DISCUSSION PAPER SERIES

DP16613

Dividend Momentum and Stock Return Predictability: A Bayesian Approach

Juan Francisco Rubio-Ramírez, Ivan Petrella and Juan Antolin-Diaz

FINANCIAL ECONOMICS

CEPR

Dividend Momentum and Stock Return Predictability: A Bayesian Approach

Juan Francisco Rubio-Ramírez, Ivan Petrella and Juan Antolin-Diaz

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Financial Economics

Dividend Momentum and Stock Return Predictability: A Bayesian Approach

Abstract

A long tradition in macro-finance studies the joint dynamics of aggregate stock returns and dividends using vector autoregressions (VARs), imposing the cross-equation restrictions implied by the Campbell-Shiller (CS) identity to sharpen inference. We take a Bayesian perspective and develop methods to draw from any posterior distribution of a VAR that encodes a priori skepticism about large amounts of return predictability while imposing the CS restrictions. In doing so, we show how a common empirical practice of omitting dividend growth from the system amounts to imposing the extra restriction that dividend growth is not persistent. We highlight that persistence in dividend growth induces a previously overlooked channel for return predictability, which we label "dividend momentum." Compared to estimation based on OLS, our restricted informative prior leads to a much more moderate, but still signi cant, degree of return predictability, with forecasts that are helpful out-of-sample and realistic asset allocation prescriptions with Sharpe ratios that out-perform common benchmarks.

Juan Francisco Rubio-Ramírez - juan.rubio-ramirez@emory.edu
Emory University and Federal Reserve Bank of Atlanta and CEPR

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

Juan Antolin-Diaz - jantolindiaz@london.edu
LBS

Dividend Momentum and Stock Return Predictability: A Bayesian Approach[*]

Juan Antolín-Díaz

London Business School

Ivan Petrella

University of Warwick

Juan F. Rubio-Ramírez[†]

Emory University

Federal Reserve Bank of Atlanta

Very Preliminary - Comments Welcome

October 4, 2021

Abstract

A long tradition in macro-finance studies the joint dynamics of aggregate stock returns and dividends using vector autoregressions (VARs), imposing the cross-equation restrictions implied by the Campbell-Shiller (CS) identity to sharpen inference. We take a Bayesian perspective and develop methods to draw from any posterior distribution of a VAR that encodes a priori skepticism about large amounts of return predictability while imposing the CS restrictions. In doing so, we show how a common empirical practice of omitting dividend growth from the system amounts to imposing the extra restriction that dividend growth is not persistent. We highlight that persistence in dividend growth induces a previously overlooked channel for return predictability, which we label "dividend momentum." Compared to estimation based on OLS, our restricted informative prior leads to a much more moderate, but still significant, degree of return predictability, with forecasts that are helpful out-of-sample and realistic asset allocation prescriptions with Sharpe ratios that out-perform common benchmarks.

JEL Classification Numbers: C32, C53, E47.

1 Introduction

A long tradition in macro-finance investigates the presence of stock return predictability by looking at the joint dynamics of aggregate dividend growth, the price-dividend ratio, and stock returns. A key insight, due to Campbell and Shiller (1988a,b), is that the definition of returns imposes cross-equation restrictions on the dynamics of these three variables that can be exploited to sharpen inference using vector autoregressions (VARs). The accepted empirical practice in imposing the restrictions is to drop one of the variables, usually dividend growth, and recover the remaining VAR coefficients from the Campbell-Shiller (CS) identity. Using this approach, Cochrane (2008) concludes that stock returns are predictable, mean-reverting over long horizons.

We start by pointing out that, contrary to widespread belief, the practice of omitting one of the variables from the identity is in general invalid. Dropping dividend growth amounts to imposing the additional restriction that this variable is not persistent after controlling for lags of the remaining variables. Intuitively, the CS identity is an intertemporal restriction; so unless dividend growth happens to be independently and identically distributed, the information contained in this variable cannot be recovered with a finite-order VAR in the price-dividend ratio and returns alone. This is true even in the case where there is no approximation error in the CS identity. We find that the extra restriction on the dynamics of dividends is not supported by annual US postwar data.

Relaxing this additional restriction does not affect Cochrane's (2008) finding that returns display mean reversion, but uncovers an additional and previously overlooked channel of return predictability, which we label "dividend momentum": following a shock that increases both returns and dividends on impact, dividend growth remains positive for many periods, and by the CS identity, future returns increase as well. This channel cannot arise in a VAR in which dividend growth is omitted from the system. The presence of dividend momentum

modifies the interpretation of a popular decomposition of return innovations into cash flow and discount rate news (see Campbell and Ammer, 1993): these two components can no longer be interpreted as permanent and transitory innovations to wealth (see Campbell and Vuolteenaho, 2004). Additionally, dividend momentum has non-trivial implications for the optimal asset allocation of long-horizon investors, as it increases the long-run variance of stock returns and reduces their intertemporal hedging demand motive. Finally, the presence of dividend momentum poses challenges to both habit-formation and long-run risk theories of asset pricing, as both of them imply that dividend growth is not persistent after controlling for the lagged price-dividend ratio.

Having established that dropping dividend growth is not an option, and that doing so has important empirical and economic consequences, we propose a Bayesian approach to inference that imposes the CS restrictions without omitting any variable. A central motivation for using Bayesian methods is to address the fact that the degree of return predictability found in the data is a priori implausible and not useful out-of-sample, and that the very high persistence of the dividend-price ratio leads to unreliable inference. Our Bayesian approach provides a tractable way to use informative priors that embody skepticism about the degree of return predictability, in the spirit of Wachter and Warusawitharana (2009, 2015), shrinking the amount of predictability and adjusting upward the persistence of the price-dividend ratio while making the CS restrictions hold. We note that our methods extend to the many applications in macroeconomics and finance where similar identities emerge, such as the ones linking bond returns and interest rates (see Campbell, Lo, and MacKinlay, 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, 2016); net foreign assets, net exports, and the return on the foreign asset portfolio (see Gourinchas and Rey, 2007), or government debt, surpluses, and interest rates (see Cochrane, 2019), all of which feature stationary but highly persistent variables as predictors of asset returns.

