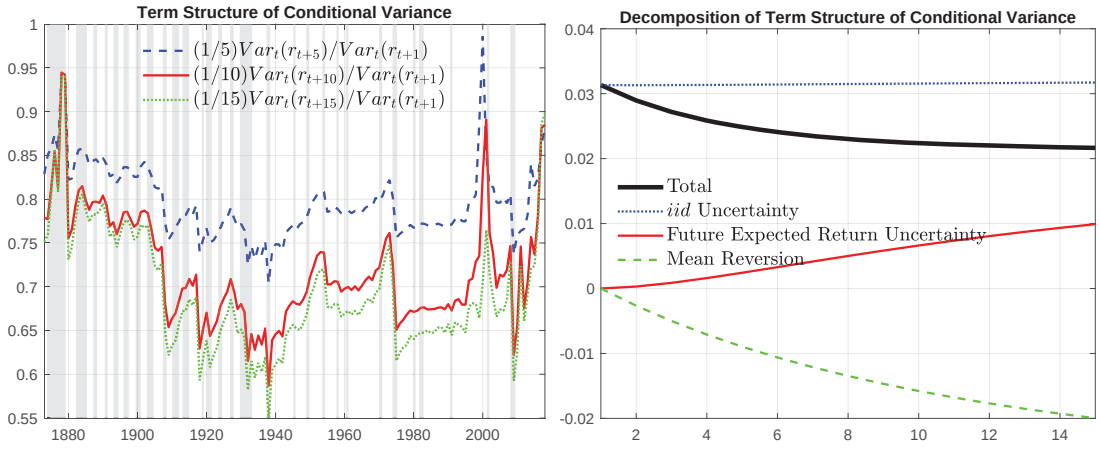
[15]Two caveats ought to be stressed. First, we are abstracting from estimation risk. This is therefore a source of downward bias in our estimates. Second, we we are disregarding associated with the movement in the trend. This could also bias downward our estimate of the conditional variance, although given the very low signal to noise ratio of stock returns, the adjustment is likely to be negligible.

Figure E.1: Conditional Volatilities and Correlations of returns



Note. The left hand panel of Figure E.1 reports the conditional volatility of returns $\mathtt{Var}_t\left(\varepsilon^r_{t+1}\right)$, whereas the left hand side reports the conditional correlation between expected returns and actual returns, $\mathtt{Corr}_t\left(\varepsilon^\mu_{t+1}, \varepsilon^r_{t+1}\right)$.

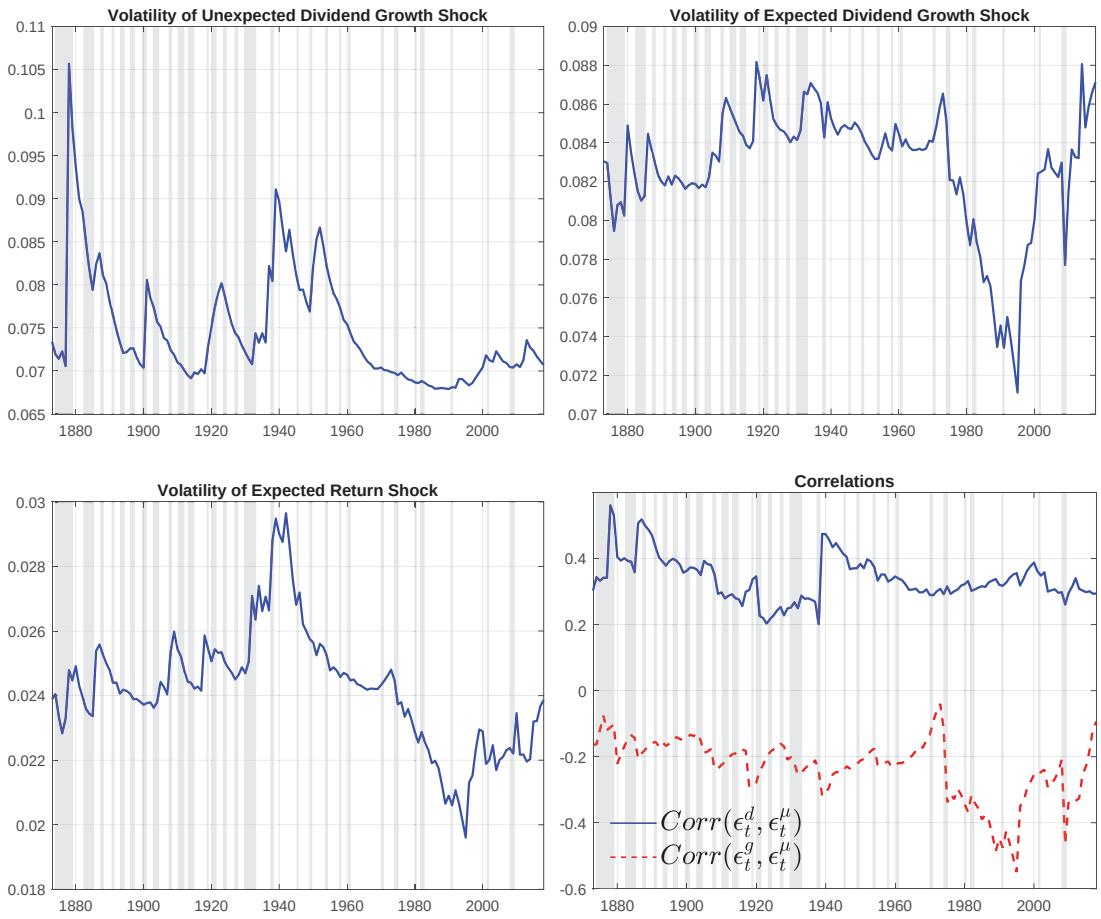
Figure E.2: Term Structure of the Returns Conditional Variance



