Research
Education
Outreach
CCA

UNIVERSITÀ
DI TORINO

Dividend momentum and stock return predictability: a Bayesian approach

Juan Antolin-Diaz
Ivan Petrella
Juan F. Rubio Ramirez

No. 26
September 2026

LTI WORKING PAPERS

www.carloalberto.org/lti/working-papers

Dividend Momentum and Stock Return Predictability: A Bayesian Approach*

Juan Antolín-Díaz
Massachusetts Institute of Technology

Ivan Petrella
Collegio Carlo Alberto
University of Turin

Juan F. Rubio-Ramírez†
Emory University
Federal Reserve Bank of Atlanta

September 9, 2025

Abstract

A long tradition in macro-finance studies the dynamics of aggregate stock returns and dividends using vector autoregressions, imposing the restrictions implied by the Campbell-Shiller (CS) identity to sharpen inference. We develop Bayesian methods that encode a priori skepticism about return predictability while imposing the restrictions. We highlight that persistence in dividend growth induces "dividend momentum," a previously overlooked channel for return predictability. By combining Bayesian shrinkage and the CS restrictions, we obtain more plausible degrees of return predictability, superior out-of-sample forecasts, and Sharpe ratios, which cannot be obtained by using either shrinkage or the CS restrictions on their own.

JEL Classification Numbers: C32, C53, G11, G12, E47.

*We are grateful to Svetlana Bryzgalova, John Y. Campbell, Gino Cenedese, John H. Cochrane, Domenico Giannone, Francisco Gomes, Marcin Kacperczyk, Jiacu Li (discussant), Pavol Povala, Hélène Rey, Vania Stavrakeva, and seminar participants at the Federal Reserve Bank of Atlanta, Federal Reserve Bank of Philadelphia, Boston College, Rice University, Bank of Canada, Toulouse School of Economics, Indiana University, University of Pennsylvania, London Business School, Warwick Business School, and Western Finance Association conference for helpful comments and suggestions. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

†Corresponding author: Juan F. Rubio-Ramírez <juan.rubio-ramirez@emory.edu>, Economics Department, Emory University, Rich Memorial Building, Room 306, Atlanta, Georgia 30322-2240.

1 Introduction

The predictability of aggregate stock returns remains one of the key empirical questions in financial economics. Early influential work by Campbell and Shiller (1988a,b) and Cochrane (2008b) employed vector autoregressions (VARs) that exploit the identity linking returns, dividend growth, and the price-dividend ratio to establish that valuation ratios forecast returns at long horizons. This finding has important implications for our understanding of time variation in discount rates. It is also critical for the optimal dynamic allocation problem faced by long-horizon investors (see Campbell and Viceira, 2002; Campbell et al., 2003). However, the practical relevance of return predictability has been undermined by its failure to produce robust out-of-sample forecasts (Viceira, 2008; Goyal and Welch, 2008). As a consequence, the literature has devoted considerable attention to Bayesian predictive regressions that incorporate prior beliefs weighted against return predictability (Kandel and Stambaugh, 1996; Stambaugh, 1999; Wachter and Warusawitharana, 2009, 2015; Pástor and Stambaugh, 2009, 2012).

In this paper, we revisit stock return predictability in a dynamic multivariate setting by proposing a Bayesian approach to inference that explicitly incorporates Campbell-Shiller (CS) cross-equation restrictions within a VAR framework that jointly models returns, dividend growth, and the price-dividend ratio. Our key methodological innovation is the combination of exact enforcement of the CS identity with Bayesian shrinkage priors that reflect skepticism about return predictability, all within a joint multivariate system. Standard Bayesian conjugate priors cannot be directly applied to models with linear restrictions on covariance matrices as implied by CS; hence, we develop a scalable importance sampling algorithm to generate posterior distributions under these constraints. This approach substantially improves inference precision, avoids the common misspecification that results from excluding dividend growth from the VAR, and provides a robust estimation framework applicable even in higher-dimensional settings. Our results show that combining empirical shrinkage with enforcement of CS constraints is a promising approach for studying predictability in small samples.

Applying our method to S&P 500 data from 1947 to 2024, we uncover a previously overlooked channel of predictability that we label "dividend momentum." We document that dividend growth displays significant autocorrelation even after controlling for valuation ratios, implying persistent

positive shocks that simultaneously boost dividends, realized returns, and future expected returns. These results hold not only for annual data but also for quarterly and monthly systems, where we explicitly model seasonality in dividend growth. Economically, dividend momentum differs fundamentally from the standard return momentum observed at higher frequencies (see, e.g., Moskowitz et al., 2012). Rather than reflecting short-term sentiment or behavioral biases, dividend momentum captures a persistent improvement in expected economic conditions that simultaneously raises current dividends, contemporaneous returns, and future expected returns. In fact, our analysis indicates that the dividend growth shock associated with the dividend momentum channel is linked to prolonged boom–bust cycles in economic activity, which first manifest as an increase in dividend growth and only later appear in broader measures of economic performance.

Crucially, we show that combining Bayesian shrinkage toward limited predictability with the exact CS restrictions significantly enhances out-of-sample forecasting performance. In stark contrast to prior literature, which generally finds negligible or negative out-of-sample predictive performance, our approach consistently outperforms naive benchmarks, achieving approximately 30 percent improvements at the five-year horizon. Moreover, in asset allocation exercises, incorporating dividend momentum into investment strategies leads to more realistic and less volatile equity allocations, resulting in economically meaningful improvements in risk-adjusted portfolio returns.

Relation to the literature Our paper contributes to the extensive literature on return predictability using valuation ratios (Fama and French, 1988; Lewellen, 2004), in particular the classic VAR analyses of Campbell and Shiller (1988a,b); Cochrane (2008b) and their implications for dynamic asset allocation (Campbell and Viceira, 1999; Campbell et al., 2003). Within a VAR framework, we uncover a novel source of return predictability—dividend momentum—stemming from autocorrelation in dividend growth, and examine its implications for portfolio choice. Our findings complement studies that document persistent expected dividend growth using latent-variable systems (Van Binsbergen and Koijen, 2010; Chen et al., 2012; Koijen and Nieuwerburgh, 2012). Unlike the latent-variable approach, which imposes additional assumptions on dividend-return dynamics and correlation structures (Cochrane, 2008a), we rely solely on the minimal restrictions implied by the CS identity, allowing dividend momentum to emerge naturally from the data. Dividend momentum also differs from dividend persistence as emphasized in the long-run risk literature (Bansal and Yaron,

2004; Schorfheide et al., 2018), where the price-dividend ratio fully captures dividend persistence. Instead, our finding aligns with consumption-CAPM frameworks that incorporate time-varying expected dividend growth, such as Menzly et al. (2004).

Methodologically, our approach is closely related to Bayesian studies that use informative priors to study return predictability (Kandel and Stambaugh, 1996; Stambaugh, 1999; Avramov, 2002; Wachter and Warusawitharana, 2009, 2015; Pástor and Stambaugh, 2009, 2012). However, following Cochrane (2008b), we depart from traditional predictive regressions and adopt a general VAR framework, analyzing returns, the price-dividend ratio, and dividend growth as a system whose dynamics are jointly constrained by the CS identity. We also build on the broader Bayesian econometrics literature on informative priors for VARs (Doan et al., 1984; Sims, 1993; Del Negro and Schorfheide, 2004; Giannone et al., 2015, 2019), extending these methods to accommodate dynamic linear restrictions that arise naturally in asset pricing (Campbell and Shiller, 1988a,b). In doing so, our methods generalize to many applications in macro-finance where similar identities appear—such as those linking bond returns and interest rates (see Campbell et al., 1997); consumption, wealth, and returns (see Campbell and Mankiw, 1989; Lettau and Ludvigson, 2001; Gourinchas and Rey, 2019); interest rate differentials and the real exchange rate (see Engel and West, 2005; Froot and Ramadorai, 2005; Engel, 2016); net foreign assets, net exports, and foreign portfolio returns (see Gourinchas and Rey, 2007); or government debt, surpluses, and interest rates (see Cochrane, 2019; Jiang et al., 2024)—all of which feature stationary but highly persistent predictors of asset returns.

2 A VAR model with Campbell-Shiller restrictions

In this section, we describe a basic macro-finance VAR model and the cross-equation restrictions among the VAR coefficients implied by CS identity for returns. We discuss a common practice of omitting one of the variables as means to impose the CS restrictions. Finally, we reconcile the VAR and the latent variables approach proposed by Van Binsbergen and Koijen (2010).

2.1 The Basic Macro-Finance VAR

The classic macro-finance VAR approach of Campbell and Shiller (1988a,b) studies the joint dynamics of log dividend growth, the log price-dividend ratio, and log returns (see also Campbell, 1991;

Cochrane, 2008b). We focus on a minimalistic VAR that contains only the three key variables above and uses only one lag. However, none of this paper's conclusions will be affected if we add more variables or lags to the VAR. The VAR(1) in $\mathbf{y}_t' = [\Delta d_{t+1}, pd_{t+1}, r_{t+1}]$ is written:

$$\underbrace{\begin{bmatrix} \Delta d_{t+1} \\ pd_{t+1} \\ r_{t+1} \end{bmatrix}}_{\mathbf{y}_{t+1}} = \underbrace{\begin{bmatrix} c^d \\ c^{pd} \\ c^r \end{bmatrix}}_{\mathbf{\Phi}_0} + \underbrace{\begin{bmatrix} \phi_{d,d} & \phi_{d,pd} & \phi_{d,r} \\ \phi_{pd,d} & \phi_{pd,pd} & \phi_{pd,r} \\ \phi_{r,d} & \phi_{r,pd} & \phi_{r,r} \end{bmatrix}}_{\mathbf{\Phi}_1} \underbrace{\begin{bmatrix} \Delta d_t \\ pd_t \\ r_t \end{bmatrix}}_{\mathbf{y}_t} + \underbrace{\begin{bmatrix} u^d_{t+1} \\ u^{pd}_{t+1} \\ u^r_{t+1} \end{bmatrix}}_{\mathbf{u}_{t+1}} \tag{1}$$

where $\mathbf{u}_{t+1}$ is Gaussian with $\mathbb{E}(\mathbf{u}_{t+1}) = \mathbf{0}_{3\times 1}$, $\mathbb{E}(\mathbf{u}_{t+1}\mathbf{u}_{t+1}') = \mathbf{\Sigma}$, $\mathbb{E}(\mathbf{u}_s\mathbf{u}_r') = \mathbf{0}_{3\times 3}$ if $r \neq s$, and $\mathbf{\Sigma}$ is a symmetric and positive semidefinite (SPD) matrix and where $\mathbf{0}_{m\times\tilde{m}}$ is a matrix of zeros of dimension $m \times \tilde{m}$. This system can be written as $\mathbf{y}_{t+1} = \mathbf{\Phi}\mathbf{z}_t' + \mathbf{u}_{t+1}$ where $\mathbf{z}_t = [1, \mathbf{y}_t']$ and $\mathbf{\Phi} = [\mathbf{\Phi_0}, \mathbf{\Phi_1}]$. We define $\boldsymbol{\mu} = (\mathbf{I}_n - \mathbf{\Phi}_1)^{-1}\mathbf{\Phi}_0$ as the unconditional mean of the variables. More compactly, and in general, $\mathbf{Y} = \mathbf{Z}\mathbf{\Phi}' + \mathbf{U}$, where T is the length of the sample, n the number of variables, and p the number of lags in the VAR, $\mathbf{Y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$ is a $T \times n$ matrix, $\mathbf{Z} = (\mathbf{z}_1', \ldots, \mathbf{z}_T')'$ is a $T \times K$ matrix, where $K = np + 1$, and $\mathbf{U} = (\mathbf{u}_1', \ldots, \mathbf{u}_T')'$ is a $T \times n$ matrix.

2.2 The Campbell-Shiller Restrictions

As first noted by Campbell and Shiller (1988a,b), if one assumes that the price-dividend ratio is stationary, log-linearizing the definition of return, $R_{t+1} = (P_{t+1} + D_{t+1})/P_t$, around the mean of the log price-dividend ratio yields an approximate identity that links log returns, log dividend growth, and changes in the log price-dividend ratio:

$$r_{t+1} \approx \kappa + \rho pd_{t+1} - pd_t + \Delta d_{t+1} \tag{2}$$

where κ and ρ are constants of approximation that depend on the steady-state log dividend-price ratio.[1] Being derived from a definition, the Campbell-Shiller (CS) identity holds very tightly in the data. It holds with equality if one adds an approximation error, denoted η_{t+1}. In our data, the log-linear approximation is accurate enough that η_{t+1} is very small, but this term can capture additional measurement error if, as is common in the literature, a smoothed price-dividend ratio

[1] We will refer to the log dividend growth, the log price-dividend ratio, and log returns as dividend growth, the price-dividend ratio, and returns, except when strictly necessary.

or a price-earnings ratio is used instead. In that case, it is easy to show that the CS identity in Equation (2) imposes the following restriction among the innovations:

$$u^r_{t+1} = u^d_{t+1} + \rho u^{pd}_{t+1} + \eta_{t+1}, \tag{3}$$

where $\mathbb{E}\left(u^d_s \eta_r\right) = \mathbb{E}\left(u^{pd}_s \eta_r\right) = 0$ for all r, s. The CS identity implies restrictions among the 3-variable VAR(1) coefficients in Equation (1). In particular, it imposes (linear) restrictions on $\mathbf{\Phi}$ and $\mathbf{\Sigma}$. The restrictions for $\mathbf{\Phi}$ are:

$$\begin{aligned} c^r &= c^d + \rho c^{pd} + \kappa, & (4)\\ \phi_{r,d} &= \phi_{d,d} + \rho\phi_{pd,d}, & (5)\\ \phi_{r,pd} &= \phi_{d,pd} + \rho\phi_{pd,pd} - 1, & (6)\\ \phi_{r,r} &= \phi_{d,r} + \rho\phi_{pd,r}, & (7) \end{aligned}$$

whereas the restrictions for $\mathbf{\Sigma}$ are:

$$\mathtt{Cov}\left(u^r_{t+1}, u^d_{t+1}\right) = \rho\mathtt{Cov}\left(u^{pd}_{t+1}, u^d_{t+1}\right) + \mathtt{Var}\left(u^d_{t+1}\right) \text{ and} \tag{8}$$

$$\mathtt{Cov}\left(u^r_{t+1}, u^{pd}_{t+1}\right) = \rho\mathtt{Var}\left(u^{pd}_{t+1}\right) + \mathtt{Cov}\left(u^d_{t+1}, u^{pd}_{t+1}\right). \tag{9}$$

If we do not consider approximation error, we will have an additional restriction for $\mathbf{\Sigma}$:

$$\mathtt{Var}\left(u^r_{t+1}\right) = \mathtt{Var}\left(u^d_{t+1}\right) + \rho^2\mathtt{Var}\left(u^{pd}_{t+1}\right) + 2\rho\mathtt{Cov}\left(u^d_{t+1}, u^{pd}_{t+1}\right). \tag{10}$$

Additionally, the derivation of the CS identity in Equation (2) requires the system to be stationary, and in particular, that the price-dividend ratio has a well-defined steady state. This is a restriction on the eigenvalues of the matrix $\mathbf{\Phi}_1$, which we write:

$$\mathbf{\Phi}_1 \in \left\{\mathbf{Z} \in \mathbb{R}^{3\times 3} : \max\{\mathtt{eig}\left(\mathbf{Z}\right)\} < 1\right\}. \tag{11}$$

This generalizes Cochrane's (2008b) requirement of an upper bound to the persistence of the price-dividend ratio to a multivariate setting. The CS restrictions (4)-(9) and the stationarity

restriction (11) are cross-equation restrictions on the 3-variable VAR(1) in Equation (1). Because the former restrictions are only valid if the latter is satisfied, from this point on, when we impose the CS restrictions (4)-(9), we impose the stationarity restriction (11) as well.

Similar log-linear identities are pervasive in the macro-finance literature, linking, e.g., bond returns and interest rates (see Campbell et al., 1997); consumption, wealth, and returns (see Campbell and Mankiw, 1989; Lettau and Ludvigson, 2001); interest rate differentials and the real exchange rate (Engel and West, 2005; Froot and Ramadorai, 2005; Engel, 2016); net foreign assets, net exports, and the return on the foreign asset portfolio (Gourinchas and Rey, 2007); or government debt, surpluses, and interest rates (Cochrane, 2019; Jiang et al., 2024). They all imply similar dynamic linear restrictions among endogenous variables, which can be expressed as linear restrictions on the coefficients and the covariance matrix of a VAR.

2.3 Omitting dividend growth

Because an identity links the three variables, one of the three *equations* in VAR (1) is redundant, as discussed, e.g., by Cochrane (2017), who states: "The definition of return means that only two of the three equations are needed, and the other one follows." This has been often interpreted in the literature as being able to drop one of the three *variables* (usually dividend growth) and estimate a 2-variable VAR(1) in returns and the price-dividend ratio only.[2] For instance, Engsted et al. (2012) criticize Chen and Zhao (2009) for excluding the dividend-price ratio from the system but state (p.1262) that "nothing is gained by modeling both returns and dividend growth in a system that also contains the dividend–price ratio." In fact, dropping one of the variables in a VAR(1) amounts to imposing extra restrictions. To see this clearly, rewrite the model as

$$\begin{bmatrix} \Delta d_{t+1} \\ \mathbf{x}_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_{d,d} & \boldsymbol{\phi}_{21} \\ \boldsymbol{\phi}_{12} & \boldsymbol{\Phi}_{11} \end{bmatrix} \begin{bmatrix} \Delta d_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} u^d_{t+1} \\ \boldsymbol{\xi}_{t+1} \end{bmatrix}, \tag{12}$$

[2]The common practice of dropping one of the variables (typically dividend growth) and then retrieve the omitted variable and associated coefficients from the CS restrictions (4)-(9) is followed by almost all studies looking at the relationship between returns, dividend growth, and the price-dividend ratio. See, for instance, Campbell and Viceira (1999); Campbell et al. (2001); Cochrane (2008b, 2011); Avramov et al. (2018). Notable exceptions are Campbell and Shiller (1988a), who find some weak evidence of persistence in dividend growth, and Larrain and Yogo (2008), who use the GMM to estimate a system without omitting variables but cite the latter as an equivalent alternative. The practice of dropping dividend growth from VAR systems featuring returns and the price-dividend ratio is also prevalent in studies featuring additional predictors of excess returns (see, e.g., Campbell, 1991; Campbell and Ammer, 1993; Campbell et al., 2003; Campbell and Vuolteenaho, 2004; Campbell et al., 2013).

where $\mathbf{x}_{t+1} = [pd_{t+1}, r_{t+1}]'$ and the constant terms are omitted without loss of generality. Assuming that the CS identity (2) holds without error, we can write $\Delta d_t = -\kappa - \rho pd_t + pd_{t-1} + r_t$, which can be replaced into (12) to obtain

$$\mathbf{x}_{t+1} = \left(\mathbf{\Phi}_{11} + \boldsymbol{\phi}_{12}(\iota_2' - \rho\iota_1')\right)\mathbf{x}_t + \left(\boldsymbol{\phi}_{12}\iota_1'\right)\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1}, \tag{13}$$

where ι_j is a selection vector corresponding to the j-th column of the identity matrix. It is easy to see that assuming that the true dynamics are represented by the VAR (1) (or (12)) together with the CS identity (2), opting for the common practice of estimating a bivariate VAR(1) on $\mathbf{x}_{t+1}$, $\mathbf{x}_{t+1} = \mathbf{A}_1\mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1}$ implies the implicit imposition of the following extra restrictions on dividend growth dynamics:

$$u^d_{t+1} = u^r_{t+1} - \rho u^{pd}_{t+1} \quad \text{(No approximation error)} \tag{14}$$

$$\boldsymbol{\phi}_{12} = \mathbf{0} \quad \text{(Dividend growth not a relevant regressor)} \tag{15}$$

$$\phi_{d,d} = \mathbf{0} \quad \text{(Follows from } \boldsymbol{\phi}_{12} = 0 \text{ and CS Restriction (6))} \tag{16}$$

We provide a formal proof with more details and implications in Appendix A. Still, another intuitive way to understand this result is to perform a simple parameter count: the autoregressive matrix $\mathbf{\Phi}_1$ of the 3-variable VAR(1) contains nine parameters. However, the three CS restrictions (5)-(7) imply that there are only six free parameters. A 2-variable VAR(1) will feature an autoregressive matrix with only four parameters, thus imposing two extra restrictions, namely $\boldsymbol{\phi}_{12} = \mathbf{0}$. Another way of saying this is that Δd_t can be dropped from the set of *regressands*, but not from the set of *regressors* from the system in Equation (1).[3]

Examining Equation (13), one might think that an alternative could be to add an extra lag and specify a 2-variable VAR(2) in pd_{t+1} and r_{t+1}. This successfully spans the information contained

[3]A related question is whether including the three variables in the system can lead to a collinearity problem. Collinearity problems only appear when using more than one lag and when the CS identity holds exactly. While we will focus our exposition on the 3-variable VAR(1) where collinearity is not a problem, collinearity is generally never a concern when using Bayesian estimation methods (see Leamer, 1973). Therefore, we will allow for systems with any lag length when discussing our estimation methods in Section 3.3. Avramov et al. (2018) also claim that the 3-variable VAR(1) "must be estimated with observation equations for only two of the [...] variables to ensure that the covariance matrix $\mathbf{\Sigma}$ is nonsingular." Yet, as highlighted by Cochrane (2008a,b), the CS identity is accurate but only approximate. Therefore, singularity is never truly a concern. Despite that, in Section 3.3, we develop inference methods that can also handle the possibility of a singular covariance matrix.

in dividend growth but features two autoregressive matrices with four parameters each, totaling eight parameters. Therefore, this approach leads to overparametrization, not imposing all available restrictions. As we shall see, there will be gains to be made, both in terms of economic interpretation and forecasting performance, by modeling the three variables as a system in the VAR in Equation (1) and imposing the CS restrictions exactly whenever dividend growth features autocorrelation after controlling for the lagged dividend-price ratio, i.e. $\phi_{d,d} \neq 0$, a phenomenon we label "dividend momentum".

2.4 Relation to the Latent Variables Approach

Before moving to the empirical analysis, we reconcile the VAR approach discussed in this section with the latent variables approach proposed by Van Binsbergen and Koijen (2010). This approach partitions returns and dividend growth into expected and unexpected components. Specifically, they write returns and dividend growth as:

$$r_{t+1} - \mu_r = f_{r,t} + e_{r,t+1}, \tag{17}$$

$$\Delta d_{t+1} - \mu_g = f_{g,t} + e_{d,t+1}, \tag{18}$$

where $\mathbf{f}_{t+1} = [f_{r,t+1}, f_{g,t+1}]'$ is the state vector with the following law of motion

$$\mathbf{f}_{t+1} = \mathbf{F}_1 \mathbf{f}_t + \boldsymbol{\upsilon}_{t+1}, \tag{19}$$

and $\boldsymbol{\upsilon}_{t+1} = [\upsilon_{r,t+1}, \upsilon_{g,t+1}]'$ and the price-dividend ratio is related to the state vector as follows

$$pd_{t+1} - \mu_{pd} = [-1, 1] \left(\mathbf{I}_2 - \rho \mathbf{F}_1\right)^{-1} \mathbf{f}_{t+1}, \tag{20}$$

We provide a novel proof in Appendix B that whenever the CS identity holds exactly, the VAR system in Equation (1) admits a unique latent present value system representation with the feedback matrix $\mathbf{F}_1$ and the covariance matrix of the innovations $[e_{r,t+1}, e_{d,t+1}, \upsilon_{r,t+1}, \upsilon_{g,t+1}]$ functions of the underlying VAR parameters. There are two important caveats: first, the matrix $\mathbf{F}_1$ will generally be not diagonal; second, the correlation between innovations to expected and

unexpected dividend growth, i.e. $corr(e_{d,t+1}, \upsilon_{g,t+1})$, need not be zero.[4] Both of these assumptions, commonly employed by the literature (Van Binsbergen and Koijen, 2010; Rytchkov, 2012), are again additional restrictions that do not come from the CS identity, and the latter in particular rules out the phenomenon of dividend momentum which we will describe in our empirical results. We thus provide a general proof of the claim made by Cochrane (2008a) that the commonly specified latent present value system imposes additional restrictions on the dynamics of returns and dividends beyond those implied solely by the CS identity.

3 Bayesian Inference under Present Value Restrictions

In this section, we describe how to perform Bayesian inference considering the CS restrictions described above. We begin by describing why we need informative priors and how to incorporate the CS restrictions into those. We finalize by describing an algorithm that allows us to draw from any posterior distribution conditional on the CS restrictions.

3.1 The Need for Informative Priors

Inference under flat priors in VAR models is problematic for many reasons. Their dense parameterization leads to unstable inference and inaccurate out-of-sample forecasts, particularly for models with many variables and lags (see, e.g., Doan et al., 1984; Giannone et al., 2015). The CS restrictions bring VARs with many lags close to multicollinearity, complicating flat-prior estimation. Moreover, one can show that flat priors imply a prior distribution over the one-period-ahead R^2 of returns that collapses to one (Giannone et al., 2021, 2022), whereas from an economic point of view, parameter combinations that imply very high degrees of return predictability should be a priori implausible (see Kandel and Stambaugh, 1996; Wachter and Warusawitharana, 2015). Thus, one needs informative priors that represent the beliefs of conservative observers who are skeptical about return predictability, in line with the proposal of Wachter and Warusawitharana (2009) and Pástor and Stambaugh (2009, 2012).

[4] Specifically, let $\mathbf{V}$ denote the diagonal matrix of the eigenvalues (in descending order) and $\mathbf{E}$ the associated eigenvectors of the matrix $\mathbf{\Phi}_1$. Then, we have $\mathbf{F}_1 = \mathbf{K}^{-1}\mathbf{J}_1\mathbf{V}\mathbf{J}_1'\mathbf{K}$, where $\mathbf{J}_1 = [\mathbf{I}_2 \ \mathbf{0}_{2\times 1}]$, and $\mathbf{K}^{-1} = \mathbf{J}_2\mathbf{E}\mathbf{V}\mathbf{J}_1'$, with $\mathbf{J}_2' = [\boldsymbol{\iota}_1 \ \boldsymbol{\iota}_3]$ and $\boldsymbol{\iota}_s$ denoting a 3×1 selection vector whose only non-zero element is 1, located in position s. Furthermore, the innovations of the system, $[e_{r,t+1}, e_{d,t+1}, \upsilon_{r,t+1}, \upsilon_{g,t+1}]$, are linear functions of the innovations of the VAR: $e_{d,t+1} = u^d_{t+1}$, $e_{r,t+1} = u^d_{t+1} + \rho u^{pd}_{t+1}$, and $[\upsilon_{g,t+1}, \upsilon_{r,t+1}]' = \mathbf{K}\mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{H}}[u^d_{t+1}, u^{pd}_{t+1}]'$, with $\widetilde{\mathbf{H}} = [1, 0, 1; 0, 1, \rho]'$.

The high persistence of the price-dividend ratio is also concerning. The presence of downward bias in OLS estimates of autoregressive parameters when these are close to unit root has been known since Hurwicz (1950) (see also Bekaert et al., 1997). Stambaugh (1999) further notes that when a persistent predictor features innovations that are highly correlated with the innovations to the predicted variable, as is the case with the price-dividend ratio and returns, this translates into OLS estimates of the regression coefficient that is biased away from zero. So not only is $\phi_{pd,pd}$ biased downward, implying strong mean reversion, but $\phi_{r,pd}$ is also biased away from zero, meaning too much return predictability. Flat priors put excessive weight on parameter combinations below one for $\phi_{pd,pd}$, and therefore on high amounts of predictability. The bias and excess predictability problems will worsen as additional persistent predictors are added.[5]

Finally, the flat priors combined with the standard use of conditional likelihood also imply that the initial values of the data are far away from their unconditional mean. This implies an implausibly good forecasting power of initial conditions and determinist components (see Sims and Uhlig, 1991; Stambaugh, 1999; Jarocinski and Marcet, 2015; Giannone et al., 2019). In practice, VAR deterministic components over-fit the low-frequency variation in the data, which worsens as lags or additional variables are included in the system.

To solve all these issues, one can consider informative priors. While our methods will work with any prior the researcher chooses, we set priors which are standard and have been widely used in macroeconomics over the last 30 years (see, e.g., Giannone et al., 2015). In particular, we combine the "Minnesota" prior originally proposed by Doan et al. (1984) with the Single Unit Root prior proposed by Sims (1993) and Sims and Zha (1998). These priors handle downward bias and excess predictability. In particular $p(\mathtt{vec}(\mathbf{\Phi}_1)|\mathbf{\Sigma}) \sim \mathcal{N}\left(\begin{bmatrix}0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\end{bmatrix}', \mathbf{\Sigma} \otimes \mathbf{\Omega}\right)$, where in the simplest case $\mathbf{\Omega} = \lambda^2(diag\left(\left[\underline{\sigma}_d^2, \underline{\sigma}_{pd}^2, \underline{\sigma}_r^2\right]\right))^{-1}$, with λ a positive scalar controlling the tightness of the prior, and $\underline{\sigma}_i^2$ an a priori estimate of the standard deviation of each variable's innovation. As desired, the Minnesota prior pushes $\phi_{pd,pd}$ toward one, and both $\phi_{r,r}$ and $\phi_{d,d}$ toward zero. The Minnesota prior is usually specified as flat for $\mathbf{\Phi}_0$. The prior for $\mathbf{\Sigma}$ is set to $p(\mathbf{\Sigma}) \sim \mathcal{IW}(diag\left(\left[\underline{\sigma}_1^2, \ldots, \underline{\sigma}_n^2\right]\right), n+2)$.

The Single Unit Root prior addresses the problem of the excessive explanatory power of initial

[5]While, in principle, the small sample bias issue could be tackled by applying bias correction to the VAR coefficient, doing that while simultaneously imposing the CS restrictions is not possible unless one is willing to follow the common approach of dropping one of the variables from the system, which is not an option as we discuss in Section 2.1.

conditions and deterministic components. Stambaugh (1999) has also emphasized the importance of properly handling the initial observation in predictive return regressions. The Single Unit Root prior is usually implemented by appending an artificial ("dummy") observation for $\mathbf{y}$ and $\mathbf{x}$: $y^*_{1\times n} \equiv \underline{\boldsymbol{\mu}}/\theta$ and $x^*_{1\times(n\cdot p+1)} \equiv [1/\theta, y^*, \ldots, y^*]$, at the beginning of the sample. Where $\underline{\boldsymbol{\mu}}$ can be used to express a prior on the unconditional mean of the variables $\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Phi}_0$, and a scalar hyperparameter θ controls the tightness of the prior. The Single Unit Root prior has the significant advantage of being a joint prior on the entire system $[r_{t+1}, pd_{t+1}, \Delta d_{t+1}]$, governing the long-run behavior of the variables in a consistent manner, influencing simultaneously the intercepts and the slope coefficients of the VAR by imposing a prior on the steady state as well as the joint stationarity of the system. The ability to impose Bayesian priors on the system as a whole is an additional reason to seek an approach of estimating the system jointly without dropping any variable or equation.

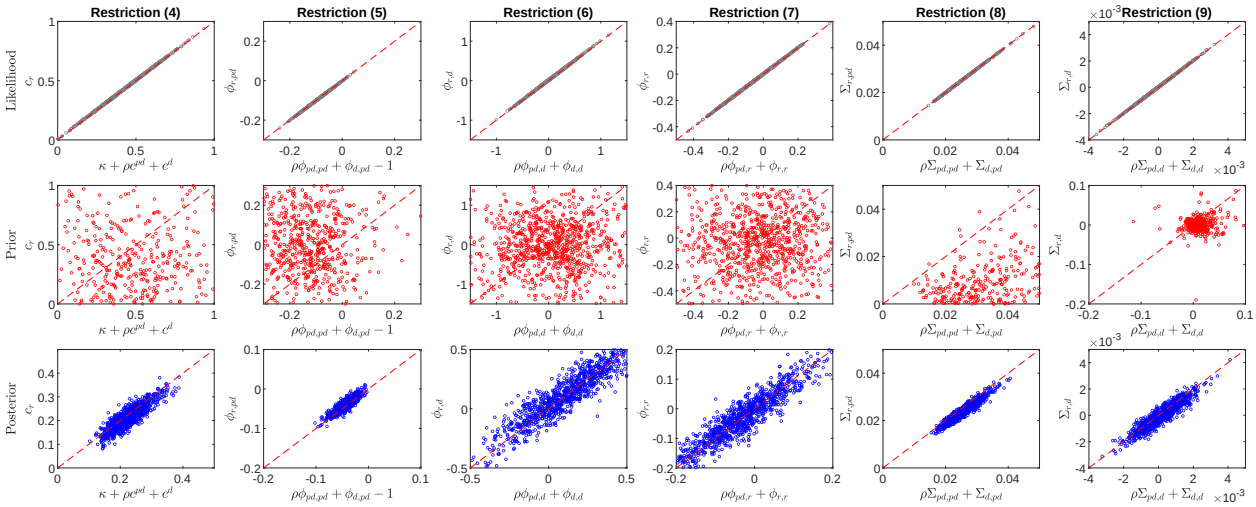
3.2 Informative Priors and the CS Restrictions

General informative priors like the one above may violate the CS restrictions, even if centered around them, since they assign positive probability to parameter values that do not satisfy the constraints. As the prior becomes more informative, the posterior is increasingly likely to breach the restrictions, even if the likelihood satisfies them. Figure 1 illustrates this using the data and priors from Section 4.1. Each column corresponds to one CS restriction (4)–(9), with rows showing draws from the likelihood, prior, and posterior, respectively. For example, the first panel plots c^r against $\rho c^{pd} + c^d$; under the CS restrictions, all points should lie on the 45-degree line. While the likelihood closely respects the restrictions, the prior and posterior do not. We now develop a method to sample from any posterior while enforcing the CS restrictions on $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$.

3.3 Bayesian Estimation under CS Restrictions

This subsection presents our main methodological contribution: a Bayesian algorithm for independently sampling from the posterior of a VAR with informative priors, while exactly imposing the CS restrictions without dropping variables or adding extra constraints. The key challenge is drawing $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ under these restrictions. We use the conjugate Normal-Inverse Wishart (NIW) framework to enable independent draws. Conjugacy and independence are especially useful as they make our approach useful for VARs with many more variables and lags, though our approach extends

Figure 1: CS Restrictions for $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ under informative priors



Note: Red represents the prior, grey is the likelihood, and blue is the posterior.

beyond NIW priors. Under the CS restrictions, the posterior of $\mathbf{\Phi}$ remains Normal, but that of $\mathbf{\Sigma}$ is no longer Inverse Wishart. Therefore, we will propose an importance sampler to draw from the restricted posterior. Readers less interested in the technical details of the algorithm may skip to the empirical results in Section 4.

The conjugate prior and posterior

We will write our algorithm independently drawing from the conjugate family of NIW posterior distributions conditional on the CS restrictions. The NIW family of distributions is the conjugate prior for this class of models. If the prior distribution over the parameters is $NIW(\underline{\nu}, \underline{\mathbf{S}}, \underline{\boldsymbol{\alpha}}, \underline{\mathbf{V}})$, then the posterior distribution over the parameters is $NIW(\overline{\nu}, \overline{\mathbf{S}}, \overline{\boldsymbol{\alpha}}, \overline{\mathbf{V}})$, where $\overline{\boldsymbol{\alpha}} = \mathtt{vec}\left(\overline{\mathbf{A}}\right)$, $\overline{\mathbf{V}} = \left(\underline{\mathbf{V}}^{-1} + \mathbf{Z}'\mathbf{Z}\right)^{-1}$, $\overline{\mathbf{A}} = \overline{\mathbf{V}}\left(\underline{\mathbf{V}}^{-1}\underline{\mathbf{A}} + \mathbf{Z}'\mathbf{Z}\hat{\mathbf{A}}\right)^{-1}$, $\hat{\mathbf{A}} = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}$, and $\overline{\mathbf{S}} = \hat{\mathbf{S}} + \underline{\mathbf{S}} + \hat{\mathbf{A}}'\mathbf{Z}'\mathbf{Z}\hat{\mathbf{A}} + \underline{\mathbf{A}}'\underline{\mathbf{V}}^{-1}\underline{\mathbf{A}} - \overline{\mathbf{A}}'\overline{\mathbf{V}}^{-1}\overline{\mathbf{A}}$, $\hat{\mathbf{S}} = \left(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{A}}\right)'\left(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{A}}\right)$, and $\overline{\nu} = T + \underline{\nu}$. The NIW posterior distributions defined above can be factored into the following conditional and marginal posterior distributions $\mathcal{N}\left(\overline{\boldsymbol{\alpha}}, \mathbf{\Sigma} \otimes \overline{\mathbf{V}}\right)$ and $\mathcal{IW}\left(\overline{\mathbf{S}}, \overline{\nu}\right)$. This structure allows us to draw from the posterior independently.