Although our methods work with any prior distribution, we use the popular class of conjugate normal-inverse Wishart priors as an illustration. While imposing linear restrictions on the normally distributed autoregressive coefficients of the VAR is straightforward, the same is not true for the restrictions imposed on the inverse Wishart distribution of the VAR variance-covariance matrix. Our technical contribution is to show that the restricted variance-covariance matrix can be linearly mapped into a block-diagonal symmetric positive semidefinite matrix. Because the volume element associated with the restricted linear transformation is constant, we can develop a very simple importance sampling algorithm to generate independent draws from any desired restricted distribution. This makes our method scalable to large systems, an advantage relative to existing alternatives, opening the door to studying return predictability in VARs with potentially hundreds of variables.

In our empirical application, we put our methods to work and explore the consequences of an informative prior that is tightly centered around no return predictability and a highly persistent price-dividend ratio but satisfies the CS restrictions. Even from this conservative starting point, using annual data for the US covering the period 1947-2018, we find that the restricted posterior distribution is consistent with an economically meaningful degree of return predictability coming from both the traditional mean-reversion and the novel dividend momentum channels. Short-run predictability in particular is attenuated relative to using an uninformative prior distribution, but this is compensated by an increase in longer-run predictability.

Turning to out-of-sample predictability, we show how the use of restricted informative priors reverts the conclusions of Goyal and Welch (2008) that any in-sample predictability is not useful out-of-sample. We obtain out-of-sample R-squared statistics that over-perform a naive benchmark by almost 30 percent at the five-year horizon, in contrast to alternatives using flat priors or ignoring the CS restrictions, which uniformly under-perform the same benchmark. To assess the economic value of these results, we consider the asset allocation problem of a long-horizon investor who can choose how much to invest in stocks each year

until retirement. If the investor estimates the VAR using flat priors, the optimal solution leads to unrealistically large swings in positions and large amounts of leverage, which ultimately under-perform a naive allocation rule. Under our restricted informative prior, the investor chooses allocations that not only are more realistic but also deliver sizable improvements in Sharpe ratios, 0.52 compared to 0.36 for the naive rule. We show that the informativeness of the no-predictability prior needs to be sufficiently strong for this result to emerge. In that sense, a conservative prior that starts from skepticism about return predictability ends up helping uncover the amount of predictability required to successfully time the market. It is important to note that the dividend momentum channel also meaningfully affects these results in the same direction, as the reduction in hedging demand for stocks makes the allocation more conservative for any given prior.

Relation to the literature Our paper is tightly connected to the vast literature that has looked at return predictability using VARs, in particular, the classic papers by Campbell and Shiller (1988a,b), Fama and French (1988), and Cochrane (2008), whose implications for asset allocation purposes have been highlighted by Campbell and Viceira (1999), Campbell, Chan, and Viceira (2003). The persistence of aggregate dividend growth in isolation has been noted by Van Binsbergen and Koijen (2010), Koijen and Nieuwerburgh (2012), and Chen, Da, and Priestley (2012). We investigate the consequences of dividend growth persistence in the canonical VAR setting and show how it is the residual autocorrelation of dividend growth, after controlling for the lagged price-dividend ratio, which gives rise to momentum in returns via the CS identity. This is distinct from the persistence in expected dividend growth emphasized by the long-run risk literature (see Bansal and Yaron, 2004; Schorfheide, Song, and Yaron, 2018), which would be fully captured by time variation in the price-dividend ratio.

Our methodological contribution shares its motivation with the literature on Bayesian learning and informative priors in the context of return predictability, as exemplified by

Wachter and Warusawitharana (2009, 2015) and Pástor and Stambaugh (2009, 2012). However, in departing from the predictive-regression approach and adopting a general VAR setting, we bring the Bayesian paradigm closer to the classic papers above, which, like ours, stress the importance of studying returns, dividends, and price-dividend ratios as a system. Our work is also broadly relevant to the Bayesian econometrics literature on appropriate priors for VARs. Early contributions in this literature go back to the work of Doan, Litterman, and Sims (1984) and Sims (1993) and have seen a renewed interest in recent years (see, e.g., Del Negro and Schorfheide, 2004; Giannone, Lenza, and Primiceri, 2015, 2019), including in the context of return predictability (see Avramov, Cederburg, and Lučivjanská, 2018).

The rest of the paper is organized as follows. Section 2 describes the basic model that we will use to illustrate our methods. Section 3 explains the econometric consequences of omitting dividend growth. Section 4 introduces dividend momentum. Section 5 provides a first exploration of the data. Section 6 describes our methods to incorporate informative priors that implement the CS restrictions, and Section 7 describes the results. Finally, Section 9 concludes.

2   A Basic Macro-Finance VAR

The classic macro-finance VAR approach of Cochrane (2008) studies the joint dynamics of log dividend growth, the log price-dividend ratio, and log returns. In this section we focus on a minimalistic VAR that contains only the three key variables above and uses only one lag. None of our conclusions, however, are affected if we add more variables to the VAR, or more

lags. The VAR(1) in $\mathbf{y}_t' = [\Delta d_{t+1}, pd_{t+1}, r_{t+1}]$ is written:

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

where $\mathbb{E}_t(\mathbf{u}_{t+1}) = \mathbf{0}_{3\times1}$, $\mathbb{E}_t(\mathbf{u}_{t+1}\mathbf{u}_{t+1}') = \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}$ is a symmetric and positive semidefinite (SPD) 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 $\mathbf{y}_{t+1} = \boldsymbol{\Phi}\mathbf{z}_t' + \mathbf{u}_{t+1}$ where $\mathbf{z}_t = [1, \mathbf{y}_t']$ and $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_0, \boldsymbol{\Phi}_1]$. We define $\boldsymbol{\mu} = \boldsymbol{\Phi}_0(\mathbf{I}_n - \boldsymbol{\Phi}_1)^{-1}$ as the unconditional mean of the variables. 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. Being derived from a definition, the Campbell-Shiller (CS) identity holds very tightly in the data.[1] It holds with equality if one adds an approximation error, denoted η_{t+1}. In the 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 or a price-earnings ratio is used instead in the VAR. In that case, it is easy to show that Equation (2) imposes the following restrictions among the VAR innovations:

$$
u_{t+1}^r = u_{t+1}^d + \rho u_{t+1}^{pd} + \eta_{t+1}, \tag{3}
$$

---

[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.