Note. The left panel of Figure E.2 reports the conditional variance of long run returns at different horizons with respect to the variance of the one year returns. On the right panel of Figure E.2 we report the average over the entire sample of the decomposition of the term structure of variance into the component associated to the returns shocks, the component of the variance that is associated with the expected returns and the mean reversion component. The black line corresponds to the average term structure of risk and is equal to the sum of the aforementioned three components.

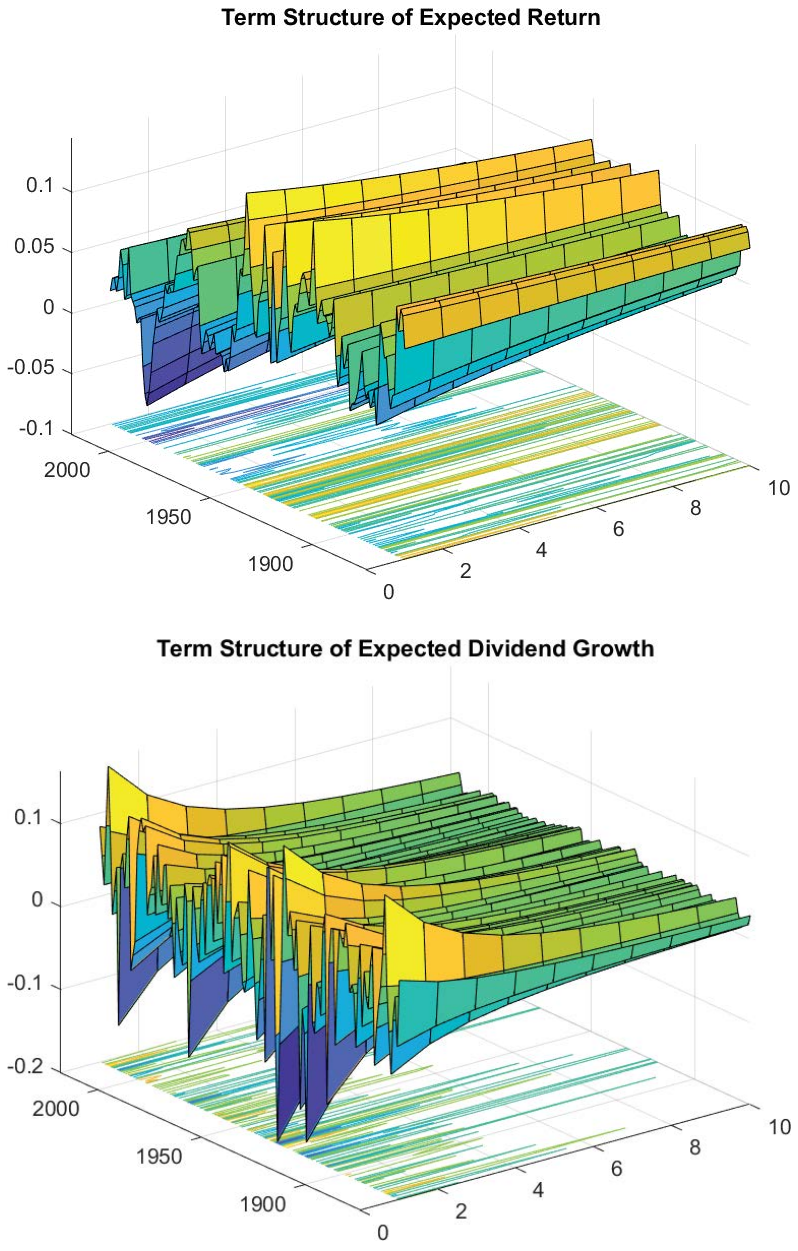
E.6 Additional Results

Figure E.3: Volatilities and Correlations



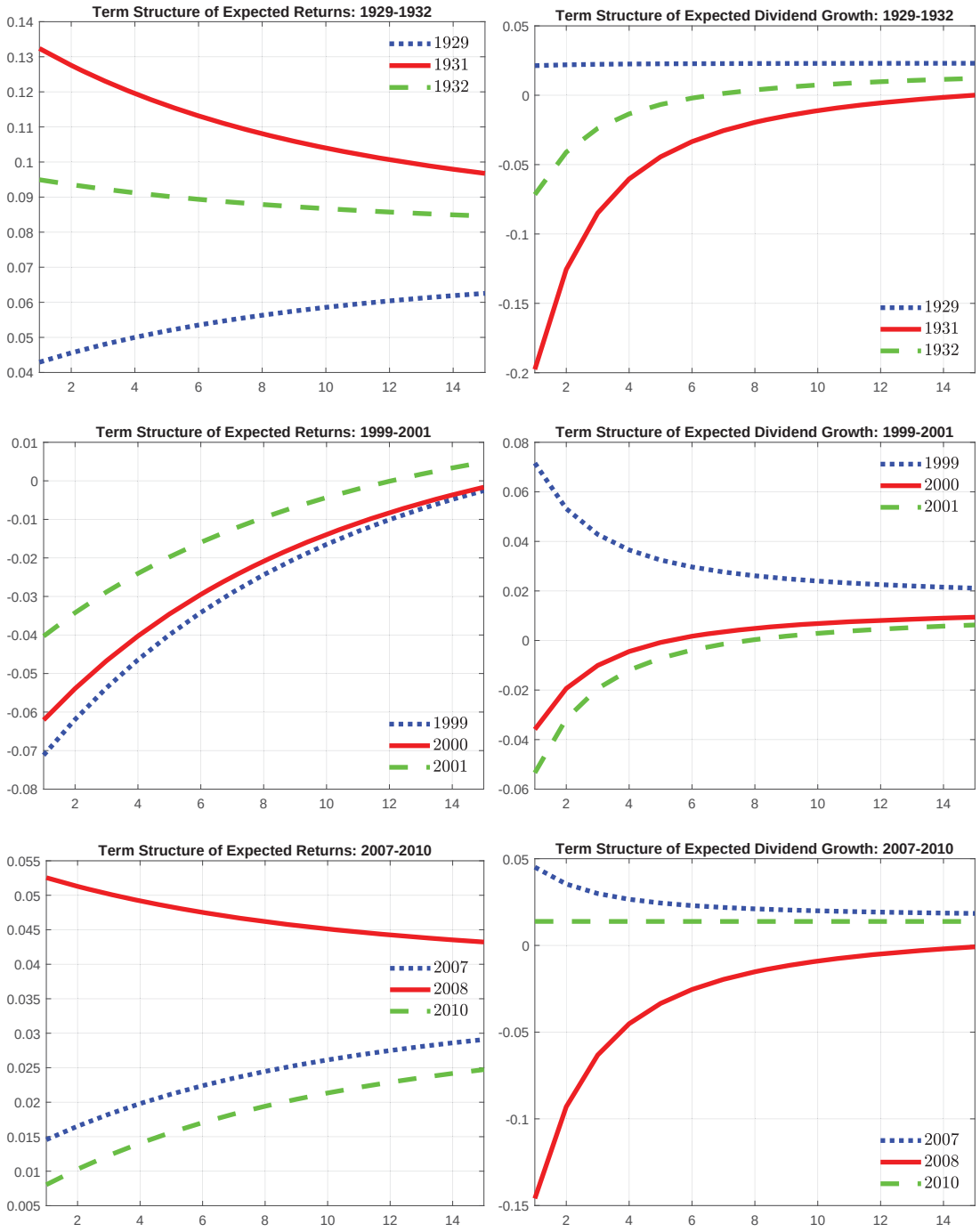
Note. Figure E.3 shows the estimated volatilities and correlations that form the matrix Q_t.

