Price Dividend Ratio and Long-Run Stock Returns: a Score Driven State Space Model

Supplementary Material

Davide Delle Monache
Bank of Italy
davide.dellemonache@bancaditalia.it

Ivan Petrella
University of Warwick and CEPR
Ivan.Petrella@wbs.ac.uk

Fabrizio Venditti
European Central Bank
fabrizio.venditti@ecb.int

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:

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

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

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

A.3 State space model in forward form

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

$$\begin{aligned}
y_t &= Z_t\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, H_t), \\
\alpha_{t+1} &= T_t\alpha_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q_t), \quad \alpha_1 \sim \mathcal{N}(a_1, P_1), \quad t = 1, ..., n.
\end{aligned} \qquad \text{(A.20)}$$

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:

$$\begin{aligned}
&v_t = y_t - Z_t a_t, && F_t = Z_t P_t Z_t' + H_t, \\
&K_t = T_t P_t Z_t' F_t^{-1}, && L_t = T_t - K_t Z_t, \\
&a_{t+1} = T_t a_t + K_t v_t && P_{t+1} = T_t P_t L_t' + Q_t, \quad t = 1, ..., n.
\end{aligned} \qquad \text{(A.21)}$$

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 $\dot{Q}_t$.

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:

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

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:

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

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 estimate of the prediction error variance (by means of v_t) from its past estimate (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. (2020) to model high-frequency multivariate financial time-series.

[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.

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 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:

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

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)

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

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 detailed in Appendix D. 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 constraints. These include the most popular priors typically considered in the Bayesian VAR literature, including the Minnesota prior, the sum of coefficients prior and the long-run prior (see, e.g., Del Negro, 2012; Kapetanios et al., 2019).

C Monte Carlo exercise

The Monte Carlo exercise is based on a battery of simple bivariate models that share a common component. 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 $f_t = a_3 + b_3 (t \geq \tau);$
Case 4: DOUBLE STEP $f_t = a_4 + b_4 I (t \geq \tau_1) + c_4 I (t \geq \tau_2);$
Case 5: RAMP $f_t = a_5 + \left(\frac{b_5}{T/c_5}\right) \mod (t);$
Case 6: AR(1) MODEL $f_t = a_6(1-b_6) + b_6 f_{t-1} + \xi_t, \quad \xi_t \sim \mathcal{N}(0, c_6);$

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.[4] We

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

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 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 C.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.[5] This result implies that our estimation method does not generate spurious time variation in the coefficients when

for DGP3 and DGP4.

[5]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.

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 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.

C.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

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

C.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 C.1-C.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 C.1: Monte Carlo Exercise

	DGP 1: time-varying LOADINGS COEFFICIENT											
	T = 250						T = 500					
	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up	RMSE	MAE	Corr.	68% Cov.	90% Cov.	# Pile up
CONSTANT	0.003	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 C.1: TIME-VARYING LOADINGS, N=250