7

The CS identity implies restrictions among the 3-variable VAR(1) coefficients in Equation (1). In particular it imposes (linear) restrictions on $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$. The restrictions for $\boldsymbol{\Phi}$ are:

$$c^r = c^d + \rho c^{pd} + \kappa, \tag{4}$$

$$\phi_{r,d} = \phi_{d,d} + \rho\phi_{pd,d}, \tag{5}$$

$$\phi_{r,pd} = \phi_{d,pd} + \rho\phi_{pd,pd} - 1, \tag{6}$$

$$\phi_{r,r} = \phi_{d,r} + \rho\phi_{pd,r}, \tag{7}$$

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

$$\texttt{Cov}\left(u_{t+1}^r, u_{t+1}^d\right) = \rho\texttt{Cov}\left(u_{t+1}^{pd}, u_{t+1}^d\right) + \texttt{Var}\left(u_{t+1}^d\right) \text{ and} \tag{8}$$

$$\texttt{Cov}\left(u_{t+1}^r, u_{t+1}^{pd}\right) = \rho\texttt{Var}\left(u_{t+1}^{pd}\right) + \texttt{Cov}\left(u_{t+1}^d, u_{t+1}^{pd}\right). \tag{9}$$

If we do not consider approximation error we will have an additional restriction for $\boldsymbol{\Sigma}$: $\texttt{Var}\left(u_{t+1}^r\right) = \texttt{Var}\left(u_{t+1}^d\right) + \rho^2\texttt{Var}\left(u_{t+1}^{pd}\right) + 2\rho\texttt{Cov}\left(u_{t+1}^d, u_{t+1}^{pd}\right)$.

Additionally, the derivation of the CS identity (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 $\boldsymbol{\Phi}_1$, which we write:

$$\boldsymbol{\Phi}_1 \in \left\{\mathbf{Z} \in \mathbb{R}^{3\times 3} : \max\{\texttt{eig}\left(\mathbf{Z}\right)\} < 1\right\}. \tag{10}$$

This generalizes Cochrane's (2008) requirement of an upper bound to the persistence of the price-dividend ratio. The CS restrictions (4)-(9) and the stationarity restriction (10) are cross-equation restrictions on the VAR described 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 (10) as well.

Similar log-linear identities are pervasive in the macro-finance literature, linking, e.g.,

bond returns and interest rates (see Campbell, Lo, and MacKinlay, 1997); consumption, wealth, and returns (see Campbell and Mankiw, 1989; Lettau and Ludvigson, 2001); net foreign assets, net exports, and the return on the foreign asset portfolio (see Gourinchas and Rey, 2007); or government debt, surpluses, and interest rates (see Cochrane, 2019). All of them imply equivalent sets of restrictions.

Because an identity links the three variables, a common belief in the literature is that one of the three can be dropped from a VAR such as the one in Equation (1). For instance, Cochrane (2017) says: "The definition of return means that only two of the three equations are needed, and the other one follows." With any two rows of coefficients, innovations, or data series, common practice is then to retrieve the omitted variable and associated coefficients from the CS identity and CS restrictions (4)-(9).[2] Campbell (2017, p.144) even states that one variable must be dropped, writing that "returns and dividend growth should not both be included in the system along with the log-price-dividend ratio, because the resulting system will have perfectly collinear variables." Avramov, Cederburg, and Lučivjanská (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 variance-covariance matrix $\boldsymbol{\Sigma}$ is nonsingular."

In fact, collinearity problems only appear when using more than one lag and in the absence of approximation error. To see this, notice that the identity links returns, dividend growth, and the price-dividend ratio with the *lagged* price-dividend ratio. With only one lag, it is clear that no explanatory variable in the VAR can be retrieved as a linear combination of the others, i.e., the matrix $\mathbf{z}_t'\mathbf{z}_t$ is full rank. For more than one lag, $\mathbf{z}_t'\mathbf{z}_t$ is still not singular because of the small approximation error or other sources of measurement error. Similarly, $\boldsymbol{\Sigma}$

---

[2]This practice is followed by most of the 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 (2008, 2011); Avramov et al. (2018). A noticeable exception is Campbell and Shiller (1988a), who find some weak evidence of persistence in dividend growth (see also Chen, Da, and Priestley, 2012). Larrain and Yogo (2008) 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 prevalent also in studies featuring additional predictors of excess returns (see, e.g., Campbell, 1991; Campbell and Ammer, 1993; Campbell, Chan, and Viceira, 2003; Campbell and Vuolteenaho, 2004).

is singular only in the absence of approximation error. Admittedly, since the approximate identity holds very closely in the data, the error is very small and 3-variable systems with more than one lag may be nearly collinear, with $\boldsymbol{\Sigma}$ close to rank deficient, and may experience numerical instabilities if estimated using ordinary least squares (OLS).

In the next section we show that the above concerns do not imply that one can omit one of the three variables from the VAR by invoking the CS identity. We will focus our exposition on the 3-variable VAR(1) in Equation (1), where collinearity is not a problem even with the CS identity holding exactly. However, collinearity will not be a concern either when using our Bayesian estimation methods (see, e.g., Leamer, 1973), which will allow for systems with any lag length, and with or without approximation error. Moreover, in Section 6 we develop inference methods that can also handle the possibility of a singular variance-covariance matrix.