Figure E.4: Term Structure of the Expected Returns and Dividend Growth



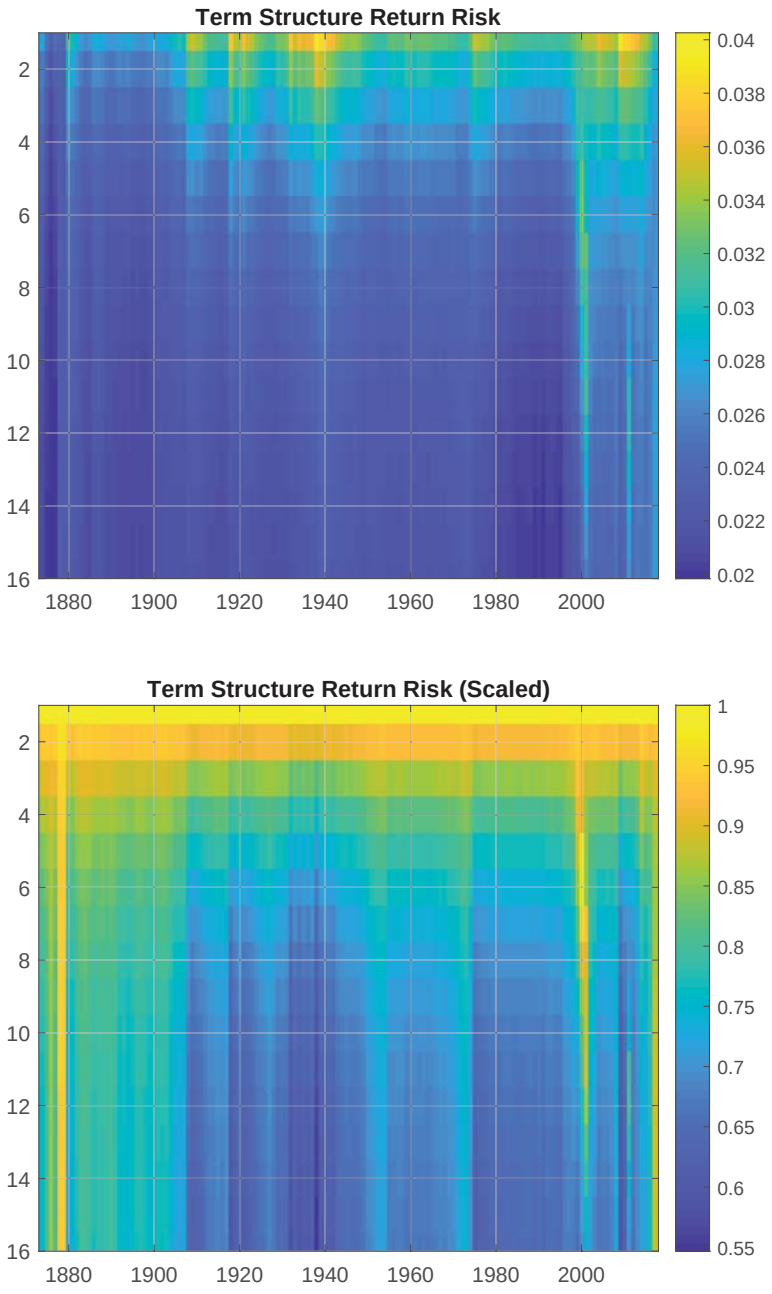
Note. Figure E.4 plots the term structure of expected return and dividend growth. Appendix E.4.3 gives details on the derivation of the term structure of expectations.