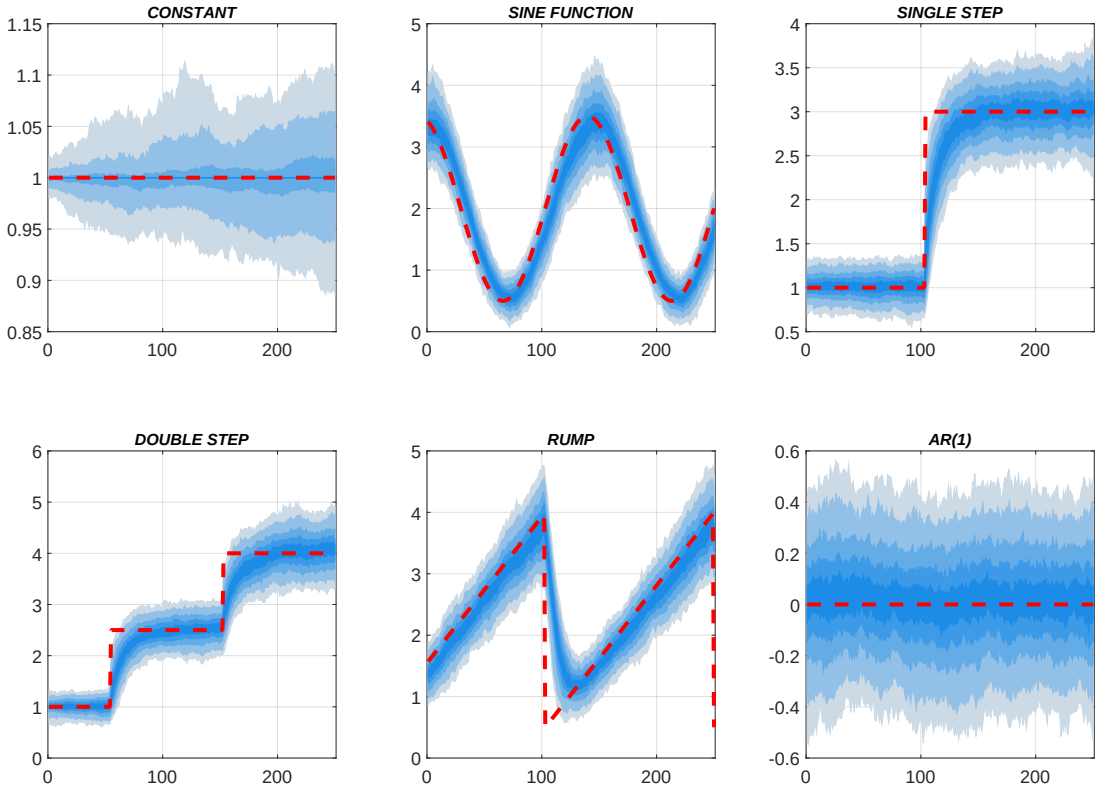


Figure C.2: TIME-VARYING LOADINGS, N=500

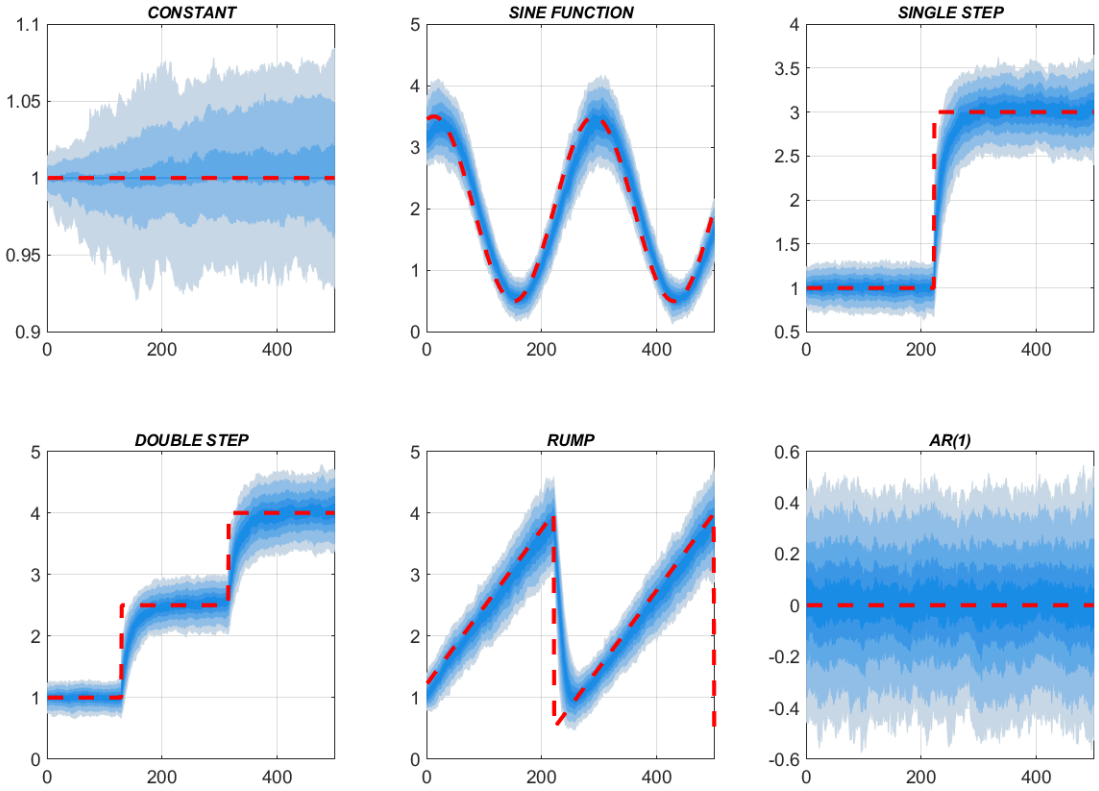


Figure C.3: TIME-VARYING AUTOREGRESSIVE COEFFICIENTS, n=250

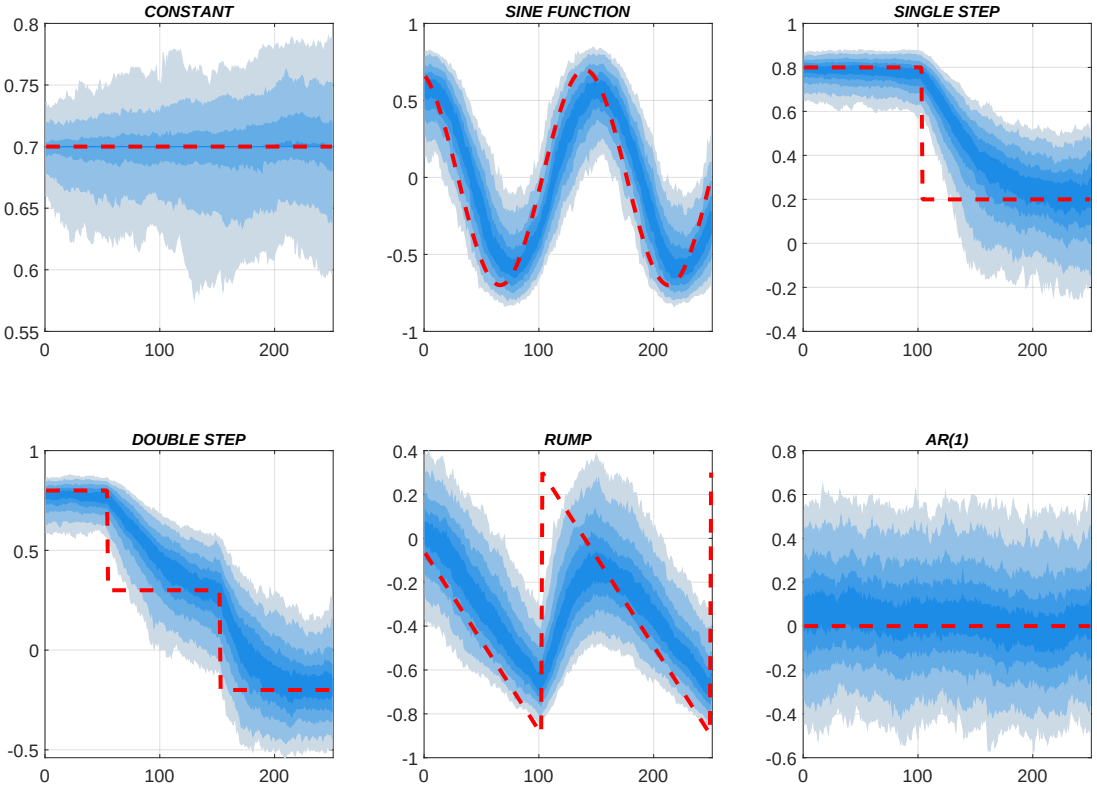


Figure C.4: TIME-VARYING AUTOREGRESSIVE COEFFICIENTS, n=500

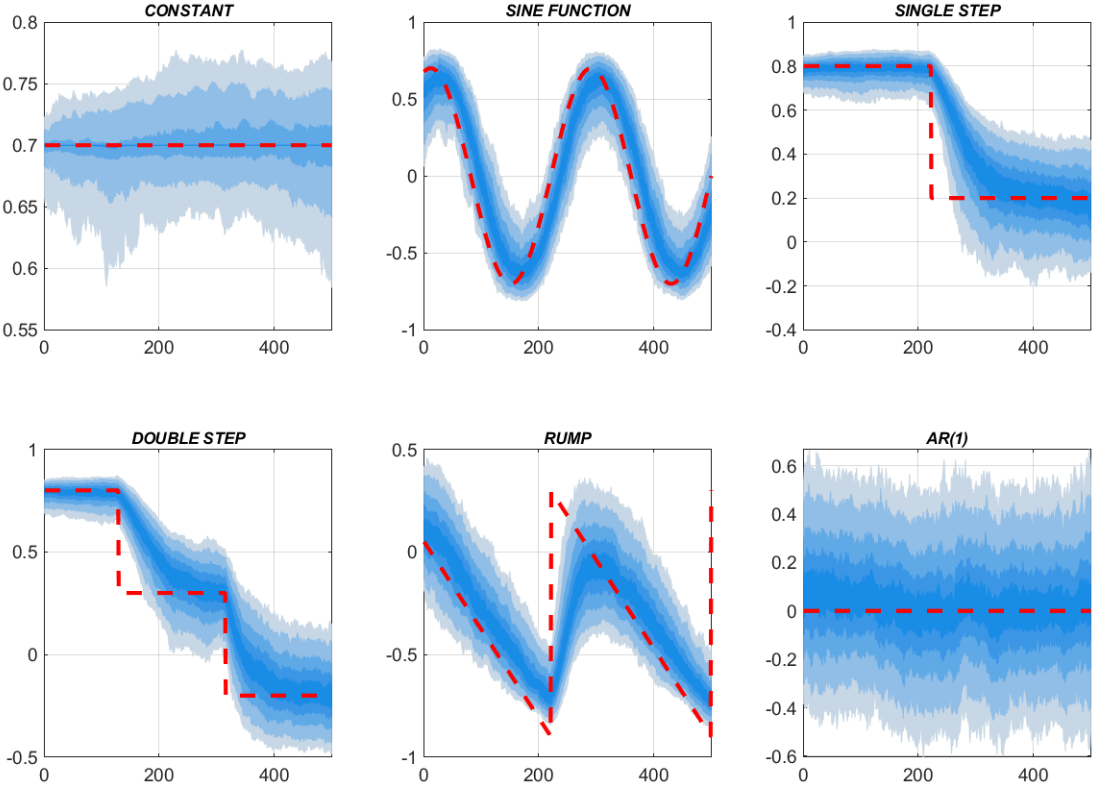


Figure C.5: TIME-VARYING MEASUREMENT EQUATION ERROR VARIANCE, n=250

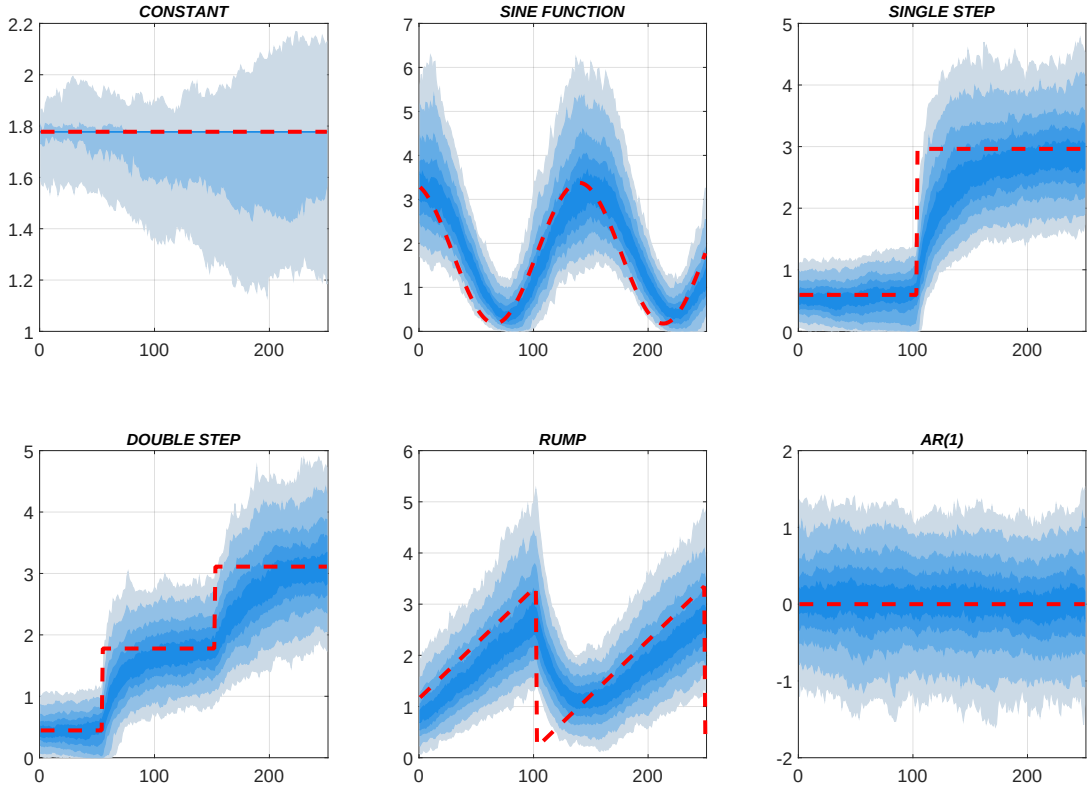


Figure C.6: TIME-VARYING MEASUREMENT EQUATION ERROR VARIANCE, n=500

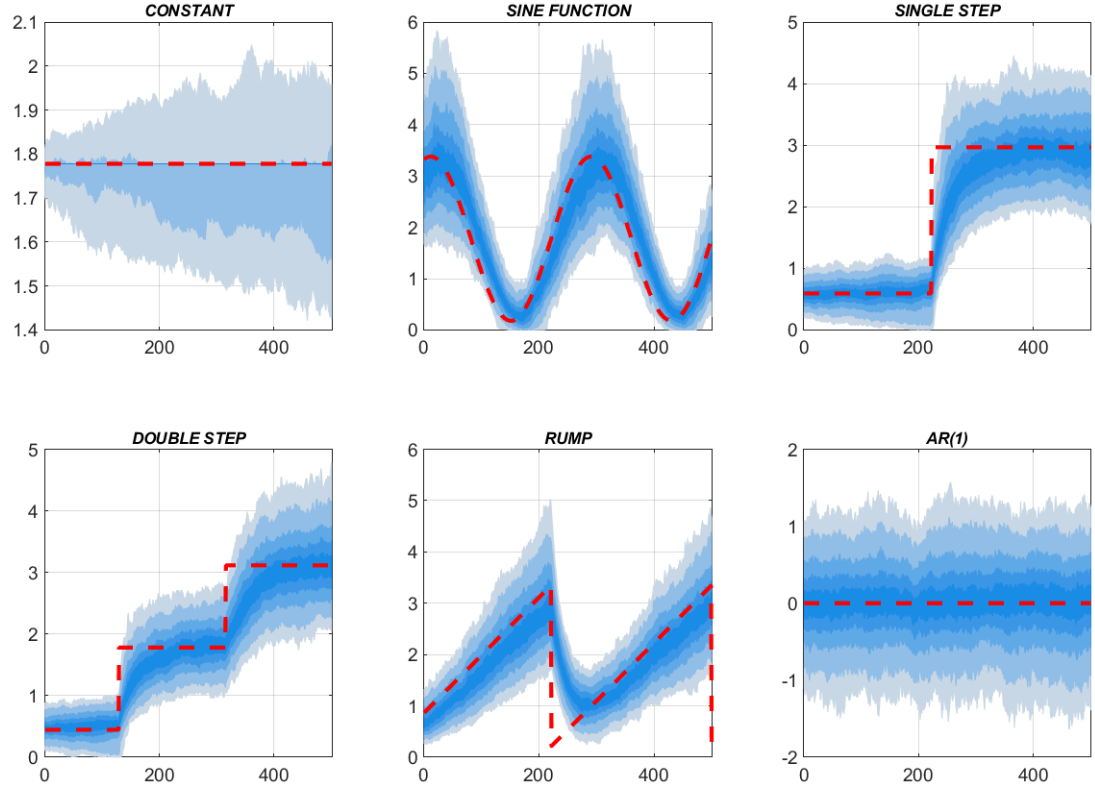


Figure C.7: TIME-VARYING TRANSITION EQUATION ERROR VARIANCE, n=250