3   Omitting Dividend Growth

We first show that, in general, one cannot drop one of the three variables in the VAR in Equation (1). Since returns are the ultimate variable of interest, the standard choice is to drop dividend growth and run a 2-variable VAR(1) on $\mathbf{x}'_{t+1} = [pd_{t+1}, r_{t+1}]$. Without loss of generality, in this section we abstract from the constant term. Let us partition the 3-variable VAR(1) in Equation (1), so as to isolate the vector $\mathbf{x}_{t+1}$:

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

where:

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

The last row of the 3-variable VAR(1) in Equation (11) implies that $\Delta d_t = (1 - \phi_{d,d}L)^{-1} [\phi_{21}\mathbf{x}_{t-1} + u_t^d]$ and the following representation for $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \left(\boldsymbol{\Phi}_{11} + \boldsymbol{\Phi}_{dd}\right)\mathbf{x}_t + \left(\boldsymbol{\phi}_{12}\boldsymbol{\phi}_{21} - \boldsymbol{\Phi}_{dd}\boldsymbol{\Phi}_{11}\right)\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1} - \boldsymbol{\Phi}_{dd}\boldsymbol{\xi}_t + \boldsymbol{\phi}_{12}u_t^d, \tag{12}$$

where $\boldsymbol{\Phi}_{dd} = \phi_{d,d}\mathbf{I}_2$, where $\mathbf{I}_m$ is an identity matrix of dimension m. Then, Equation (9) implies $u_t^d = [-\rho, 1]\boldsymbol{\xi}_t - \eta_t$. Therefore Equation (12) collapses to:

$$\mathbf{x}_{t+1} = \mathbf{G}_1\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1} + \mathbf{B}_1\boldsymbol{\xi}_t - \phi_{12}\eta_t, \tag{13}$$

where $\mathbb{E}_t\left(\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}_{t+1}'\right) = \boldsymbol{\Omega}_\xi$, $\mathbf{G}_1 = \boldsymbol{\Phi}_{11} + \boldsymbol{\Phi}_{dd}$, $\mathbf{G}_2 = \boldsymbol{\phi}_{12}\boldsymbol{\phi}_{21} - \boldsymbol{\Phi}_{dd}\boldsymbol{\Phi}_{11}$, and $\mathbf{B}_1 = \phi_{12}[-\rho, 1] - \boldsymbol{\Phi}_{dd}$. Using the results in Appendix A, Equation (13) implies the following VARMA(2,1) representation of $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{G}_1\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1\mathbf{e}_t, \tag{14}$$

where $\mathbb{E}_t\left(\mathbf{e}_{t+1}\mathbf{e}_{t+1}'\right) = \boldsymbol{\Omega}_e$ and we have defined $\boldsymbol{\eta}_{t+1} = \eta_{t+1}$, $\boldsymbol{\Omega}_\eta = \sigma_\eta^2$, $\mathbf{M}_0 = -\boldsymbol{\phi}_{12}$, and $\boldsymbol{\Omega}_{\xi\eta} = \begin{bmatrix} 0 & \sigma_\eta^2 \end{bmatrix}'$. In what follows we assume that the approximation error, η_{t+1}, is not correlated with u_{t+1}^{pd} or u_{t+1}^d. In this case we can write $\boldsymbol{\Omega}_{\xi\eta} = \begin{bmatrix} 0 & \sigma_\eta^2 \end{bmatrix}'$. Specify a 2-variable VAR(1) for $\mathbf{x}_{t+1}$:

$$\mathbf{x}_{t+1} = \mathbf{A}_1\mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1}, \tag{15}$$

where $\mathbb{E}_t\left(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}'\right) = \boldsymbol{\Omega}_\varepsilon$. Because $\mathbf{A}_1$ satisfies the following moment condition $\mathbb{E}\left[\left(\mathbf{x}_{t+1} - \mathbf{A}_1\mathbf{x}_t\right)\mathbf{x}_t'\right] = \mathbf{0}_{2\times2}$, we have that $\mathbf{G}_1\boldsymbol{\Gamma}_0 + \mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e - \mathbf{A}_1\boldsymbol{\Gamma}_0 = \mathbf{0}_{2\times2}$, where $\boldsymbol{\Gamma}_j$ is the j-th autocovariance of $\mathbf{x}_t$. Rearranging the solution above highlights the link between $\mathbf{A}_1$ and the parameters in the 3-variable VAR(1) in Equation (1), specifically:

$$\mathbf{A}_1 = \mathbf{G}_1 + \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)\boldsymbol{\Gamma}_0^{-1}. \tag{16}$$

Equation (16) highlights that, unless $\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e = \mathbf{0}_{2\times2}$, it is the case that $\mathbf{A}_1 \neq \mathbf{G}_1$.

11

But it should be clear that even then we have $\mathbf{A}_1 \neq \mathbf{\Phi}_{11}$; indeed $\phi_{d,d}$ always affects $\mathbf{A}_1$. This is another example of the argument in Cochrane (2008); the dynamics of dividend growth will affect our estimates of $\mathbf{A}_1$. Only in the case that $\phi_{pd,d} = \phi_{r,d} = 0$, and therefore $\phi_{d,d} = 0$, we have that $\mathbf{A}_1 = \mathbf{\Phi}_{11}$. The next theorem formalizes this claim.

Theorem 1. *The VARMA(2,1) in Equation* (14) *will have* $\mathbf{G}_1 = \mathbf{\Phi}_{11}$, $\mathbf{G}_2 = \mathbf{0}_{2\times2}$, *and* $\mathbf{D}_1 = \mathbf{0}_{2\times2}$ *if and only if* $\phi_{pd,d} = \phi_{r,d} = 0$, *with* $\phi_{d,d} = 0$ *following from CS restriction* (5).