For the case of the NIW posterior, we want to draw from the restricted normal posterior distribution of $\boldsymbol{\alpha}$ conditional on $\mathbf{\Sigma}$ and from the restricted inverse-Wishart (IW) posterior distribution of $\mathbf{\Sigma}$. By the restricted normal posterior distribution of $\boldsymbol{\alpha}$ conditional on $\mathbf{\Sigma}$ we mean the distribution $\mathcal{N}\left(\overline{\boldsymbol{\alpha}}, \mathbf{\Sigma} \otimes \overline{\mathbf{V}}\right)$ conditional on the CS restrictions on $\mathbf{\Phi}$ and by the restricted IW posterior distribution

of $\boldsymbol{\Sigma}$ we mean the distribution $\mathcal{IW}\left(\overline{\mathbf{S}}, \overline{\nu}\right)$ conditional on the CS restrictions on $\boldsymbol{\Sigma}$. In the same spirit, we will call the distribution $NIW\left(\overline{\nu}, \overline{\mathbf{S}}, \overline{\boldsymbol{\alpha}}, \overline{\mathbf{V}}\right)$ conditional on the CS restrictions on $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ the restricted NIW posterior of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$. As we will see, an analytical expression exists for the restricted normal posterior distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\Sigma}$. This is not true for the restricted IW posterior distribution of $\boldsymbol{\Sigma}$. We will present a numerical algorithm to draw from it independently.

The Restricted Normal Posterior of $\boldsymbol{\alpha}$

The CS restrictions on $\boldsymbol{\Phi}$ are linear restrictions on $\boldsymbol{\alpha}$. As with any linear restriction on $\boldsymbol{\alpha}$, they can be written as $\mathbf{R}_{\Phi}\boldsymbol{\alpha} = \mathbf{r}_{\Phi}$. We follow Antolín-Díaz et al. (2021) and draw from the restricted normal posterior distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\Sigma}$, i.e. $\mathcal{N}\left(\tilde{\boldsymbol{\alpha}}, \widetilde{\mathbf{V}}\right)$ where $\tilde{\boldsymbol{\alpha}} = \overline{\boldsymbol{\alpha}} - \mathbf{F}\left(\mathbf{R}_{\Phi}\overline{\boldsymbol{\alpha}} - \mathbf{r}_{\Phi}\right)$, $\widetilde{\mathbf{V}} = \left(\mathbf{I}_{nK} - \mathbf{F}\mathbf{R}_{\Phi}\right)\left(\boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)\left(\mathbf{I}_{nK} - \mathbf{F}\mathbf{R}_{\Phi}\right)'$, and $\mathbf{F} = \left(\boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)\mathbf{R}_{\Phi}'\left(\mathbf{R}_{\Phi}\left(\boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)\mathbf{R}_{\Phi}'\right)^{-1}$. If $\mathcal{U}_{\Phi}$ denotes the set of all $\boldsymbol{\alpha}$ that satisfy the CS restrictions on $\boldsymbol{\Phi}$, any draw from $\mathcal{N}\left(\tilde{\boldsymbol{\alpha}}, \widetilde{\mathbf{V}}\right)$ will belong to $\mathcal{U}_{\Phi}$. As emphasized in Antolín-Díaz et al. (2021), this specific way of imposing linear restrictions on the vector of Gaussian parameters minimizes the Kullback-Leibler divergence from the unrestricted distribution while adhering to the constraints. In other words, this approach ensures that the restrictions are imposed while deviating minimally from the estimates obtained from an unrestricted system. We implement restriction (11) by discarding draws of the posterior that do not satisfy the stationary restriction. This truncates and re-normalizes the restricted normal posterior.

The Restricted IW Posterior of $\boldsymbol{\Sigma}$

We rely on simulation to independently draw from the restricted IW posterior distribution of $\boldsymbol{\Sigma}$. In this section, we describe the methods that we will use to accomplish that. We will show that the CS restrictions map to a set of orthogonality restrictions between the approximation error associated with the restrictions and a set of the original residuals. This allows us to design a simple algorithm that draws from the set of $\boldsymbol{\Sigma}$'s that satisfy the CS restrictions. However, the resulting draws are not from the desired restricted IW posterior distribution of $\boldsymbol{\Sigma}$. Therefore, we will use an importance sampler to accomplish our objective.

Although up to now we have considered only stock returns, with one associated CS identity, in this section we consider the general case of k assets (as in, e.g. Campbell et al., 2003), in which case there would be k associated CS identities. In this case, the CS restrictions on the $n \times 1$ vector of

innovations, $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{\Sigma})$, can be represented as k linear restrictions and $n-k$ orthogonality restrictions. The k linear restrictions are:

$$\mathbf{L}\mathbf{u}_t = \boldsymbol{\eta}_t \sim \mathcal{N}\left(\mathbf{0}_{(n-k)\times 1}, \mathbf{\Omega}\right) \tag{21}$$

where $\mathbf{L}$ is a given $k \times n$ matrix. Equation (21) identifies the $k \times 1$ vector of approximation errors associated with the CS restrictions, $\boldsymbol{\eta}_t$. For instance, for the case considered in Section 2.1, $k = 1$ and the CS restriction on the innovations represented by Equation (3) can be mapped into Equation (21) by specifying $\mathbf{L} = [-1, -\rho, 1]$.

Moreover, there are $n-k$ innovations that are orthogonal to the approximation errors:

$$\mathbb{E}\left(\mathbf{\Xi}\mathbf{u}_t\boldsymbol{\eta}_t'\right) = \mathbf{0}_{(n-k)\times k} \tag{22}$$

where $\mathbf{\Xi}$ is a given $(n-k)\times n$ selection matrix. For the case considered in Section 2.1 we have that $\mathbf{\Xi} = [\mathbf{I}_2, \mathbf{0}_{2\times 1}]$. Putting Equations (21) and (22) together, we obtain the result that the CS restrictions on $\mathbf{\Sigma}$ can be represented as $\mathbf{\Xi}\mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right)\mathbf{L}' = \mathbf{\Xi}\mathbf{\Sigma}\mathbf{L}' = \mathbf{0}_{(n-k)\times k}$. Vectorizing this equation implies the following linear restrictions:

$$\mathbf{R}_\Sigma \operatorname{vec}(\mathbf{\Sigma}) = \mathbf{0}_{(n-k)k\times 1} \tag{23}$$

where $\mathbf{R}_\Sigma = \mathbf{L} \otimes \mathbf{\Xi}$. Equation (23) appropriately imposes the CS restrictions on $\mathbf{\Sigma}$. For instance, for the case considered in Section 2.1, these are represented by Equations (8) and (9).

Define now $\mathbf{H} = [\mathbf{\Xi}', \mathbf{L}']'$, the linear transformation of the original innovations $\mathbf{H}\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \boldsymbol{W})$, and the following mapping between a $n \times n$ SPD matrix $\mathbf{W}$ and $\mathbf{\Sigma}$:

$$\mathbf{W} = \mathbf{H}\mathbf{\Sigma}\mathbf{H}', \tag{24}$$

where:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{12}' & \mathbf{W}_{22} \end{bmatrix}.$$

The above mapping shows that the CS restrictions on $\mathbf{\Sigma}$ hold if and only if $\mathbf{W}$ is block diagonal.

To see this, notice that on the one hand, the mapping implies that $\mathbf{W}_{12} = \mathbf{\Xi\Sigma L}'$. Hence, if Equation (23) holds, it is the case that $\mathbf{W}_{12} = \mathbf{0}_{(n-k)\times k}$. On the other hand, the inverse mapping implies that $\mathbf{R}_\Sigma \operatorname{vec}(\mathbf{\Sigma}) = \mathbf{R}_\Sigma \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \operatorname{vec}(\mathbf{W})$. Then, one can show that:

$$\mathbf{R}_\Sigma \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) = \left[\mathbf{0}_{k\times(n-k)} \otimes \mathbf{\Xi H}^{-1}, \mathbf{I}_k \otimes \mathbf{\Xi H}^{-1}\right] = \left[\mathbf{0}_{(n-k)k\times(n-k)n}, \mathbf{I}_{(n-k)k}, \mathbf{0}_{(n-k)k\times k^2}\right],$$

and $\mathbf{R}_\Sigma \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \operatorname{vec}(\mathbf{W}) = \operatorname{vec}(\mathbf{W}_{12})$. Thus, if $\mathbf{W}_{12} = \mathbf{0}_{(n-k)\times k}$, it is the case that Equation (23) holds. The result above is key to making independent draws from the set of all $\mathbf{\Sigma}$ satisfying the CS restrictions using the following algorithm.

Algorithm 1. *The following makes independent draws from a distribution over* $\mathbf{\Sigma}$ *conditional on the CS restrictions.*

1. *Draw* $\mathbf{W}_{11}$ *independently from the* $\mathcal{IW}\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}\right)$ *distribution.*

2. *Draw* $\mathbf{\Omega}$ *independently from the* $\mathcal{IW}\left(\mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}}\right)$ *distribution.*

3. *Set*

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{0}_{(n-k)\times k} \\ \mathbf{0}_{k\times(n-k)} & \mathbf{\Omega} \end{bmatrix}$$

   and define $\mathbf{\Sigma} = \mathbf{H^{-1}W(H^{-1})}'$.

4. *Return to Step 1 until the required number of draws has been obtained.*

Algorithm 1 draws from a distribution over $\mathbf{W}$ conditional on the CS restrictions on $\mathbf{\Sigma}$ and transforms the draws into $\mathbf{\Sigma}$. The independent draws of $\mathbf{\Sigma}$ will not be from the desired restricted IW posterior distribution of $\mathbf{\Sigma}$. Because both the mapping and the CS restrictions on $\mathbf{\Sigma}$ are linear on $\operatorname{vec}(\mathbf{\Sigma})$, applying the change of variable theorem outlined in Arias et al. (2018) implies that the density implied by Algorithm 1 involves a volume element that is independent of $\mathbf{\Sigma}$. Hence, the volume element will be irrelevant to the importance sampler we derive. We will call $\pi_{\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}}\right)}(\mathbf{\Sigma}) = \mathcal{IW}_{\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}\right)}(\mathbf{W}_{11})\, \mathcal{IW}_{\left(\mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}}\right)}(\mathbf{\Omega})$, where $\mathbf{W}_{11} = \mathbf{\Xi\Sigma\Xi}'$ and $\mathbf{\Omega} = \mathbf{L\Sigma L}'$, the density implied by Algorithm 1 and $\pi\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}}\right)$ its distribution. A natural choice for $\mathbf{S}_{\mathbf{W}_{11}}$ and $\mathbf{S}_{\mathbf{\Omega}}$ is $\mathbf{S}_{\mathbf{W}_{11}} = \mathbf{\Xi S\Xi}'$ and $\mathbf{S}_{\mathbf{\Omega}} = \mathbf{LSL}'$, while one could choose $\nu_{\mathbf{W}_{11}} = \nu_{\mathbf{\Omega}} = \nu$.

It is relatively easy to adapt Algorithm 1 for the case where there is no approximation error, and hence $\mathbf{\Sigma}$ is singular; Appendix C gives the details. This case is relevant for applications in which the restrictions hold exactly. Examples include the decomposition of foreign currency returns, where interest rate differentials are multiplicative (rather than additive, as in the case of dividends), see, e.g. Engel and West (2005); Froot and Ramadorai (2005), and, similarly, the alternative exact decomposition of stock market returns proposed by Gao and Martin (2021), which considers $\log(1 + P_t/D_t)$ instead of $\log(P_t/D_t)$.

Since our objective is to independently draw from the $\mathcal{IW}(\mathbf{S}, \nu)$ conditional on the CS restrictions on $\mathbf{\Sigma}$, the results above justify the following importance sampler algorithm.

Algorithm 2. *Let a scalar $\nu \geq n$ and $\mathbf{S}$ be an $n \times n$ SPD matrix. The following algorithm independently draws from the $\mathcal{IW}(\mathbf{S}, \nu)$ conditional on the CS restrictions on $\mathbf{\Sigma}$.*

1. *Use Algorithm 1 to independently draw $\mathbf{\Sigma}$ from $\pi\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}}\right)$.*

2. *Set its importance weight to*
$$\frac{\mathcal{IW}_{(\mathbf{S},\nu)}(\mathbf{\Sigma})}{\pi_{\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}}\right)}(\mathbf{\Sigma})}.$$

3. *Return to Step 1 until the required number of draws has been obtained.*

4. *Re-sample with replacement using the importance weights.*

Algorithm 2 shows how to independently draw from a $\mathcal{IW}(\mathbf{S}, \nu)$ conditional on the CS restrictions on $\mathbf{\Sigma}$ for general $(\mathbf{S}, \nu)$. If we want to draw from the restricted IW posterior distribution of $\mathbf{\Sigma}$, we need to set $\mathbf{S} = \overline{\mathbf{S}}$ and $\nu = \overline{\nu}$. It is easy to modify the algorithm for the case of no approximation error; we need to make independent draws of $\mathbf{\Sigma}$ from $\pi\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}\right)$ in Step 1 and change the importance weights accordingly.

3.4 Drawing Any Posterior

The results above can be used to independently draw from the restricted NIW posterior of $\boldsymbol{\alpha}$ and $\mathbf{\Sigma}$. In particular, we have the following algorithm:

Algorithm 3. *The following algorithm independently draws from the posterior distribution $NIW\left(\overline{\nu}, \overline{\mathbf{S}}, \overline{\boldsymbol{\alpha}}, \overline{\mathbf{V}}\right)$ conditional on the CS restrictions on $\mathbf{\Phi}$ and $\mathbf{\Sigma}$.*

1. *Use Algorithm 2 to draw* $\boldsymbol{\Sigma}$ *from* $\mathcal{IW}\left(\overline{\nu}, \overline{\mathbf{S}}\right)$ *conditional on the CS restrictions on* $\boldsymbol{\Sigma}$.

2. *Draw* $\boldsymbol{\alpha}$ *from* $\mathcal{N}\left(\overline{\boldsymbol{\alpha}}, \overline{\mathbf{V}}\right)$ *conditional on the draw of* $\boldsymbol{\Sigma}$ *and the CS restrictions on* $\boldsymbol{\Phi}$.

3. *Discard the draws that do not satisfy the stationarity restriction.*

4. *Return to Step 1 until the required number of draws has been obtained.*

3.5 An Alternative Algorithm

Algorithm 3 enables posterior sampling using informative priors while exactly imposing the Campbell-Shiller (CS) restrictions draw by draw, without adding extra restrictions or assuming away the measurement error. This is achieved by placing a prior on the full system and then applying the CS restrictions. An alternative, simpler method, which works only assuming no measurement error, would impose the CS restrictions on a reduced system (returns and price-dividend ratio, with dividend growth as a regressor) and back out the dividend equation. However, this approach is not equivalent: it prevents use of the Single Unit Root (SUR) prior and the stationarity restriction in Equation (11). The SUR prior is important because, together with the CS restriction, it links the prior on the unconditional equity premium to those of dividend growth and the price-dividend ratio (Fama and French, 2002). We also show in Appendix D that this alternative approach implies relatively flat priors on the dividend growth coefficients and, empirically, performs much worse in forecasting compared to our baseline method.

4 Return Predictability and Dividend Momentum

We now present the empirical results using our methods. After analyzing the restricted posterior, we will focus on return predictability, dividend momentum, its implications for cash flow and discount rate news, and the variance of long-run returns. All the results in this section will be in-sample. We will analyze out-of-sample results and implications for asset allocation in the next section.

4.1 Data

We use annual US postwar data for the VAR(1) above, between 1947 and 2024, taking the S&P 500 index as our measure of the stock market.[6] Annual returns conform closely to the homoskedastic Gaussian assumption above. Moreover, because dividend payments are known to be highly seasonal, the focus on annual data ensures that seasonal patterns do not simply drive any dividend growth autocorrelation we find. In Section 6 we will examine the robustness of our results to using quarterly or monthly data and modeling seasonality explicitly. An important issue is how to treat the reinvestment of dividends. The annual series traditionally used in the literature implicitly measures dividends after reinvestment in the stock market each month within the year. Van Binsbergen and Koijen (2010) and Koijen and Nieuwerburgh (2012) convincingly argue that this assumption induces spurious distortions that amount to mismeasurement of dividend growth and its time series properties. Moreover, they show how a VAR(1) on reinvested dividends would be misspecified if the cash dividends followed an autoregressive process. For this reason, we measure dividends with no reinvestment.[7]

4.2 A First Look at the Data

For comparison, we give a first look at the data by specifying a flat prior without taking into account the CS restrictions. In this case, the Bayesian posterior means are centered around the OLS estimates, and the Bayesian high posterior density intervals coincide with the classical confidence intervals.[8] Table 1 reports the posterior of $\boldsymbol{\mu} \equiv (\mathbf{I}_n - \boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Phi}_0$, $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Sigma}$ in Equation (1) under a flat prior. The estimates of $\boldsymbol{\mu}$ are consistent with a steady-state 5.7% log dividend growth rate, a steady-state log dividend price ratio of 3.5, and a 9% steady state log return; all these coefficients, however, are estimated relatively imprecisely. In line with Cochrane's (2008b) conclusions, we also find that the price-dividend ratio is highly persistent, $\phi_{pd,pd} = 0.906$, and that $\phi_{r,pd} = -0.171$, significantly below zero, whereas $\phi_{d,pd}$ is approximately zero. Therefore, the price-dividend ratio forecasts returns and does not forecast dividend growth. There is little evidence of serial correlation

[6] Results using CRSP market returns are very similar and available upon request.

[7] Results using dividends reinvested at the risk-free rate are very similar and available upon request.

[8] Following Uhlig (2005), the flat prior is obtained as a special case, letting $\underline{\nu} = 0$ and $\underline{\mathbf{V}}^{-1} = \mathbf{0}_{K\times K}$ with $\underline{\mathbf{S}}$ and $\underline{\boldsymbol{\alpha}}$ arbitrary. All the results in this section are obtained from 5,000 draws of the posterior distribution, and we have that $n = 3$, $p = 1$, and $T = 72$.

Table 1: Posterior Distribution of $\boldsymbol{\mu}$, $\boldsymbol{\Phi}_1$, and $\boldsymbol{\Sigma}$ under Flat Priors

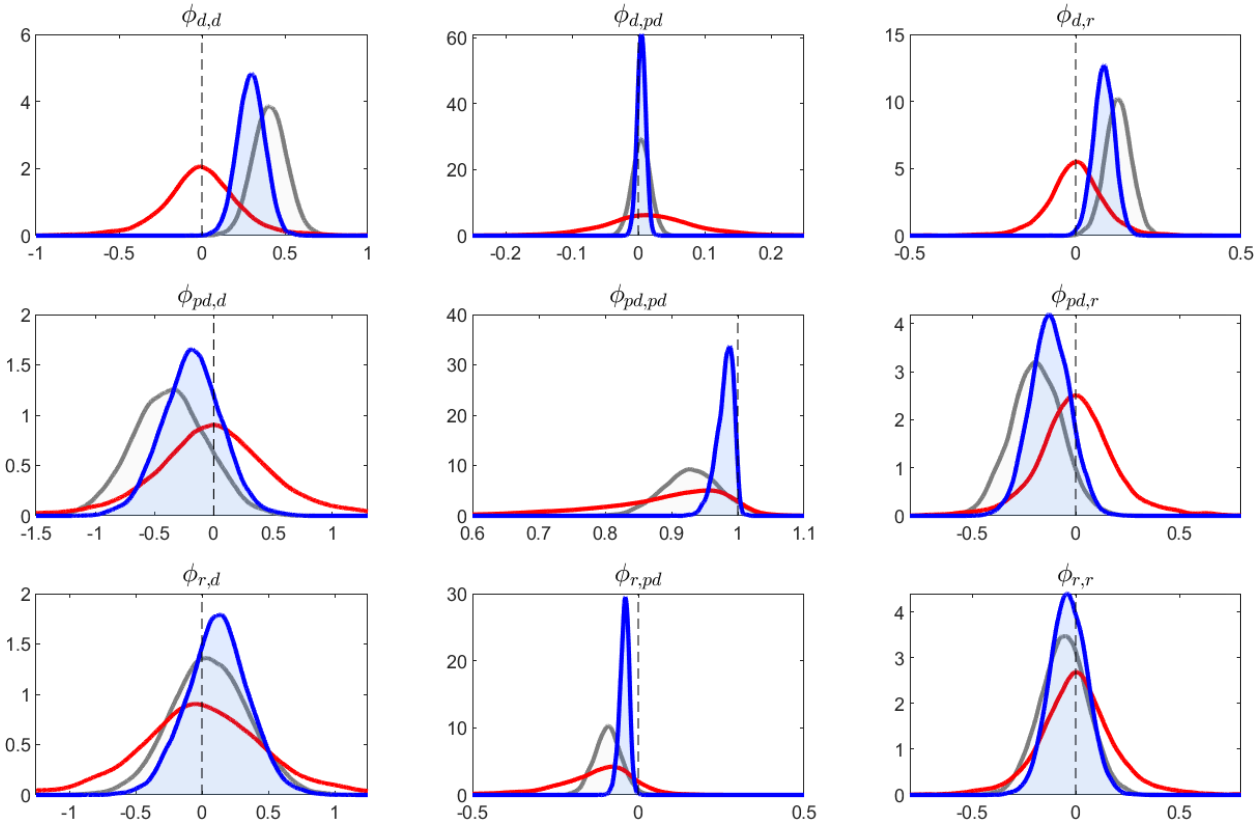
	$\boldsymbol{\mu}$	$\boldsymbol{\Phi}_1$		
		Δd_t	pd_t	r_t
Δd_{t+1}	0.056	0.409	0.004	0.128
	$[0.034, 0.073]$	$[0.310, 0.510]$	$[-0.009, 0.018]$	$[0.090, 0.167]$
pd_{t+1}	3.846	−0.371	0.930	−0.190
	$[3.447, 4.469]$	$[-0.682, -0.058]$	$[0.887, 0.973]$	$[-0.313, -0.069]$
r_{t+1}	0.079	0.056	−0.092	−0.055
	$[0.042, 0.103]$	$[-0.223, 0.342]$	$[-0.131, -0.053]$	$[-0.169, 0.057]$
		$\boldsymbol{\Sigma}$ (corr/std)		
		u^d_{t+1}	u^{pd}_{t+1}	u^r_{t+1}
u^d_{t+1}		0.053		
		$[0.049, 0.058]$		
u^{pd}_{t+1}		−0.329	0.168	
		$[-0.429, -0.227]$	$[0.156, 0.183]$	
u^r_{t+1}		−0.003	0.945	0.154
		$[-0.114, 0.110]$	$[0.932, 0.957]$	$[0.143, 0.168]$

Note: The table shows the parameter estimates under a flat prior for a first-order VAR model including a constant, the log dividend growth (Δd_{t+1}), the price-dividend ratio (pd_{t+1}), and the log market return (r_{t+1}). For each coefficient, the first line reports the posterior median value, and the second line reports the 68^{th} posterior credible intervals in square brackets. The table also reports the parameters of the correlation matrix of the innovations with innovation standard deviations on the diagonal, labeled "corr/std," instead of the parameters of the covariance matrix.

in returns after we control for dividend growth and the price-dividend ratio, as $\phi_{r,r}$ is centered around zero (with a posterior mean value of 0.03). Moreover, the mean of the posterior of $\phi_{r,d}$ is centered around zero (with a posterior mean value of 0.02), meaning that lagged dividends do not directly forecast one-period-ahead returns either.

A central finding is that dividend growth is strongly predictable in ways often overlooked. The posterior mean of $\phi_{d,d}$ is 0.41, with 95% of the posterior above 0.22, indicating highly significant and economically meaningful persistence: for comparison, consider the autocorrelation in annual US real GDP, at 0.14. Crucially, this persistence holds even after controlling for lagged returns and valuation ratios. We also find $\phi_{d,r} = 0.14$, implying returns help predict dividend growth, and $\phi_{pd,d} < 0$, so dividend growth negatively forecasts the next period's price-dividend ratio. While dividend growth doesn't forecast short-term returns, it may still shape expected returns at longer horizons.

Figure 2: Bayesian restricted prior and posterior of $\mathbf{\Phi}_1$



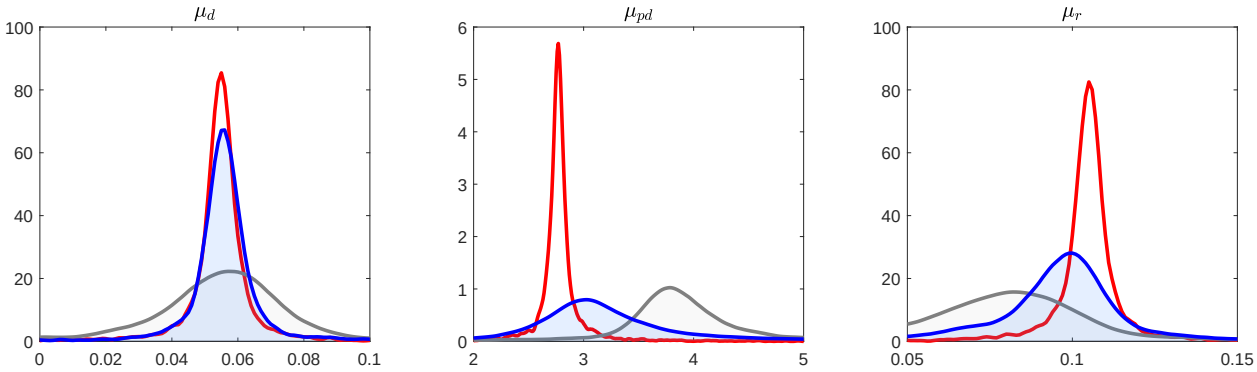
Note: Red represents the prior, grey is the likelihood, and blue is the posterior.

4.3 The Restricted Posterior

Figure 2 shows the restricted informative prior, the likelihood, and the restricted informative posterior for $\mathbf{\Phi_1}$.[9] The likelihood (gray p.d.f.) mean matches the flat-prior estimates from Table 1. Posterior distributions are generally more concentrated, reflecting the combination of prior and data. In the first column, $\phi_{d,d}$ falls from 0.41 to 0.3 in the posterior but remains clearly positive, confirming persistent dividend growth. Coefficients $\phi_{pd,d}$ and $\phi_{r,d}$ shrink slightly toward zero with tighter uncertainty. In the second column, $\phi_{pd,pd}$ rises and becomes tightly concentrated around 0.98, indicating slower mean reversion than the likelihood estimate of 0.91. This reflects both the stationarity and CS restrictions, which the likelihood alone does not impose. The posterior for $\phi_{d,pd}$

[9]The hyperparameters $\lambda = 0.17$ and $\theta = 0.05$ are chosen to maximize the marginal likelihood (Giannone et al., 2015). We set $\underline{\sigma}_i^2$ to the residual variance from AR(1) models: $\underline{\sigma}_d^2 = 0.0034$, $\underline{\sigma}_{pd}^2 = 0.0284$, $\underline{\sigma}_r^2 = 0.0254$. The prior mean return $\underline{\mu}_r$ is 10.5%, consistent with a 4% risk-free rate and 6.5% equity premium; $\underline{\mu}_d$ is 5.5%, consistent with long-run U.S. nominal GDP growth. The CS restriction (4) implies $\underline{\mu}_{pd} \approx 2.8$.

Figure 3: Bayesian restricted prior and posterior of $\boldsymbol{\mu}$



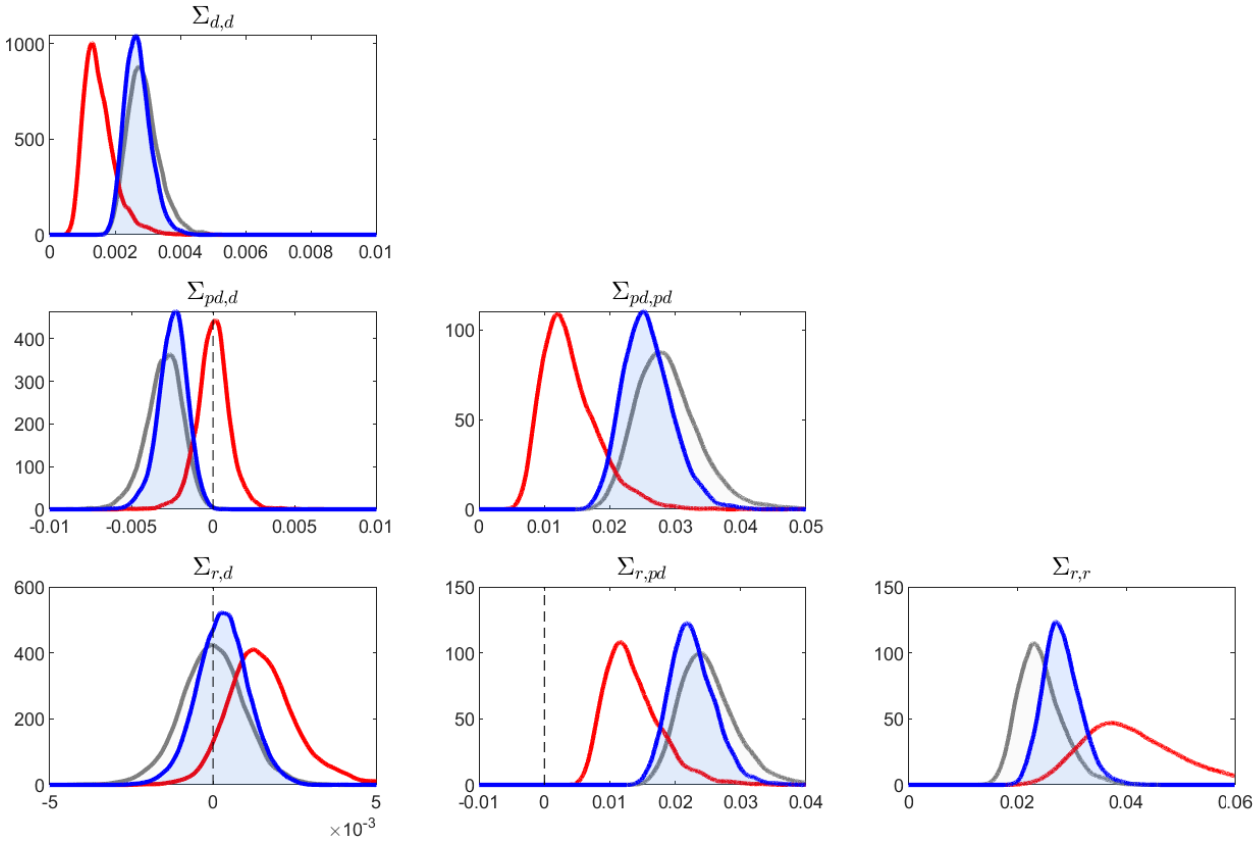
Note: Red represents the prior, grey is the likelihood, and blue is the posterior.

centers near zero, consistent with Cochrane (2008b)'s conclusion that the price-dividend ratio does not forecast dividend growth. Given the high persistence of pd_t and CS identity (9), this implies $\phi_{r,pd}$ must be negative. Thus, greater persistence in pd_t leads to less return predictability and slower mean reversion. Finally, the third column shows posterior distributions more precise and shrunk toward zero, but $\phi_{d,r}$ remains significantly positive, amplifying the dividend momentum channel despite prior shrinkage.

Figure 3 looks at the same distributions for the unconditional mean of the variables, $\boldsymbol{\mu}$. The likelihood alone is very uncertain about the value of these parameters, so, not surprisingly, we see the restricted posterior distribution moving closely toward the restricted informative prior distribution. The parameter μ_d mostly gains precision. For the case of μ_{pd}, the posterior moves closer to its initial condition, consistent with a price-dividend ratio close to non-stationarity. For μ_r, the posterior is consistent with a markedly higher and less uncertain unconditional equity return in nominal terms. Incorporating reasonable prior distributions over the system's steady state is a key advantage of the Bayesian approach, given the difficulty of accurately estimating these parameters from the likelihood alone.

Figure 4 reports the distribution of the covariance elements. The CS restrictions push the prior covariance of u_t^r with u_t^d and u_t^{pd} away from zero. The posterior of $\Sigma_{pd,d}$ is centered around negative values, so the CS restrictions imply a downward revision of $\Sigma_{r,d}$. Instead, the posterior estimate is associated with an upward revision of $\Sigma_{r,pd}$. This reflects the upward revision of the posterior estimate of $\Sigma_{pd,pd}$ and the tight link between $\Sigma_{r,pd}$ and $\Sigma_{pd,pd}$ imposed by the CS restriction.

Figure 4: Bayesian restricted prior and posterior of $\boldsymbol{\Sigma}$



Note: Red represents the prior, grey is the likelihood, and blue is the posterior.

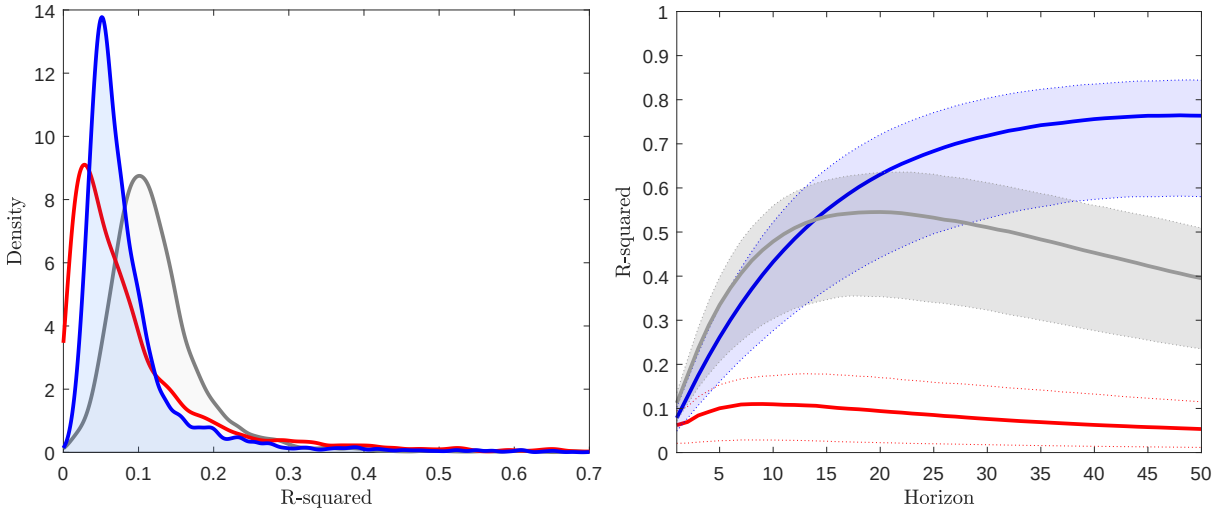
4.4 Return Predictability

Figure 5 looks at the R^2 of the return equation both at one period ahead and as a function of the horizon. Appendix E derives how to compute R^2 for multiple-period returns. Panel (a) draws the density associated with prior, likelihood, and posterior for the one-period-ahead R^2 of the return equation. Compared with the likelihood, both the restricted informative prior and the associated posterior offer a much more skeptical view of the one-period-ahead predictability of returns.