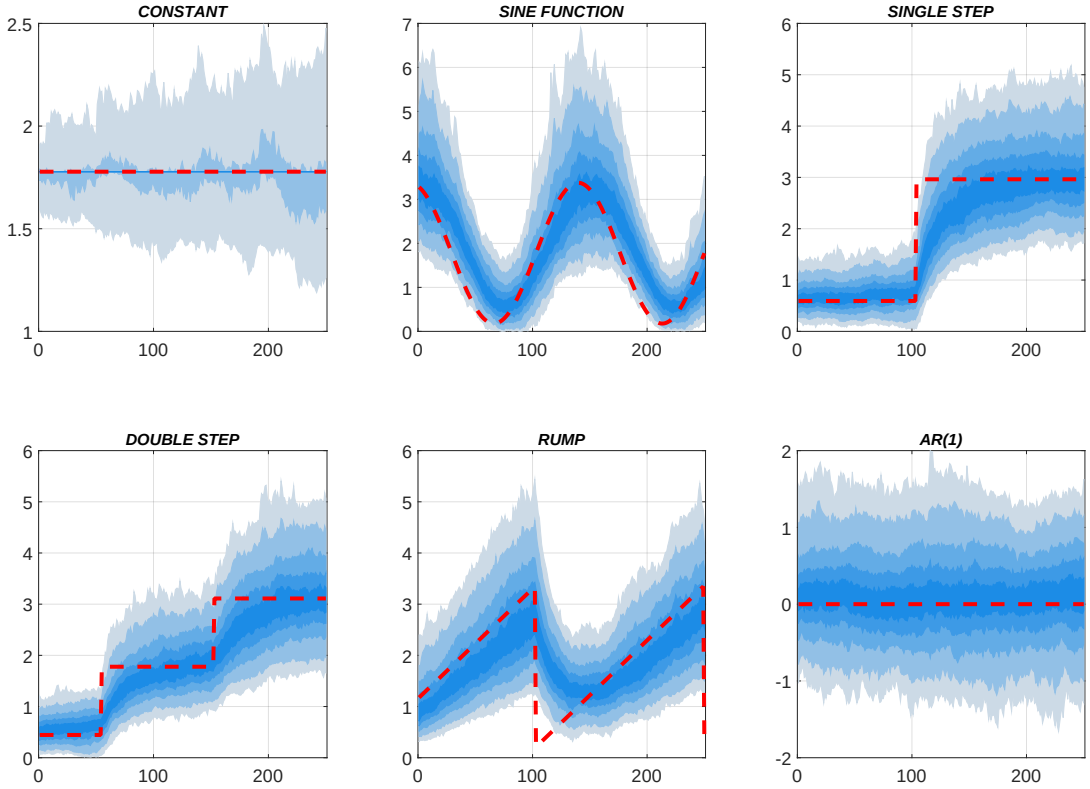
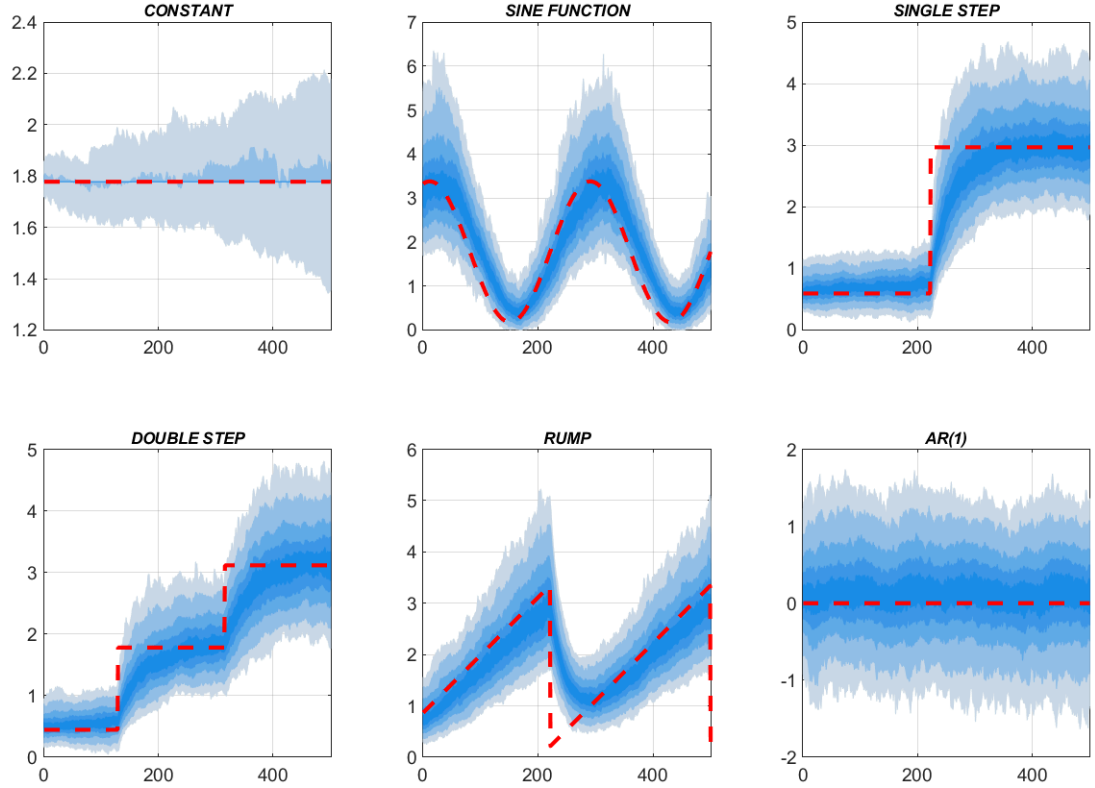


Figure C.8: TIME-VARYING TRANSITION EQUATION ERROR VARIANCE, n=500



D Shrinking the vector of parameters by the 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). Following (Theil and Goldberger, 1961), priors can be incorporated into the state space model (1), with the score driven system matrices described by (4), augmenting the model using the following Gaussian constraints:

$$r_t = R_t f_t + u_t \quad u_t \sim \mathcal{N}(0, \Sigma_t), \tag{D.1}$$

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:

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