Proof. See Appendix B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Theorem 1 implies that if we run the 2-variable VAR(1) in Equation (15), $\mathbf{A}_1 = \mathbf{\Phi}_{11}$ if and only if $\phi_{pd,d} = \phi_{r,d} = 0$. This additional restriction does not follow from the CS identity. Notice that were these extra restrictions true, the return identity would also imply that $\phi_{d,d} = 0$.[3] 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 (15). Doing so, while assuming $\mathbf{A}_1 = \mathbf{\Phi}_{11}$, amounts to imposing the additional restrictions $\phi_{pd,d} = \phi_{r,d} = 0$. These claims are formalized in the following corollary.

Corollary 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 (16) is that the innovations of the 2-variable VAR(1) are predictable with lagged information. In fact, it follows from Equations (15) and (16) that:

$$\boldsymbol{\varepsilon}_{t+1} = -\left(\mathbf{G}_2\mathbf{\Gamma}_1' + \mathbf{D}_1\mathbf{\Omega}_e\right)\mathbf{\Gamma}_0^{-1}\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1\mathbf{e}_t. \tag{17}$$

In Appendix C we show how to calculate the variance of $\boldsymbol{\varepsilon}_{t+1}$:

$$\Omega_\varepsilon = \mathbf{\Omega}_e + \mathbf{D}_1\mathbf{\Omega}_e\mathbf{D}_1' + \mathbf{G}_2\mathbf{\Gamma}_0\mathbf{G}_2' - \left(\mathbf{G}_2\mathbf{\Gamma}_1' + \mathbf{D}_1\mathbf{\Omega}_e\right)\mathbf{\Gamma}_0^{-1}\left(\mathbf{G}_2\mathbf{\Gamma}_1' + \mathbf{D}_1\mathbf{\Omega}_e\right)'. \tag{18}$$

---

[3]By a similar argument, one can show that a 2-variable VAR(1) on $[\Delta d_{t+1}, pd_{t+1}]'$ would impose the extra restrictions $\phi_{pd,r} = \phi_{d,r} = 0$, with $\phi_{r,r} = 0$ following from CS restriction (7).

With these ingredients we can write the following corollary:

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

Clearly, Corollary 2 implies that the innovations of the 2-variable VAR(1) in Equation (15) 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_{t+1}^d because of the approximation error. It is also useful to note an interesting special case of Theorem 1, by which it is possible to recover the return innovation, u_{t+1}^r:

Corollary 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_{t+1}^r if and only if $\phi_{r,d} = 0$.*

Corollary 3 implies that if there is no direct predictability from dividends to returns, both the one-step-ahead forecast and its prediction error coincide with those of the 3-variable VAR. Therefore, results that depend only on those quantities will be recovered correctly.

3.1 No Approximation Error

It is important to also notice that neither Theorem 1 nor Corollary 2 depend on the existence of approximation error. If we now consider the case with no approximation error, the VARMA(2,1) representation of $\mathbf{x}_{t+1}$ is $\mathbf{x}_{t+1} = \mathbf{G}_1\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1} + \mathbf{B}_1\boldsymbol{\xi}_t$; thus, $\mathbf{D}_1 = \mathbf{B}_1$ and $\mathbf{e}_{t+1} = \boldsymbol{\xi}_{t+1}$ and Equations (16)-(18) become $\mathbf{A}_1 = \mathbf{G}_1 + \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{B}_1\boldsymbol{\Omega}_\xi\right)\boldsymbol{\Gamma}_0^{-1}$, $\boldsymbol{\varepsilon}_{t+1} = -\left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{B}_1\boldsymbol{\Omega}_\xi\right)\boldsymbol{\Gamma}_0^{-1}\mathbf{x}_t + \mathbf{G}_2\mathbf{x}_{t-1} + \boldsymbol{\xi}_{t+1} + \mathbf{B}_1\boldsymbol{\xi}_t$, and $\boldsymbol{\Omega}_\varepsilon = \boldsymbol{\Omega}_\xi + \mathbf{B}_1\boldsymbol{\Omega}_\xi\mathbf{B}_1' + \mathbf{G}_2\boldsymbol{\Gamma}_0\mathbf{G}_2' - \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{B}_1\boldsymbol{\Omega}_\xi\right)\boldsymbol{\Gamma}_0^{-1}\left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{B}_1\boldsymbol{\Omega}_\xi\right)'$. Appendix B shows that Theorem 1 and Corollary 2 hold in the absence of approximation error.

4 Dividend Momentum

Section 3 emphasizes the econometric consequences of $\phi_{d,d} > 0$ for the practice of omitting dividend growth. In this section we highlight its economic implications. In particular, we

highlight that whenever $\phi_{d,d} > 0$ and $\phi_{r,pd} < 0$, a previously overlooked channel of return predictability arises, 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. We will finish this section by highlighting the consequences for portfolio choice.

Iterating forward the CS identity, applying expectations, and imposing the transversality condition $\lim_{T\to\infty} \rho^T \mathbb{E}_t pd_{t+T} = 0$, we obtain the result that 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:

$$r_{t+1} - \mathbb{E}_t r_{t+1} = \underbrace{(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=0}^{\infty} \rho^j\ \Delta d_{t+j+1}}_{NCF_{t+1}} - \underbrace{(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+j+1}}_{NDR_{t+1}} \tag{19}$$

An unexpected positive return today must come from positive revisions to the discounted sum of current and expected future dividend growth (news about cash flows, or NCF_{t+1}) or negative revisions to the discounted sum of expected future returns (news about discount rates, or NDR_{t+1}). NCF_{t+1} and NDR_{t+1} are useful to define mean reversion and momentum. On the one hand, there is mean reversion if a shock exists such that it causes a positive revision to current return and a negative revision to expected future returns. This implies that $r_{t+1} - \mathbb{E}_t r_{t+1}$ moves in the opposite direction to NDR_{t+1}. On the other hand, momentum arises whenever a shock exists such that it causes a positive unexpected return today and a positive revision in expected future returns. In other words, $r_{t+1} - \mathbb{E}_t r_{t+1}$ and NDR_{t+1} move in the same direction. Whenever momentum is associated with positive revisions to both current and expected future dividend growth, we have dividend-induced momentum in returns, which we call "dividend momentum."[4] Campbell and Vuolteenaho (2004) interpret surprises to NCF_{t+1}

---

[4]We therefore consider dividend momentum as a specific case of return momentum. Dividend momentum in returns is different than positive serial correlation in dividend growth. Indeed, dividend growth could be

and NDR_{t+1} approximately as permanent and transitory shocks to wealth, respectively. With dividend momentum this decomposition is not valid; news about discount rates is correlated with surprises about future expected dividend growth and accumulates over time.