Panel (b) draws the median and the 68^{th} interval associated with the restricted informative prior, likelihood, and restricted informative posterior for the R^2 of the return equation as a function of the horizon.[10] The figure clearly shows the points raised in Cochrane (2009, p.228). For the case of the likelihood, the R^2 initially increases before decaying. Instead, our restricted informative posterior features a shift of return predictability toward longer horizons. This is seen from an R^2 that increases

[10]We thank John H. Cochrane for suggesting Panel (b).

Figure 5: Return Equation R-squared



(a) One-Period-Ahead R^2 (b) R^2 as a function of horizon

Note: Red represents the prior, grey is the likelihood, and blue is the posterior. In Panel (b), the solid line is the median value, while the shadow area represents the 68^{th} posterior credible intervals.

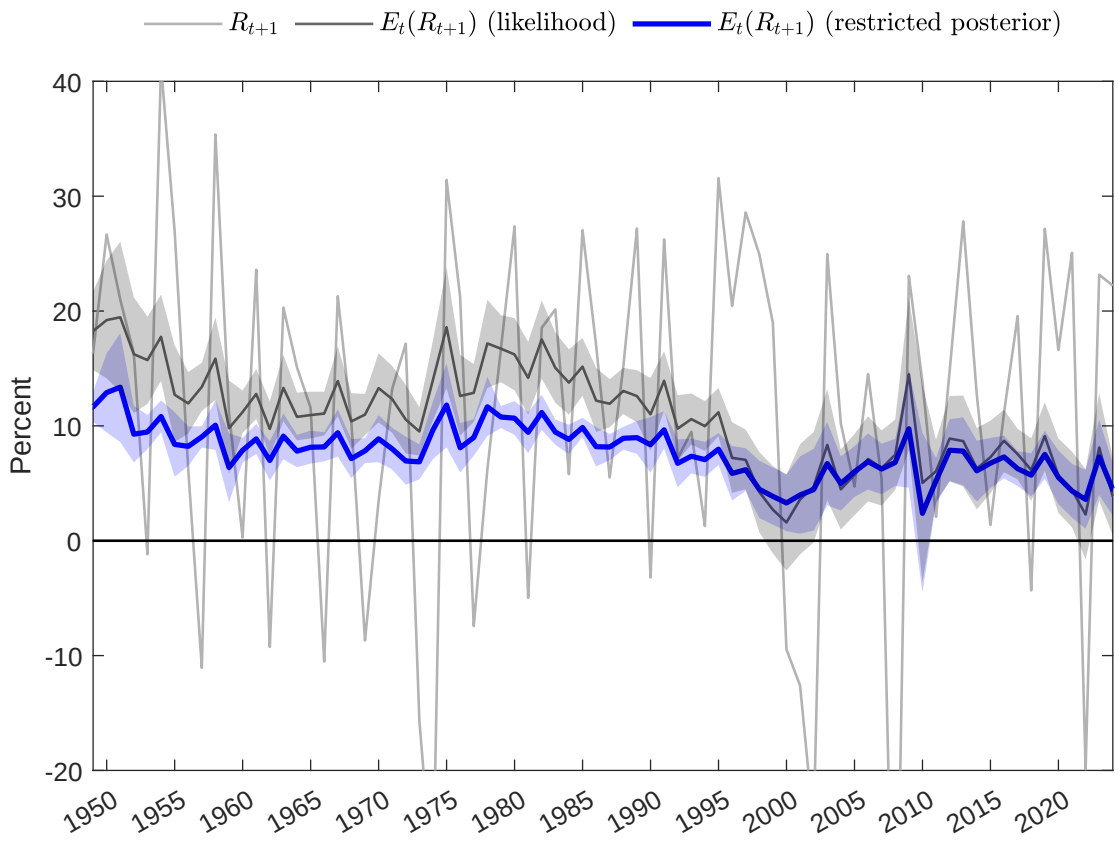
more slowly but reaches higher values for longer horizons. This is due to the combination of a lower degree of short-term predictability and a higher persistence of the price-dividend ratio. Note that, contrary to the concerns in Boudoukh et al. (2008), this result is not hard-wired by choosing a prior for a highly persistent price-dividend ratio: our restricted informative prior is centered around low values for the short-term predictive coefficients; hence, the a priori R^2 is low at all horizons.

Figure 6 plots the time series of the one-period-ahead expected return implied by the likelihood and restricted informative posterior. The figure clearly shows how the restricted informative prior reduces the one-period-ahead return predictability, reinforcing the message from Panel (a) of Figure 5. The restricted informative posterior also shows less uncertainty. Notably, the median of the restricted informative posterior estimate never becomes negative, a desirable property as emphasized by Campbell and Thompson (2008) and Pettenuzzo et al. (2014). Nevertheless, there is still time variation in expected returns in the restricted informative posterior.

4.5 Dividend Momentum

Our empirical results show that dividend growth is positively autocorrelated after controlling for the price-dividend ratio, i.e., $\phi_{d,d} > 0$. We now highlight how this fact implies a previously overlooked

Figure 6: One-Period-Ahead Expected Return



Note: The solid line is the median value, while the shadow area represents the 68^{th} posterior credible intervals.

channel of return predictability, which we call dividend momentum. We first define dividend momentum using news about cash flows and discount rates and then show how it manifests in the impulse response functions (IRFs) and the correlation between cash flow and discount rate news.

Cashflow and Discount Rate News

Iterating forward the CS identity, applying expectations, and imposing the transversality condition $\lim_{T\to\infty} \rho^T \mathbb{E}_t pd_{t+T} = 0$, the price-dividend ratio is equal to the expected discounted sum of future dividend growth minus the expected discounted sum of future returns. Combining this result with the CS identity, Campbell and Ammer (1993) derive a decomposition of unexpected returns:

$$r_{t+1} - \mathbb{E}_t r_{t+1} = \underbrace{(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=0}^{\infty} \rho^j \ \Delta d_{t+j+1}}_{NCF_{t+1}} - \underbrace{(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+j+1}}_{NDR_{t+1}} \tag{25}$$

An unexpected positive return must reflect either upward revisions to current and future dividend growth (NCF_{t+1}, cash-flow news) or downward revisions to expected future returns (NDR_{t+1}, discount-rate news). These components help define mean reversion and momentum: mean reversion occurs when returns rise today but expected future returns fall, so $r_{t+1} - \mathbb{E}_t r_{t+1}$ moves opposite to NDR_{t+1}; momentum arises when both move in the same direction. If both current and expected dividend growth increase alongside expected returns, we observe dividend-induced momentum in returns, or "dividend momentum."[11] While the literature often interprets NCF_{t+1} and NDR_{t+1} as permanent and transitory shocks to wealth, respectively, the presence of dividend momentum implies that discount-rate news are correlated with revisions to expected dividend growth and accumulates over time.

Dividend momentum will generally arise whenever dividends are persistent, in a way not fully captured by the lagged price-dividend ratio (i.e., $\phi_{d,d} > 0$). This follows from the restrictions implied by the CS identity, even if the price-dividend ratio is the only variable directly predicting returns. To see this, consider the following simplified system:

$$\begin{bmatrix} \Delta d_{t+1} \\ pd_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_{d,d} \\ \phi_{pd,d} \\ 0 \end{bmatrix} \Delta d_t + \begin{bmatrix} 0 \\ \phi_{pd,pd} \\ \phi_{r,pd} \end{bmatrix} pd_t + \begin{bmatrix} u^d_{t+1} \\ u^{pd}_{t+1} \\ u^r_{t+1} \end{bmatrix} \tag{26}$$

The zero coefficients for $\phi_{r,d}$ (lagged dividends do not forecast returns) and $\phi_{d,pd}$ (the lagged price-dividend ratio does not forecast dividends) are approximately true in the data, as seen in Figure 2. The omission of the third column is for expositional purposes only and does not affect any of the economic implications.[12] In this simplified setting, the CS restrictions (5) and (6) imply $\phi_{pd,d} = -\phi_{d,d}/\rho$ and $\phi_{r,pd} = \rho\phi_{pd,pd} - 1$. To further simplify the discussion we also assume that $\texttt{Cov}\left(u^{pd}_{t+1}, u^d_{t+1}\right) = 0$. This assumption is not required but simplifies the exposition, allowing an interpretation of u^{pd}_{t+1} and u^d_{t+1} as distinct shocks that we call price-dividend ratio and dividend growth shocks. The rest of the covariances between innovations are backed out from the CS restrictions (8) and (9), yielding $\texttt{Cov}\left(u^r_{t+1}, u^d_{t+1}\right) = \texttt{Var}\left(u^d_{t+1}\right)$ and $\texttt{Cov}\left(u^r_{t+1}, u^{pd}_{t+1}\right) = \rho\texttt{Var}\left(u^{pd}_{t+1}\right)$.[13]

[11] Dividend momentum is a special case of return momentum, distinct from serial correlation in dividend growth. If the price-dividend ratio captures all persistence, as in Bansal and Yaron (2004), NDR_{t+1} remains unaffected and no return momentum arises.

[12] Our estimates show that $\phi_{d,r} > 0$, strengthening the basic intuitions in this section.

[13] In Appendix B.4, we show that this simplified setting leads to a latent variable model along the lines of

In this case, the decomposition in Equation (25) becomes:

$$r_{t+1} - \mathbb{E}_t r_{t+1} = \underbrace{(1+\Psi)\, u^d_{t+1}}_{NCF_{t+1}} - \underbrace{\left(-\rho u^{pd}_{t+1} + \Psi u^d_{t+1}\right)}_{NDR_{t+1}},$$

with $\Psi = \frac{\rho\phi_{d,d}}{1-\rho\phi_{d,d}}$. Notice that the two terms include Ψu^d_{t+1} with opposite sign, so they cancel out and the expression for $r_{t+1} - \mathbb{E}_t r_{t+1} = u^d_{t+1} + \rho u^{pd}_{t+1}$ is equal to, as per the CS identity, u^r_{t+1}, which does not depend on $\phi_{d,d}$. While the sum of the two components is unaltered irrespective of $\phi_{d,d}$, when $\phi_{d,d} > 0$, the NCF_{t+1} term is now bigger by the factor Ψ, and NDR_{t+1} is affected by both u^{pd}_{t+1} and u^d_{t+1}. Importantly, a positive dividend growth shock will lead to positive unexpected returns and dividend growth today, a positive revision to the discounted sum of expected future dividend growth, and a positive revision to the discounted sum of expected future returns (NDR_{t+1}); this shock generates dividend momentum, and it contributes to the (positive) correlation of NCF_{t+1} and NDR_{t+1}. If instead $\phi_{d,d} = 0$, the dividend growth shock does not affect NDR_{t+1} and, thus, it does not generate dividend momentum.

Another method to detect dividend momentum is to look at the contribution of a dividend growth shock to the variance of NCF_{t+1} and NDR_{t+1} and the correlation between them. In our simplified model, the variance of the unexpected return is:

$$\mathtt{Var}(u^r_{t+1}) = \underbrace{(1+\Psi)^2\,\mathtt{Var}(u^d_{t+1})}_{\mathtt{Var}(NCF_{t+1})} + \underbrace{\rho^2\mathtt{Var}(u^{pd}_{t+1}) + (\Psi)^2\,\mathtt{Var}(u^d_{t+1})}_{\mathtt{Var}(NDR_{t+1})} - \underbrace{2(1+\Psi)\Psi\mathtt{Var}(u^d_{t+1})}_{2\mathtt{Cov}(NCF_{t+1},NDR_{t+1})}.$$

When $\phi_{d,d} > 0$, dividend momentum implies that an increase in u^d_{t+1} leads to a positive revision of current and expected future cash flows and discount rates. As a consequence, $\mathtt{Var}(NCF_{t+1}) > \mathtt{Var}(u^d_{t+1})$, $\mathtt{Var}(NDR_{t+1}) > \rho^2\mathtt{Var}(u^{pd}_{t+1})$, and $\mathtt{Corr}(NCF_{t+1}, NDR_{t+1}) = \rho\phi_{d,d}\sqrt{\mathtt{Var}(NCF_{t+1})/\mathtt{Var}(NDR_{t+1})} > 0$. Therefore, if dividend momentum is present in the data, one should see an increased contribution of dividend growth shocks to the volatility of both NCF_{t+1} and NDR_{t+1} and the correlation between the two.

Van Binsbergen and Koijen (2010), where expected dividend growth and expected returns follow a VAR(1), where revisions in expected dividend growth lead to revisions of future expected returns whenever $\phi_{d,d} \neq 0$. Additionally, we emphasize that assuming correlation zero (rather than one) between current and expected dividend growth, as done by Van Binsbergen and Koijen (2010), precludes the existence of dividend momentum. The latter, as discussed in Section 4.5, is built on the premise that a shock to dividends will have a prolonged effect on future dividend growth and, consequently, on expected future returns.

Table 2: Shock's Contribution to NCF_{t+1} and NDR_{t+1}

	Total	u^{pd}_{t+1}	u^{d}_{t+1}
$Var(NDR_{t+1})$	0.023 [0.013, 0.035]	97.1% [93.9%, 98.7%]	2.9% [1.3%, 6.1%]
$Var(NCF_{t+1})$	0.007 [0.005, 0.010]	8.5% [0.7%, 32.5%]	91.3% [67.3%, 99.0%]
$Corr(NDR_{t+1},\ NCF_{t+1})$	16.1% [−25.9%, 55.1%]	−1.9% [−43.6%, 38.9%]	15.9% [10.5%, 22.9%]

Note: We report the posterior median and the 68^{th} posterior credible intervals. The "Total" column reflects the posterior of moments, while the u^{pd}_{t+1} and u^{d}_{t+1} columns reflect the posterior contribution of the two shocks.

Whenever NCF_{t+1} and NDR_{t+1} are correlated, it is useful to separate the orthogonal shocks driving both. In our simple model, these correspond to the uncorrelated innovations u^{pd}_{t+1} and u^{d}_{t+1}. In more general settings, they must be orthogonalized to identify price-dividend and dividend growth shocks. A natural approach is to identify one shock that fully explains the unexpected variance of the price-dividend ratio on impact, and a second shock, orthogonal to the first, that leaves the price-dividend ratio unchanged but, together with the first, explains all of the unexpected variance in dividend growth. This can be achieved via a Cholesky decomposition with variable ordering $[pd_{t+1}, \Delta d_{t+1}, r_{t+1}]$, which recovers u^{pd}_{t+1} and u^{d}_{t+1} in the simple case.[14]

Table 2 analyzes how both price-dividend ratio and dividend growth shocks contribute to NDR_{t+1} and NCF_{t+1} in the data using the restricted informative posterior, with the two shocks identified in the way just described. The table shows that NDR_{t+1} and NCF_{t+1} are correlated, as expected, whenever dividend momentum is present, and most of the correlation comes from the dividend momentum associated with dividend growth shocks.

Impulse Response Functions

A helpful way to understand dividend momentum is to compute the IRFs for each of the two shocks. Figure 7 plots our simple model's discounted cumulative IRFs for returns and dividend growth. We plot the IRFs discounting by ρ^h and then cumulating; so, for returns, the IRF converges to the sum of the initial unexpected return and NDR_{t+1}, whereas, for dividends, the IRF converges to NCF_{t+1}.

[14] A third shock affecting only contemporaneous returns corresponds to the CS approximation error η_{t+1}, which vanishes if the identity holds exactly. Alternative orthogonalizations exist. For instance, Campbell et al. (2013) orders returns first and the price-earnings ratio second, so the first shock explains all unexpected return variance, and the second affects the price-dividend ratio.

We would observe a perfectly flat IRF beyond the initial jump if a variable were unpredictable. After an initial positive impact, downward-sloping IRFs for returns indicate mean reversion, whereas upward-sloping ones are a sign of momentum. We will have dividend momentum if a shock with an initial positive impact and an upward-sloping IRF for dividend growth causes momentum.

Panel (a) of the figure plots the case $\phi_{d,d} = 0$, whereas Panel (b) has $\phi_{d,d} > 0$. For a price-dividend ratio shock, the IRFs are the same in both panels: returns jump on impact but have a negative slope after that. This is the classic mean reversion effect in returns, caused by mean reversion of the price-dividend ratio, as documented by Fama and French (1988) and Campbell and Shiller (1988a). On the contrary, the IRF for dividend growth is zero every period, so neither contemporaneous nor expected future dividend growth changes.

Consider now a dividend growth shock. By the CS identity, returns and dividend growth initially jump by the same amount regardless of the $\phi_{d,d}$ value. However, the value of $\phi_{d,d}$ is relevant beyond impact. If $\phi_{d,d} = 0$, future expected returns and dividend growth are unaffected: the slope of both IRFs is flat, and there is no dividend momentum. If instead $\phi_{d,d} > 0$, expectations of future returns and dividend growth increase, the slope of both IRFs is positive, inducing dividend momentum. Panel (c) plots the IRFs implied by the data as captured by the restricted informative posterior, again identifying the shocks using the Cholesky approach described above. The downward slope for returns in the left panel indicates mean reversion after a price-dividend shock, and the upward-sloping IRFs of both dividends and returns in the right panel indicate dividend momentum after a dividend growth shock.

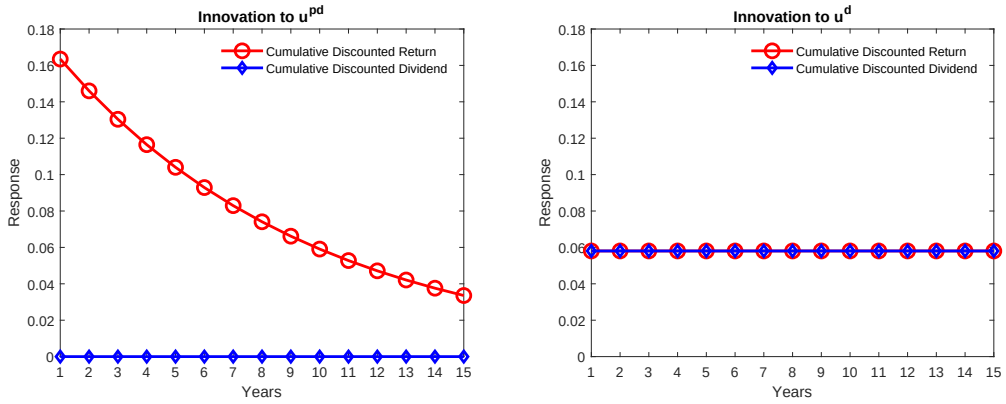
Appendix F compares the empirical IRFs in Panel (c) with those estimated under flat priors, which reveals how, in the latter case, return predictability is exaggerated, implying steeper slopes for the return IRFs in response to both shocks and is estimated more imprecisely. It also discusses how omitting dividend growth from the VAR leads to biased estimates that make detecting the dividend momentum result impossible.

Variance of Long-Horizon Returns and Implications for Portfolio Choice

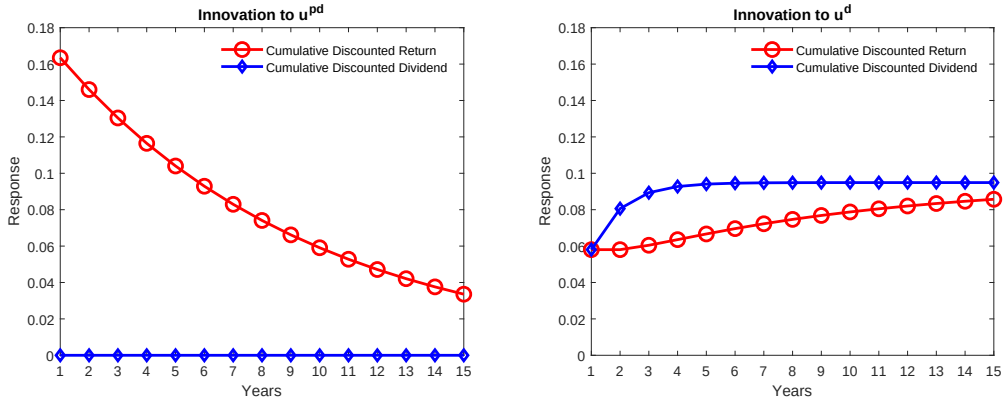
Dividend momentum affects the portfolio choice of the investor who cares about long-run returns, $r_{t,t+k} = \sum_{j=1}^{k} r_{t+j}$. The risk for the long-run investor is a function of the unpredictable component

Figure 7: Dividend Momentum and IRFs

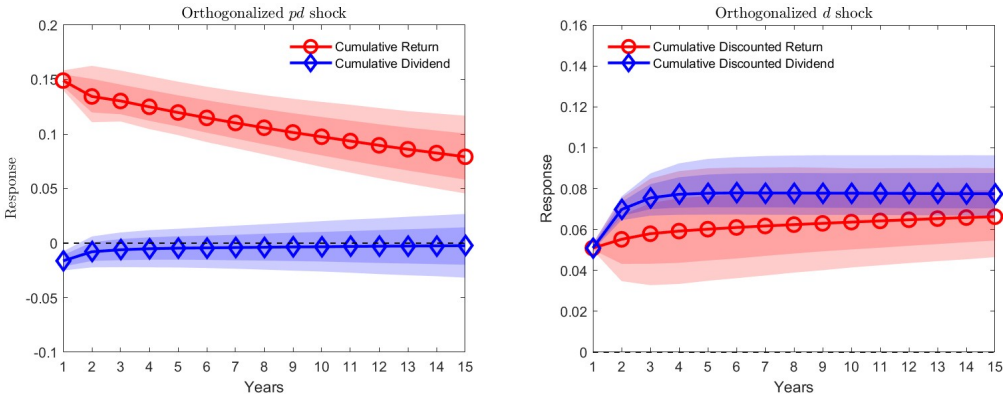
(a) Simplified Model: No Dividend Momentum ($\phi_{d,d} = 0$)



(b) Simplified Model: With Dividend Momentum ($\phi_{d,d} = 0.4$)



(c) Data: Restricted Informative Posterior



Note: Panel (a): $\phi_{d,d} = 0.4$. Panel (b): $\phi_{d,d} = 0$. In addition, we use the following numerical values $\phi_{pd,pd} = 0.92$, $\texttt{Var}\left(u^{d}_{t+1}\right) = 0.003$, and $\texttt{Var}\left(u^{pd}_{t+1}\right) = 0.028$. $\rho = 0.971$. Panel (c): The solid lines represent the median posterior response. The darker shadow area represents the 68^{th} posterior credible intervals, while the lighter shadow area represents the 95^{th} posterior credible intervals.

for long-run returns, $r_{t,t+k} - \mathbb{E}_t r_{t,t+k}$. For the simplified model in (26), we have that

$$r_{t,t+k} - \mathbb{E}_t r_{t,t+k} = \sum_{j=1}^{k} u^r_{t+j} + \sum_{j=1}^{k-1} \left[a_j u^{pd}_{t+k-j} + b_j u^d_{t+k-j} \right], \tag{27}$$

where $a_j = \frac{1-\phi^j_{pd,pd}}{1-\phi_{pd,pd}} \left(\rho\phi_{pd,pd} - 1 \right)$, $b_j = -\frac{\phi_{d,d}}{\rho} \frac{\left(\rho\phi_{pd,pd}-1\right)}{\left(\phi_{d,d}-\phi_{pd,pd}\right)} \left[\frac{\left(1-\phi^j_{d,d}\right)}{1-\phi_{d,d}} - \frac{\left(1-\phi^j_{pd,pd}\right)}{1-\phi_{pd,pd}} \right]$, and $u^r_{t+1} = u^d_{t+1} + \rho u^{pd}_{t+1}$ as implied by the CS identity. Notice that $b_j = 0$ if $\phi_{d,d} = 0$. The unpredictable component of long-run returns reflects the innovations to the one-period-ahead returns and the contribution of shocks to the price-dividend ratio and dividend growth that will occur during the investment horizon and lead to revisions in expected returns. In the presence of dividend momentum, i.e., when $\phi_{d,d} > 0$, dividend growth shocks affect the long-run investor's risk beyond the direct impact on the innovation on the one-period-ahead returns through its effect on expected returns.

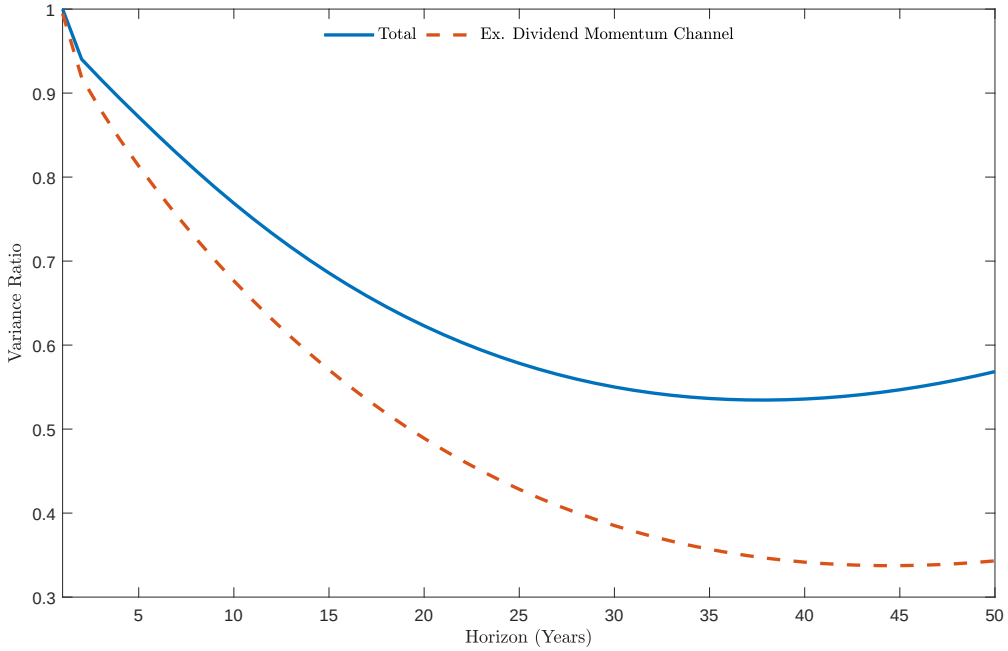
The variance of the long-horizon returns for our simplified system can be decomposed into:[15]

$$\begin{aligned} \texttt{Var}\left(r_{t,t+k}\right) = k\texttt{Var}\left(u^r_{t+1}\right) &+ \sum_{j=1}^{k-1} \left\{ a_j^2 \texttt{Var}\left(u^{pd}_{t+1}\right) + b_j^2 \texttt{Var}\left(u^d_{t+1}\right) \right\} \\ &+ 2\sum_{j=1}^{k-1} \left\{ a_j \texttt{Cov}\left(u^r_{t+1}, u^{pd}_{t+1}\right) + b_j \texttt{Cov}\left(u^r_{t+1}, u^d_{t+1}\right) \right\}, \end{aligned} \tag{28}$$

The first term in Equation (28) reflects the uncertainty coming from the innovations to one-period-ahead returns, often labeled the "i.i.d." component of uncertainty since this will be present even in the case where stock returns are unpredictable. The second term reflects the uncertainty associated with the effects of price-dividend ratio and dividend growth shocks on revisions to future expected returns. The last component reflects the covariance between the one-period-ahead return innovations and the revisions to future expected returns over the investment horizon. Investing in stocks is perceived as less risky for the long-run investor whenever $\texttt{Var}_t\left(r_{t,t+k}\right) < k\texttt{Var}\left(u^r_{t+1}\right)$, which is only possible if the last component is negative. If $0 < \phi_{pd,pd} < 1$ we have that $a_j < 0$ for all j, and since $\texttt{Cov}\left(u^r_{t+1}, u^{pd}_{t+1}\right) = \rho\texttt{Var}\left(u^{pd}_{t+1}\right)$ the mean reversion effects of price-dividend ratio shocks generate negative covariance, reducing the risk associated with long-run investment and generating a positive hedging demand motive for holding stocks for long-run risk-averse investors (see Campbell and

[15] The derivations for the general VAR case are presented in Appendix G. In this section, we are neglecting the estimation uncertainty that will be present any time one estimates a model to predict future returns (see Avramov, 2002; Pástor and Stambaugh, 2012).

Figure 8: Variance Ratio



Note: The blue solid line reflects the ratio when both mean reversion and dividend momentum are present. The orange dashed line reflects the ratio when dividend momentum is excluded.

Viceira, 1999). The presence of dividend momentum generates additional sources of risk for the long-run investor. Shocks to dividend growth increase the uncertainty of future expected returns. Moreover, if $0 < \phi_{d,d} < 1$ we have that $b_j > 0$ for all j, and since $\texttt{Cov}\left(u^r_{t+1}, u^d_{t+1}\right) = \texttt{Var}\left(u^d_{t+1}\right)$, the last component increases the variance of long-run returns, reflecting the positive correlation between shocks to future returns and shocks to future dividend growth.

Figure 8 reports the ratio between $\texttt{Var}_T(r_{T,T+k})$ and $k\texttt{Var}_T(r_{T,T+1})$ obtained from the data using the restricted informative posterior, as a function of the horizon k and evaluates the contribution of dividend momentum to its shape. The solid blue line reflects the ratio when both mean reversion and dividend momentum are present. The orange dashed line reflects the ratio when dividend momentum is excluded. We eliminate dividend momentum by canceling any effect of dividend growth shocks after impact: dividend momentum increases this ratio for all reported horizons.

The presence of dividend momentum implies that dividend growth shocks increase current and future expected returns. Therefore, these shocks increase the variance of long-run returns, $\texttt{Var}_T(r_{T,T+k})$, by both increasing the uncertainty about future expected returns and inducing

positive co-movement between one-period returns and future expected returns surprises; thus dividend momentum contributes quite substantially to the risk faced by the long-run investor. Without dividend momentum, the variance ratio would be roughly one-third smaller. Therefore, the presence of dividend momentum generates a negative hedging demand component that partially offsets the traditional positive hedging demand arising from the mean reversion in returns (see Koijen et al., 2009). As we will see in the next section, this negative hedging demand is empirically essential and affects the portfolio choice of long-horizon investors trying to time the market.

Implications for theory

From a theoretical standpoint, dividend momentum has important implications for asset pricing theories that seek to explain fluctuations in expected returns. These models typically seek to explain the counter-cyclical variation in expected returns relative to dividend and/or consumption growth (see, e.g., Cochrane, 2017). The presence of dividend momentum does not challenge the idea that such mechanisms are the primary drivers of return fluctuations. Instead, it points to an additional channel that generates procyclical variation in expected returns, linked to the persistence of dividend growth. This channel naturally arises in a consumption-based asset pricing model when expectations of future dividend growth are time-varying. As emphasized in the theoretical work of Menzly et al. (2004), in such a setting, an increase in expected dividend growth implies that the asset delivers more of its cash flows further in the future, making its price more sensitive to shocks in the aggregate discount rate—i.e., to fluctuations in investors' risk preferences. To the extent that this added volatility is—at least partially—correlated with changes in risk appetite, it must be priced, leading to a higher risk premium. More broadly, the presence of dividend momentum induces a richer connection between expected returns and macroeconomic dynamics: in the short run, expected returns may exhibit procyclical behavior tied to persistent dividend growth and its drivers, while in the long run, they remain countercyclical, reflecting the tendency of prices to revert toward the persistent component of dividends.

5 Out of Sample Results: Forecasting and Asset Allocation

Given the degree of return predictability found in-sample, even for the case of the restricted informative prior, individual investors would find it optimal to time the market aggressively (Kandel and Stambaugh, 1996). This is even more the case for long-run investors (see Barberis, 2000; Campbell and Viceira, 1999). However, Goyal and Welch (2008), among many others, have pointed out that in-sample return predictability often leads to disappointing results out-of-sample and that portfolio allocation strategies attempting to take advantage of in-sample return predictability rarely outperform simple benchmarks. In this section, we investigate the out-of-sample performance of investment strategies where the investor chooses between investing in cash and equities using the VAR with our restricted informative prior. For this purpose, we must consider a more general 4-variable model that includes the risk-free rate. In Appendix H, we describe this more general model and the CS restrictions associated with it. We compare the performance of an investor who uses our restricted informative prior with that of five other investors who use different priors. Two of the investors use the 4-variable VAR(1) described above. Of those, one will use flat priors and another informative priors without imposing the CS restrictions.[16] To analyze the importance of modeling dividends explicitly in the VAR, we will consider two more investors who follow the standard approach in the literature and drop dividend growth from the VAR; they run a 3-variable VAR(1) with the risk-free rate, the price-dividend ratio, and the excess stock return using either flat priors or informative priors. Finally, a naive investor uses the historical average returns, computed yearly as new data becomes available. Each year, the five investors estimate the parameters of their models and produce forecasts using only information available at each point in time using an expanding window. The out-of-sample period spans from 1973:Q1 to 2024:Q4.

5.1 Out-of-sample forecasting

As a first step, Table 3 evaluates out-of-sample the forecasts made by the different investors, looking at the cumulative excess return $\sum_{i=1}^{h}(r_{t+h} - r^f_{t+h})$ at horizons $h = 1, \ldots, 10$. The table reports the "Out-of-Sample R^2" as in Campbell and Thompson (2008), defined as the percentage improvement in the out-of-sample fit of each investor with respect to the naive investor. A negative value implies

[16] Appendix I describes in detail the flat and the informative prior parameterizations used in this section.

Table 3: Out-of-Sample Return Equation R-squared

	Flat priors		**Informative Priors**		
	Omitting Divs.	Including Divs.	Unrestricted (Inc. Divs.)	Unrestricted (Omit. Divs.)	**Baseline**
$h = 1$	-9.7%	-12.0%	-1.1%	-0.4%	5.9%
$h = 2$	-7.9%	-10.5%	0.2%	0.8%	14.5%
$h = 3$	-26.9%	-28.8%	-0.9%	0.2%	21.5%
$h = 4$	-39.9%	-39.1%	-4.8%	-3.3%	26.2%
$h = 5$	-44.8%	-41.7%	-13.6%	-12.3%	28.8%
$h = 6$	-82.5%	-76.0%	-37.2%	-36.2%	27.0%
$h = 7$	-126.3%	-115.3%	-60.2%	-59.2%	22.3%
$h = 8$	-161.1%	-145.9%	-80.8%	-79.7%	19.7%
$h = 9$	-210.4%	-190.4%	-104.6%	-103.5%	15.5%
$h = 10$	-292.4%	-264.9%	-138.5%	-137.0%	8.5%

Note: Percentage improvement in out-of-sample fit of each investor with respect to the naive investor. Out-of-sample period: 1973:Q1 to 2024:Q4.

that it under-performs the naive investor. Not surprisingly, both investors with flat priors obtain much worse out-of-sample results than the naive investor and adding dividend growth to the VAR with flat priors leads to a moderate improvement at horizons greater than or equal to 4 years.

All investors using informative priors improve their results with respect to those of both investors using flat priors. Thus, shrinkage can improve out-of-sample forecasting performance. Including or omitting dividend growth from the system when introducing informative priors makes little difference to the out-of-sample performance. Neither of the two approaches improves upon the naïve benchmark. Only the investor using the restricted informative priors beats the naive investor out-of-sample. This highlights that (a) there is enough persistence in dividend growth to alter predictions for returns at multiple horizons substantially, and (b) the information encoded into the CS restrictions is essential to leverage the information contained in the dynamics of dividends to sustain out-of-sample forecasting gains. The gains peak at around five years but remain economically significant up to 10 years ahead. The main conclusion of this analysis is that one critically needs to combine Bayesian shrinkage *and* the CS restrictions to obtain a good out-of-sample performance.

In Appendix D, we also consider the out-of-sample forecasting performance of two alternative methods for incorporating the information from lagged dividend growth in predicting returns. The first strategy, discussed in Section 2.1, omits Δd_{t+1} from the model but considers a 3-variable VAR(2) having the additional lags to proxy for the dynamics of (lagged) dividend growth (see Appendix A.4

for further details). This latter approach is correctly specified without approximation but does not impose the CS restrictions (which would require additional restrictions in the coefficients for the second lags of the VAR). The results for this strategy are reported in Table D.2. The informative prior is set following the steps described in Appendix I, accommodating for the omitted variable and the extra lag. The second option is the one detailed in Section 3.5, which essentially allows dividend growth to act as a predictor for $[r_{t+1}, pd_{t+1}]$ without directly modeling the dynamics of d_{t+1}.