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

$$s_t^p = (\mathcal{I}_t^p)^{-1} \nabla_t^p = (\mathcal{I}_t + R_t' \Sigma_t^{-1} R_t)^{-1} (\nabla_t + R_t' \Sigma_t^{-1} u_t). \tag{D.3}$$

Using the Woodbury identity we obtain[6]:

$$(\mathcal{I}_t^p)^{-1} = (\mathcal{I}_t + R_t' \Sigma_t^{-1} R_t)^{-1} = \mathcal{I}_t^{-1} - \mathcal{I}_t^{-1} R_t' (R_t \mathcal{I}_t^{-1} R_t' + \Sigma_t)^{-1} R_t \mathcal{I}_t^{-1} = (I - \Upsilon_t R_t) \mathcal{I}_t^{-1}, \tag{D.4}$$

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:

$$s_t^p = (I - \Upsilon_t R_t) s_t + \Upsilon_t u_t, \tag{D.5}$$

From the 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 define $R_t = I$, $r_t = \bar{f}$, and $\Sigma_t = \frac{1}{\lambda} I$ so that the constraints reduce 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, that is $f_{t+1} = f_t + B s_t^p$, and the information matrix equals to identity

[6] 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.

matrix, the resulting regularized score-driven filter is:[7]

$$f_{t+1} = \frac{\mathrm{B}\lambda}{1+\lambda}\bar{f} + \left(I - \frac{\mathrm{B}\lambda}{1+\lambda}\right) f_t + \frac{\mathrm{B}}{1+\lambda} s_t. \tag{D.6}$$

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.