As we highlight below, if dividends are persistent, in a way not fully captured by the lagged price-dividend ratio, (i.e., $\phi_{d,d} > 0$), dividend momentum will generally arise. This follows from the restrictions implied by the CS identity, even if the price-dividend ratio is the only variable that directly predicts returns.

To see this, consider the following simplified system:

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

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 we will show in Section 7. The omission of the third column is for expositional purposes only and does not affect any of the economic implications.[5] 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 $\text{Cov}\left(u_{t+1}^{pd}, u_{t+1}^d\right) = 0$. This assumption is not required but simplifies the exposition allowing an interpretation of u_{t+1}^{pd} and u_{t+1}^d 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 $\text{Cov}\left(u_{t+1}^r, u_{t+1}^d\right) = \text{Var}\left(u_{t+1}^d\right)$ and $\text{Cov}\left(u_{t+1}^r, u_{t+1}^{pd}\right) = \rho\text{Var}\left(u_{t+1}^{pd}\right)$.

---

persistent, but if that persistence were captured by the lagged price-dividend ratio, there would be no impact on NDR_{t+1} and therefore no dividend momentum in returns. The latter situation arises, for example, in the long-run risk model of Bansal and Yaron (2004), where dividend growth has a persistent component but is predicted by pd_t.

[5] Our estimates will show that $\phi_{d,r} > 0$, which strengthens the basic intuitions laid out in this section.

In this case the decomposition in Equation (19) 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 return 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.

A useful way to detect dividend momentum is to compute the IRFs to each of the two shocks. Figure 1 plots the discounted cumulative IRFs for returns and dividend growth for our simple model. 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}. If a variable were unpredictable, we would observe a perfectly flat IRF beyond the initial jump. After an initial positive impact, downward-sloping IRFs for returns are indicative of mean reversion, whereas upward-loping ones are a sign of momentum. If momentum is caused by a shock with an initial positive impact and an upward-sloping IRF for dividend growth, we will have dividend momentum.
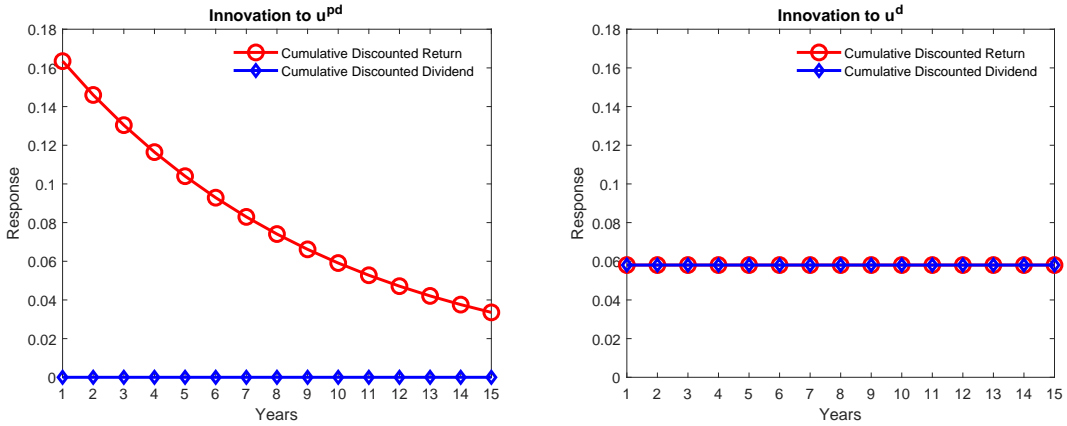
Panel (a) of the figure plots the case $\phi_{d,d} > 0$, whereas Panel (b) has $\phi_{d,d} = 0$. For a price-dividend ratio shock, the IRFs are the same in both panels: returns jump on impact but have a negative slope thereafter. This is the classic mean reversion effect in returns,

16

Figure 1: Dividend Momentum and IRFs

(a) Dividend Momentum ($\phi_{d,d} = 0.4$)
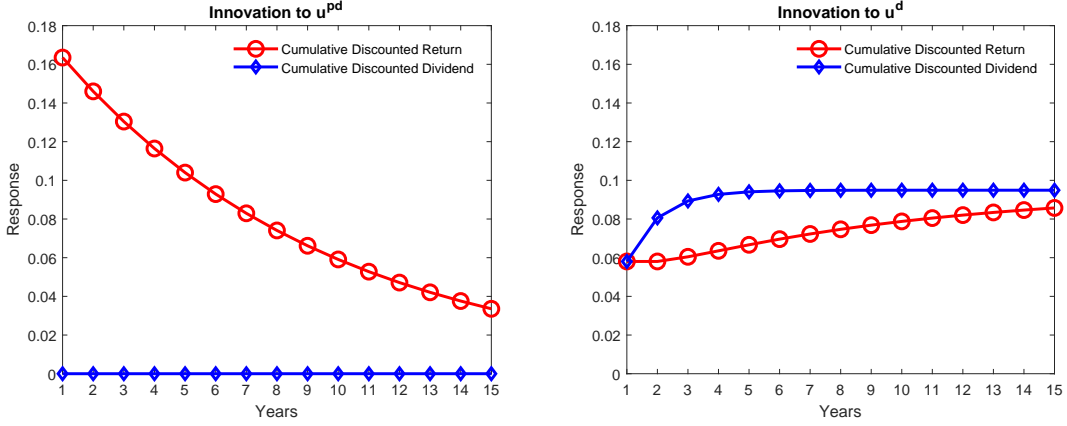
(b) No Dividend Momentum ($\phi_{d,d} = 0$)

Note: Panel (a): $\phi_{d,d} = 0.4$. Panel (b): $\phi_{d,d} = 0$. In addition, we use the following numerical values $\phi_{pd,pd} = 0.92$, $\mathtt{Var}\left(u_{t+1}^d\right) = 0.003$, and $\mathtt{Var}\left(u_{t+1}^{pd}\right) = 0.028$, which are justified from the results in Section 5. For the rest of the paper, we will use $\rho = 0.971$.