Neither of these alternatives yields positive R^2 statistics in out-of-sample results, and both produce forecasts that are systematically inferior to our preferred model. These results underscore the main conclusion of our paper: it is crucial not only to have a correctly specified VAR but also to combine Bayesian shrinkage *and* the CS restrictions to achieve good out-of-sample performance.

5.2 Asset Allocation

We now discuss optimal long-horizon allocations.[17] The investors maximize the expected utility of the terminal value of wealth under Constant Relative Risk Aversion (CRRA) preferences.[18] Each investor chooses the allocation between the risk-free asset and the stock return for one year using different forecasts. To compute the allocations, we use the solution derived by Jurek and Viceira (2011) where the investors re-balance every year, considering how many years are left until retirement. The investors start with a planning horizon of 45 years beginning at the end of 1973 and can invest in equity and a risk-free cash instrument. We will use the posterior mean for each period. We consider a risk aversion coefficient of $\gamma = 5$.

Figure 9 displays the results. Panel (a) displays the wealth accumulation profiles for each of the four investors. To facilitate comparisons across portfolios, each curve is scaled by its ex-post volatility.[19] Panel (b), in turn, presents the history of the weights for the risky asset. Two points

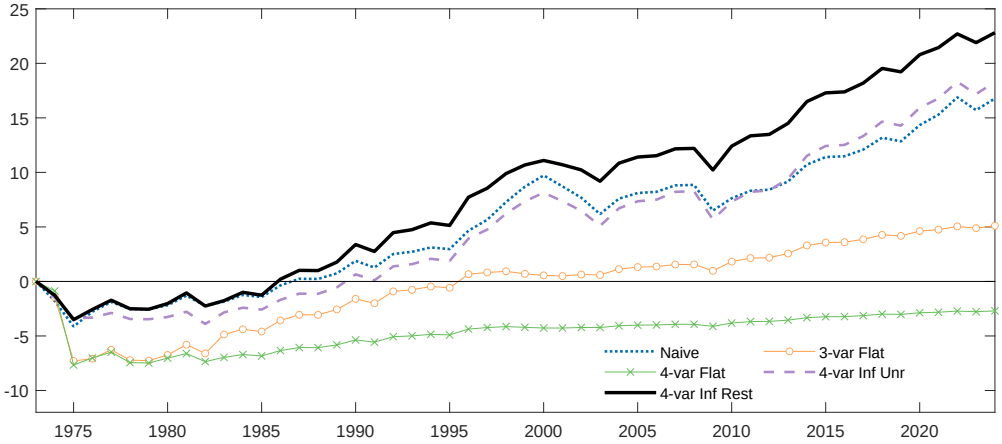
[17] In the subsequent part of this section, we drop the investor from the alternative model set who uses informative priors but excludes dividend growth from their information set. This investor's allocations and related performance closely resemble those of the investor who uses informative priors and includes dividend growth in VAR but does not impose the CS restrictions. As a result, the allocations replicate the similarities highlighted by the forecasting exercise reported in Table 3.

[18] This problem resembles one of the target-date funds, whose dramatic increase in importance in the last decade has been documented by Parker et al. (2020).

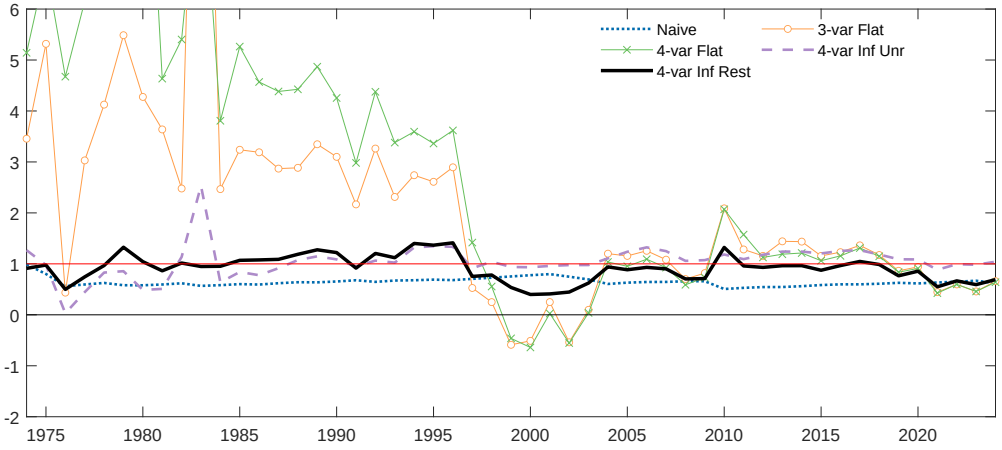
[19] Specifically, we compute the cumulative log excess return earned per unit of risk, where risk is measured by the volatility of log excess returns. For each period T from 1973:Q2 to 2024:Q4, the cumulative log excess return is calculated as $\sum_{t=1973:Q1}^{T} \log\left(\frac{1+r_t^p}{1+r_t^f}\right)$, which represents the log of the final wealth of an investor who holds a portfolio yielding return r_t^p while borrowing at the risk-free rate r_t^f. This cumulative return is then divided by the standard

Figure 9: Asset Allocation Results: Long-Term Strategy

(a) Risk-Adjusted Log Wealth



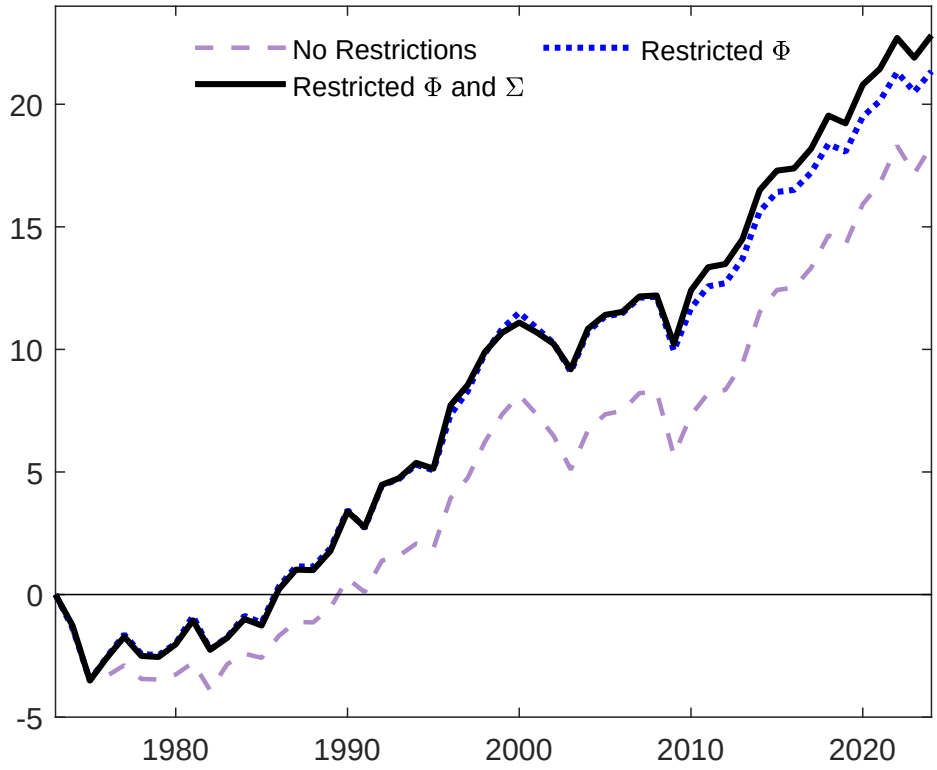
(b) Portfolio Weight to Equities (%)



are worth noticing. First, the investor using the restricted informative prior is the only one who systematically improves upon the naive approach. This is in line with the out-of-sample performance reported above. Second, both investors using flat priors use massive amounts of leverage (up to 800 percent), which leads to large draw-downs, particularly in the first half of the sample, i.e. when the investment horizon is long. Both investors using informative priors have more moderate weight profiles. Still, the restricted informative prior leads to weights leveraged only occasionally and by marginal amounts, and the investors never take short positions. These two desirable properties lead to the superior performance in Panel (a). Although the profile of weights of the naive investor is the most stable, it performs worse (in terms of risk-adjusted wealth) than the profiles of both

deviation of $\log\left(\frac{1+r_t^p}{1+r_t^f}\right)$, computed over the entire sample period.

Figure 10: Risk-Adjusted Log Wealth Breakdown



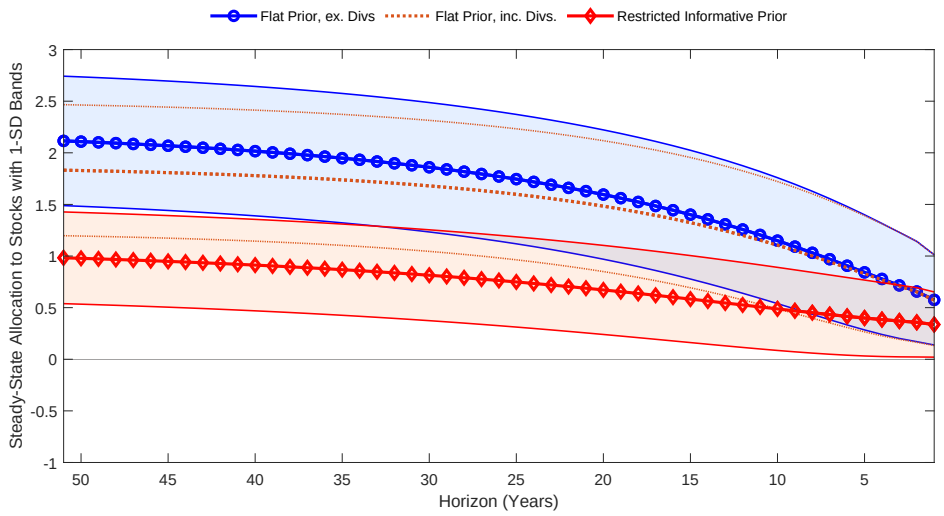
investors using informative priors since it does not take advantage of the moderate degree of return predictability estimated in the data.

Figure 10 illustrates how the risk-adjusted wealth gains of an investor employing restricted informative priors can be decomposed into contributions from the imposition of the CS restrictions on (1) $\mathbf{\Phi}$ and (2) $\mathbf{\Sigma}$. Approximately two-thirds of the cumulative gains are driven by the CS restrictions on $\mathbf{\Phi}$, while the remaining one-third is attributable to those imposed on $\mathbf{\Sigma}$.

Why does the model with restricted informative priors lead to less leverage and less aggressive timing? Figure 11 plots the mean allocations as a function of investment horizon, together with bands representing the expected standard deviation of the allocations using the posterior mean using data until 2024. Thus, the central lines with markers represent the average allocation to stocks that an investor would expect to hold if the variables of the VAR were at their unconditional mean, and the width of the bands represents the amount of market timing that the investor expects to engage in, in response to changes in investment opportunities.

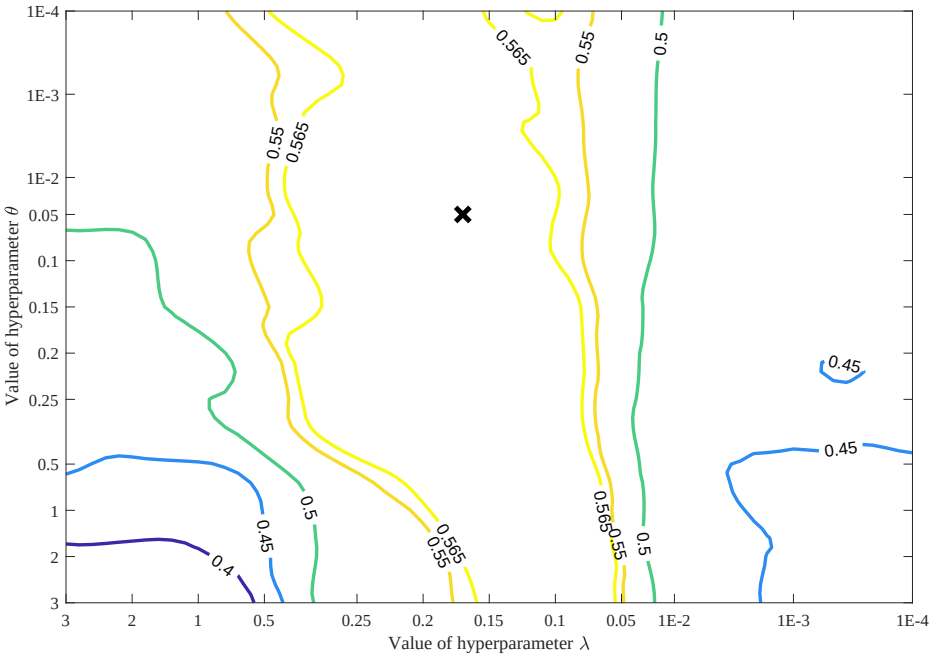
The first model considered is the one with flat priors excluding dividend growth. This model

Figure 11: STEADY STATE ALLOCATION TO STOCKS UNDER DIFFERENT MODELS



assumes a large degree of predictability, which has two effects. First, it implies ample amounts of return mean reversion, which, combined with the fact that this model rules out dividend momentum, implies a strong hedging demand for stocks (large declining slope and highly leveraged average position). Second, it shows excellent timing around the average position (wide bands). The resulting large and volatile allocations are in line with the classic findings of Campbell and Viceira (1999) but have been challenged by Viceira (2008) concerning their applicability in comparison to the real-world practices of the finance industry. The addition of dividends, while maintaining flat priors in the second model, allows the investor to incorporate dividend momentum. This leads to a reduction in average allocations and slope, consistent with the fact that dividend momentum makes long-term returns riskier, as explained in Section 4.5. This model still has wide bands associated with the large degree of return predictability implied by flat priors. The restricted informative prior, in turn, leads to an additional decline in both the slope and the average long-term allocation by tilting the model parameters toward values consistent with a lower degree of return predictability. This leads to positions close to 100 percent stocks at the beginning of the planning horizon, close to 50 percent when the investor nears retirement, and positions with a modest amount of market timing. It is worth noting that these average allocation prescriptions resemble those of real-world target-date funds, whereas the ones based on flat priors are unrealistic from this point of view. Viceira (2008) notes that existing investment advice is inconsistent with the quantitative results of portfolio choice problems. Still, we show that priors that encode shrinkage of return predictability, together with

Figure 12: Sharpe Ratio as a function of Hyperparameters



Note: The hyperparameter values used throughout this section are marked with a cross in the figure.

accounting for the increase in risk for the long-run investor associated with the presence of dividend momentum, may be able to reconcile the two.

5.3 Choosing the Prior Tightness

Our restricted informative priors embody a degree of skepticism about the existence of return predictability. This is governed by two hyperparameters that we choose to maximize the value of the marginal likelihood, as proposed by Giannone et al. (2015), using only data up to 1973. The previous results highlight that imposing informative priors delivers clear gains with respect to any alternative prior when measured in terms of risk-adjusted wealth at the end of the sample. Figure 12 explores the sensitivity of the results above to the tightness of the restricted informative prior.

The figure plots the contours of the Sharpe ratio at the end of the sample under different choices of hyperparameters. The Sharpe ratios are calculated with out-of-sample moments. The "**X**" in the graph represents our baseline choice of hyperparameters, resulting from maximizing the value of the marginal likelihood in the pre-sample up to 1973. The figure shows that the ex-ante procedure by Giannone et al. (2015) selects hyperparameters close to those that ex-post maximize the Sharpe

ratio. In particular, the hyperparameter λ, which governs the overall tightness of the Minnesota prior, needs to be sufficiently tight, between a value of 0.2 and 0.05, to obtain the gains reported above.[20] Of course, if the parameter is tighter and tighter, the gains start to decline as the degree of predictability is dogmatically driven to zero, and the model converges to the naive strategy. The Sharpe ratio is less sensitive to the choice of the hyperparameter θ, which governs the overall tightness of the Single Unit Root prior.

5.4 The Myopic Investor

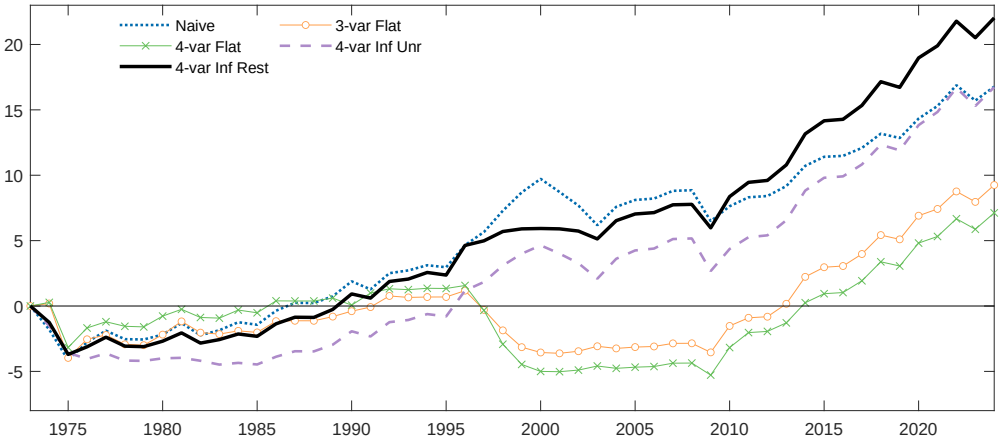
Let us now consider the problem of myopic investors. These investors act as if they were a one-period investor every year, not taking into account that they will continue investing for many years to come. Figure 13 replicates Figure 9 for the case of myopic investors.

Again, two points are worth noticing. First, the investor using the restricted informative prior is the only one who systematically improves upon the investor using the naive approach. This is in line with the out-of-sample performance reported above. Second, the profile of weights of the naive investor is the most stable. Both investors using flat priors take large short positions, leading to large draw-downs, particularly in the last half of the 1990s. Both investors using informative priors have more moderate weight profiles. However, the restricted informative prior leads to weights short only occasionally and by marginal amounts and never require leverage. These two desirable properties lead to superior performance. Figure 14 replicates Figure 12 for the case of myopic investors. As before, the figure shows that to maximize the Sharpe ratios, one needs to center the prior around non-predictability and with a sufficiently high degree of tightness for this type of investor. Interestingly, in this case the hyperparameter θ needs to be sufficiently tight as well. Moreover, the combination selected a priori by maximizing the marginal likelihood is once again very close to the ex-post optimal choice.

[20] In the macroeconomics literature, the value of 0.2 is usually considered a benchmark.

Figure 13: Asset Allocation Results: Myopic Strategy

(a) Risk-Adjusted Log Wealth



(b) Portfolio Weight to Equities (%)

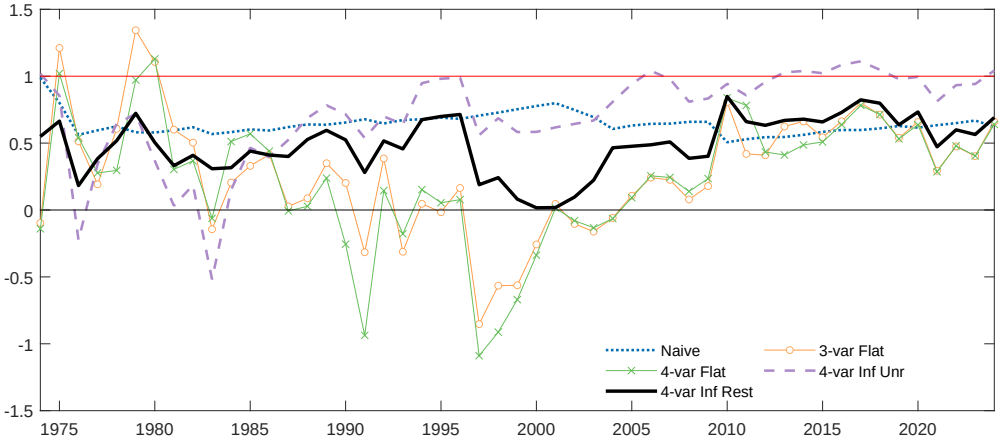
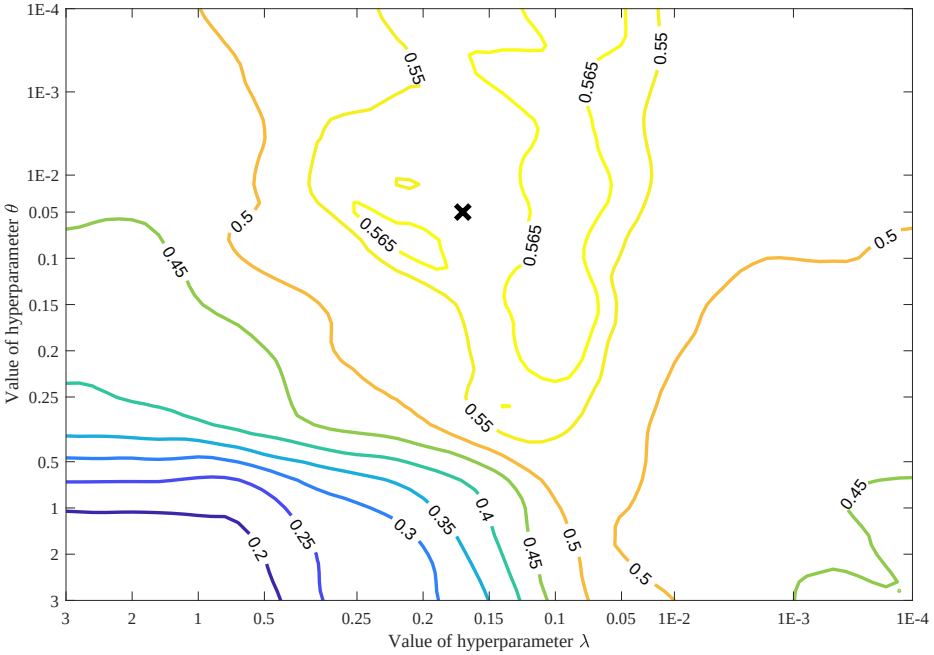


Figure 14: Sharpe Ratio as a function of Hyperparameters



Note: The hyperparameter values used throughout this section are marked with a cross in the figure.

6 Dividend Momentum at Higher Frequency

In this section, we investigate the presence of dividend momentum using quarterly data. Specifically, we use S&P500 data on dividends and prices for the sample period 1946:Q1 to 2024:Q4. The data and the prior are constructed consistently as we described in Section 4.1. Working with data at a quarterly frequency requires addressing the presence of seasonality in dividends. A common approach to eliminate seasonality is to "smooth" dividends using a moving average of dividends paid over the preceding year. Taking a moving average of dividends implies that dividend growth could exhibit some mechanical persistence. An alternative approach is to use smoothed dividends but treat the unsmoothed quarterly growth rate in dividends as a latent factor (see Schorfheide et al., 2018). While this approach has the appealing feature of avoiding the need to explicitly model seasonality in the data, it does not allow for restricting that the seasonality of the price-dividend ratio and that of dividend growth are intrinsically related, as required by the CS identity.

We assume that the price-dividend ratio and dividend growth can be decomposed into seasonal and seasonally adjusted components, whereas returns do not exhibit any seasonality. Therefore:

$$\begin{aligned} \Delta d_{t+1} &= \Delta d^{s}_{t+1} + \Delta d^{sa}_{t+1} \\ pd_{t+1} &= pd^{s}_{t+1} + pd^{sa}_{t+1} \end{aligned} \tag{29}$$

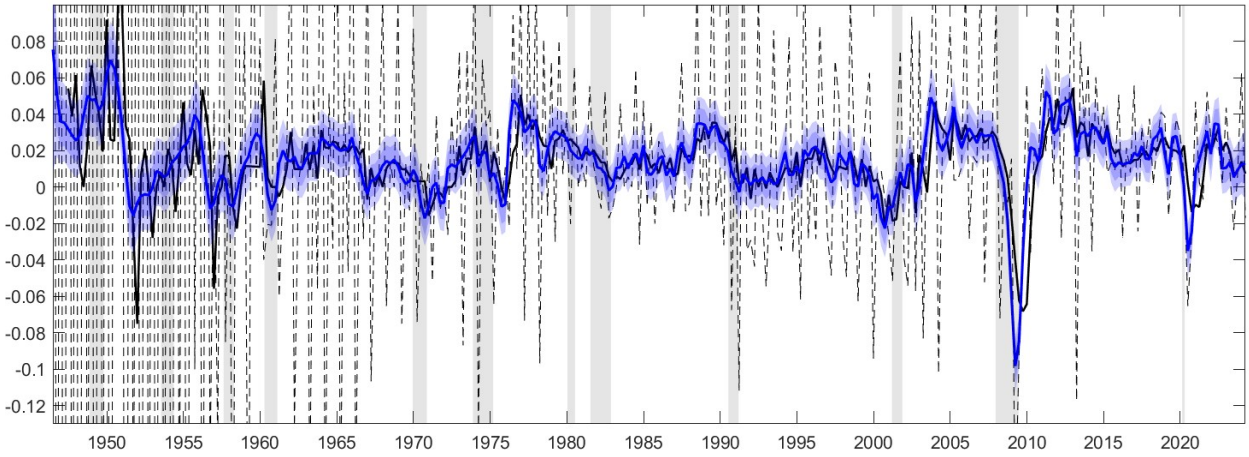
where the superscript "s" denotes the seasonal component and "sa" denotes the seasonally adjusted component. If returns do not exhibit any seasonal pattern, the CS identity implies that the seasonal components of dividends and the price-dividend ratio are related, both deriving from the seasonality in dividends:

$$\Delta d^{s}_{t+1} = -\rho pd^{s}_{t+1} + pd^{s}_{t} \tag{30}$$

In line with Harvey and Todd (1983), we allow the seasonal variation in the data to change slowly over time, using a mechanism that ensures the sum of the seasonal components over any three consecutive quarters has an expected value of zero:

$$\sum_{j=0}^{3} pd^{s}_{t-j} = \sigma^{s}_{t} e^{s}_{t} \tag{31}$$

Figure 15: Seasonally Adjusted Quarterly Dividend Growth



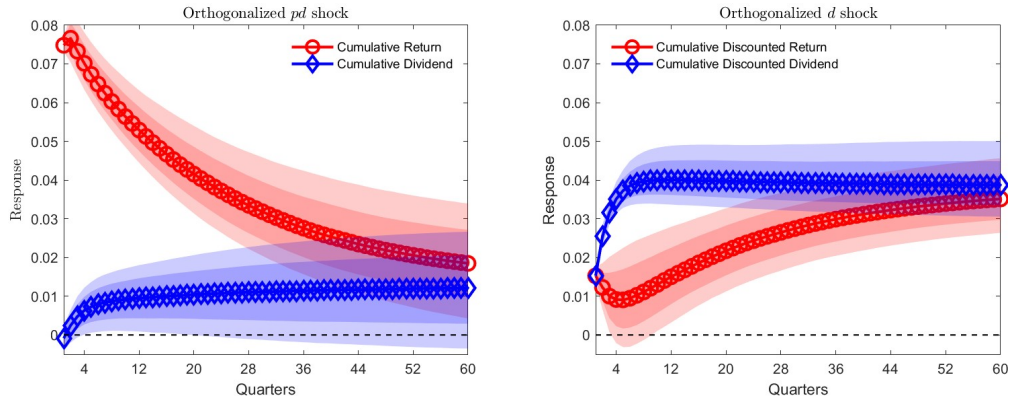
Note: Black broken line denotes the raw data, black continuous line denotes the data constructed using the average of the past 4 quarters of dividends. The blue line denotes the seasonally adjusted data from our model. The darker shadow area represents the 68^{th} posterior credible intervals, while the lighter shadow area represents the 95^{th} posterior credible intervals.

where $e_t^s \sim$ i.i.d.$\mathcal{N}(0,1)$. Compared to the classic formulation of a seasonal model, we incorporate stochastic volatility. Specifically, we assume that $\log \sigma_t^s = \log \sigma_{t-1}^s + \omega_{t+1}$, where $\omega_t \sim$ i.i.d. $\mathcal{N}(0, \sigma_\omega^2)$. This allows us to capture the impact of the evolution in dividend payout policies over the sample. In the early part of the sample, dividends were often concentrated in specific months, whereas in the later part of the sample, they became more evenly distributed. Indeed, the first two decades of the sample exhibit significantly larger seasonality compared to the remainder of the period. Lastly, we assume that the (partially unobserved) seasonally adjusted data (i.e., $[r_t, \Delta d_t^{sa}, pd_t^{sa}]$) follow a 3-variable VAR(1), with the restrictions implied by the CS identity imposed. The model can be cast in state space, allowing recovery of the latent (seasonal and seasonally adjusted) dividends and dividend growth. We set priors following the steps described in Section 3.1. For the stochastic volatility, we follow Omori et al. (2007). Parameter estimates can be fully recovered using a Gibbs sampler. For more details, see Appendix J.

Figure 15 reports the retrieved estimates of seasonally adjusted dividend growth alongside the seasonally unadjusted data and dividend growth computed on moving average quarterly dividends over the past year. Dividends display pronounced seasonality in the early part of the sample, which declines rapidly starting from the 1960s.[21] As expected, series constructed using smoothed dividends

[21]We estimate that the volatility in the early part of the sample is roughly eight times larger than the average level observed since the mid-1960s. In fact, in the earlier part of the sample, seasonal variations are also observable in the

Figure 16: Dividend Momentum with Quarterly Data



Note: The solid lines represent the median posterior response. The darker shadow area represents the 68^{th} posterior credible intervals, while the lighter shadow area represents the 95^{th} posterior credible intervals.

tend to lag our estimates, especially around major turning points. Most importantly, dividends exhibit pronounced cyclicality even at a quarterly frequency. Lagged dividend growth remains a relevant predictor of both dividend growth and the price-dividend ratio, even after accounting for the predictability of the lagged price-dividend ratio and past returns.

Figure 16 displays the response of cumulative dividends and returns following a shock to the price-dividend ratio and a dividend growth shock, identified as discussed in Section 4.5. Dividend momentum remains a significant feature of the data, even at a quarterly frequency. Dividend shocks result in an immediate increase in dividends, followed by a persistent movement in expected dividend growth and expected returns in the same direction.

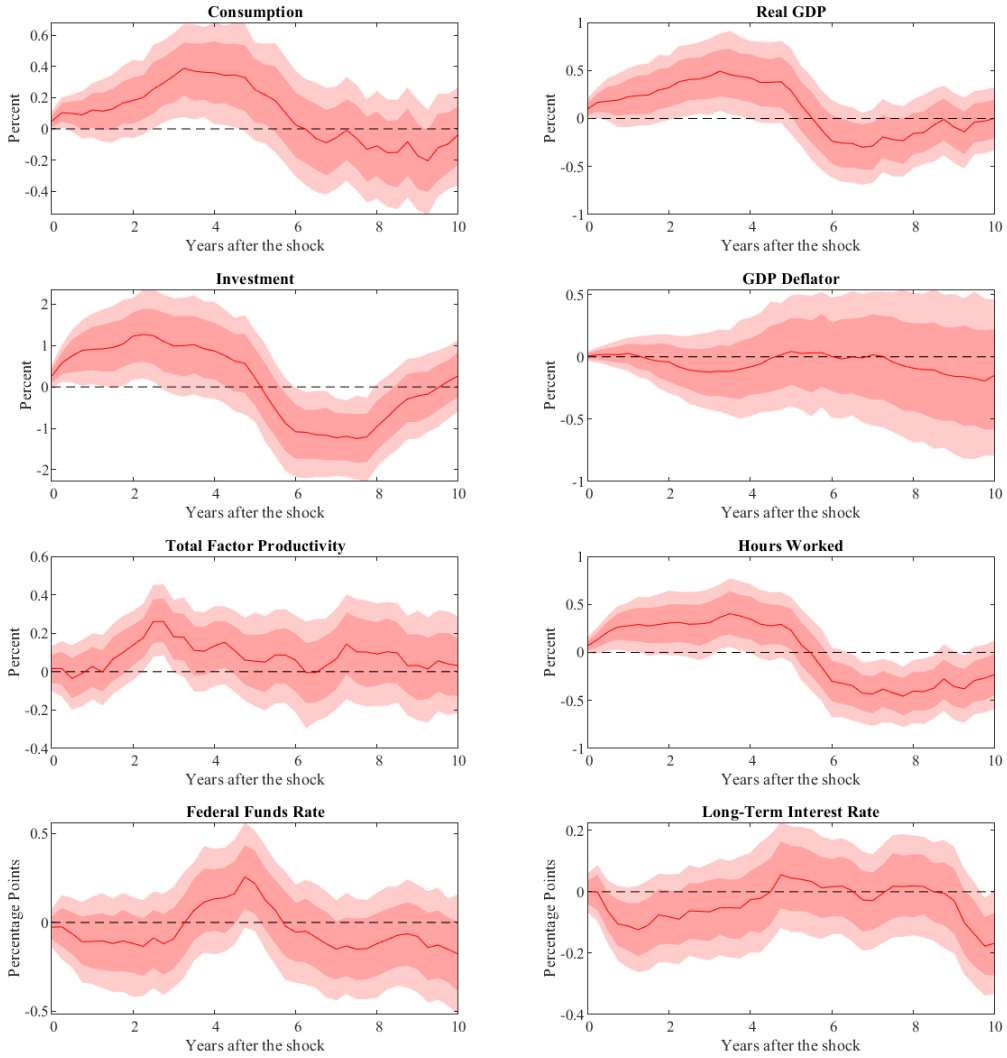
One of the advantages of our quarterly specification is that we can use the identified shocks in Local Projections (LPs, Jordà, 2005) onto macroeconomic aggregates to gain insights into the economic interpretation of the dividend momentum channel. Figure 17 displays such LPs using the dividend growth shocks. To deal with the fact that the shock from the VAR is a generated regressor, we sample 1000 time series of the shock from the posterior distribution and report 68% and 90% confidence intervals that integrate out the uncertainty around VAR coefficients. We also control for four lags of the outcome and impulse variables in the LPs. The figure shows that such shocks are associated with significant effects macroeconomic variables at frequencies somewhat longer than the usual business cycles: while the response is muted on impact, it builds out over time, reaching a peak

price-dividend ratio, even when smoothed dividends are used. In the post-1960s period, the estimated time-varying volatility of the seasonal component of the data is relatively stable.

between three and four years after impact–somewhat later than the peak of the dividend growth IRF seen in Figure 16. Interestingly, the responses of most variables turn negative after about five years, with those of investment and hours worked significantly so. The responses of prices and interest rates are not significant for most of the projection horizon. In Appendix J.4 we plot the counterpart of this figure for the price-dividend shock. The results show that this type of shock is associated with responses of macroeconomic variables that are in the same direction but are larger and peak at around the one-year horizon, and with a significant increase in the federal funds rate–more similar to the standard business cycle fluctuations. Therefore, while both shocks are associated with similar movements in economic variables, it appears that they are capturing phenomena occurring at different points in the frequency spectrum. While the price-dividend shock is associated with rapid responses in macroeconomic variables, the results suggest that the dividend growth shock leading to dividend momentum is connected to prolonged boom-bust cycles in economic activity that materialize first in an increase in dividend growth and only later on measures of broad economic activity. We stress that these results, which rely on a particular recursive ordering of the shocks, are merely suggestive. As the vast literature on identification in macroeconomics has established (Arias et al., 2018, see, e.g.) even beyond the possible orderings of the variables, there is an infinite number of potential orthogonalizations that are equally consistent with the reduced-form covariance matrix. Nevertheless, they are informative about an association between dividend growth and persistent movements in economic variables at business cycle frequencies and beyond. Understanding the deep causal drivers of these fluctuations is an important avenue for further research.