E 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.[8] Given this reparameterization, the measurement equation in (1) is modified as follows:

$$W_t y_t = W_t Z_t \alpha_t + W_t \varepsilon_t, \quad W_t \epsilon_t \sim \mathcal{N}(0, W_t H_t W_t'), \tag{E.1}$$

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

$$\begin{aligned} v_t &= W_t(y_t - Z_t a_t), & F_t &= W_t(Z_t P_t Z_t' + H_t)W_t', \\ a_{t|t} &= a_t + P_t Z_t' W_t' F_t^{-1} v_t, & P_{t|t} &= P_t - P_t Z_t' W_t' F_t^{-1} W_t Z_t P_t. \end{aligned} \tag{E.2}$$

The expressions in (7) become

$$\begin{aligned} \dot{V}_t &= -[(a_t' \otimes W_t)\dot{Z}_t + (a_{t-1|t-1}' \otimes W_t Z_t)\dot{T}_t], \\ \dot{F}_t &= 2N_{N_t}(W_t Z_t P_t \otimes W_t)\dot{Z}_t + (W_t Z_t \otimes W_t Z_t)2N_m(T_t P_{t-1|t-1} \otimes I_m)\dot{T}_t \\ &\quad + (W_t \otimes W_t)\dot{H}_t + (W_t Z_t \otimes W_t Z_t)\dot{Q}_t. \end{aligned} \tag{E.3}$$

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

[7] 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.

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

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).

F 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:

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

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

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

G Empirical application

G.1 Identification of the model

In this section, we consider the identification of the model presented in section 3. Let us start by looking at the constant parameter version of the model:

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

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 (G.1) is equivalent to one estimated by Binsbergen and Koijen (2010), whose identification issue is discussed at lenght in Cochrane (2008), with the key difference being that we have added the measurement error in the $\mathtt{pd}_t$ equation. To be more precise, the specification for Δd_{t+1} implies an ARMA(1,1) process, while the model for $\mathtt{pd}_t$ 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 correlation to zero 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_ν^2 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 (21)-(22). 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.

G.2 Modelling the correlation matrix by 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 and j. To satisfy those bounds on the partial correlations π_{ij} we use the Fisher transformation, that is $\pi_{ij} = \tanh(\gamma_{ij})$, so that we model $\gamma_{ij} = \text{atanh}\pi_{ij}$. The function mapping the elements

[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).

of R into the elements of Π is:[11]

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

Thus, we perform two transformations:

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

where $\psi_{r,1}(\cdot) = \tanh(\cdot)$, $\psi_{r,2}(\cdot)$ is defined in (G.2). 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: If the partial correlations π_{ij} are bounded using the cosine function, i.e. $\psi_{r,1}(\cdot) = \cos\gamma_{ij}$, the transformation (G.3) turns out to be the same as the *hyperspherical coordinates* used in, e.g., Creal et al. (2011) and Buccheri et al. (2020). This means that the use of hyperspherical coordinates implies modelling inverse consine of the partial correlations. 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

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

G.3 State space, score driven vector and jacobians

The model in section 3 can be cast easily in state space form:

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

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

where

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

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

$Q_t = S_\eta \Omega_t S_\eta'$, $\Omega_t = D_t R_t D_t$, with D_t contains the standard deviations, and R_t denotes the correlation matrix. 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. The resulting matrices are is:

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

The vector of time-varying parameters is:

$$f_t = \begin{bmatrix} \varphi_t \\ \delta_t \\ \gamma_t \end{bmatrix}, \quad \varphi_t = \begin{bmatrix} \overline{\mu}_t, \\ \overline{g}_t \end{bmatrix}, \quad \delta_t = \begin{bmatrix} \log \sigma_{d,t} \\ \log \sigma_{g,t} \\ \log \sigma_{\mu,t} \end{bmatrix}, \quad \gamma_t = \begin{bmatrix} \operatorname{atanh} \pi_{d\mu,t} \\ \operatorname{atanh} \pi_{g\mu,t} \end{bmatrix}.$$

The vector f_t follows the score driven model discussed in section 2, with the following specification of the static parameters:

$$\begin{aligned} \mathrm{c} &= [0, 0, c_{\sigma_d}, c_{\sigma_g}, c_{\sigma_\mu}, c_{\pi_{d,\mu}}, c_{\pi_{g,\mu}}]', \\ \mathrm{A} &= \mathrm{diag}([1, 1, a_{\sigma_d}, a_{\sigma_g}, a_{\sigma_\mu}, a_{\pi_{d,\mu}}, a_{\pi_{g,\mu}}]), \\ \mathrm{B} &= \mathrm{diag}([b_\mu, b_g, b_{\sigma_d}, b_{\sigma_g}, b_{\sigma_\mu}, b_{\pi_{d,\mu}}, b_{\pi_{g,\mu}}]). \end{aligned}$$