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, both returns and dividend growth initially jump by the same amount regardless of the value of $\phi_{d,d}$. However, the value of $\phi_{d,d}$

is relevant beyond impact. If $\phi_{d,d} = 0$ is zero, 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.

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:

$$\text{Var}(u_{t+1}^r) = \underbrace{(1 + \Psi)^2 \text{Var}(u_{t+1}^d)}_{\text{Var}(NCF_{t+1})} + \underbrace{\rho^2 \text{Var}(u_{t+1}^{pd}) + (\Psi)^2 \text{Var}(u_{t+1}^d)}_{\text{Var}(NDR_{t+1})} - \underbrace{2(1 + \Psi)\Psi \text{Var}(u_{t+1}^d)}_{2\text{Cov}(NCF_{t+1}, NDR_{t+1})}.$$

When $\phi_{d,d} > 0$, dividend momentum implies that an increase in u_{t+1}^d leads to a positive revision of current and expected future cash flows and discount rates. As a consequence, $\text{Var}(NCF_{t+1}) > \text{Var}(u_{t+1}^d)$, $\text{Var}(NDR_{t+1}) > \rho^2\text{Var}(u_{t+1}^{pd})$, and $\text{Corr}(NCF_{t+1}, NDR_{t+1}) = \rho\phi_{d,d}\sqrt{\text{Var}(NCF_{t+1})/\text{Var}(NDR_{t+1})} > 0$. 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.

Finally, it is important to note that with correlated NCF_{t+1} and NDR_{t+1}, one may also need to separate the orthogonal shocks that may be simultaneously affecting both components. In our simple model above, these naturally correspond to the uncorrelated u_{t+1}^{pd} and u_{t+1}^d innovations. In a more empirically relevant case in which u_{t+1}^{pd} and u_{t+1}^d are correlated, the innovations need to be orthogonalized to obtain the price-dividend ratio and dividend growth shocks. While there may be many valid orthogonalizations, a particularly simple and intuitive option is to consider two uncorrelated shocks with the following properties: a shock that explains all of the unexpected variance of the price-dividend ratio at horizon zero and another shock, orthogonal to the first one, which does not move the price-dividend ratio on impact, and together with the first, explains the entirety of the unexpected variance of dividend

18

growth on impact. These can be retrieved by means of a Cholesky decomposition in which variables are ordered $[pd_{t+1}, \Delta d_{t+1}, r_{t+1}]$, and it is easy to see that in the simple model above we would retrieve the innovations u_{t+1}^{pd} and u_{t+1}^{d}.[6,7]

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

$$r_{t,t+k} - \mathbb{E}_t r_{t,t+k} = \sum_{j=1}^{k} u_{t+j}^{r} + \sum_{j=1}^{k-1} \left[a_j u_{t+k-j}^{pd} + b_j u_{t+k-j}^{d} \right], \tag{20}$$

where $a_j = \frac{1-\phi_{pd,pd}^{j}}{1-\phi_{pd,pd}} \left(\rho \phi_{pd,pd} - 1 \right)$, $b_j = -\frac{\phi_{d,d}}{\rho} \frac{\left(\rho \phi_{pd,pd} - 1 \right)}{\left(\phi_{d,d} - \phi_{pd,pd} \right)} \left[\frac{\left(1 - \phi_{d,d}^{j} \right)}{1-\phi_{d,d}} - \frac{\left(1 - \phi_{pd,pd}^{j} \right)}{1-\phi_{pd,pd}} \right]$, and $u_{t+1}^{r} = u_{t+1}^{d} + \rho u_{t+1}^{pd}$ 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, as well as 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 $\phi_{d,d} > 0$, shocks to dividend growth affect the risk for the long-run investor 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

---

[6]A third shock, which affects only contemporaneous returns, would correspond to the approximation error η_{t+1}, and would have zero variance in the case in which the CS identity holds exactly.

[7]This is not the only orthogonalization possible. In fact, Campbell et al. (2013) use an alternative Cholesky decomposition with return ordered first and the price-earnings ratio second. In this case, the first shock would explain all of the unexpected variance of returns at horizon zero, while the second would affect the price-dividend ratio on impact and both would explain the entirety of the unexpected variance of returns on impact.

into three terms:[8]

$$\text{Var}\left(r_{t,t+k}\right) = k\text{Var}\left(u_{t+1}^r\right) + \sum_{j=1}^{k-1}\left\{a_j^2\text{Var}\left(u_{t+1}^{pd}\right) + b_j^2\text{Var}\left(u_{t+1}^d\right)\right\} \qquad (21)$$
$$+ 2\sum_{j=1}^{k-1}\left\{a_j\text{Cov}\left(u_{t+1}^r, u_{t+1}^{pd}\right) + b_j\text{Cov}\left(u_{t+1}^r, u_{t+1}^d\right)\right\},$$

The first term in Equation (21) reflects the uncertainty coming from the innovations to one-period-ahead returns, often labelled 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 $\text{Var}_t\left(r_{t,t+k}\right) < k\text{Var}\left(u_{t+1}^r\right)$, which is only possible if the last component is negative. In fact, if $0 < \phi_{pd,pd} < 1$ we have that $a_j < 0 \forall j$, and since $\text{Cov}\left(u_{t+1}^r, u_{t+1}^{pd}\right) = \rho\text{Var}\left(u_{t+1}^{pd}\right)$ 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 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 $b_j > 0 \forall j$, and since $\text{Cov}\left(u_{t+1}^r, u_{t+1}^d\right) = \text{Var}\left(u_{t+1}^d\right)$, 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. Therefore, the presence of dividend momentum increases the variance of long-run returns and generates a negative hedging demand for long-run investors. This fact will be connected to the results on

---

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

portfolio choice in Section .