The difficulty of predicting returns out-of-sample with quarterly data is well documented (see, e.g., Goyal and Welch, 2008). We examine whether our model can outperform the trailing mean benchmark in this setting. Starting from 1973:Q1, we generate real-time forecasts at various horizons using the 3-variable VAR(1) model with restricted informative priors described above. These forecasts are then used to compute cumulative returns over the forecast window. The model consistently outperforms the naïve benchmark across all horizons (up to five years). The "Out-of-Sample R^2" for one-quarter-ahead forecasts is 0.6%, which aligns with the economically plausible gains according to Campbell and Thompson (2008). This metric increases to 3.1% at the one-year horizon and reaches 17% for five-year-ahead forecasts. For comparison, note that the "Out-of-Sample R^2" values are negative for both a 2-variable VAR(1) using $[r_t, pd_t]$ and a 3-variable

Figure 17: Local Projections of Macroeconomic Variables on Dividend Growth Shock



Note: The solid lines represent the median posterior response. The darker shadow area represents the 68^{th} posterior credible intervals, while the lighter shadow area represents the 90^{th} posterior credible intervals.

VAR(1) including $[r_t, pd_t, \Delta d_t]$ estimated using flat priors (in both cases using smoothed dividends). For instance, the 3-variable VAR(1) yields "Out-of-Sample R^2" values of -1%, -5.5%, and -8.7% at the one-quarter, one-year, and five-year horizons, respectively.

In Appendix J, we report additional robustness analyses: (a) re-estimating the model while assuming constant volatility for the seasonal component of the data and (b) estimating the quarterly model using smoothed dividend data, as is customary in the literature. In both cases, the results are consistent with those reported in this section, providing strong evidence supporting the presence of dividend momentum. Lastly, we investigate the presence of dividend momentum using monthly

data and find evidence that supports the existence of this channel even at this frequency.

7 Conclusion and Implications for Future Research

In this paper, we have proposed a Bayesian approach to VAR inference that starts from informative priors that embody skepticism about the degree of predictability in stock returns, address the high persistence of the price-dividend ratio, and impose the cross-equation restrictions implied by the CS identity. Adopting a Bayesian perspective with a conservative prior belief regarding the predictability of future returns leads to substantial gains in out-of-sample forecasting performance and asset allocation recommendations. We highlight the importance of including dividend growth in a VAR with the price-dividend ratio and returns and note that a common practice of omitting dividend growth amounts to imposing the additional restriction that this variable is not persistent after controlling for lags of the remaining variables. Using annual postwar data and relaxing this additional restriction uncovers an additional and previously overlooked channel of return predictability, which we label "dividend momentum." We show how dividend momentum has non-trivial implications for interpreting cash flow and discount rate news and the optimal asset allocation of long-horizon investors.

These results offer valuable insights for future research areas that remain unexplored within this paper. On the empirical front, there is burgeoning literature that uses large data sets to search for evidence of predictability in aggregate stock returns (see Kelly and Pruitt, 2013; Rapach and Zhou, 2020). Shrinkage or regularization of a potentially large parameter and predictor space becomes essential in this context. However, when focusing on shrinkage, it is easy to forget the lessons of the classic papers by Campbell and Shiller (1988a,b) and Cochrane (2008b): the price-dividend ratio and aggregate dividend growth are not *any* predictors: they are fundamentally linked to returns by the CS identity. A corollary of our results in the context of larger data sets is that any predictor of dividend growth will indirectly forecast returns through the dividend momentum channel.The logic of the CS identity means that any predictor of dividends, other than the price-dividend ratio itself, will either predict returns or the price-dividend ratio (with the opposite sign), therefore predicting returns with a lag. Thus, shocks to this predictor will also induce dividend momentum. Therefore, our approach opens the door to using large Bayesian VARs, which are highly successful in forecasting

macroeconomic and financial variables with hundreds of predictors (see Koop, 2013; Carriero et al., 2019), for the task of uncovering return predictability while respecting the cross-equation restrictions implied by the CS type of identities. More broadly, as mentioned at the outset of the paper, our methods extend to the many applications in macroeconomics and finance where identities equivalent to the CS one emerge. In those applications, a highly persistent ratio or spread is the key long-run predictor of an asset's return. Exploring the implications of priors that push higher the persistence of such predictors and shrink toward zero the coefficients related to predictability would be a fruitful avenue of research.

From a theoretical standpoint, dividend momentum carries implications for asset pricing theories that try to explain fluctuations in expected returns. In the habit model of Campbell and Cochrane (1999) and the prospect theory of Barberis et al. (1999), dividend growth is modeled as independently and identically distributed, and all of the predictability of returns comes from variation in discount rates. In the long-run risk model of Bansal and Yaron (2004), dividend growth features a persistent, low-frequency component but is independently and identically distributed after controlling for the lagged price-dividend ratio. These theories are concerned with explaining the variation in expected returns that is counter-cyclical with respect to dividend and/or consumption growth. The presence of dividend momentum does not negate that such mechanisms are the largest driver of fluctuations in expected returns. Instead, it points to at least one shock that generates pro-cyclical variation in expected returns through channels that the literature has so far not explored.

References

Antolín-Díaz, J., I. Petrella, and J. F. Rubio-Ramírez (2021): "Structural scenario analysis with SVARs," *Journal of Monetary Economics*, 117, 798–815.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018): "Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685–720.

Avramov, D. (2002): "Stock return predictability and model uncertainty," *Journal of Financial Economics*, 64, 423–458.

Avramov, D., S. Cederburg, and K. Lučivjanská (2018): "Are Stocks Riskier over the Long Run? Taking Cues from Economic Theory," *Review of Financial Studies*, 31, 556–594.

Bansal, R. and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *The Journal of Finance*, 59, 1481–1509.

Barberis, N. (2000): "Investing for the Long Run When Returns Are Predictable," *The Journal of Finance*, 55, 225–264.

Barberis, N., M. Huang, and T. Santos (1999): "Prospect Theory and Asset Prices," *Quarterly Journal of Economics*, 116, 1–53.

Bekaert, G., R. J. Hodrick, and D. A. Marshall (1997): "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Financial Economics*, 44, 309–348.

Boudoukh, J., M. Richardson, and R. F. Whitelaw (2008): "The Myth of Long-Horizon Predictability," *The Review of Financial Studies*, 21, 1577–1605.

Campbell, J. Y. (1991): "A Variance Decomposition for Stock Returns," *Economic Journal*, 101, 157–179.

Campbell, J. Y. and J. Ammer (1993): "What moves the stock and bond markets? A variance decomposition for long-term asset returns," *The Journal of Finance*, 48, 3–37.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira (2003): "A Multivariate Model of Strategic Asset Allocation," *Journal of Financial Economics*, 67, 41–80.

Campbell, J. Y., J. Cocco, F. Gomes, P. J. Maenhout, and L. M. Viceira (2001): "Stock Market Mean Reversion and the Optimal Equity Allocation of a Long-Lived Investor," *Review of Finance*, 5, 269–292.

Campbell, J. Y. and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

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

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

Campbell, J. Y. and N. G. Mankiw (1989): "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," *NBER Macroeconomics Annual*, 4, 185–216.

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

——— (1988b): "Stock prices, earnings, and expected dividends," *The Journal of Finance*, 43, 661–676.

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

Campbell, J. Y. and L. M. Viceira (1999): "Consumption and Portfolio Decisions when Expected Returns are Time Varying," *The Quarterly Journal of Economics*, 114, 433–495.

——— (2002): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Clarendon Lectures in Economic.

Campbell, J. Y. and T. Vuolteenaho (2004): "Bad Beta, Good Beta," *American Economic Review*, 94, 1249–1275.

Carriero, A., T. E. Clark, and M. Marcellino (2019): "Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors," *Journal of Econometrics*, 212, 137–154.

Chen, L., Z. Da, and R. Priestley (2012): "Dividend Smoothing and Predictability," *Management Science*, 58, 1834–1853.

Chen, L. and X. Zhao (2009): "Return decomposition," *The Review of Financial Studies*, 22, 5213–5249.

Cochrane, J. H. (2008a): "State-Space vs. VAR Models for Stock Returns," *Unpublished Manuscript*.

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

——— (2009): *Asset Pricing: Revised Edition*, Princeton University Press.

——— (2011): "Discount Rates," *Journal of Finance*, 66, 1047–1108.

——— (2017): "Macro-finance," *Review of Finance*, 21, 945–985.

——— (2019): "The Fiscal Roots of Inflation," *National Bureau of Economic Research*.

Del Negro, M. and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARS," *International Economic Review*, 45, 643–673.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and Conditional Projection using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Engel, C. (2016): "Exchange Rates, Interest Rates, and the Risk Premium," *American Economic Review*, 106, 436–474.

Engel, C. and K. D. West (2005): "Exchange Rates and Fundamentals," *Journal of Political Economy*, 113, 485–517.

Engsted, T., T. Q. Pedersen, and C. Tanggaard (2012): "Pitfalls in VAR based return decompositions: A clarification," *Journal of Banking & Finance*, 36, 1255–1265.

Fama, E. F. and K. R. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

——— (2002): "The Equity Premium," *Journal of Finance*, 57, 637–659.

Froot, K. A. and T. Ramadorai (2005): "Currency Returns, Intrinsic Value, and Institutional-Investor Flows," *Journal of Finance*, 60, 1535–1566.

Gao, C. and I. W. R. Martin (2021): "Volatility, Valuation Ratios, and Bubbles: An Empirical Measure of Market Sentiment," *Journal of Finance*, 76, 3211–3254.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

——— (2019): "Priors for the Long Run," *Journal of the American Statistical Association*, 114, 565–580.

——— (2021): "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica*, 89, 2409–2437.

——— (2022): "Rethinking Instrumental Variables," Tech. rep.

Gourinchas, P.-O. and H. Rey (2007): "International Financial Adjustment," *Journal of Political Economy*, 115, 665–703.

——— (2019): "Global Real Rates: A Secular Approach," BIS Working Papers 793, Bank for International Settlements.

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

Harvey, A. C. and P. H. J. Todd (1983): "Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study," *Journal of Business & Economic Statistics*, 1, 299–307.

Hurwicz, L. (1950): "Least Squares Bias in Time Series," in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, Cowles Commission Monograph number 10, New York: Wiley.

Jarocinski, M. and A. Marcet (2015): "Contrasting Bayesian and Frequentist Approaches to Autoregressions: The Role of the Initial Condition," Tech. rep., Working Papers 776, Barcelona Graduate School of Economics.

Jiang, Z., H. Lustig, S. V. Nieuwerburgh, and M. Z. Xiaolan (2024): "What Drives Variation in the U.S. Debt-to-Output Ratio? The Dogs that Did not Bark," *Journal of Finance*, 79, 2603–2665.

Jordà, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American economic review*, 95, 161–182.

Jurek, J. W. and L. M. Viceira (2011): "Optimal Value and Growth Tilts in Long-Horizon Portfolios," *Review of Finance*, 15, 29–74.

Kandel, S. and R. F. Stambaugh (1996): "On the Predictability of Stock Returns: An Asset-Allocation Perspective," *The Journal of Finance*, 51, 385–424.

Kelly, B. and S. Pruitt (2013): "Market Expectations in the Cross-Section of Present Values," *The Journal of Finance*, 68, 1721–1756.

Koijen, R. and S. Nieuwerburgh (2012): "Predictability of Returns and Cash Flows," *Annual Review of Financial Economics*, 3.

Koijen, R. S. J., J. C. Rodríguez, and A. Sbuelz (2009): "Momentum and Mean Reversion in Strategic Asset Allocation," *Management Science*, 55, 1199–1213.

Koop, G. M. (2013): "Forecasting with Medium and Large Bayesian VARs," *Journal of Applied Econometrics*, 28, 177–203.

Larrain, B. and M. Yogo (2008): "Does Firm value Move Too Much to Be Justified by Subsequent Changes in Cash Flow?" *Journal of Financial Economics*, 87, 200–226.

Leamer, E. E. (1973): "Multicollinearity: A Bayesian Interpretation," *The Review of Economics and Statistics*, 55, 371–380.

Lettau, M. and S. Ludvigson (2001): "Consumption, Aggregate Wealth, and Expected Stock Returns," *The Journal of Finance*, 56, 815–849.

Lewellen, J. (2004): "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, 74, 209–235.

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

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012): "Time series momentum," *Journal of Financial Economics*, 104, 228–250.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425–449.

Parker, J. A., A. Schoar, and Y. Sun (2020): "Retail Financial Innovation and Stock Market Dynamics: The Case of Target Date Funds," Tech. rep., National Bureau of Economic Research.

Pástor, L. and R. F. Stambaugh (2009): "Predictive Systems: Living with Imperfect Predictors," *The Journal of Finance*, 64, 1583–1628.

——— (2012): "Are Stocks Really Less Volatile in the Long Run?" *The Journal of Finance*, 67, 431–478.

Pettenuzzo, D., A. Timmermann, and R. Valkanov (2014): "Forecasting stock returns under economic constraints," *Journal of Financial Economics*, 114, 517–553.

Rapach, D. E. and G. Zhou (2020): "Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods," *Machine Learning for Asset Management: New Developments and Financial Applications*, 1–33.

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

Schorfheide, F., D. Song, and A. Yaron (2018): "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach," *Econometrica*, 86, 617–654.

Sims, C. and H. Uhlig (1991): "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, 59, 1591–99.

Sims, C. A. (1993): "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators, and Forecasting*, University of Chicago Press, 179–212.

Sims, C. A. and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 949–968.

Stambaugh, R. F. (1999): "Predictive regressions," *Journal of Financial Economics*, 54, 375–421.

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381 – 419.

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

Viceira, L. (2008): *Life-Cycle Funds*, University of Chicago Press.

Wachter, J. A. and M. Warusawitharana (2009): "Predictable Returns and Asset Allocation: Should a Skeptical Investor Time the Market?" *Journal of Econometrics*, 148, 162–178.

——— (2015): "What is the Chance that the Equity Premium Varies over Time? Evidence from Regressions on the Dividend-Price Ratio," *Journal of Econometrics*, 186, 74–93.

Internet Appendix to

"Dividend Momentum and Stock Return Predictability: A Bayesian Approach"

Appendix A Omitting Dividend Growth

Since returns are the ultimate variable of interest, the standard choice is to drop dividend growth and run a 2-variable VAR(1) on $\mathbf{x}'_{t+1} = [pd_{t+1}, r_{t+1}]$. Let us partition the 3-variable VAR(1) in Equation (1), to isolate the vector $\mathbf{x}_{t+1}$:[22]

$$\begin{bmatrix} \Delta d_{t+1} \\ \mathbf{x}_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_{d,d} & \boldsymbol{\phi}_{21} \\ \boldsymbol{\phi}_{12} & \boldsymbol{\Phi}_{11} \end{bmatrix} \begin{bmatrix} \Delta d_t \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} u^d_{t+1} \\ \boldsymbol{\xi}_{t+1} \end{bmatrix}, \tag{A.1}$$

where:

$$\boldsymbol{\Phi}_{11} = \begin{bmatrix} \phi_{pd,pd} & \phi_{pd,r} \\ \phi_{r,pd} & \phi_{r,r} \end{bmatrix}, \ \boldsymbol{\phi}_{12} = \left[\phi_{pd,d}, \phi_{r,d}\right]', \ \boldsymbol{\phi}_{21} = \left[\phi_{d,pd}, \phi_{d,r}\right], \text{ and } \boldsymbol{\xi}_{t+1} = \left[u^{pd}_{t+1}, u^{r}_{t+1}\right]'.$$

At this stage, the 3-variable VAR(1) is unrestricted and does not impose the CS restrictions. However, the CS identity in Equation (2) implies that $\Delta d_t = r_t - \rho pd_t + pd_{t-1} + \eta_t$. Substituting this into Equation (A.1) one gets the following representation for $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{G}_1 \mathbf{x}_t + \mathbf{G}_2 \mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1} - \boldsymbol{\phi}_{12} \eta_t, \tag{A.2}$$

where $\mathbb{E}\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}'_{t+1}\right) = \boldsymbol{\Omega}_{\boldsymbol{\xi}}$, $\mathbf{G}_1 = \boldsymbol{\Phi}_{11} + \boldsymbol{\phi}_{12}\left[-\rho, 1\right]$, $\mathbf{G}_2 = \left[\boldsymbol{\phi}_{12}, \mathbf{0}_{2\times 1}\right]$. Using the results in Appendix A.1, Equation (A.2) implies the following VARMA(2,1) representation of $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{G}_1 \mathbf{x}_t + \mathbf{G}_2 \mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1 \mathbf{e}_t, \tag{A.3}$$

Where $\mathbf{e}_{t+1}$ is Gaussian with $\mathbb{E}\left(\mathbf{e}_{t+1}\right) = \mathbf{0}_{2\times 1}$, $\mathbb{E}\left(\mathbf{e}_{t+1}\mathbf{e}'_{t+1}\right) = \boldsymbol{\Omega}_{\mathbf{e}}$, $\mathbb{E}\left(\mathbf{e}_s\mathbf{e}'_r\right) = \mathbf{0}_{2\times 2}$ if $r \neq s$, and $\boldsymbol{\Omega}_{\mathbf{e}}$ is an SPD matrix, where $\boldsymbol{\Omega}_{\mathbf{e}}$ and $\mathbf{D}_1$ are nonlinear functions of the original parameters of the model,

[22] Without loss of generality, in this section we abstract from the constant term.

therefore, the autoregressive parameters of the VARMA(2,1) are linearly related to the ones in the 3-variable VAR(1) in Equation (1), and the presence of the moving average component arises from the approximation error in the CS identity.[23]

Let us assume that we want to follow the common practice and specify a 2-variable VAR(1) for $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{A}_1 \mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1}, \tag{A.4}$$

where $\boldsymbol{\varepsilon}_{t+1}$ is Gaussian with $\mathbb{E}(\boldsymbol{\varepsilon}_{t+1}) = \mathbf{0}_{2\times 1}$, $\mathbb{E}(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}') = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon}}$, $\mathbb{E}(\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_r') = \mathbf{0}_{2\times 2}$ if $r \neq s$, and $\boldsymbol{\Omega}_{\boldsymbol{\varepsilon}}$ is an SPD matrix. $\mathbf{A}_1$ satisfies the following moment condition $\mathbb{E}[(\mathbf{x}_{t+1} - \mathbf{A}_1\mathbf{x}_t)\mathbf{x}_t'] = \mathbf{0}_{2\times 2}$, which implies that $\mathbf{G}_1\boldsymbol{\Gamma}_0 + \mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_{\mathbf{e}} - \mathbf{A}_1\boldsymbol{\Gamma}_0 = \mathbf{0}_{2\times 2}$, where $\boldsymbol{\Gamma}_j$ is the j-th autocovariance of $\mathbf{x}_t$, hence:

$$\mathbf{A}_1 = \mathbf{G}_1 + \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)\boldsymbol{\Gamma}_0^{-1}. \tag{A.5}$$

Equation (A.5) highlights the link between $\mathbf{A}_1$ and the parameters in the 3-variable VAR(1) in Equation (1). Whether the 2-variable VAR(1) in Equation (A.4) is a misspecified representation of the joint dynamics of returns and the price-dividend ratio depends on the dynamics of dividend growth. The next theorem formalizes this link.

Theorem A.1. *The VARMA(2,1) in Equation (A.3) will have* $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$, $\mathbf{G}_2 = \mathbf{0}_{2\times 2}$, *and* $\mathbf{D}_1 = \mathbf{0}_{2\times 2}$ *if and only if* $\phi_{pd,d} = \phi_{r,d} = 0$, *with* $\phi_{d,d} = 0$ *following from CS restriction (5).*

Proof. First, let us assume that $\phi_{r,d} = \phi_{pd,d} = 0$; then, from CS restriction (5) $\phi_{d,d} = 0$. Since $\boldsymbol{\phi}_{12} = \mathbf{0}_{2\times 1}$, $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$ and $\mathbf{G}_2 = \mathbf{0}_{2\times 2}$. Moreover, since $\mathbf{D}_1$ solves the following moment restriction $-\phi_{12}\mathbb{E}(\eta_t\mathbf{x}_t') = \mathbf{D}_1\boldsymbol{\Omega}_{\mathbf{e}}$, $\mathbf{D}_1 = \mathbf{0}_{2\times 2}$ if $\boldsymbol{\phi}_{12} = \mathbf{0}_{2\times 1}$. Second, any of the following conditions $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$, $\mathbf{G}_2 = \mathbf{0}_{2\times 2}$, or $\mathbf{D}_1 = \mathbf{0}_{2\times 2}$ requires that $\boldsymbol{\phi}_{12} = \mathbf{0}_{2\times 1}$ and hence $\phi_{d,d} = 0$ because of CS restriction (5). □

Theorem A.1 implies that if we run the 2-variable VAR(1) in Equation (A.4), $\mathbf{A}_1 = \boldsymbol{\Phi}_{11}$ if and only if $\phi_{d,d} = 0$. This additional restriction does not follow from the CS identity. The punchline is clear: the CS identity does not allow one to drop one of the three variables and run the 2-variable VAR(1) in Equation (A.4). Doing so, while assuming $\mathbf{A}_1 = \boldsymbol{\Phi}_{11}$, amounts to imposing the additional

[23] In Appendix A.3 we show that a VARMA(2,1) representation for $\mathbf{x}_{t+1}$ also arises from assuming that the price-dividend ratio is stationary and price and dividend follow a VECM with one lag specification.

restrictions $\phi_{d,d} = \phi_{pd,d} = \phi_{r,d} = 0$. These claims are formalized in the following corollary.

Corollary A.1. *Assuming* $\mathbf{A}_1 = \mathbf{\Phi}_{11}$ *implicitly imposes that* $\phi_{pd,d} = \phi_{r,d} = 0$*, with* $\phi_{d,d} = 0$ *following from CS restriction* (5)*.*

An implication of Equation (A.5) is that the innovations of the 2-variable VAR(1) are predictable with lagged information. In fact, it follows from Equations (A.4) and (A.5) that:

$$\boldsymbol{\varepsilon}_{t+1} = -\left(\mathbf{G}_2\mathbf{\Gamma}_1' + \mathbf{D}_1\mathbf{\Omega}_e\right)\mathbf{\Gamma}_0^{-1}\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1\mathbf{e}_t. \tag{A.6}$$

Clearly, Equation (A.6) implies that $\mathbf{\Omega}_{\boldsymbol{\varepsilon}} \neq \mathbf{\Omega}_{\mathbf{e}}$. With these ingredients, we can write the following corollary:

Corollary A.2. *The innovations of the 2-variable VAR(1),* $\boldsymbol{\varepsilon}_{t+1}$*, equals* $\boldsymbol{\xi}_{t+1}$ *if and only if* $\phi_{pd,d} = \phi_{r,d} = 0$*, with* $\phi_{d,d} = 0$ *following from CS restriction* (5)*.*

Clearly, Corollary A.2 implies that the innovations of the 2-variable VAR(1) in Equation (A.4) are not predictable if and only if $\phi_{pd,d} = \phi_{r,d} = 0$. Interestingly, even when $\boldsymbol{\varepsilon}_{t+1} = \boldsymbol{\xi}_{t+1}$, one cannot recover u^d_{t+1} because of the approximation error. It is also useful to note an interesting special case of Theorem A.1, by which it is possible to recover the return innovation, u^r_{t+1}:

Corollary A.3. *The innovation to the return equation in the 2-variable VAR(1), i.e., the last element in* $\boldsymbol{\varepsilon}_{t+1}$*, is equal to the return innovation* u^r_{t+1} *if and only if* $\phi_{r,d} = 0$*.*

Corollary A.3 implies that if there is no direct predictability from dividends to returns, both the one-step-ahead return forecast and innovation coincide with those of the 3-variable VAR(1). Yet, even in this case, the multi-step-ahead forecasts and innovations will not be retrieved correctly.

To sum up, we have shown that considering a 2-variable VAR(1) is not a valid way to impose the CS restrictions. Instead, we will need to consider a 3-variable VAR(1) and explicitly impose the restrictions. We will explain how to do this in Section 3.3. The results derived in this section carry over to settings with multiple lags. Specifically, assuming a 3-variable VAR(p) implies that omitting dividend growth leads to a VARMA(p+1,p) representation for $\mathbf{x}_{t+1}$, so that any finite order VAR specification for $\mathbf{x}_{t+1}$ leads to a misspecification of the dynamics of the system.

A.1 VARMA(2,1) Mapping

The results in the appendix are general, so they can be used to obtain the VARMA(2,1) representations in both the previous section and Appendix H. Let us consider the following model of the $n_x \times 1$ vector $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{G}_1 \mathbf{x}_t + \mathbf{G}_2 \mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1} + \mathbf{M}_0 \boldsymbol{\eta}_t. \tag{A.7}$$

where $\boldsymbol{\xi}_{t+1}$ is a Gaussian distributed $n_x \times 1$ vector, $\boldsymbol{\eta}_{t+1}$ is a Gaussian distributed $n_\eta \times 1$ vector, $\mathbb{E}\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}_{t+1}'\right) = \boldsymbol{\Omega}_{\boldsymbol{\xi}}$, $\mathbb{E}\left(\boldsymbol{\eta}_{t+1}\boldsymbol{\eta}_{t+1}'\right) = \boldsymbol{\Omega}_{\boldsymbol{\eta}}$ $\mathbb{E}\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\eta}_{t+1}'\right) = \boldsymbol{\Omega}_{\boldsymbol{\xi}\boldsymbol{\eta}}$, $\mathtt{E}\left(\boldsymbol{\xi}_s\boldsymbol{\xi}_r'\right) = \mathbf{0}_{n_x \times n_x}$, $\mathtt{E}\left(\boldsymbol{\eta}_s\boldsymbol{\eta}_r'\right) = \mathbf{0}_{n_\eta \times n_\eta}$, and $\mathtt{E}\left(\boldsymbol{\xi}_s\boldsymbol{\eta}_r'\right) = \mathbf{0}_{n_x \times n_\eta}$ for all $s \neq r$, where $\mathbb{E}\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\eta}_{t+1}'\right) = \boldsymbol{\Omega}_{\boldsymbol{\xi}\boldsymbol{\eta}}$ reflects the fact that the approximation error $\boldsymbol{\eta}_{t+1}$ may be contemporaneously correlated with $\boldsymbol{\xi}_{t+1}$.[24]

We can then compute the autocovariance of the model in Equation (A.7). Define with $\boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_{-j}' = \mathbb{E}\left(\mathbf{x}_{t+1}\mathbf{x}_{t+1-j}'\right)$, post multiplying Equation (A.7) by $\mathbf{x}_{t+1-j}$ and taking expectations gives:

$$\boldsymbol{\Gamma}_0 = \mathbf{G}_1\boldsymbol{\Gamma}_{-1} + \mathbf{G}_2\boldsymbol{\Gamma}_{-2} + \boldsymbol{\Psi}_0, \tag{A.8}$$

$$\boldsymbol{\Gamma}_1 = \mathbf{G}_1\boldsymbol{\Gamma}_0 + \mathbf{G}_2\boldsymbol{\Gamma}_{-1} + \boldsymbol{\Psi}_1, \text{ and} \tag{A.9}$$

$$\boldsymbol{\Gamma}_j = \mathbf{G}_1\boldsymbol{\Gamma}_{j-1} + \mathbf{G}_2\boldsymbol{\Gamma}_{j-2} \text{ for } j > 1, \tag{A.10}$$

where

$$\begin{aligned}
\boldsymbol{\Psi}_0 &= \mathbb{E}\left(\boldsymbol{\xi}_{t+1}\mathbf{x}_{t+1}'\right) + \mathbf{M}_0\mathbb{E}\left(\boldsymbol{\eta}_t\mathbf{x}_{t+1}'\right), \qquad (A.11)\\
&= \mathbb{E}\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}_{t+1}'\right) + \mathbf{M}_0\mathbb{E}\left(\boldsymbol{\eta}_t\mathbf{x}_t'\right)\mathbf{G}_1' + \mathbf{M}_0\mathbb{E}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right)\mathbf{M}_0',\\
&= \mathbb{E}\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}_{t+1}'\right) + \mathbf{M}_0\mathbb{E}\left(\boldsymbol{\eta}_t\boldsymbol{\xi}_t'\right)\mathbf{G}_1' + \mathbf{M}_0\mathbb{E}\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right)\mathbf{M}_0',\\
&= \boldsymbol{\Omega}_{\boldsymbol{\xi}} + \mathbf{M}_0\boldsymbol{\Omega}_{\boldsymbol{\xi}\boldsymbol{\eta}}\mathbf{G}_1' + \mathbf{M}_0\boldsymbol{\Omega}_{\boldsymbol{\eta}}\mathbf{M}_0',\\
\boldsymbol{\Psi}_1 &= \mathbf{M}_0\mathbb{E}\left(\boldsymbol{\eta}_t\mathbf{x}_t'\right), \qquad (A.12)\\
&= \mathbf{M}_0\boldsymbol{\Omega}_{\boldsymbol{\xi}\boldsymbol{\eta}},
\end{aligned}$$

Given the shape of the autocovariance function in Equations (A.8)-(A.10), it is possible to show that a VARMA(2,1) can replicate the statistical properties of the model in Equation (A.7). Specifically,

[24] Equation (A.7) boils down to Equation (A.3) by defining $\boldsymbol{\eta}_{t+1} = \eta_{t+1}$, $\boldsymbol{\Omega}_{\boldsymbol{\eta}} = \sigma_\eta^2$, $\mathbf{M}_0 = -\boldsymbol{\phi}_{12}$, and $\boldsymbol{\Omega}_{\boldsymbol{\xi}\boldsymbol{\eta}} = \begin{bmatrix}0 & \sigma_\eta^2\end{bmatrix}'$.

the VARMA(2,1) representation of $\mathbf{x}_t$ is:

$$\mathbf{x}_{t+1} = \mathbf{G}_1\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1\mathbf{e}_t, \tag{A.13}$$

where $\mathbb{E}(\mathbf{e}_{t+1}\mathbf{e}'_{t+1}) = \mathbf{\Omega}_\mathbf{e}$, $\mathbb{E}(\mathbf{e}_s\mathbf{e}'_r) = \mathbf{0}_{n_x \times n_x}$ for all $s \neq r$, and $\mathbf{D}_1$ and $\mathbf{\Omega}_\mathbf{e}$ are chosen so that the autocovariance associated with the moving average component of Equation (A.13):

$$\tilde{\mathbf{\Psi}}_0 = \mathbf{\Omega}_\mathbf{e} + \mathbf{D}_1\mathbf{\Omega}_\mathbf{e}\mathbf{G}'_1 + \mathbf{D}_1\mathbf{\Omega}_\mathbf{e}\mathbf{D}'_1 \tag{A.14}$$

$$\tilde{\mathbf{\Psi}}_1 = \mathbf{D}_1\mathbf{\Omega}_\mathbf{e}. \tag{A.15}$$

matches the ones in Equations (A.11)-(A.12). Therefore $\mathbf{D}_1$ and $\mathbf{\Omega}_\mathbf{e}$ are functions of $\mathbf{G}_1$, $\mathbf{M}_0$, $\mathbf{\Omega}_\xi$, $\mathbf{\Omega}_\eta$, and $\mathbf{\Omega}_{\xi\eta}$. Imposing $\tilde{\mathbf{\Psi}}_1 = \mathbf{\Psi}_1$, from Equation (A.15) we have that $\mathbf{\Omega}_\mathbf{e} = \mathbf{D}_1^{-1}\mathbf{\Psi}_1$. Substituting this expression into Equation (A.14) and imposing that $\tilde{\mathbf{\Psi}}_0 = \mathbf{\Psi}_0$ we get that:

$$\mathbf{\Psi}_0 = \mathbf{D}_1^{-1}\mathbf{\Psi}_1 + \mathbf{D}_1\mathbf{D}_1^{-1}\mathbf{\Psi}_1\mathbf{G}'_1 + \mathbf{D}_1\mathbf{D}_1^{-1}\mathbf{\Psi}_1\mathbf{D}'_1 = \mathbf{D}_1^{-1}\mathbf{\Psi}_1 + \mathbf{\Psi}_1\left(\mathbf{G}'_1 + \mathbf{D}'_1\right),$$

which implies that $\mathbf{D}_1$ needs to solve the following quadratic matrix equation:

$$\mathbf{D}_1\mathbf{\Psi}_1\mathbf{D}'_1 + \mathbf{D}_1(\mathbf{\Psi}_1\mathbf{G}'_1 - \mathbf{\Psi}_0) + \mathbf{\Psi}_1 = \mathbf{0}. \tag{A.16}$$

The solution to Equation (A.16) can be easily found numerically. Starting with a guess for $\mathbf{D}_1$ (e.g., $\mathbf{D}_1^0 = \mathbf{0}$), one can find the solution iterating over $\mathbf{D}_1^{k+1} = \mathbf{\Psi}_1\left(\mathbf{\Psi}_0 - \mathbf{\Psi}_1\left(\mathbf{G}'_1 + (\mathbf{D}_1^k)'\right)\right)^{-1}$.

A.2 No Approximation Error

It is important to also notice that neither Theorem A.1 nor Corollary A.2 depends on the existence of approximation error. The VARMA(2,1) representation of $\mathbf{x}_t$ becomes a 2-variable VAR(2) and equals $\mathbf{x}_{t+1} = \mathbf{G}_1\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1}$, while Equations (A.5) and (A.6) become $\mathbf{A}_1 = \mathbf{G}_1 + \mathbf{G}_2\mathbf{\Gamma}'_1\mathbf{\Gamma}_0^{-1}$ and $\boldsymbol{\varepsilon}_{t+1} = -\mathbf{G}_2\mathbf{\Gamma}'_1\mathbf{\Gamma}_0^{-1}\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1}$; hence $\mathbf{\Omega}_{\boldsymbol{\varepsilon}} \neq \mathbf{\Omega}_{\boldsymbol{\xi}}$. Again, $\mathbf{G}_1 = \mathbf{\Phi}_{11} + \phi_{12}[-\rho, 1]$, and $\mathbf{G}_2 = [\phi_{12}, \mathbf{0}_{2\times 1}]$. Therefore, a 2-variable VAR(1) on $\mathbf{x}_{t+1}$ will be misspecified even in the absence of approximation error unless the conditions of Theorem A.1 are satisfied, i.e., $\phi_{d,d} = 0$.

A.3 A VECM Approach

In this appendix, we show that assuming that the price-dividend ratio is stationary and that the dynamics of prices and dividends can be characterized as a VECM as in Campbell and Shiller (1988) implies a VARMA(2,1) model for $\mathbf{x}_{t+1} = [pd_{t+1}, r_{t+1}]'$, which simplifies to a restricted VAR(2) in the absence of an approximation error in the CS identity. Therefore, the conclusions of Section A extend to this alternative setting.

Let us consider the following VECM with one lag:

$$\begin{bmatrix} \Delta d_{t+1} \\ \Delta p_{t+1} \end{bmatrix} = \boldsymbol{\lambda} pd_t + \mathbf{B} \begin{bmatrix} \Delta d_t \\ \Delta p_t \end{bmatrix} + \boldsymbol{\upsilon}_{t+1}, \tag{A.17}$$

where $\mathbf{B}$ is an unrestricted 2×2 matrix and $\boldsymbol{\lambda}$ is an unrestricted 2×1 vector. where $\mathbb{E}(\boldsymbol{\upsilon}_{t+1}) = \mathbf{0}_{2\times 1}$, $\mathbb{E}(\boldsymbol{\upsilon}_{t+1}\boldsymbol{\upsilon}_{t+1}') = \boldsymbol{\Omega}_{\boldsymbol{\upsilon}}$, $\mathbb{E}(\boldsymbol{\upsilon}_s\boldsymbol{\upsilon}_r') = \mathbf{0}_{2\times 2}$ if $r \neq s$, and $\boldsymbol{\Omega}_{\boldsymbol{\upsilon}}$ is an SPD. Since $\Delta(pd_t) = \Delta p_t - \Delta d_t$, the model in Equation (A.17) can always be rewritten as a restricted 2-variable VAR(2) for $[pd_{t+1}, \Delta p_{t+1}]$. Specifically, define:

$$\mathbf{M} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix},$$

then

$$\begin{bmatrix} \Delta(pd_{t+1}) \\ \Delta p_{t+1} \end{bmatrix} = \mathbf{M}\boldsymbol{\lambda} pd_t + \mathbf{M}\mathbf{B}\mathbf{M}^{-1} \begin{bmatrix} \Delta(pd_t) \\ \Delta p_t \end{bmatrix} + \mathbf{M}\boldsymbol{\upsilon}_{t+1}.$$

Therefore

$$\begin{bmatrix} pd_{t+1} \\ \Delta p_{t+1} \end{bmatrix} = \boldsymbol{\Lambda}_1 \begin{bmatrix} pd_t \\ \Delta p_t \end{bmatrix} + \boldsymbol{\Lambda}_2 \begin{bmatrix} pd_{t-1} \\ \Delta p_{t-1} \end{bmatrix} + \mathbf{M}\boldsymbol{\upsilon}_{t+1}, \tag{A.18}$$

where $\boldsymbol{\Lambda}_1 = \mathbf{M}\mathbf{B}\mathbf{M}^{-1} + [\boldsymbol{\iota}_1 + \mathbf{M}\boldsymbol{\lambda}, \mathbf{0}_{2\times 1}]$, $\boldsymbol{\Lambda}_2 = \left[-\mathbf{M}\mathbf{B}\mathbf{M}^{-1}\boldsymbol{\iota}_1, \mathbf{0}_{2\times 1}\right]$, and $\boldsymbol{\iota}_1 = [1, 0]'$. Therefore the coefficient matrix on the second lag, $\boldsymbol{\Lambda}_2$, is restricted so that it is only the column associated with the price-dividend that features non-zero coefficients.