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

$$\mathrm{vec}(Z_t) = S_{0,z} + S_{1,z}\psi_z\left(S_{2,z} f_t\right),$$

where

$$S_{0,z} = \begin{bmatrix} 0_{6\times 1} \\ 1 \\ 0 \\ 1 \\ 0_{5\times 1} \end{bmatrix}, \quad S_{1,z} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ & & 0_{8\times 4} & \end{bmatrix}, \quad S'_{2,z} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

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

The Jacobian matrix is:

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

Time variation in the Q matrix. 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:

$$\dot{Q}_t = (S_\eta \otimes S_\eta)\left[(D_tR_t \otimes I + I \otimes D_tR_t)\dot{D}_t + (D_t \otimes D_t)\dot{R}_t\right].$$

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

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

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

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

The functions $\psi_d(\delta_t)$ and $\psi_r(\gamma_t)$ and their Jacobians are described in section G.2. Specifically, we have that:

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

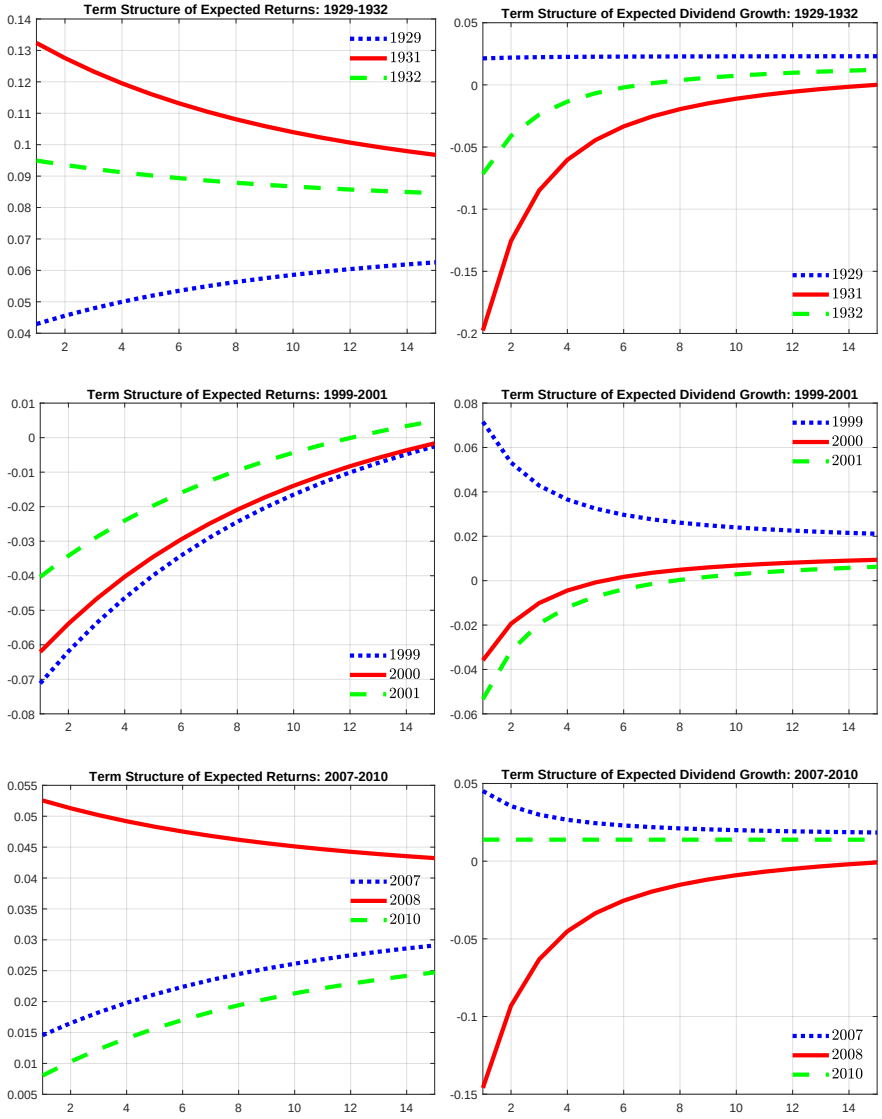
where

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

H Term structure of expected returns and dividend growth in recessions

In Figure H.1 we plot the whole term structure of expected returns and expected dividend growth for three historical episodes. In particular, we look at the year before the recession, the peak of the recession and the year after the recession. We find that discount 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 narrative in Campbell et al. (2013).