Figure E.5: Expected Returns and Dividend Growth: Selected Episodes



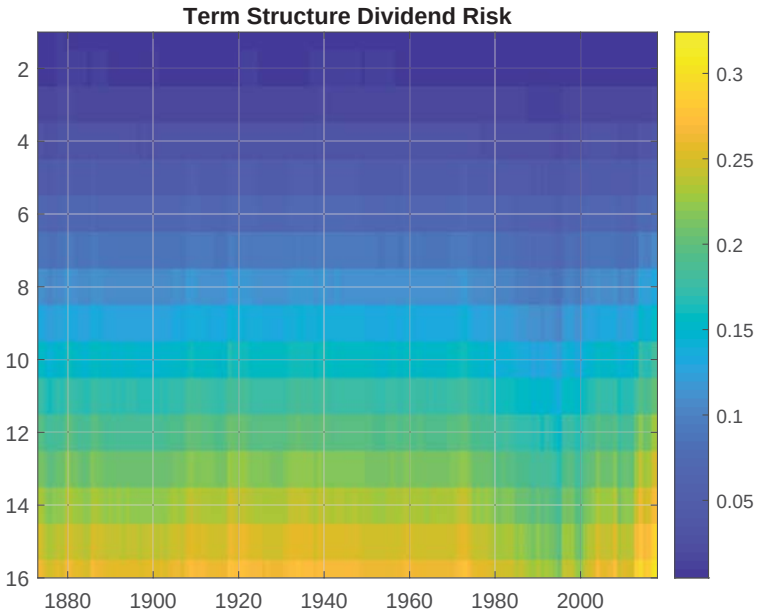
Note. Figure E.5 plots the term structure of expected return (left panel) and dividend growth (right panel) around some specific events. In particular, the upper panel looks at the Great Depression, the middle panel looks at the years around the 2001 recession and the bottom panel looks at the years of the Great Recession.

Figure E.6: Term Structure of the Returns Risk



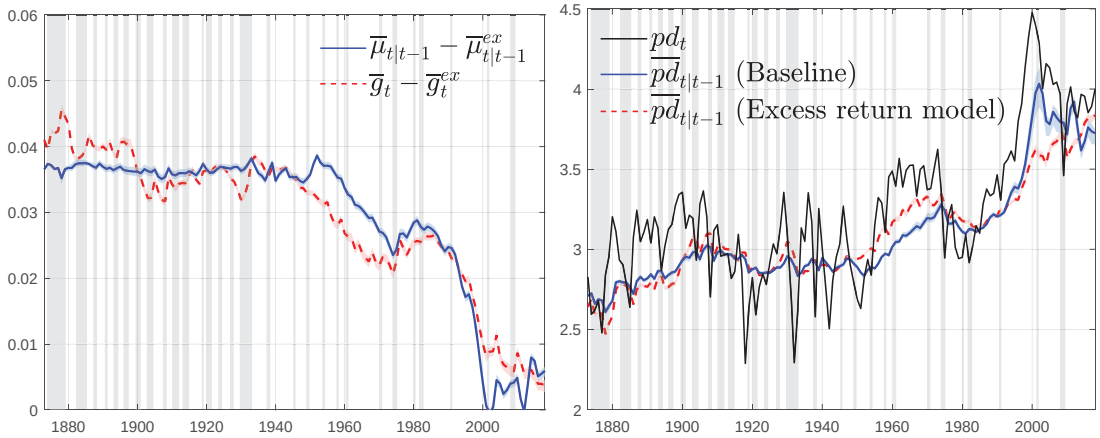
Note. The upper panel of Figure E.6 reports the term structure of the volatility of returns. In the lower panel the same term structure is normalized to 1 for the 1 year head investment. Appendix E.4.3 gives details on the derivation of the term structure of volatility.