To map the dynamics of this system to the one of $\mathbf{x}_{t+1}$ we can use the the CS identity in

Equation (2); this implies that $r_{t+1} = (\rho - 1)\, pd_{t+1} + \Delta p_{t+1} + \eta_{t+1}$. Therefore defining:

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ \rho - 1 & 1 \end{bmatrix},$$

one has that

$$\mathbf{x}_{t+1} = \widetilde{\mathbf{G}}_1 \mathbf{x}_t + \widetilde{\mathbf{G}}_2 \mathbf{x}_{t-1} + \widetilde{\mathbf{u}}_{t+1} + \mathbf{M}_0 \eta_{t+1} + \mathbf{M}_1 \eta_t + \mathbf{M}_2 \eta_{t-1}, \tag{A.19}$$

where $\widetilde{\mathbf{u}}_{t+1} = \mathbf{W}\mathbf{M}\boldsymbol{\upsilon}_{t+1}$, $\widetilde{\mathbf{G}}_1 = \mathbf{W}\boldsymbol{\Lambda}_1\mathbf{W}^{-1}$, and $\widetilde{\mathbf{G}}_2 = \mathbf{W}\boldsymbol{\Lambda}_2\mathbf{W}^{-1}$. Also $\mathbf{M}_0 = -\boldsymbol{\iota}_2$, $\mathbf{M}_1 = \widetilde{\mathbf{G}}_1\boldsymbol{\iota}_2$ and $\mathbf{M}_2 = \widetilde{\mathbf{G}}_2\boldsymbol{\iota}_2$, with $\boldsymbol{\iota}_2 = [0, 1]'$. Since $\boldsymbol{\Lambda}_2\boldsymbol{\iota}_2 = \mathbf{0}_{2\times 1}$, it is easy to show that $\widetilde{\mathbf{G}}_2\boldsymbol{\iota}_2 = \mathbf{0}_{2\times 1}$. Therefore, $\widetilde{\mathbf{G}}_2$ is a restricted matrix with non-zero elements associated only with the column of the price-dividend ratio. Moreover, $\mathbf{M}_2 = \mathbf{0}_{2\times 1}$. So that Equation (A.19) becomes:

$$\mathbf{x}_{t+1} = \widetilde{\mathbf{G}}_1 \mathbf{x}_t + \widetilde{\mathbf{G}}_2 \mathbf{x}_{t-1} + \widetilde{\mathbf{u}}_{t+1} + \mathbf{M}_0 \eta_{t+1} + \mathbf{M}_1 \eta_t. \tag{A.20}$$

Looking at the autocovariance function implied by (A.20) one can see that this takes the same structure as the one in Equations (A.8)-(A.10). Therefore the system (A.20) implies a VARMA(2,1) representation:

$$\mathbf{x}_{t+1} = \widetilde{\mathbf{G}}_1 \mathbf{x}_t + \widetilde{\mathbf{G}}_2 \mathbf{x}_{t-1} + \widetilde{\mathbf{e}}_{t+1} + \widetilde{\mathbf{D}}_1 \widetilde{\mathbf{e}}_t, \tag{A.21}$$

where $\mathbb{E}\left(\widetilde{\mathbf{e}}_{t+1}\widetilde{\mathbf{e}}'_{t+1}\right) = \boldsymbol{\Omega}_{\widetilde{\mathbf{e}}}$, $\mathbb{E}\left(\widetilde{\mathbf{e}}_s\widetilde{\mathbf{e}}'_r\right) = \mathbf{0}_{2\times 2}$ for all $s \neq r$ and where $\widetilde{\mathbf{D}}_1$ and $\boldsymbol{\Omega}_{\widetilde{\mathbf{e}}}$ can be derived following the same steps detailed in Appendix A.1.

The moving average term in Equation (A.21) arises from the presence of an approximation error in Equation (A.20) (i.e., $\widetilde{\mathbf{D}}_1 = \mathbf{0}$ if $\sigma^2_\eta = 0$). Therefore, assuming there is no approximation error associated with the CS identity, a VECM with one lag representation for price and dividends with the price-dividend ratio being stationary implies a restricted 2-variable VAR(2) representation for $\mathbf{x}_{t+1}$, with the coefficients associated with the second lag of returns being zero as discussed in Section A.2.

A.4 Adding Lags

If one insists on omitting dividend growth from the system, after reading the previous subsection it might be tempting to ignore the approximation error and run an unrestricted 2-variable VAR(2) on

$\mathbf{x}_{t+1}$.[25] While this VAR would be correctly specified in the absence of approximation or measurement error, this approach does not automatically impose the CS restrictions any more than considering a 3-variable VAR(1). To gain understanding, one can again resort to a simple parameter count. As discussed, the autoregressive matrix $\mathbf{\Phi}_1$ of the 3-variable VAR(1) contains nine parameters whereas the CS restrictions imply that only six of them are free parameters. The autoregressive matrices of a 2-variable VAR(2) have eight free parameters. Hence, this model, while correctly specified, is overparametrized. In particular, the restrictions on $\mathbf{G}_2$ derived above are not explicitly imposed. While one will always be able to recover the true parameter asymptotically, in small samples, not imposing these restrictions can lead to inefficient estimates and, as a result, a deteriorated performance of the model, as we will show in the empirical application.

This result will also carry over to a setting in which the data-generating process features more lags. Specifically, assuming that the model in Equation (1) is a 3-variable VAR(p) implies that omitting dividends and assuming no measurement error leads to a 2-variable VAR(p+1) representation for $\mathbf{x}_{t+1}$, where the coefficient matrix associated with the last lag is restricted so that only the price-dividend ratio takes non-zero coefficients.

A.5 Omitting Other Variables

Our results so far indicate that the CS identity does not justify dropping dividend growth from the VAR unless the conditions from Theorem A.1 are satisfied, namely, that the column corresponding to dividend growth is entirely composed of zeros. This result would apply symmetrically to omitting stock returns from the system. This would be a valid strategy if $\phi_{d,r} = \phi_{pd,r} = 0$, with $\phi_{r,r} = 0$ following from CS restriction (7). It is immediately obvious that dropping the price-dividend ratio is never acceptable, as the three coefficients cannot all be zero without violating CS restriction (5). Thus, Engsted et al. (2012)'s criticism of Chen and Zhao (2009) for excluding the price-dividend ratio is justified. In any case Engsted et al. (2012) still claim that one can drop either returns or dividend growth from the system as long as we include "additional state variables that capture part of the predictive variability of [the omitted variable] not captured by the dividend–price ratio" (p.1262). Our analysis shows that this is not enough: the additional state variables need to *perfectly*

[25] VARs with more than a single lag are very seldomly considered in the empirical literature. Moreover, many of the papers surveyed used a smoothed price-to-earnings ratio rather than the price-dividend ratio, which introduces additional measurement error.

span the omitted variable. In such a case, it seems reasonable to include the omitted variable directly. Otherwise, the CS restriction corresponding to the column of the omitted variable would apply to a linear combination of the additional state variables.

A.6 Implications for the Calculation of NCF_{t+1} and NDR_{t+1}

Chen and Zhao (2009) and Engsted et al. (2012) consider the question of whether omitting either returns or dividend growth affects the computation of the news terms. Our basic VAR in Equation (1) implies that $NDR_{t+1} = \mathbf{e}'_{3,3}\mathbf{\Lambda}\mathbf{u}'_{t+1}$ and $NCF_{t+1} = \mathbf{e}'_{1,3}\mathbf{\Lambda}\mathbf{u}'_{t+1} + \mathbf{e}'_{1,3}\mathbf{u}'_{t+1}$, where $\mathbf{\Lambda} = \rho\mathbf{\Phi}_1(\mathbf{I}_3 - \rho\mathbf{\Phi}_1)^{-1}$ and $\mathbf{e}_{j,n}$ is the j^{th} column of the matrix $\mathbf{I}_n$. If dividend growth is omitted, and one runs a 2-variable VAR(1) on $[pd_{t+1}, r_{t+1}]$, NCF_{t+1} cannot be calculated directly, but using the CS identity one can obtain it residually from $NCF_{t+1} = NDR_{t+1} + \tilde{u}^r_{t+1}$ instead. In this case $NDR_{t+1} = \mathbf{e}'_{2,2}\tilde{\mathbf{\Lambda}}\tilde{\boldsymbol{\varepsilon}}'_{t+1}$ and $\tilde{u}^r_{t+1} = \mathbf{e}'_{2,2}\tilde{\boldsymbol{\varepsilon}}'_{t+1}$, where $\tilde{\mathbf{\Lambda}} = \rho\tilde{\mathbf{A}}_1(\mathbf{I}_2 - \rho\tilde{\mathbf{A}}_1)^{-1}$ and $\tilde{\mathbf{A}}_1$ and $\tilde{\boldsymbol{\varepsilon}}_{t+1}$ are the autoregressive coefficient matrix and residuals of the 2-variable VAR(1) on $[pd_{t+1}, r_{t+1}]$ derived in Equations (A.5) and (A.6), respectively. Analogously, if one omits returns and runs a 2-variable VAR(1) on $[\Delta d_{t+1}, pd_{t+1}]$, NDR_{t+1} has to be calculated as a residual using the following formula $NDR_{t+1} = NCF_{t+1} + \hat{u}^r_{t+1}$. In this case $NCF_{t+1} = \mathbf{e}'_{1,2}\hat{\mathbf{\Lambda}}\hat{\boldsymbol{\varepsilon}}'_{t+1} + \mathbf{e}'_{1,2}\hat{\boldsymbol{\varepsilon}}'_{t+1}$ and $\hat{u}^r_{t+1} = \hat{u}^d_{t+1} + \rho\hat{u}^{pd}_{t+1} = \mathbf{e}'_{1,2}\hat{\boldsymbol{\varepsilon}}'_{t+1} + \rho\mathbf{e}'_{2,2}\hat{\boldsymbol{\varepsilon}}'_{t+1}$, where $\hat{\mathbf{\Lambda}} = \rho\hat{\mathbf{A}}_1(\mathbf{I}_2 - \rho\hat{\mathbf{A}}_1)^{-1}$ and $\hat{\mathbf{A}}_1$ and $\hat{\boldsymbol{\varepsilon}}_{t+1}$ are the autoregressive coefficient matrix and residuals of the 2-variable VAR(1) on $[\Delta d_{t+1}, pd_{t+1}]$, respectively.[26]

What our results from the previous section highlight is that, unless the conditions of Theorem A.1 are satisfied, $\hat{\mathbf{\Lambda}} \neq \tilde{\mathbf{\Lambda}} \neq \mathbf{\Lambda}$ and $\hat{\boldsymbol{\varepsilon}}_{t+1} \neq \tilde{\boldsymbol{\varepsilon}}_{t+1} \neq \mathbf{u}_{t+1}$. Therefore, omitting either returns or dividend growth from the VAR and backing up either NDR_{t+1} or NCF_{t+1} indirectly will affect the results. This is true for the cases with and without approximation error and even in the special cases where u^r_{t+1} is retrieved correctly, (see Corollary A.3, for example), as $\hat{\mathbf{\Lambda}} \neq \tilde{\mathbf{\Lambda}} \neq \mathbf{\Lambda}$. In Appendix A.7 we show how omitting dividend growth leads to underestimation of the variance of NCF_{t+1}, overestimation of the variance of NDR_{t+1}, and underestimation of the correlation between the two. Omitting returns instead in this case leads to correct calculations, as the column of returns in the simplified model is assumed to be zeros. If the matrix $\mathbf{\Phi}_1$ is such that neither column is zero, which we shall see in the next sections is the case for US postwar data, omitting either returns or dividend growth will lead to different, and incorrect, results.

[26] Omitting returns assumes that there is no approximation error.

A.7 Estimation of NDR and NCF with Omitted Variables

Chen and Zhao (2009) and Engsted et al. (2012) consider the question of whether the computation of NDR_{t+1} and NCF_{t+1} in a VAR is affected by whether either of them is treated as a residual in the decomposition in Equation (25), repeated here for convenience:

$$r_{t+1} - \mathbb{E}_t r_{t+1} = \underbrace{(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=0}^{\infty} \rho^j \ \Delta d_{t+j+1}}_{NCF_{t+1}} - \underbrace{(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+j+1}}_{NDR_{t+1}} \tag{A.22}$$

Recall that $r_{t+1} - \mathbb{E}_t r_{t+1} \equiv u^r_{t+1}$. Our basic VAR in Equation (1) implies that:

$$NDR_{t+1} = \mathbf{e}'_{3,3} \mathbf{\Lambda} \mathbf{u}'_{t+1} \tag{A.23}$$

$$NCF_{t+1} = \mathbf{e}'_{2,3} \mathbf{\Lambda} \mathbf{u}'_{t+1} + \mathbf{e}'_{2,3} \mathbf{u}'_{t+1} \tag{A.24}$$

where $\mathbf{\Lambda} = \rho \mathbf{\Phi}_1 (\mathbf{I}_3 - \rho \mathbf{\Phi}_1)$ and $\mathbf{e}_{j,n}$ is the j-th column of the matrix $\mathbf{I}_n$. Equation (A.22) also implies that:

$$NCF_{t+1} = NDR_{t+1} + u^r_{t+1} = (\mathbf{e}'_3 + \mathbf{e}'_3 \mathbf{\Lambda}) \mathbf{u}'_{t+1} \tag{A.25}$$

Does it matter whether we compute NDR_{t+1} and NCF_{t+1} directly from (A.23) and (A.24), or whether we compute one of them directly and the other as a residual from (A.25)? Engsted et al. (2012) claim that for the estimation of NDR_{t+1} and NCF_{t+1}, "nothing is gained by modeling both returns and dividend growth in a system that also contains the dividend–price ratio."

If dividend growth is omitted, and one runs a 2-variable VAR(1) on $[pd_{t+1}, r_{t+1}]$, NCF_{t+1} cannot be calculated directly, but using the CS identity one can obtain it residually from $NCF_{t+1} = NDR_{t+1} + \tilde{u}^r_{t+1}$ instead. In this case $NDR_{t+1} = \mathbf{e}'_{2,2} \tilde{\mathbf{\Lambda}} \tilde{\boldsymbol{\varepsilon}}'_{t+1}$ and $\tilde{u}^r_{t+1} = \mathbf{e}'_{2,2} \tilde{\boldsymbol{\varepsilon}}'_{t+1}$, where $\tilde{\mathbf{\Lambda}} = \rho \tilde{\mathbf{A}}_1 (\mathbf{I}_2 - \rho \tilde{\mathbf{A}}_1)^{-1}$ and $\tilde{\mathbf{A}}_1$ and $\tilde{\boldsymbol{\varepsilon}}_{t+1}$ are the autoregressive coefficient matrix and residuals of the 2-variable VAR(1)on $[pd_{t+1}, r_{t+1}]$ derived in Equations (A.5) and (A.6), respectively. Analogously, if one omits returns and runs a 2-variable VAR(1) on $[\Delta d_{t+1}, pd_{t+1}]$, NDR_{t+1} has to be calculated as a residual using the following formula: $NDR_{t+1} = NCF_{t+1} + \hat{u}^r_{t+1}$. In this case $NCF_{t+1} = \mathbf{e}'_{1,2} \hat{\mathbf{\Lambda}} \hat{\boldsymbol{\varepsilon}}'_{t+1} + \mathbf{e}'_{1,2} \hat{\boldsymbol{\varepsilon}}'_{t+1}$ and $\hat{u}^r_{t+1} = \hat{u}^d_{t+1} + \rho \hat{u}^{pd}_{t+1} = \mathbf{e}'_{1,2} \hat{\boldsymbol{\varepsilon}}'_{t+1} + \rho \mathbf{e}'_{2,2} \hat{\boldsymbol{\varepsilon}}'_{t+1}$, where $\hat{\mathbf{\Lambda}} = \rho \hat{\mathbf{A}}_1 (\mathbf{I}_2 - \rho \hat{\mathbf{A}}_1)^{-1}$ and $\hat{\mathbf{A}}_1$ and $\hat{\boldsymbol{\varepsilon}}_{t+1}$ are the autoregressive coefficient matrix and residuals of the 2-variable VAR(1)on $[\Delta d_{t+1}, pd_{t+1}]$, respectively.

Omitting returns assumes that there is no approximation error.

More importantly, the results will generally differ if either returns or dividend growth is omitted from the system and the conditions of Theorem A.1 are not satisfied. Using a simple framework we now show how the claims in Engsted et al. (2012) do not hold. The simplified VAR is:

$$\begin{bmatrix} \Delta d_{t+1} \\ pd_{t+1} \\ r_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_{d,d} \\ \phi_{pd,d} \\ 0 \end{bmatrix} \Delta d_t + \begin{bmatrix} 0 \\ \phi_{pd,pd} \\ \phi_{r,pd} \end{bmatrix} pd_t + \begin{bmatrix} \phi_{d,r} \\ \phi_{pd,r} \\ 0 \end{bmatrix} r_t + \begin{bmatrix} u^d_{t+1} \\ u^{pd}_{t+1} \\ u^r_{t+1} \end{bmatrix}$$

Notice that the column of $\mathbf{\Phi}_1$ corresponding to dividend growth is not composed entirely of zeros but the one for returns is; so according to Theorem A.1, a 2-variable VAR(1) on $[pd_{t+1}, r_{t+1}]$ would be misspecified. Let us consider first a case where the $\phi_{d,r} = \phi_{pd,r} = 0$. As explained a 2-variable VAR(1) on $[pd_{t+1}, \Delta d_{t+1}]$ would be correctly specified because the column of $\mathbf{\Phi}_1$ corresponding to returns is composed entirely of zeros.

Assume the following numerical values $\phi_{pd,pd} = 0.92$, $\texttt{Var}\left(u^d_{t+1}\right) = 0.003$, $\texttt{Var}\left(u^{pd}_{t+1}\right) = 0.028$, $\texttt{Cov}\left(u^{pd}_{t+1}, u^{d}_{t+1}\right) = 0$ and $\rho = 0.971$. We will consider values for $\phi_{d,d} \in \{0, 0.4, 0.8\}$. The rest of the parameters are backed out from the CS restrictions (4)-(10), since we assume that there is no approximation error. Panel A of Table A.1 reports the variances of NDR_{t+1} and NCF_{t+1}, as well as twice the covariance between the two news components, all divided by the variance of the return innovation. The sum of the first two rows minus the third adds up to one due to the CS restrictions. We report these numbers for the 3-variable VAR(1)and the two 2-variable VAR(1) for the three values of $\phi_{d,d}$ we consider. We observe the following results:

1. As $\phi_{d,d}$ becomes larger, the variance of both news components in the true model increases, but the variance of NCF_{t+1} increases by more. Moreover, the correlation between the two components increases.

2. The 2-variable VAR(1)model omitting Δd_{t+1} is misspecified according to Theorem A.1 whenever $\phi_{d,d} > 0$. We can see how, as $\phi_{d,d}$ increases, this VAR tends to overestimate the variance of NDR_{t+1}, underestimate the variance of NCF_{t+1}, and underestimate the covariance between the two.

3. The 2-variable VAR(1)model omitting r_{t+1} is correctly specified independently of the value of

$\phi_{d,d}$. As we can see, this VAR correctly estimates all the moments of interest.

Let us now consider a case where the column on the effects of r_t is not zero. Panel B of Table A.1 shows results when we choose the same values as before but $\phi_{d,d} = 0.4$ and $\phi_{d,r} = 0.14$. The rest of the parameters are backed out from the CS restrictions (4)-(10), since we assume that there is no approximation error. In this case both 2-variable VAR(1) on $[pd_{t+1}, r_{t+1}]$ and $[pd_{t+1}, \Delta d_{t+1}]$ would be misspecified. We observe the following results:

1. Both 2-variable VARs provide different and incorrect results.

2. The 2-variable VAR(1)omitting dividend growth tends to slightly overestimate the variance of NCF_{t+1} and underestimate the variance of NDR_{t+1}, whereas the 2-variable VAR(1)omitting returns underestimates both variances.

3. Both 2-variable VARs underestimate the covariance between NDR_{t+1} and NCF_{t+1}.

4. This is even though in the true model, the price-dividend ratio is the only relevant predictor of returns (hence, $\phi_{r,d} = \phi_{r,r} = 0$), and therefore, the return innovation, u^r_{t+1}, and its variance are estimated correctly.

As Panel C of Table A.1 shows results based on a more realistic calibration, related to the $\mathbf{\Phi}_1$ estimated in Section 4. The results are almost identical to the ones in Panel B. For this last case we used $\phi_{d,d} = 0.29$ and $\phi_{d,r} = 0.1$.

Table A.1: RESULTS

Panel A: Based on $\mathbf{\Phi}_1$ from Simplified Model with Column of r_t equal to Zero.

	True model			2-variable VAR(1)omitting Δd_{t+1}			2-variable VAR(1)omitting r_{t+1}		
Value of $\phi_{d,d}$	0	0.4	0.8	0	0.4	0.8	0	0.4	0.8
$\frac{\mathtt{Var}(NDR)}{\mathtt{Var}(u^r)}$	0.888	0.933	2.240	0.888	0.985	2.826	0.888	0.932	2.240
$\frac{\mathtt{Var}(NCF)}{\mathtt{Var}(u^r)}$	0.112	0.299	2.242	0.112	0.147	1.004	0.112	0.299	2.242
$\frac{2\mathtt{Cov}(NDR,NCF)}{\mathtt{Var}(u^r)}$	0.000	0.233	3.483	0.000	0.132	2.848	0.000	0.232	3.502

Panel A: Based on $\mathbf{\Phi}_1$ from Simplified Model with Column of r_t not equal to Zero.

	True model	2-variable VAR(1)omitting Δd_{t+1}	2-variable VAR(1)omitting r_{t+1}
$\frac{\mathtt{Var}(NDR)}{\mathtt{Var}(u^r)}$	1.024	1.031	0.816
$\frac{\mathtt{Var}(NCF)}{\mathtt{Var}(u^r)}$	0.495	0.213	0.430
$\frac{2\mathtt{Cov}(NDR,NCF)}{\mathtt{Var}(u^r)}$	0.519	0.244	0.245

Panel C: Based on $\mathbf{\Phi}_1$ estimated in Section 4

	True model	2-variable VAR(1)omitting Δd_{t+1}	2-variable VAR(1)omitting r_{t+1}
$\frac{\mathtt{Var}(NDR)}{\mathtt{Var}(u^r)}$	0.930	0.932	0.953
$\frac{\mathtt{Var}(NCF)}{\mathtt{Var}(u^r)}$	0.290	0.162	0.297
$\frac{2\mathtt{Cov}(NDR,NCF)}{\mathtt{Var}(u^r)}$	0.220	0.095	0.250

Appendix B Latent Present Value System

In this section, we show results that relate our VAR(1) to the Latent Present Value System popularized by Van Binsbergen and Koijen (2010).

B.1 Mapping of VAR(1) to a Latent Present Value System

Let us define the variables as deviations from their unconditional mean as $\widetilde{\mathbf{y}}_t = \mathbf{y}_t - \boldsymbol{\mu}$, where $\boldsymbol{\mu} \equiv (\mathbf{I}_n - \mathbf{\Phi}_1)^{-1}\mathbf{\Phi}_0$. Working with a demeaned system simplifies the derivations without any loss of generality. The VAR(1) in Equation (1) can be rewritten as:

$$\widetilde{\mathbf{y}}_{t+1} = \mathbf{\Phi}_1\widetilde{\mathbf{y}}_t + \mathbf{u}_{t+1}. \tag{B.1}$$

When the CS identity holds exactly, $\mathbf{\Phi}_1$ and $\mathbf{\Sigma}$ are rank deficient. Therefore $\mathbf{u}_{t+1} = \widetilde{\mathbf{H}}\widetilde{\mathbf{u}}_{t+1}$, where $\widetilde{\mathbf{u}}_{t+1} = [u^d_{t+1}, u^{pd}_{t+1}]$ and $\widetilde{\mathbf{H}}$ is a 3×2 matrix, imposing restrictions (8)-(10),

$$\widetilde{\mathbf{H}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & \rho \end{bmatrix}$$

and $\widetilde{\mathbf{u}}_{t+1} \sim N(\mathbf{0}, \mathbf{W})$ with $\mathbf{W}$ a 2×2 full rank covariance matrix. Let the eigenvalue decomposition of $\mathbf{\Phi}_1 = \mathbf{EVE}^{-1}$, where $\mathbf{V}$ and $\mathbf{E}$ denote the diagonal matrix with the eigenvalues (ordered in decreasing values) and the associated matrix of eigenvectors. Rank deficiency implies that the last element of $\mathbf{V}$ is a zero.[27]

Therefore, introducing the selection matrix $\mathbf{J}_1 = [\mathbf{I}_2 \ \mathbf{0}_{2\times 1}]$, the system in Equation (B.1) implies that the dynamics of the expected values of the vector of observables, $E_t(\widetilde{\mathbf{y}}_{t+1})$, is only function of a two-dimensional state vector, $\widetilde{\mathbf{x}}_t = \mathbf{J}_1\mathbf{E}^{-1}\mathbf{y}_t$:

$$E_t(\widetilde{\mathbf{y}}_{t+1}) = \mathbf{\Phi}_1\widetilde{\mathbf{y}}_t = \mathbf{EVE}^{-1}\widetilde{\mathbf{y}}_t = \mathbf{EVJ}_1'\mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{y}}_t = \mathbf{EVJ}_1'\widetilde{\mathbf{x}}_t \tag{B.2}$$

since $\mathbf{V} = \mathbf{VJ}_1'\mathbf{J}_1$. Moreover, each element of the follows an AR(1) process:

$$\widetilde{\mathbf{x}}_{t+1} = \mathbf{J}_1\mathbf{VJ}_1'\widetilde{\mathbf{x}}_t + \widetilde{\upsilon}_t \tag{B.3}$$

with correlated innovations, $\widetilde{\upsilon}_t = \mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{H}}\widetilde{\mathbf{u}}_{t+1}$. To see that note that starting from Equation (B.1) and using the eigenvalue decomposition of the matrix of coefficients $\mathbf{\Phi}_1 = \mathbf{EVE}^{-1}$, and noting that $\mathbf{V} = \mathbf{VJ}_1'\mathbf{J}_1$, we get that

$$\begin{aligned} \mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{y}}_{t+1} &= \mathbf{J}_1\mathbf{E}^{-1}\mathbf{\Phi}_1\widetilde{\mathbf{y}}_t + \mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{H}}\widetilde{\mathbf{u}}_{t+1} = \mathbf{J}_1\mathbf{VJ}_1'\mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{y}}_t + \mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{H}}\widetilde{\mathbf{u}}_{t+1} \\ \widetilde{\mathbf{x}}_{t+1} &= \mathbf{J}_1\mathbf{VJ}_1'\widetilde{\mathbf{x}}_t + \widetilde{\upsilon}_t \end{aligned} \tag{B.4}$$

The state vector can be rotated so that one can interpret the rotated latent state vector as

[27] The CS identity implies that only one of the eigenvalues is zero. Imposing additional restrictions, e.g. that the price dividend ratio is the only predictor in the system (as in, e.g., Cochrane, 2008b) implies that two eigenvalues are zero. In that case, all derivations below are still valid redefining the selection matrix $\mathbf{J}_1 = [1\ 0\ 0]$.

(demeaned) expected dividend growth and expected returns. Specifically, defining the selection matrix

$$\mathbf{J}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

defining the rotation matrix $\mathbf{K} = \mathbf{J}_2\mathbf{E}\mathbf{V}\mathbf{J}_1'$, we get $\mathbf{f}_t = \mathbf{K}\widetilde{\mathbf{x}}_t$, and Equation (B.2) implies that

$$\mathbf{J}_2 E_t\left(\widetilde{\mathbf{y}}_{t+1}\right) = \mathbf{J}_2\mathbf{E}\mathbf{V}\mathbf{J}_1'\widetilde{\mathbf{x}}_t = \mathbf{J}_2\mathbf{E}\mathbf{V}\mathbf{J}_1'\mathbf{K}^{-1}\mathbf{K}\widetilde{\mathbf{x}}_t = \mathbf{f}_t$$

Therefore,

$$\Delta d_{t+1} - \mu_g = f_{g,t} + e_{d,t+1}, \tag{B.5}$$

$$r_{t+1} - \mu_r = f_{r,t} + e_{r,t+1}, \tag{B.6}$$

where $\mathbf{f}_t = [f_{g,t}, f_{r,t}]'$ and

$$\left[e_{d,t+1}, e_{r,t+1}\right]' = \mathbf{J}_2\widetilde{\mathbf{H}}\widetilde{\mathbf{u}}_{t+1} = \begin{bmatrix} u^d_{t+1} \\ u^d_{t+1} + \rho u^{pd}_{t+1} \end{bmatrix} \tag{B.7}$$

As for the latent state dynamics, Equation (B.4) implies that

$$\mathbf{f}_{t+1} = \mathbf{F}_1\mathbf{f}_t + \upsilon_t \tag{B.8}$$

where $\mathbf{F}_1 = \mathbf{K}\mathbf{J}_1\mathbf{V}\mathbf{J}_1'\mathbf{K}^{-1}$ and

$$\upsilon_t = \left[\upsilon_{g,t+1}, \upsilon_{r,t+1}\right]' = \mathbf{K}\mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{H}}\widetilde{\mathbf{u}}_{t+1}. \tag{B.9}$$

To see this, note that pre-multiplying Equation (B.4) by $\mathbf{K}$ we have that $\mathbf{K}\widetilde{\mathbf{x}}_{t+1} = \mathbf{K}\mathbf{J}_1\mathbf{V}\mathbf{J}_1'\mathbf{K}^{-1}\mathbf{K}\widetilde{\mathbf{x}}_t + \mathbf{K}\widetilde{\upsilon}_t$.

Last, the CS restrictions guarantee that the (demeaned) price dividend ratio reflects the difference between future cash flows and discount rates (in deviations from their mean). Given the dynamics

of the latent state variables in Equation (B.8), this implies that:

$$pd_{t+1} - \mu_{pd} = [-1, 1] \left(\mathbf{I}_2 - \rho \mathbf{F}_1\right)^{-1} \mathbf{f}_{t+1}. \tag{B.10}$$

These results echo and qualify some of the comments made by Cochrane (2008a) about latent present value systems. Specifically, he notes (p.10) that *"there is no reason that expected returns and expected dividend growth should evolve [...] in AR(1) fashion. [...] And, while tempting, there is no economic reason to impose a particular correlation structure on the shocks"*. He goes on to highlight (p.32) that *"we have no economic reason to impose the particular structure of [the expected returns and dividend growth] rather than the time-series structure of expected returns and dividend growth that results from an arbitrary VAR"*.

The VAR(1) in Equation (1) always admits a state space representation where the dynamics of the observable vector, $\mathbf{y}_{t+1}$, reflect time variation in expected returns and dividend growth, $\mathbf{f}_{t+1}$, whose dynamics follow the VAR(1) process in Equation (B.8). The feedback matrix of the states, $\mathbf{F}_1$, is generally not diagonal unless $\mathbf{K}$ is diagonal. So imposing *ad-hoc* restrictions on $\mathbf{F}_1$, e.g. assuming that this matrix is diagonal as is commonly done in the literature (see, e.g., Van Binsbergen and Koijen, 2010; Rytchkov, 2012), imposes additional restrictions, beyond the ones deriving from the CS identity. Moreover, the innovations of the system are all linear functions of the innovations in dividend growth and the price dividend (see Equation (B.7) and Equation (B.9)). The shocks to dividends and expected dividends are generally correlated. So restricting that shocks to dividends and expected dividends are uncorrelated (see, e.g., Van Binsbergen and Koijen, 2010) again implies imposing additional restrictions, beyond those deriving from the CS identity.

B.2 Covariance Identification

The last section has derived the latent state space model implied by the VAR(1) used in this paper. Van Binsbergen and Koijen (2010) directly describe such a system without deriving from a VAR(1) plus the CS identity. In their description, they highlight that one needs to model the full covariance

of the following set of shocks:

$$\mathbf{e}_{t+1} = \begin{bmatrix} e_{d,t+1} \\ \upsilon_{g,t+1} \\ \upsilon_{r,t+1} \end{bmatrix} = \begin{bmatrix} \Delta d_{t+1} - E_t\left(\Delta d_{t+1}\right) \\ f_{g,t+1} - E_t\left(f_{g,t+1}\right) \\ f_{r,t+1} - E_t\left(f_{r,t+1}\right) \end{bmatrix} \tag{B.11}$$

because not all elements of the covariance can be identified. Therefore they impose the restriction $E\left(e_{d,t+1}\upsilon_{g,t+1}\right) = 0$. As mentioned above, if one derives the latent state space system from the VAR(1) plus the CS identity all elements of the covariance can be identified and one does not need to impose any additional restriction.

This restriction rules out, by definition dividend momentum as we have defined in the paper. This is one of many alternative possible identifications. To see that, note that, in general, one can specify the innovation to dividends as follows

$$e_{d,t+1} = \lambda_g \upsilon_{g,t+1} + \lambda_r \upsilon_{r,t+1} + \widetilde{e}_{d,t+1} \tag{B.12}$$

where $\widetilde{e}_{d,t+1} \sim N\left(0, \sigma_d^2\right)$ denotes the surprise in dividends that is orthogonal to the update of the expected dividend and expected return component. Only 2 of the 3 parameters $\left(\lambda_g, \lambda_r, \sigma_d^2\right)$ can be identified. Van Binsbergen and Koijen (2010) set $\lambda_g = 0$ and focus on $\left(\lambda_r, \sigma_d^2\right)$.

The VAR model implicitly sets the identification restriction $\sigma_d^2 = 0$, so that effectively there are only two shocks in the model. This implies that $e_{d,t+1}$ can always be recovered as a linear combination of the other shocks, and all of the shocks are nothing more than a linear combination to the independent innovations of the VAR, $\widetilde{\mathbf{u}}_{t+1} = [u_{t+1}^d, u_{t+1}^{pd}]'$. Therefore the VAR restricts the surprise to dividend to be a linear function of the remaining shocks of the latent state space model, with exact restrictions pinned down by the VAR estimates. Specifically, from Equation (B.9) we have that $[\lambda_g, \lambda_r] = [1, 0](\mathbf{K}\mathbf{J}_1\mathbf{E}^{-1}\widetilde{\mathbf{H}})^{-1}$.

Van Binsbergen and Koijen (2010) emphasizes that the fit of the model is invariant to the identification restriction. Therefore, alternative identification assumptions produce the same prediction errors. Yet other transformations and statistics, such as filtering and smoothing of the state vector as well as the forecasts produced, are not invariant to these alternative identifications (see, e.g., Harvey, 1990). As such, those are not innocuous.