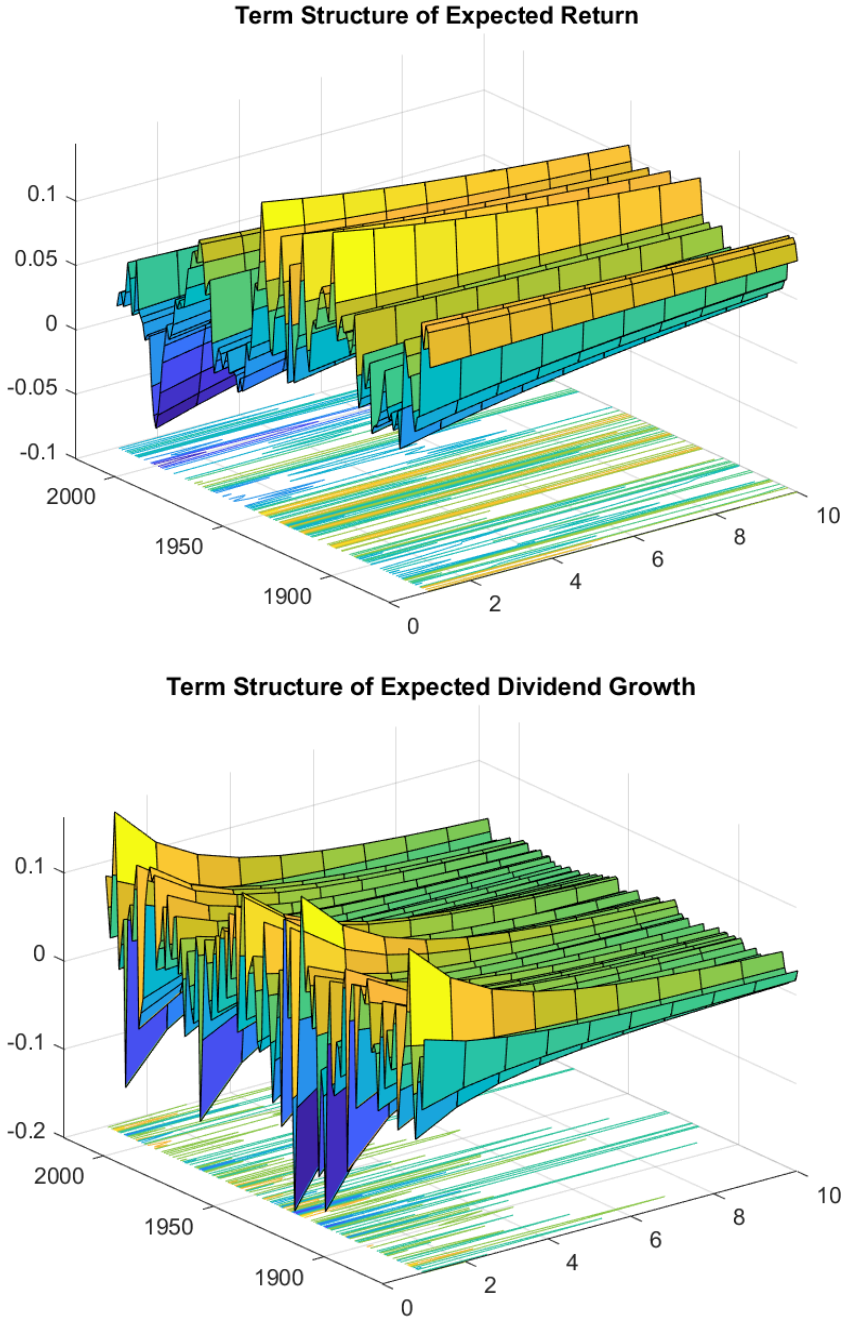
Figure H.1: Expected Returns and Dividend Growth: Selected Episodes



Note. Figure H.1 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.

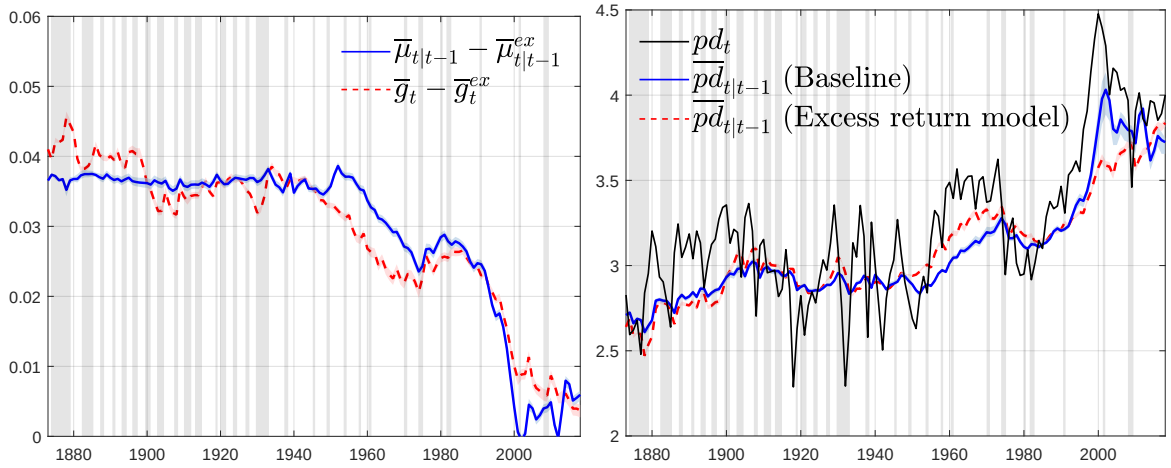
I Additional Results

Figure I.1: Term Structure of Expected Returns and Dividend Growth



Note. Figure I.1 plots the term structure of expected return and dividend growth.

Figure I.2: Steady State Comparisons



Note. The left panel of Figure I.2 reports two alternative measures of the long-run riskless real rate that we recover from the estimates in section 5. 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 I.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 I.1 reports parameter estimates for the model estimated in section 5. First column: autoregressive coefficients of expected returns and expected dividend growth (ϕ_μ and ϕ_g) and 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. σ^2_ν is the 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 (4) where A and B are diagonal matrices, 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*, 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., 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.

Buccheri, G., Bormetti, G., Corsi, F., and Lillo, F. (2020). A Score-Driven Conditional Correlation Model for Noisy and Asynchronous Data: an Application to High-Frequency Covariance Dynamics. *Journal of Business Economic Statistics*, forthcoming.

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

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

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.

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.

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 timevarying parameter VARs: A nonparametric approach. *Journal of Applied Econometrics*, 34(7):1027–1049.

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

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.

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.

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.

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