Figure E.7: Term Structure of the Dividend Growth Risk



Note. Figure E.7 reports the term structure of the volatility of dividend growth. Appendix E.4.3 gives details on the derivation of the term structure of volatility.

Figure E.8: Steady State Comparisons



Note. The left panel of Figure E.8 reports two alternative measures of the long-run riskless real rate that we recover from the estimates in section 6. Specifically, $\overline{\mu}_{t|t-1} - \overline{\mu^{ex}}_{t|t-1}$ and $\overline{g}_{t|t-1} - \overline{g}^{ex}_{t|t-1}$. The estimates of $\overline{\mathtt{pd}}_{t|t-1}$ from the two models are reported in the panel on the right together with the (log) level of the price dividend ratio.

Table E.1: Excess Return Model: Estimation Results

ϕ_μ	0.863			b_μ	0.081
	[0.007]				[0.010]
ϕ_g	0.355			b_g	0.053
	[0.011]				[0.007]
$\bar{\sigma}_d$	0.062	a_{σ_d}	0.858	b_{σ_d}	0.016
	[0.061; 0.065]		[0.023]		[0.001]
$\bar{\sigma}_g$	0.097	a_{σ_g}	0.765	b_{σ_g}	0.012
	[0.096; 0.104]		[0.052]		[0.003]
$\bar{\sigma}_\mu$	0.019	a_{σ_μ}	0.847	b_{σ_μ}	0.015
	[0.018; 0.024]		[0.051]		[0.003]
$\bar{\rho}_{d,\mu}$	0.888	$a_{\pi_{d,\mu}}$	0.980	$b_{\pi_{d,\mu}}$	0.025
	[0.660; 0.892]		[0.010]		[0.010]
$\bar{\rho}_{g,\mu}$	-0.001	$a_{\pi_{g,\mu}}$	0.820	$b_{\pi_{g,\mu}}$	0.025
	[-0.026; -0.018]		[0.047]		[0.005]
σ^2_ν	0.008			κ_h	0.020
	[0.0002]				[0.0001]
Log Lik.	322.232				

Note. Table E.1 reports parameter estimates for our baseline model. In the first two entries of the first column we report the estimates of the autoregressive coefficients of expected returns and expected dividend growth (ϕ_μ and ϕ_g). Then we show the average (over the whole sample) estimates of the volatilities ($\bar{\sigma}_d$, $\bar{\sigma}_g$ and $\bar{\sigma}_\mu$) and correlations ($\bar{\rho}_{d,\mu}$ and $\bar{\rho}_{g,\mu}$) that form the matrix Q_t in the state space model. Finally, we report the estimated volatility of the measurement error for the price dividend ratio. The second and third columns show the estimates of the coefficients that enter the law of motion of the score driven time-varying processes: $f_{t+1} = \mathrm{c} + \mathrm{A} f_t + \mathrm{B} s_t$ where A and B are diagonal matrices. Recall that the first two elements of f_t are martingales. This implies that the first two elements of the diagonal of A are zeros. Estimates for the remaining five entries are reported in the second column of this table. In the third column we report the seven elements that form the diagonal of B, that is the loadings on the likelihood score, and the smoothing coefficient applied to the Hessian term (κ_h). For each coefficient we report in square brackets the associated standard error. For the average volatilities and correlations in the first column we report the 68% confidence interval from 1000 simulations of the model (calculated as in Blasques et al., 2016).

References

Abadir, K. and Magnus, J. (2005). *Matrix Algebra*. Cambridge University Press, Cambridge, UK.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley series in probability and mathematical statistic.

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). *Now-Casting and the Real-Time Data Flow*, volume 2 of *Handbook of Economic Forecasting*, chapter 0, pages 195–237. Elsevier.

Binsbergen, J. H. V. and Koijen, R. S. J. (2010). Predictive Regressions: A Present Value Approach. *Journal of Finance*, 65(4):1439–1471.

Blasques, F., Koopman, S. J., and Lucas, A. (2014). Optimal Formulations for Nonlinear Autoregressive Processes. Tinbergen Institute Discussion Papers 14-103/III, Tinbergen Institute.