B.3 VAR(1), VAR(∞) and VARMA

Van Binsbergen and Koijen (2010) show that their model that returns and price dividend depends on an infinite lag of price dividend and dividend growth. Similarly Rytchkov (2012) shows that the model implies a VAR(∞) representation for returns and dividend growth or equivalently a VARMA(1,1). So in general it has been shown that any pair of variables admits a dynamic representation that requires an infinite amount of lags. This is not incompatible with our starting point a VAR(1) for the 3 variables implies that any two variables can be written as a VARMA system (see, e.g., Zellner and Palm, 1974). Therefore a (rank deficient) VAR(1) for the three variables maps into a dynamic representation with only two variables that require an infinite amount of lags.

In fact, should one specify two univariate processes (with correlated innovations) for expected dividend growth and expected returns? The VAR(1) process in Equation (B.8), with a full feedback matrix of coefficients, $\mathbf{F}_1$, can always be rewritten as two separate ARMA processes with correlated innovations (see, e.g., Zellner and Palm, 1974).

B.4 Example in Section 4.5

We consider the simplified setting in Section 4.5 and derive the latent present value system representation of that model. Recall that in this specific case:

$$\begin{bmatrix} \widetilde{\Delta d}_{t+1} \\ \widetilde{pd}_{t+1} \\ \widetilde{r}_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_{d,d} & 0 & 0 \\ \phi_{pd,d} & \phi_{pd,pd} & 0 \\ 0 & \phi_{r,pd} & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta d}_{t} \\ \widetilde{pd}_{t} \\ \widetilde{r}_{t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & \rho \end{bmatrix} \begin{bmatrix} u^d_{t+1} \\ u^{pd}_{t+1} \end{bmatrix}. \tag{B.13}$$

Therefore expected dividends $\mathbb{E}_t(\Delta d_{t+1}) = f_{g,t} = \phi_{d,d}\Delta d_t$ follow an AR(1) process:

$$f_{g,t} = \phi_{d,d} f_{g,t-1} + \phi_{d,d} u^d_t, \tag{B.14}$$

whereas the last equation of the VAR implies that $\mathbb{E}_t(r_{t+1}) = f_{r,t} = \phi_{r,pd} pd_t$, therefore, it is easy to show that expected returns is predicted by its lag as well as a lag in expected dividend:

$$f_{r,t} = \left(\frac{1}{\rho} - \phi_{pd,pd}\right) f_{g,t-1} + \phi_{pd,pd} f_{r,t-1} + \left(\rho\phi_{pd,pd} - 1\right) u^{pd}_t \tag{B.15}$$

since $\phi_{pd,d} = -\phi_{d,d}/\rho$ and $\phi_{r,pd} = \rho\phi_{pd,pd} - 1$. Therefore the VAR(1) in Equation (B.13) admits a VAR(1) representation for expected dividends and expected returns:

$$\begin{bmatrix} f_{g,t} \\ f_{r,t} \end{bmatrix} = \begin{bmatrix} \phi_{d,d} & 0 \\ \left(\frac{1}{\rho} - \phi_{pd,pd}\right) & \phi_{pd,pd} \end{bmatrix} \begin{bmatrix} f_{g,t-1} \\ f_{r,t-1} \end{bmatrix} + \begin{bmatrix} u_t^g \\ u_t^\mu \end{bmatrix} \tag{B.16}$$

where $u_t^g = \phi_{d,d} u_t^d$ and $u_t^\mu = \left(\rho\phi_{pd,pd} - 1\right) u_t^{pd}$. Therefore, for this specific case, the matrix of the feedback coefficients, $\mathbf{F}_1$, is lower triangular as a result of the restrictions to the VAR imposed in Equation (B.13). Moreover, the VAR(1) in Equation (B.13) imposes the restriction that $\texttt{Corr}\left(u_t^g, u_t^d\right) = 1$.

Last, it is worth noticing that Equation (B.16) implies that expected returns, $f_{r,t}$, follows an ARMA(2,1) specification:

$$\left(1 - \phi_{pd,pd}L\right)\left(1 - \phi_{d,d}L\right) f_{r,t} = \left(\frac{1}{\rho} - \phi_{pd,pd}\right)\phi_{d,d} u_{t-1}^d + \left(1 - \phi_{d,d}L\right)\left(\rho\phi_{pd,pd} - 1\right) u_t^{pd} \tag{B.17}$$

as opposed to a more restrictive AR(1) specification (as in, e.g., Van Binsbergen and Koijen, 2010).

Appendix C No Approximation Error

There is no approximation error when $\mathbf{\Omega} = \mathbf{0}_{k\times k}$ in Equation (21). This restriction implies that $\mathbf{L}\mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right)\mathbf{L}' = \mathbf{L}\mathbf{\Sigma}\mathbf{L}' = \mathbf{\Omega} = \mathbf{0}_{k\times k}$, which, because the covariance matrix is symmetric, leads to $\frac{(k+1)k}{2}$ additional restrictions:

$$\widetilde{\mathbf{R}}_\Sigma \operatorname{vec}(\mathbf{\Sigma}) = \mathbf{0}_{\frac{(k+1)k}{2}} \tag{C.1}$$

where $\widetilde{\mathbf{R}}_\Sigma = \mathbf{D}_k^+ \left(\mathbf{L} \otimes \mathbf{L}\right)$ and $\mathbf{D}_k^+$ is a $\frac{(k+1)k}{2} \times k^2$ selection matrix, defined as the Moore-Penrose inverse of the duplication matrix, $\mathbf{D}_k$, so that for any k-dimensional symmetric matrix $\mathbf{A}$, $\mathbf{D}_k^+ \operatorname{vec}(\mathbf{A}) = \operatorname{vech}(\mathbf{A})$ (see Abadir and Magnus, 2005, Ch. 11). Using the mapping described in Equation (24), one can show that the CS restrictions on $\mathbf{\Sigma}$ hold if and only if $\mathbf{W}$ has the following form:

$$\mathbf{W} = \begin{bmatrix} \mathbf{\Xi}\mathbf{\Sigma}\mathbf{\Xi}' & \mathbf{0}_{(n-k)\times k} \\ \mathbf{0}_{k\times(n-k)} & \mathbf{0}_{k\times k} \end{bmatrix}.$$

when the CS restrictions on $\boldsymbol{\Sigma}$ hold. To see this, notice that on the one hand, the mapping implies that $\mathbf{W}_{12} = \boldsymbol{\Xi}\boldsymbol{\Sigma}\mathbf{L}'$ and $\mathbf{W}_{22} = \mathbf{L}\boldsymbol{\Sigma}\mathbf{L}'$. Hence, if Equations (23) and (C.1) hold, it is the case that $\mathbf{W}_{12} = \mathbf{0}_{(n-k)\times k}$ and $\mathbf{W}_{22} = \mathbf{0}_{k\times k}$. On the other hand, the inverse mapping implies that $\left[\mathbf{R}_\Sigma', \widetilde{\mathbf{R}}_\Sigma'\right]' \text{vec}(\boldsymbol{\Sigma}) = \left[\mathbf{R}_\Sigma', \widetilde{\mathbf{R}}_\Sigma'\right]' \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \text{vec}(\mathbf{W})$. It is easy to show that:

$$\widetilde{\mathbf{R}}_\Sigma \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) = \mathbf{D}_k^+ \left(\mathbf{L} \otimes \mathbf{L}\right) \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) = \mathbf{D}_k^+ \left(\mathbf{L}\mathbf{H}^{-1} \otimes \mathbf{L}\mathbf{H}^{-1}\right) = [\mathbf{0}_{\frac{(k+1)k}{2}\times(n^2-k^2)}, \mathbf{D}_k^+],$$

therefore $\left[\mathbf{R}_\Sigma', \widetilde{\mathbf{R}}_\Sigma'\right]' \left(\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}\right) \text{vec}(\mathbf{W}) = [\text{vec}(\mathbf{W}_{12})', \text{vech}(\mathbf{W}_{22})']'$. Thus, if $\mathbf{W}_{12} = \mathbf{0}_{(n-k)\times k}$ and $\mathbf{W}_{22} = \mathbf{0}_{k\times k}$, it is the case that Equations (23) and (C.1) hold.

This result highlights that one can easily modify Algorithm 1 for the case of no approximation error. In this case one simply skips Step 2 and fixes $\boldsymbol{\Omega} = \mathbf{0}_{k\times k}$. In this case, we will call $\pi_{(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}})}(\boldsymbol{\Sigma}) = \mathcal{IW}_{(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}})}(\mathbf{W}_{11})$, where $\mathbf{W}_{11} = \boldsymbol{\Xi}\boldsymbol{\Sigma}\boldsymbol{\Xi}'$ is the density implied by the algorithm and $\pi(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}})$ its distribution. As before, a natural choice for $\mathbf{S}_{\mathbf{W}_{11}}$ is $\mathbf{S}_{\mathbf{W}_{11}} = \boldsymbol{\Xi}\mathbf{S}\boldsymbol{\Xi}'$, while one could choose $\nu_{\mathbf{W}_{11}} = \nu$.

Appendix D Results for Alternative Approaches

In this appendix, we report the out-of-sample forecasting performance of both: the strategy discussed in Section 2.1 that omits Δd_{t+1} from the model but considers a 3-variable VAR(2) having the additional lags to proxy for the dynamics of (lagged) dividend growth and the simpler alternative procedures discussed in Section 3.5. In both cases, we compare the performance to our baseline results.

Table D.1 reports the out-of-sample forecasting performance of the alternative 3-variable VAR(2) and compare it with the rest of the models considered in Section 5.

In the case of the simpler strategy discussed in Section 3.5, starting with a prior on the entire system and imposing the linear restrictions is not equivalent to imposing a marginalized prior on the lower two rows and retrieving the first row of the parameters from the restrictions, and it has implications for out-of-sample forecasting performance. In Figure D.1, we compare the implied prior on the VAR coefficients for the two procedures starting from the (same) standard Minnesota prior. The red lines represent the priors implied by our procedure, while the blue dashed lines represent

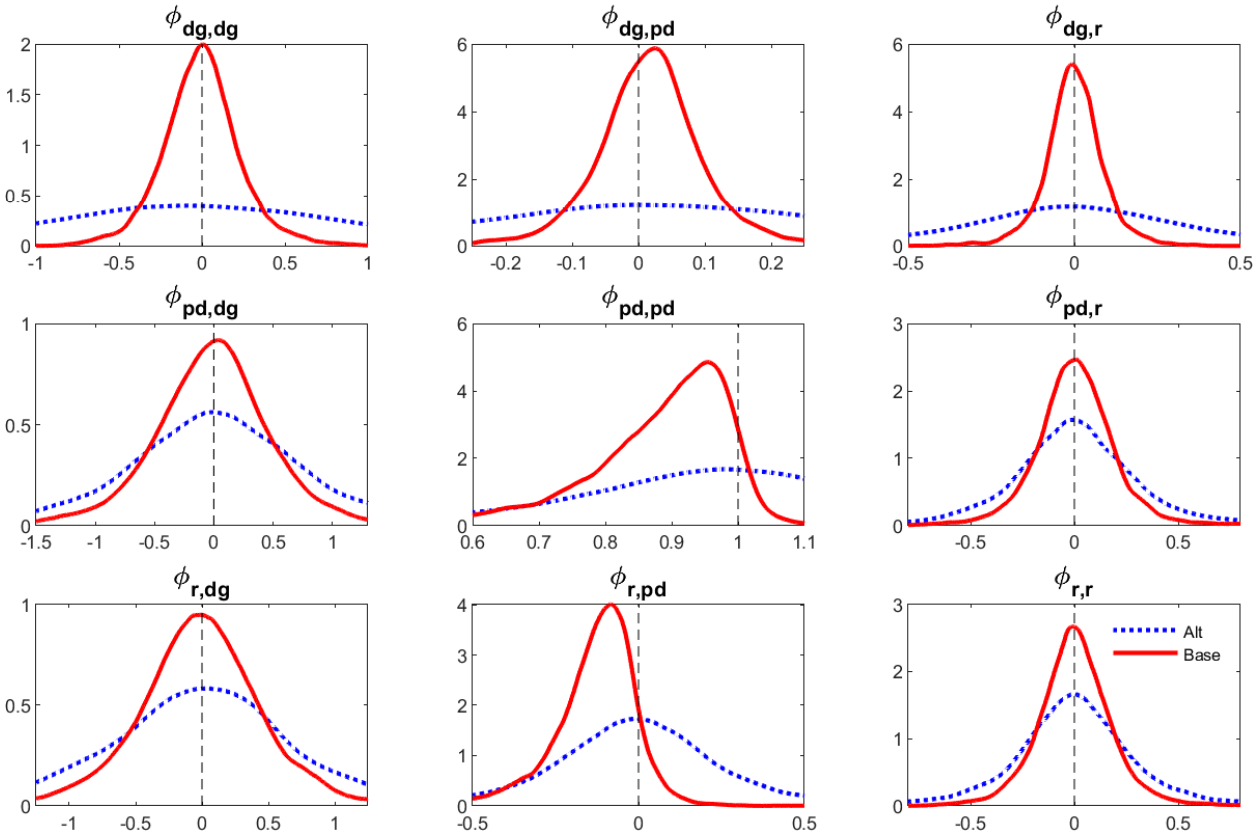
Table D.1: Out-of-Sample Return Equation R-squared: VAR(2)

	Flat priors			**Informative Priors**		
	Unrestricted (Ex. Divs; Two lags)	Unrestricted (Ex. Divs; One lag)	Unrestricted (Inc. Divs., One lag)	Unrestricted (Inc. Divs., One lag)	Unrestricted (Ex. Divs; Two lags)	Restricted (Inc. Divs., One lag)
$h = 1$	-0.4%	-9.7%	-12.0%	-1.1%	-0.4%	5.9%
$h = 2$	1.6%	-7.9%	-10.5%	0.2%	0.8%	14.5%
$h = 3$	0.0%	-26.9%	-28.8%	-0.9%	0.2%	21.5%
$h = 4$	-3.2%	-39.9%	-39.1%	-4.8%	-3.3%	26.2%
$h = 5$	-12.3%	-44.8%	-41.7%	-13.6%	-12.3%	28.8%
$h = 6$	-37.6%	-82.5%	-76.0%	-37.2%	-36.2%	27.0%
$h = 7$	-61.5%	-126.3%	-115.3%	-60.2%	-59.2%	22.3%
$h = 8$	-82.3%	-161.1%	-145.9%	-80.8%	-79.7%	19.7%
$h = 9$	-107.3%	-210.4%	-190.4%	-104.6%	-103.5%	15.5%
$h = 10$	-142.2%	-292.4%	-264.9%	-138.5%	-137.0%	8.5%

Note: Percentage improvement in out-of-sample fit of each investor with respect to the naive investor.

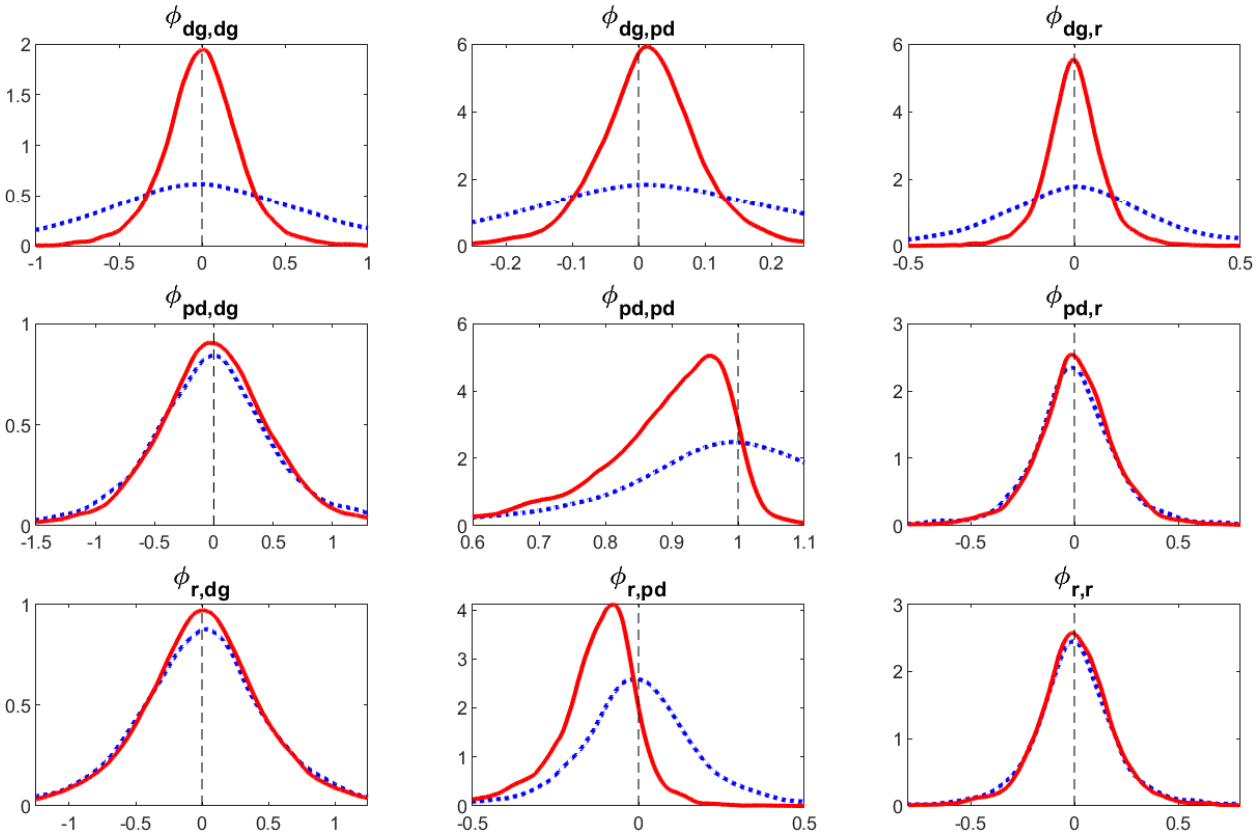
the prior implied by the simpler alternative. In the simpler alternative, the prior for the bottom two rows is set to the standard Minnesota prior, and the top row is recovered from the Campbell-Shiller restrictions. As can be seen, the prior implied by the simpler alternative is generally much flatter, particularly for the dividend growth equation. Due to the differences in priors, the posteriors will also differ, thus affecting the out-of-sample forecasting performance. In the first column of Table D.2, show that indeed the prior implied by the "simpler" alternative leads to worse out-of-sample forecasting performance. Therefore, using our procedure pays off. After observing Figure D.1, one could argue that the problem is the prior over the bottom two rows being "too" flat, resulting in the prior over the top row always being "too" flat. To address this issue, one could tighten the Minnesota prior in the case of the "simpler" alternative to make the bottom two rows as similar as possible. The results can be seen in Figure D.2. However, the prior for the coefficients of the dividend growth equation implied by this alternative procedure are still much flatter than our baseline joint procedure. In the second column of Table D.2, we show that indeed this "matched variance" prior also leads to worse out-of-sample forecasting performance than the joint procedure.

Figure D.1: Bayesian restricted prior and posterior of $\mathbf{\Phi}_1$: Baseline joint prior vs simple alternative procedure



Note: Red represents the baseline prior and blue the prior implied by the simple procedure discussed in Section 3.5.

Figure D.2: Bayesian restricted prior and posterior of $\mathbf{\Phi}_1$: Baseline joint prior vs simple alternative procedure (with Matched Variance)



Note: Red represents the baseline prior and blue the prior implied by the simple procedure discussed in Section 3.5.

Table D.2: Out-of-Sample Return Equation R-squared: "Simpler" procedure that drops Δd_{t+1} equation

	Informative Priors		
	"Simpler" Alternative (Standard)	**"Simpler" Alternative (Matched Variance)**	**Baseline**
$h = 1$	-1.9%	1.2%	5.9%
$h = 2$	-1.2%	2.8%	14.5%
$h = 3$	-13.0%	-4.7%	21.5%
$h = 4$	-17.0%	-6.4%	26.2%
$h = 5$	-14.3%	-3.0%	28.8%
$h = 6$	-28.7%	-11.3%	27.0%
$h = 7$	-43.9%	-20.7%	22.3%
$h = 8$	-52.1%	-25.7%	19.7%
$h = 9$	-64.8%	-34.0%	15.5%
$h = 10$	-86.6%	-48.5%	8.5%

Note: Percentage improvement in out-of-sample fit of each investor with respect to the naive investor.

Appendix E Derivations of the R^2 for Multiple-Period Returns

Consider the VAR in Equation (1) and let $\tilde{\mathbf{\Gamma}}_j = \mathbb{E}_t\left(\mathbf{y}_{t+1}\mathbf{y}'_{t+1+j}\right)$. The R^2 associated with next-period returns can be computed as:

$$R^2(1) = \frac{\mathtt{Var}\left[\mathbb{E}_t\left(r_{t+1}\right)\right]}{\mathtt{Var}\left(r_{t+1}\right)} = \frac{\mathbf{s}_r\mathbf{\Phi}_1\tilde{\mathbf{\Gamma}}_0\mathbf{\Phi}'_1\mathbf{s}'_r}{\mathbf{s}_r\tilde{\mathbf{\Gamma}}_0\mathbf{s}'_r}.$$

For multiple-period returns, we have that:

$$R^2(k) = \frac{\mathtt{Var}\left[\mathbb{E}_t\left(r_{t,t+k}\right)\right]}{\mathtt{Var}\left(r_{t,t+k}\right)} = \frac{\mathbf{s}_r\left(\sum_{j=1}^{k}\mathbf{\Phi}_1^j\right)\tilde{\mathbf{\Gamma}}_0\left(\sum_{j=1}^{k}\mathbf{\Phi}_1^j\right)'\mathbf{s}'_r}{\mathbf{s}_r\left[k\tilde{\mathbf{\Gamma}}_0 + \sum_{j=1}^{k-1}(k-j)\left(\tilde{\mathbf{\Gamma}}_j + \tilde{\mathbf{\Gamma}}'_j\right)\right]\mathbf{s}'_r}.$$

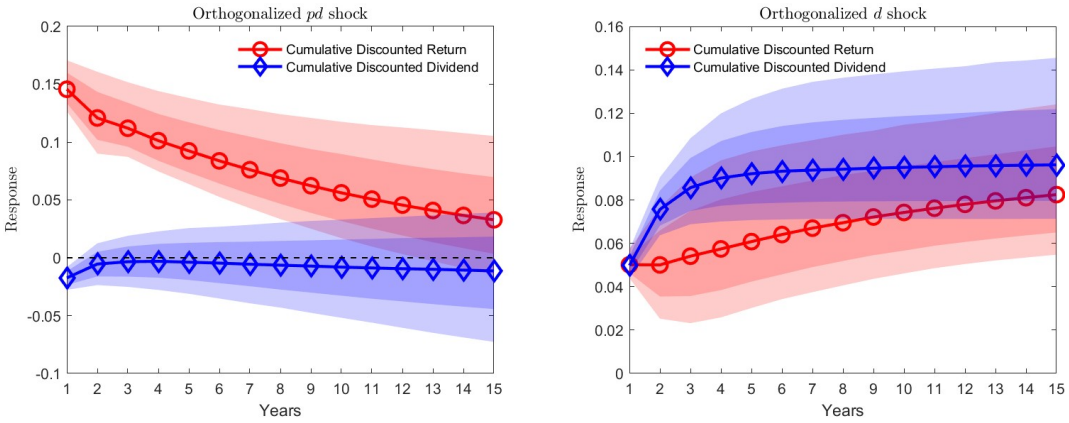
Appendix F Omitting Dividend Growth under Flat Priors

Panel (a) of Figure F.3 reproduces the IRFs in Panel (a) of Figure 7 for flat priors. Because the mean posterior correlation between u_{t+1}^{pd} and u_{t+1}^{d} is -0.32, we obtain the price-dividend ratio and the dividend growth shocks by orthogonalizing the innovations using a Cholesky decomposition in which variables are ordered $[pd_{t+1}, \Delta d_{t+1}, r_{t+1}]$. Panel (a) shows that the price-dividend ratio shock displays the traditional mean reversion channel. Following a positive price-dividend ratio shock returns jump, but the IRF falls over the subsequent 10 years, converging to zero. Since the IRF for cumulative discounted returns converges to the sum of impact effects on returns and NDR_{t+1}, the fact that the IRF converges to zero implies that the impact effect on NDR_{t+1} is strong and negative. Dividend growth is essentially unaffected by the price-dividend ratio shock, implying that the impact effect of NCF_{t+1} is negligible. For a dividend growth shock, we observe the dividend momentum effect: after an initial identical jump in returns and dividend growth, both IRFs show a positive slope, with returns lagging and only catching up gradually.

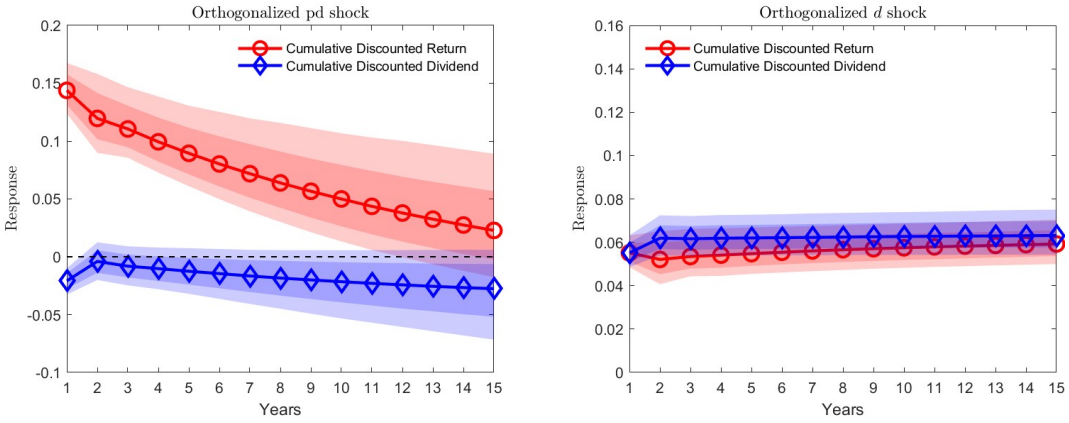
What are the economic consequences if we had followed the common practice of dropping dividend growth from the VAR and backed out the remaining coefficients from the CS restrictions? Panel (b) of Figure F.3 displays the results. The price-dividend ratio shock looks very similar to the one in the 3-variable VAR. As for the dividend growth shock, the IRFs of both returns and dividends are almost perfectly flat after the initial positive jump; estimating the 2-variable VAR(1)

Figure F.3: Impulse Response Functions Under Flat Priors

(a) 3-variable VAR



(b) 2-variable VAR(1) omitting dividend growth



Note: The solid lines represent the median posterior response. The darker shadow area represents the 68^{th} posterior credible intervals, while the lighter shadow area represents the 95^{th} posterior credible intervals.

will not find dividend momentum and leads to similar results as in Panel (b) of Figure 7. This result highlights that dropping dividend growth effectively imposes $\phi_{d,d} = 0$ and arbitrarily rules out dividend momentum. Table F.3 analyzes the contribution of the two shocks to the variances and correlation of NCF_{t+1} and NDR_{t+1}. Because dividend growth shocks lead to positive revisions to current and future cash flows in the 3-variable VAR, the contribution of this shock to the variance of both NCF_{t+1} and NDR_{t+1}, and their correlation is meaningfully larger.

Table F.3: Shock Contribution to NCF_{t+1} and NDR_{t+1} under Flat Priors

	2-variable VAR(1) omitting dividend growth			3-variable VAR		
	Total	u^{pd}_{t+1}	u^{d}_{t+1}	Total	u^{pd}_{t+1}	u^{d}_{t+1}
$Var(NDR_{t+1})$	0.032 [0.022, 0.052]	99.7% [99.4%, 99.9%]	0.3% [0.1%, 0.6%]	0.030 [0.018, 0.059]	89.8% [78.6%, 95.3%]	10.2% [4.7%, 21.4%]
$Var(NCF_{t+1})$	0.006 [0.004, 0.013]	26.1% [4.3%, 60.7%]	73.9% [39.3%, 95.7%]	0.013 [0.008, 0.029]	7.9% [0.7%, 34.2%]	92.0% [65.7%, 99.2%]
$Corr(NDR_{t+1},\ NCF_{t+1})$	55.4% [20.3%, 81.0%]	50.3% [14.9%, 77.5%]	4.4% [2.4%, 6.7%]	51.7% [17.4%, 79.9%]	16.9% [−16.2%, 51.6%]	29.3% [19.0%, 43.1%]

Note: We report the posterior median and the 68^{th} posterior credible intervals. For each model, the "Total" column reflects the posterior of moments, while the u^{pd}_{t+1} and u^{d}_{t+1} columns reflect the posterior contribution of the two shocks.

Appendix G Variance of Long-Horizon Returns

In this section, we derive the variance of long-horizon returns for a system such as the one described in Equation (1), although the calculation is also valid for more general models. Defining the multiple-period returns as $r_{T,T+k} = \sum_{j=1}^{k} r_{T+j}$, the variance of long-horizon returns can be decomposed as:

$$\mathtt{Var}_T\left(r_{T,T+k}\right) = \underbrace{\mathbb{E}_T\left[\mathtt{Var}_T\left(r_{T,T+k}|\mathbf{\Theta}\right)\right]}_{\text{expected variance of long-horizon returns}} + \underbrace{\mathtt{Var}_T\left[\mathbb{E}_T\left(r_{T,T+k}|\mathbf{\Theta}\right)\right]}_{\text{estimation risk}} \tag{G.1}$$

where $\mathbf{\Theta} = \{\mathbf{\Phi}, \mathbf{\Sigma}\}$ denotes the parameters in the VAR. The first term in Equation (G.1) corresponds to the expected variance of long-horizon returns, and since we have assumed constant volatility in the VAR, $\mathtt{Var}_T\left(r_{T,T+k}|\mathbf{\Theta}\right) = \mathtt{Var}\left(r_{T,T+k}|\mathbf{\Theta}\right)$. The estimation risk component instead reflects the parameter uncertainty.

Let us compute the estimation risk component. Let $\mathbf{s}_r$ be a $1 \times n$ vector with 1 corresponding to the position of the returns within the vector of state variables $\mathbf{y}_{t+1}$ (and 0 otherwise), then r_{t+k} can be written as:

$$r_{t+k} = \mathbf{s}_r \left[\left(\mathbf{I}_n - \mathbf{\Phi}_1\right)^{-1} \left(\mathbf{I}_n - \mathbf{\Phi}_1^k\right) \mathbf{\Phi}_0 + \mathbf{\Phi}_1^k \mathbf{y}_t + \mathbf{u}_{t+k} + \sum_{j=1}^{k-1} \mathbf{\Phi}_1^j \mathbf{u}_{t+k-j} \right]$$

Therefore, the expected long-run return can be re-written as:

$$\mathbb{E}_T\left(r_{T,T+k}|\mathbf{\Theta}\right) = \mathbf{s}_r \left(\mathbf{I}_n - \mathbf{\Phi}_1\right)^{-1} \left\{ \left[k\mathbf{I}_n - \left(\mathbf{I}_n - \mathbf{\Phi}_1\right)^{-1} \mathbf{\Phi}_1 \left(\mathbf{I}_n - \mathbf{\Phi}_1^k\right) \right] \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \left(\mathbf{I}_n - \mathbf{\Phi}_1^k\right) \mathbf{y}_T \right\}.$$

This implies that the estimation risk component can be computed as:

$$\texttt{Var}_T\left[\mathbb{E}_T\left(r_{T,T+k}|\boldsymbol{\Theta}\right)\right] = \texttt{Var}_T\left[\mathbf{s}_r\left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1}\left\{\left[k\mathbf{I}_n - \left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1}\boldsymbol{\Phi}_1\left(\mathbf{I}_n - \boldsymbol{\Phi}_1^k\right)\right]\boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1\left(\mathbf{I}_n - \boldsymbol{\Phi}_1^k\right)\mathbf{y}_T\right\}\right].$$

Let us now calculate the expected variance of long-horizon returns. The unpredictable component of the $t+k$ returns is:

$$r_{t+k} - \mathbb{E}_t\left(r_{t+k}|\boldsymbol{\Theta}\right) = \mathbf{s}_r\mathbf{u}_{t+k} + \mathbf{s}_r\sum_{j=1}^{k-1}\boldsymbol{\Phi}_1^j\mathbf{u}_{t+k-j}.$$

and for multiple-period returns:

$$r_{t,t+k} - \mathbb{E}_t\left(r_{t,t+k}|\boldsymbol{\Theta}\right) = \mathbf{s}_r\sum_{j=1}^{k}\mathbf{u}_{t+j} + \mathbf{s}_r\sum_{j=1}^{k-1}\Phi_1\left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right)\mathbf{u}_{t+k-j}$$

Therefore the variance of long-horizon returns consists of three sources of uncertainty: the i.i.d. uncertainty reflecting the accumulated uncertainty of the one-period returns, the future expected return uncertainty, and a component reflecting the covariance between the one-period return uncertainty and the revisions to the future expected returns:

$$\begin{aligned}\texttt{Var}\left(r_{T,T+k}|\boldsymbol{\Theta}\right) \quad = \quad & \underbrace{k\mathbf{s}_r\boldsymbol{\Sigma}\mathbf{s}_r'}_{\text{iid uncertainty}} \quad \underbrace{+\,2\mathbf{s}_r\boldsymbol{\Phi}_1\textstyle\sum_{j=1}^{k-1}\left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right)\boldsymbol{\Sigma}\mathbf{s}_r'}_{\text{covariance component}} + \\ & \underbrace{+\textstyle\sum_{j=1}^{k-1}\left[\mathbf{s}_r\boldsymbol{\Phi}_1\left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right)\right]\boldsymbol{\Sigma}\left[\mathbf{s}_r\boldsymbol{\Phi}_1\left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1}\left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right)\right]'}_{\text{future expected return uncertainty}}\end{aligned} \tag{G.2}$$

The expected variance of long-horizon returns is the expectation of Equation (G.2). The covariance component is generally labeled as "mean reversion." This is because, in the absence of dividend momentum, this covariance term tends to be dominated by the negative co-movement arising between next-period futures and multiple-period returns following a shock to the price-dividend ratio.

G.1 Shock Decomposition of the Variance of Long-Horizon Returns

More generally, one can decompose the expected variance of long-horizon returns into the contribution of each of the structural shocks. Specifically, let $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}'$ where $\mathbf{B}$ corresponds to the matrix capturing the IRF coefficients associated with the structural shocks on impact. We have that

$\boldsymbol{\Sigma} = \sum_{i=1}^{n} \mathbf{B}_{(:,i)}\mathbf{B}'_{(:,i)}$ where $\mathbf{B}_{(:,i)}$ denotes the i-th column of the matrix $\mathbf{B}$. This means that we can retrieve the contribution of the i-th structural shocks to the three sources of uncertainty in Equation (G.2) substituting $\mathbf{B}_{(:,i)}\mathbf{B}'_{(:,i)}$ for $\boldsymbol{\Sigma}$ for each i. Dividend momentum will affect the expected variance of long-horizon returns as long as dividend growth shocks affect the covariance component and future expected uncertainty about returns. To compute the variance of long-horizon returns without dividend momentum, we substitute $\mathbf{B}_{(:,1)}\mathbf{B}'_{(:,1)} + \mathbf{B}_{(:,3)}\mathbf{B}'_{(:,3)}$ for $\boldsymbol{\Sigma}$ in the covariance component and future expected return uncertainty terms of Equation (G.2), since the dividend growth shock is the second shock in our simplified model.

Appendix H Results for a More General Model

The basic macro-finance VAR of Sections 2.1 and A is useful to analyze the basic insights of dropping dividend growth in VAR(1) but one may consider more general models. To do so we expand the model in Equation (1) with the risk-free rate r^f_{t+1} and a vector of dimension $n_w \times 1$ of additional external predictors $\mathbf{w}_{t+1}$. Thus, consider the vector of endogenous variables $\mathbf{y}'_t = \left[\Delta d_{t+1}, \mathbf{w}'_{t+1}, r^f_{t+1}, pd_{t+1}, r_{t+1} - r^f_{t+1}\right]$ of dimension $(n_w + 4) \times 1$ and assume it follows a VAR(1) structure:

$$\underbrace{\begin{bmatrix} \Delta d_{t+1} \\ \mathbf{w}_{t+1} \\ r^f_{t+1} \\ pd_{t+1} \\ r_{t+1} - r^f_{t+1} \end{bmatrix}}_{\mathbf{y}_{t+1}} = \underbrace{\begin{bmatrix} c^d \\ \mathbf{c}^w \\ c^{r^f} \\ c^{pd} \\ c^r \end{bmatrix}}_{\boldsymbol{\Phi}_0} + \underbrace{\begin{bmatrix} \phi_{d,d} & \boldsymbol{\phi}_{d,w} & \phi_{d,r^f} & \phi_{d,pd} & \phi_{d,r} \\ \boldsymbol{\phi}_{w,d} & \boldsymbol{\phi}_{w,w} & \boldsymbol{\phi}_{w,r^f} & \boldsymbol{\phi}_{w,pd} & \boldsymbol{\phi}_{w,r} \\ \phi_{r^f,d} & \boldsymbol{\phi}_{r^f,w} & \phi_{r^f,r^f} & \phi_{r^f,pd} & \phi_{r^f,r} \\ \phi_{pd,d} & \boldsymbol{\phi}_{pd,w} & \phi_{pd,r^f} & \phi_{pd,pd} & \phi_{pd,r} \\ \phi_{r,d} & \boldsymbol{\phi}_{r,w} & \phi_{r,r^f} & \phi_{r,pd} & \phi_{r,r} \end{bmatrix}}_{\boldsymbol{\Phi}_1} \underbrace{\begin{bmatrix} \Delta d_t \\ \mathbf{w}_t \\ r^f_t \\ pd_t \\ r_t - r^f_t \end{bmatrix}}_{\mathbf{y}_t} + \underbrace{\begin{bmatrix} u^d_{t+1} \\ \mathbf{u}^w_{t+1} \\ u^{r^f}_{t+1} \\ u^{pd}_{t+1} \\ u^r_{t+1} \end{bmatrix}}_{\mathbf{u}_{t+1}}. \tag{H.1}$$

As before, this system can be written compactly as $\mathbf{y}_{t+1} = \boldsymbol{\Phi}\mathbf{z}'_t + \mathbf{u}_{t+1}$ where $\mathbf{z}_t = [1, \mathbf{y}'_t]$ and $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_\mathbf{0}, \boldsymbol{\Phi}_\mathbf{1}]$.

In this case, the CS identity implies a relationship between excess returns, dividend growth, changes in the log price-dividend ratio, and the risk-free rate:

$$r_{t+1} - r^f_{t+1} \approx \kappa + \rho pd_{t+1} - pd_t + \Delta d_{t+1} - r^f_{t+1}. \tag{H.2}$$

Equation (H.2) imposes the following restrictions among the innovations:

$$u^r_{t+1} = u^d_{t+1} - u^{r^f}_{t+1} + \rho u^{pd}_{t+1} + \eta_{t+1},$$

the following restrictions on $\mathbf{\Phi}$:

$$\begin{aligned}
c^r &= c^d - c^{r^f} + \rho c^{pd} + \kappa, && \text{(H.3)}\\
\phi_{r,d} &= \phi_{d,d} - \phi_{r^f,d} + \rho\phi_{pd,d}, && \text{(H.4)}\\
\boldsymbol{\phi}'_{r,w} &= \boldsymbol{\phi}'_{d,w} - \boldsymbol{\phi}'_{r^f,w} + \rho\boldsymbol{\phi}'_{pd,w}, && \text{(H.5)}\\
\phi_{r,r^f} &= \phi_{d,r^f} - \phi_{r^f,r^f} + \rho\phi_{pd,r^f}, && \text{(H.6)}\\
\phi_{r,pd} &= \phi_{d,pd} - \phi_{r^f,pd} + \rho\phi_{pd,pd} - 1, \text{ and} && \text{(H.7)}\\
\phi_{r,r} &= \phi_{d,r} - \phi_{r^f,r} + \rho\phi_{pd,r} && \text{(H.8)}
\end{aligned}$$

and the following restrictions on $\mathbf{\Sigma}$:

$$\mathtt{Cov}\left(u^d_{t+1}, u^r_{t+1}\right) = \rho\mathtt{Cov}\left(u^d_{t+1}, u^{pd}_{t+1}\right) + \mathtt{Var}\left(u^d_{t+1}\right) - \mathtt{Cov}\left(u^d_{t+1}, u^{r^f}_{t+1}\right),$$

$$\mathtt{Cov}\left(\mathbf{u}^w_{t+1}, u^r_{t+1}\right) = \rho\mathtt{Cov}\left(\mathbf{u}^w_{t+1}, u^{pd}_{t+1}\right) + \mathtt{Cov}\left(\mathbf{u}^w_{t+1}, u^d_{t+1}\right) - \mathtt{Cov}\left(\mathbf{u}^w_{t+1}, u^{r^f}_{t+1}\right),$$

$$\mathtt{Cov}\left(u^{r^f}_{t+1}, u^r_{t+1}\right) = \rho\mathtt{Cov}\left(u^{r^f}_{t+1}, u^{pd}_{t+1}\right) + \mathtt{Cov}\left(u^{r^f}_{t}, u^d_{t+1}\right) - \mathtt{Var}\left(u^{r^f}_{t+1}\right), \text{ and}$$

$$\mathtt{Cov}\left(u^{pd}_{t+1}, u^r_{t+1}\right) = \rho\mathtt{Var}\left(u^{pd}_{t+1}\right) + \mathtt{Cov}\left(u^{pd}_{t+1}, u^d_{t+1}\right) - \mathtt{Cov}\left(u^{pd}_{t+1}, u^{r^f}_{t+1}\right)$$

respectively. The only difference in the absence of approximation error is that an additional restriction in the covariance matrix $\mathbf{\Sigma}$ linking the variance of u^r_{t+1} with variances and covariances of u^d_{t+1}, $u^{r^f}_{t+1}$, and u^{pd}_{t+1} is needed. In particular, we would have the extra restriction:

$$\begin{aligned}
\mathtt{Var}\left(u^r_{t+1}\right) = \mathtt{Var}\left(u^d_{t+1}\right) + \mathtt{Var}\left(u^{r^f}_{t+1}\right) + \rho^2\mathtt{Var}\left(u^{pd}_{t+1}\right) +\\
2\rho\mathtt{Cov}\left(u^{pd}_{t+1}, u^d_{t+1}\right) - 2\rho\mathtt{Cov}\left(u^{pd}_{t+1}, u^{r^f}_{t+1}\right) - 2\mathtt{Cov}\left(u^d_{t+1}, u^{r^f}_{t+1}\right).
\end{aligned}$$

The stationarity restriction is now $\mathbf{\Phi}_1 \in \left\{\mathbf{Z} \in \mathbb{R}^{(n_w+4)\times(n_w+4)} : \max\{\mathtt{eig}\left(\mathbf{Z}\right)\} < 1\right\}$.

Using the results in Appendix A.1, it can be shown that if we drop dividend growth and run a

VAR(1) on $\mathbf{x}'_{t+1} = \left[\mathbf{w}'_{t+1}, r^f_{t+1}, pd_{t+1}, r_{t+1} - r^f_{t+1}\right]$ we can obtain the same VARMA(2,1) representation of $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{G}_1\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1\mathbf{e}_t, \tag{H.9}$$

where

$$\mathbf{\Phi}_{11} = \begin{bmatrix} \boldsymbol{\phi}_{w,w} & \boldsymbol{\phi}_{w,r^f} & \boldsymbol{\phi}_{w,pd} & \boldsymbol{\phi}_{w,r} \\ \boldsymbol{\phi}_{r^f,w} & \phi_{r^f,r^f} & \phi_{r^f,pd} & \phi_{r^f,r} \\ \boldsymbol{\phi}_{dp,w} & \phi_{dp,r^f} & \phi_{dp,pd} & \phi_{dp,r} \\ \boldsymbol{\phi}_{r,w} & \phi_{r,r^f} & \phi_{r,pd} & \phi_{r,r} \end{bmatrix}, \boldsymbol{\phi}_{12} = \begin{bmatrix} \boldsymbol{\phi}_{w,d} \\ \phi_{r^f,d} \\ \phi_{pd,d} \\ \phi_{r,d} \end{bmatrix}, \boldsymbol{\phi}'_{21} = \begin{bmatrix} \boldsymbol{\phi}'_{d,w} \\ \phi_{d,r^f} \\ \phi_{d,pd} \\ \phi_{d,r} \end{bmatrix}, \text{ and } \boldsymbol{\xi}_{t+1} = \begin{bmatrix} \mathbf{u}^z_{t+1} \\ u^{r^f}_{t+1} \\ u^{pd}_{t+1} \\ u^r_{t+1} \end{bmatrix}.$$

As before, without loss of generality, we abstract from the constant term. Consider now specifying a VAR(1) for $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{A}_1\mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1}. \tag{H.10}$$

The matrix $\mathbf{A}_1$ follows the expression in Equation (A.5); thus the link between $\mathbf{A}_1$ and the parameters in the VAR(1) in Equation (H.1) highlighted in Section A also exists here. The next theorem replicates Theorem A.1 in this more general set up.

Theorem H.1. *The VARMA(2,1) in Equation (H.9) will have* $\mathbf{G}_1 = \mathbf{\Phi}_{11}$, $\mathbf{G}_2 = \mathbf{0}_{(n+3)\times(n+3)}$, *and* $\mathbf{D}_1 = \mathbf{0}_{(n+3)\times(n+3)}$ *if and only if* $\boldsymbol{\phi}_{w,d} = \mathbf{0}_n$ *and* $\phi_{r^f,d} = \phi_{pd,d} = \phi_{r,d} = 0$, *with* $\phi_{d,d} = 0$ *following from Equation (H.4).*

Proof. The proof of the theorem follows easily using the same steps as in the proof of Theorem A.1. □

Of course, Theorem H.1 implies that Corollary A.1 also holds for the more general model described in this section. The innovations in the VAR(1) specified in Equation (H.10) also follow the expression in Equation (A.6) and as a consequence $\mathbf{\Omega}_{\boldsymbol{\varepsilon}} \neq \mathbf{\Omega}_{\mathbf{e}}$. As expected, it is also the case that Corollary A.2 holds here and the results in Section A.2 when the approximation error is not present also hold in this more general case.

Appendix I Priors for Section 5

In this section we describe the prior parameterizations used in Section 5.

I.1 The Flat Prior

We follow Uhlig (2005) and set $\underline{\nu} = 0$ and $\underline{\mathbf{V}}^{-1} = \mathbf{0}_{5\times5}$ and let $\underline{\mathbf{S}}$ and $\underline{\boldsymbol{\alpha}}$ be arbitrary. This means that the posterior means are centered around the OLS estimates and the Bayesian high posterior density intervals coincide with the classical confidence intervals.

I.2 The Informative Prior

The Minnesota priors in this section can be written:

$$p(\mathtt{vec}(\mathbf{\Phi}_1)|\mathbf{\Sigma}) \sim \mathcal{N}\left(\mathtt{vec}\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}, \mathbf{\Sigma}\otimes\mathbf{\Omega}\right)$$

where in the simplest case $\mathbf{\Omega} = \lambda^2(diag\left([\underline{\sigma}_d^2, \underline{\sigma}_{rf}^2, \underline{\sigma}_{pd}^2, \underline{\sigma}_r^2]\right))^{-1}$, with $\lambda = 0.185$. We follow the common practice of setting $\underline{\sigma}_i^2$ to the residual variance of an AR(1) model. For $\mathbf{\Phi}_0$, the Minnesota prior is usually specified as flat.

We choose a value of $\theta = 0.05$, which controls the tightness of the Single Unit Root prior. As before, we follow Giannone et al. (2015) when choosing λ and θ. In particular, we use the starting sample finishing in 1973 to choose them. For the mean of the excess return, $\underline{\mu}_r$ we chose a value of 6.5 percent; the mean of the risk-free rate $\underline{\mu}_{rf}$ is chosen to be 4 percent; and for the mean nominal dividend growth, $\underline{\mu}_d$ we chose a value of 5.5 percent, consistent with long-run nominal GDP growth in the United States. Given these values we can back out the implied $\underline{\mu}_{pd}$ from CS restriction (4) to be about 2.8, which is close to the value of the log price-dividend ratio at the beginning of the postwar sample. The hyperparameters λ and θ are chosen to maximize the value of the marginal likelihood, as proposed by Giannone et al. (2015).

Appendix J Additional Details on the Analysis of Data at Higher Frequency

In this appendix, we provide the state-space representation of the model with seasonality in dividends introduced in Section 6, along with additional details on its estimation and robustness results.

J.1 State-Space Representation of the Model

The model has the following state-space representation:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t \\ \mathbf{x}_t &= \mathbf{a} + \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}_t\mathbf{e}_t \end{aligned}$$

where $\mathbf{y}_t = [\Delta d_t, pd_t, r_t]$ is the vector of observed data, and $\mathbf{x}_t$ is a vector of partially unobserved state variables. Specifically, we let $\mathbf{x}_t = [\Delta d_t^{sa}, pd_t^{sa}, r_t, pd_t^s, pd_{t-1}^s, pd_{t-2}^s]$, thus:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & -\rho & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{J.1}$$

The vector $\mathbf{a} = [\mathbf{\Phi}_0', \mathbf{0}_{1\times 3}]'$,

$$\mathbf{F} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{0}_{3\times 3} \\ \mathbf{0}_{3\times 3} & \mathbf{\Phi}_s \end{bmatrix}, \ \mathbf{\Phi}_s = \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{J.2}$$

while:

$$\mathbf{G}_t = \begin{bmatrix} \mathbf{I}_3 & 0_{3\times 1} \\ 0_{1\times 3} & \sigma_t^s \end{bmatrix} \tag{J.3}$$

and $\mathbf{e}_t = [\varepsilon_t', e_t^s]'$, where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \mathbf{\Sigma})$ and $e_t^s \sim$ i.i.d. $\mathcal{N}(0,1)$.

J.2 Gibbs Sampler Algorithm

Posterior estimates are obtained using the following Gibbs sampler algorithm:

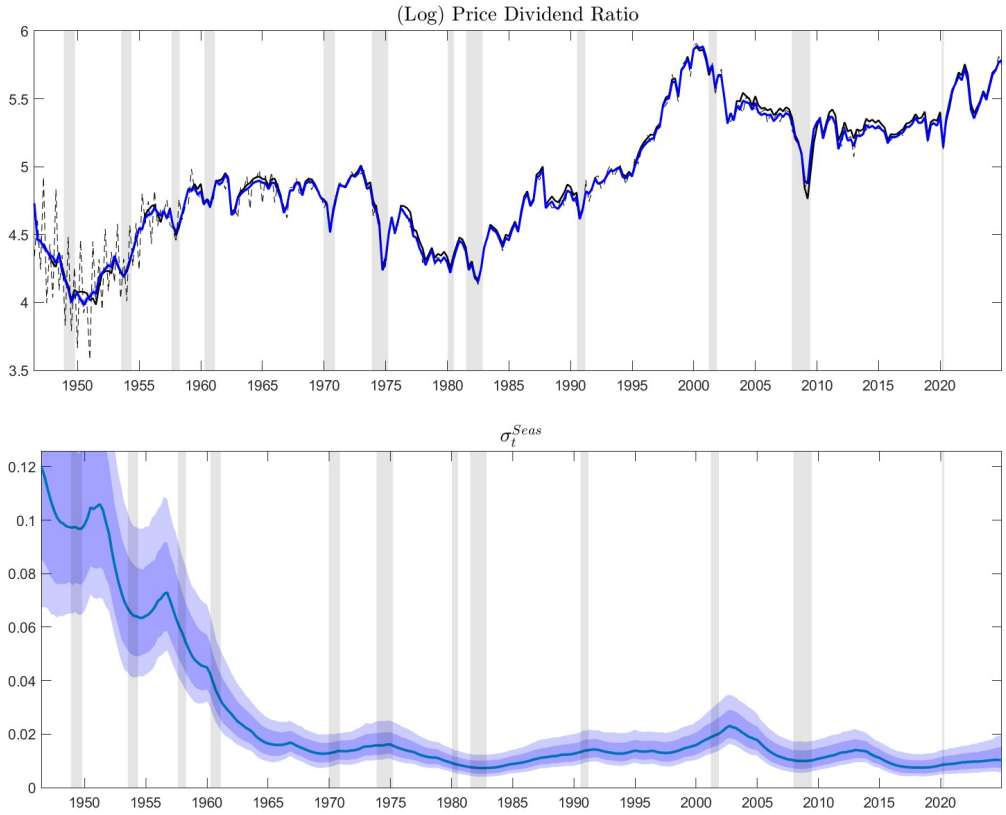
- **Step 0:** Initialize the parameters of the model $(\mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Sigma}, \{\sigma_t^s\}_{t=0}^T, \sigma_\omega^2)$.
- **Step 1:** Conditional on $(\mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Sigma}, \{\sigma_t^s\}_{t=0}^T)$, draw $\mathbf{x}_t$ from $p\left(\mathbf{x}_t | \mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Sigma}, \{\sigma_t^s\}_{t=0}^T\right)$ using the Kalman filter and simulation smoother (see Durbin and Koopman, 2001).
- **Step 2:** Conditional on $\mathbf{x}_t$, follow the algorithm in Section 3 to draw $(\mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Sigma})$ using the seasonally adjusted data defined as $\widetilde{\mathbf{y}}_t = [\mathbf{I}_3, \mathbf{0}_{3\times 3}]\,\mathbf{x}_t = [\Delta d_t^{sa}, pd_t^{sa}, r_t]$.
- **Step 3:** Conditional on $\mathbf{x}_t$, define $u_t = \sum_{j=0}^3 pd_{t-j}^s$, where $pd_t^s = [\mathbf{0}_{1\times 3}, 1, \mathbf{0}_{1\times 2}]\,\mathbf{x}_t$. Obtain a draw of the stochastic volatilities of the innovations to the seasonal component, $\{\sigma_t^s\}_{t=0}^T$, from $p\left(\{\log \sigma_t^s\}_{t=0}^T \,|\, \{u_t\}_{t=0}^T, \sigma_\omega^2\right)$ using the algorithm of Omori et al. (2007), which employs a mixture of normal random variables to approximate the elements of the log-variance.
- **Step 4:** Conditional on $\{\log \sigma_t^s\}_{t=0}^T$, draw σ_ω^2 from an inverse gamma distribution, assuming an inverse-gamma prior $p(\sigma_\omega^2) \sim IG(S_\omega, v_\omega)$, such that the conditional posterior of σ_ω^2 is also drawn from an inverse-gamma distribution. We choose the scale $S_\omega = 10^{-4}$ and degrees of freedom $v_\omega = 1$.
- Repeat from **Step 1** until convergence is achieved.

In practice, retain 5,000 draws from the posterior distribution after discarding the first 1,000 as burn-in.

J.3 Additional Results and Robustness

Figure J.4 presents additional results for the baseline model using quarterly data. Specifically, the upper panel of Figure J.4 compares pd_t^s with the original data, where the price-dividend ratio is constructed without any transformation of the original dividend data. It also includes the (log) price-dividend ratio constructed using the sum of dividends over the past year as a method to address the seasonality in the dividend data. The lower panel displays the posterior estimates of $\log \sigma_t^s$.

Figure J.4: Additional Results from Quarterly Model with Seasonality

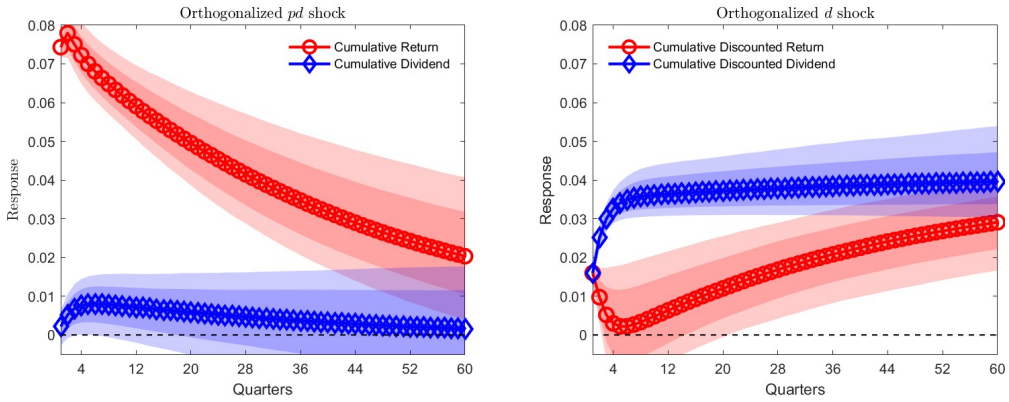


Note: The upper panel plots the (log) price-dividend ratio. The black broken line represents the raw data, while the black continuous line corresponds to the data constructed using a moving average of the past four quarters of dividends. The blue line shows the seasonally adjusted data from our model. The lower panel displays the posterior estimates of σ_t^s. The darker shaded area represents the 68th posterior credible interval, while the lighter shaded area represents the 95th posterior credible interval.

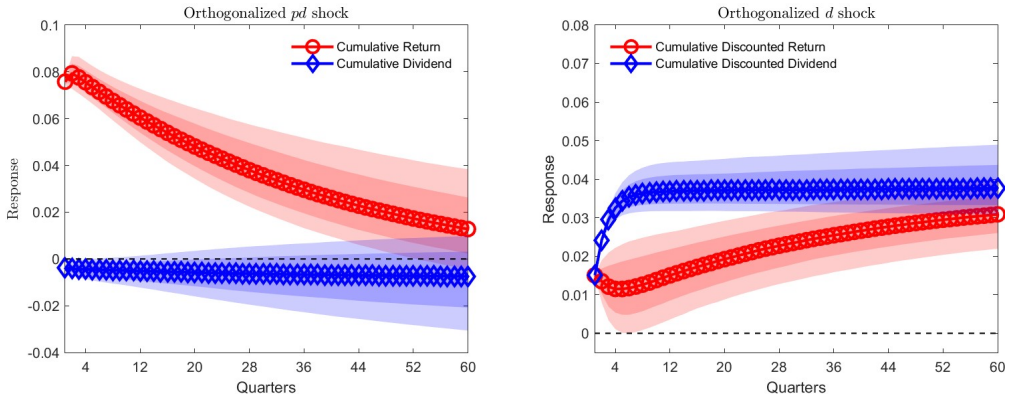
Finally, we report the impulse responses associated with orthogonalized shocks to u_{pd} and u_d for two alternative versions of the model. The upper panel of Figure J.5 shows the results for a version of the model that includes seasonality but assumes constant volatility for the seasonal component. The lower panel of Figure J.5 displays the corresponding results for a VAR(1) model estimated on data where quarterly dividends are measured as the sum of the past four quarters of dividends. In both cases, the impulse response functions (IRFs) provide clear evidence of dividend momentum, with results closely resembling those presented in Section 6.

Figure J.5: Dividend Momentum with Quarterly Data – Alternative Specifications

(a) No stochastic volatility in seasonal component



(b) Model estimated using smoothed dividends



Note: The solid lines represent the median posterior response. The darker shadow area represents the 68th posterior credible intervals, while the lighter shadow area represents the 95th posterior credible intervals.

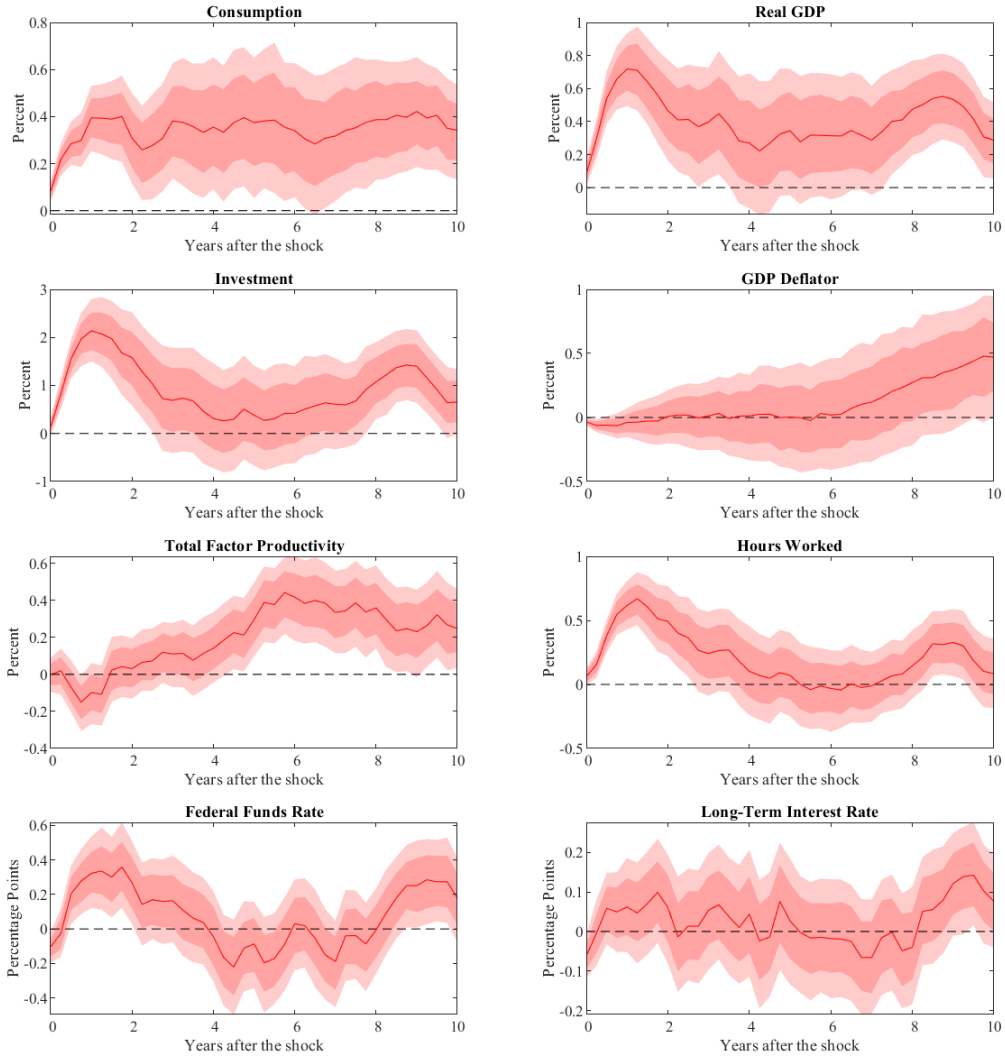
J.4 Local Projections of price-dividend shock on Macroeconomic Aggregates

In Figure J.6, we present the local projections showing the responses of key macroeconomic indicators to a price-dividend (PD) shock. A brief summary of the results is provided in Section 6.

J.5 Analysis of Monthly Data

In this final section, we investigate the presence of dividend momentum at monthly frequency. Dividends are backed out from S&P's daily series of total returns and returns excluding dividends, available since January 1988. These inferred dividends are aggregated to a monthly frequency, producing a series of non-reinvested cash dividends.

Figure J.6: Local Projections of Macroeconomic Aggregates on PD Shock



Note: The solid lines represent the median posterior response. The darker shadow area represents the 68^{th} posterior credible intervals, while the lighter shadow area represents the 90^{th} posterior credible intervals.

We adopt the same model described in Section 6. As before, both dividend growth and the (log) price-dividend ratio exhibit seasonal variation, modeled as the sum of seasonal and non-seasonal components (see Eq. 29). For consistency with the lower-frequency specifications, we continue to assume that the dynamics of the non-seasonal component are captured by a VAR(1).[28] Assuming no seasonal variation in returns, the CS restriction implies that any seasonality in dividend growth must be mirrored in the seasonality of the (log) price-dividend ratio (see Eq. 30).

We allow the seasonal pattern in the data to evolve slowly over time, using a mechanism that

[28] Note that with monthly data, one would naturally favor a VAR specification with more than a single lag.

ensures the expected value of the seasonal component over any twelve consecutive months is zero:

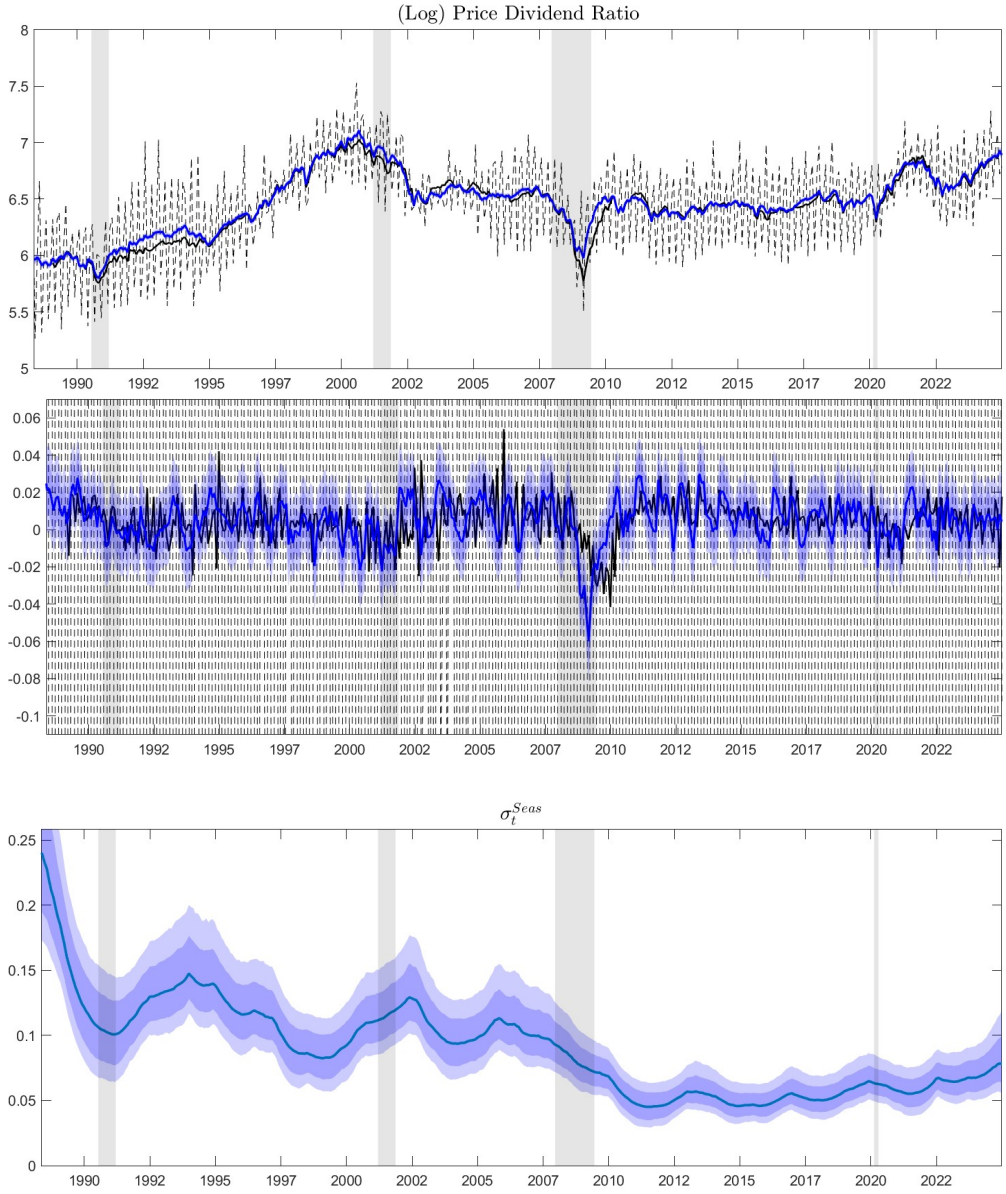
$$\sum_{j=0}^{11} pd_{t-j}^{s} = \sigma_t^s e_t^s \tag{J.4}$$

where $e_t^s \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\log \sigma_t^s = \log \sigma_{t-1}^s + \omega_{t+1}$, with $\omega_t \sim$ i.i.d. $\mathcal{N}(0, \sigma_\omega^2)$.

The model is estimated following the procedure described in the previous section. Figure J.7 presents results from the baseline monthly-frequency model. The top panel compares pd_t^s with the raw (log) price-dividend ratio and a smoothed version constructed using trailing twelve-month dividends to mitigate seasonality. The middle panel shows the same comparison for dividend growth. The range is truncated to prevent the seasonal component from overwhelming the cyclical variability in dividend growth. The seasonally adjusted dividend growth aligns more closely with the business cycle, exhibiting clearer reversion around turning points—most notably during the Great Recession and again in 2020–2021. In contrast, the smoothed series mechanically lags. The bottom panel displays the posterior estimates of $\log \sigma_t^s$, reflecting the time-varying nature of the seasonal component's volatility—clearly visible, for example, in the upper panel of Figure J.7.

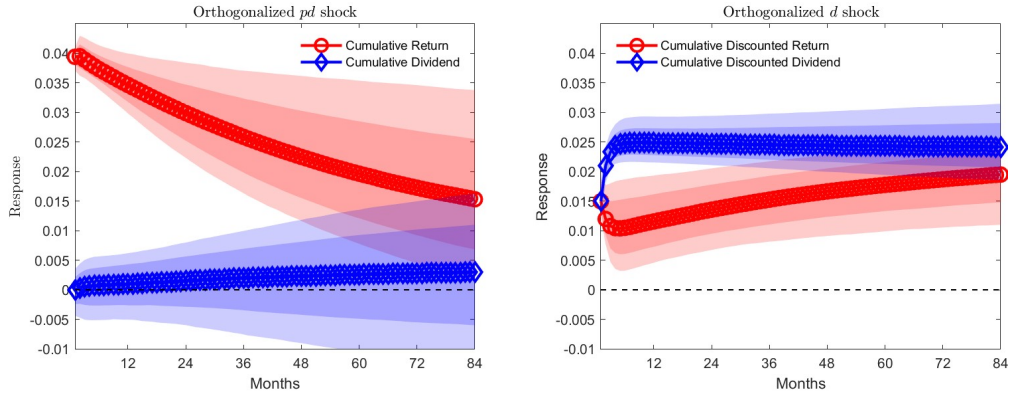
Figure J.8 shows the responses of cumulative dividends and returns following shocks to the price-dividend ratio and dividend growth, identified as in Section 4.5. Even at monthly frequency, dividend momentum emerges as a prominent feature. Dividend shocks lead to an immediate rise in dividends, followed by a persistent increase in expected dividend growth and expected returns.

Figure J.7: Additional Results from Monthly Model with Seasonality



Note: The upper and middle panels display the (log) price-dividend ratio and dividend growth, respectively. The dashed black line represents the raw data; the solid black line shows the smoothed data based on a moving average of the past four quarters; and the blue line displays the seasonally adjusted data from our model. The lower panel presents the posterior estimates of σ_t^s. The darker shaded area denotes the 68% credible interval, while the lighter shaded area shows the 95% credible interval.

Figure J.8: Dividend Momentum with Monthly Data



Note: Solid lines show median posterior responses. The darker shaded region represents the 68% credible interval, and the lighter shaded area denotes the 95% credible interval.

Appendix References

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

Campbell, J. Y. and R. J. Shiller (1988): "Interpreting cointegrated models," *Journal of Economic Dynamics and Control*, 12, 505–522.

Chen, L. and X. Zhao (2009): "Return decomposition," *The Review of Financial Studies*, 22, 5213–5249.

Cochrane, J. H. (2008a): "State-Space vs. VAR Models for Stock Returns," *Unpublished Manuscript*.

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

Durbin, J. and S. J. Koopman (2001): *Time Series Analysis by State Space Methods*, Oxford University Press.

Engsted, T., T. Q. Pedersen, and C. Tanggaard (2012): "Pitfalls in VAR based return decompositions: A clarification," *Journal of Banking & Finance*, 36, 1255–1265.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

Harvey, A. C. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007): "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425–449.

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

Uhlig, H. (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, 52, 381 – 419.

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

Zellner, A. and F. Palm (1974): "Time series analysis and simultaneous equation econometric models," *Journal of Econometrics*, 2, 17–54.