Buccheri, G., Bormetti, G., Corsi, F., and Lillo, F. (2018). A Score-Driven Conditional Correlation Model for Noisy and Asynchronous Data: an Application to High-Frequency Covariance Dynamics. Papers 1803.04894, arXiv.org.

Budden, M., Hadavas, P., and Hoffman, L. (2008). On the generation of correlation matrices. *Applied Mathematics E-Notes*, 8:279–282.

Campbell, J. Y. and Viceira, L. (2005). The Term Structure of the Risk-Return Tradeoff. NBER Working Papers 11119, National Bureau of Economic Research, Inc.

Campbell, J. Y. and Viceira, L. M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long- Term Investors*. Oxford University Press, Oxford, UK.

Carvalho, C. M., Lopes, H. F., and McCulloch, R. E. (2018). On the long-run volatility of stocks. *Journal of the American Statistical Association*, 113(523):1050–1069.

Cochrane, J. H. (2008). State-Space vs. VAR Models for Stock Returns. Unpublished manuscript.

Creal, D., Koopman, S. J., and Lucas, A. (2008). A General Framework for Observation Driven Time-Varying Parameter Models. Tinbergen Institute Discussion Papers 08-108/4, Tinbergen Institute.

Creal, D., Koopman, S. J., and Lucas, A. (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations. *Journal of Business & Economic Statistics*, 29(4):552–563.

Daniels, M. and Pourahmadi, M. (2009). Modeling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis*, 100(10):2352–2363.

Del Negro, M. (2012). Bayesian Macroeconometrics. In *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.

Delle Monache, D. and Petrella, I. (2017). Adaptive models and heavy tails with an application to inflation forecasting. *International Journal of Forecasting*, 33(2):482–501.

Doan, T., Litterman, R. B., and Sims, C. A. (1986). Forecasting and conditional projection using realistic prior distribution. Staff Report 93, Federal Reserve Bank of Minneapolis.

Giannone, D., Lenza, M., and Primiceri, G. E. (2019). Priors for the Long Run. *Journal of the American Statistical Association*, 114(526):565–580.

Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97(10):2177 – 2189.

Kapetanios, G., Marcellino, M., and Venditti, F. (2019). Large time-varying parameter VARs: a non-parametric approach. *Journal of Applied Econometrics, forthcoming*.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198.

Kreps, D. M. (1998). Anticipated Utility and Dynamic Choice. In Jacobs, Kalai, and Kamien, editors, *Frontiers of Research in Economic Theory*. Cambridge University Press.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989 – 2001.

Litterman, R. B. (1979). Techniques of forecasting using vector autoregressions. Working Papers 115, Federal Reserve Bank of Minneapolis.

Lucas, A., Opschoor, A., and Schaumburg, J. (2016). Accounting for missing values in score-driven time-varying parameter models. *Economics Letters*, 148(C):96–98.

Pastor, L. and Stambaugh, R. F. (2012). Are Stocks Really Less Volatile in the Long Run? *Journal of Finance*, 67(2):431–478.

Rytchkov, O. (2012). Filtering Out Expected Dividends and Expected Returns. *Quarterly Journal of Finance*, 2(03):1–56.

Siegel, J. J. (2008). *Stocks for the Long Run*. McGraw Hill, New York, NY.

Stock, J. H. and Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39(s1):3–33.

Yule, G. and Kendall, M. (1965). *An introduction to the theory of statistics*. C. Griffin & Co., Belmont, California; 14th ed.

Acknowledgements

The views expressed in this paper belong to the authors and are not necessarily shared by the Bank of Italy or by the European Central Bank.

Davide Delle Monache
Bank of Italy, Rome, Italy; email: davide.dellemonache@bancaditalia.it

Ivan Petrella
University of Warwick, Coventry, United Kingdom; CEPR; email: ivan.petrella@wbs.ac.uk

Fabrizio Venditti
European Central Bank, Frankfurt am Main, Germany; email: fabrizio.venditti@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-4012-2 ISSN 1725-2806 doi:10.2866/699185 QB-AR-20-021-EN-N