5 A First Look at the Data

In this section we take a first look at the data using the 3-variable model in Section 2 and using flat priors. We will then document the presence of dividend momentum in the data. We will highlight the drawbacks of using flat priors and the need to use informative priors. We close the section by highlighting that standard techniques to draw informative priors cannot be used to impose the CS restrictions.

5.1 Bayesian Estimation and Data

In general, any VAR model can be written in matrix form as $\mathbf{Y} = \mathbf{Z}\boldsymbol{\Phi}' + \mathbf{U}$. Denoting T as 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. The vector of innovations $\mathbf{u}_t$ is assumed to be independently and identically distributed $\mathcal{N}(0, \boldsymbol{\Sigma})$. The NIW family of distributions is conjugate for this class of models.[9] 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 $\overline{\boldsymbol{\alpha}} = \text{vec}(\overline{\mathbf{A}})$, $\overline{\mathbf{V}} = (\underline{\mathbf{V}}^{-1} + \mathbf{Z}'\mathbf{Z})^{-1}$, $\overline{\mathbf{A}} = \overline{\mathbf{V}}(\underline{\mathbf{V}}^{-1}\underline{\mathbf{A}} + \mathbf{Z}'\mathbf{Z}\hat{\mathbf{A}})^{-1}$, $\hat{\mathbf{A}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$, and $\overline{\mathbf{S}} = \hat{\mathbf{S}} + \underline{\mathbf{S}} + \hat{\mathbf{A}}'\mathbf{Z}'\mathbf{Z}\hat{\mathbf{A}} + \underline{\mathbf{A}}'\underline{\mathbf{V}}^{-1}\underline{\mathbf{A}} - \overline{\mathbf{A}}'\overline{\mathbf{V}}^{-1}\overline{\mathbf{A}}$, $\hat{\mathbf{S}} = (\mathbf{Y} - \mathbf{Z}\hat{\mathbf{A}})'(\mathbf{Y} - \mathbf{Z}\hat{\mathbf{A}})$, and $\overline{\nu} = T + \underline{\nu}$. The conjugate prior is convenient for its analytical tractability and is amenable to efficient

---

[9]We denote a normal distribution with mean $\boldsymbol{\zeta}$ and variance-covariance matrix $\boldsymbol{\Upsilon}$ by $\mathcal{N}(\boldsymbol{\zeta}, \boldsymbol{\Upsilon})$ and denote its density evaluated at $\boldsymbol{\beta}$ by $\mathcal{N}_{(\boldsymbol{\zeta}, \boldsymbol{\Upsilon})}(\boldsymbol{\beta})$. We denote the inverse-Wishart distribution with parameters $\boldsymbol{\Gamma}$ and π by $\mathcal{IW}(\boldsymbol{\Gamma}, \pi)$ and denote its density evaluated at $\boldsymbol{\iota}$ by $\mathcal{IW}_{(\boldsymbol{\Gamma}, \pi)}(\boldsymbol{\iota})$. A normal-inverse-Wishart distribution is characterized by four parameters: a scalar $\nu \geq n$, an $n \times n$ SPD matrix $\mathbf{S}$, an $Kn \times 1$ vector $\boldsymbol{\Psi}$, and an $K \times K$ SPD matrix $\boldsymbol{\Omega}$. We denote this distribution by $NIW(\nu, \mathbf{S}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$ and its density by $NIW_{(\nu, \mathbf{S}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$. Furthermore,

$$NIW_{(\nu, \mathbf{S}, \boldsymbol{\Psi}, \boldsymbol{\Omega})}(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) \propto \underbrace{|\det(\boldsymbol{\Sigma})|^{-\frac{\nu+n+1}{2}} e^{-\frac{1}{2}\text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})}}_{\text{inverse-Wishart}} \underbrace{|\det(\boldsymbol{\Sigma})|^{-\frac{K}{2}} e^{-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\Psi})'(\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega})^{-1}(\boldsymbol{\alpha}-\boldsymbol{\Psi})}}_{\text{conditionally normal}}.$$

sampling. We will use this prior to illustrate our methods, but they can be adapted to any posterior distribution by means of an importance sampling algorithm. The NIW posterior distributions defined above can be factored into the following conditional and marginal posterior distributions $\mathcal{N}\left(\overline{\boldsymbol{\alpha}}, \boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)$ and $p(\boldsymbol{\Sigma}|\mathbf{y}) \sim \mathcal{IW}\left(\overline{\mathbf{S}}, \overline{\nu}\right)$. This structure allows us to independently draw from the posterior.

We use data for the log dividend growth, the log price-dividend ratio, and log returns. Our data choices closely follow Van Binsbergen and Koijen (2010) and Koijen and Nieuwerburgh (2012). We use postwar annual data between 1947 and 2018 for the S&P 500 index.[10] Because dividend payments are known to be highly seasonal, we focus on annual data so as to ensure that any dividend growth autocorrelation we find is not simply driven by seasonal patterns. A more important issue is how to treat the reinvestment of dividends. The annual series traditionally used in the literature implicitly measure dividends after reinvestment at 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.[11] 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 $T = 72$.

5.2 The Flat Prior and Dividend Momentum

Table 1 reports the the posterior of $\boldsymbol{\mu}$, $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Sigma}$ in Equation (1) under a flat prior. In this case the posterior means are centered around the OLS estimates and the Bayesian high posterior density intervals coincide with the classical confidence intervals. The most important message is that the posterior mean of the coefficient $\phi_{d,d}$ is 0.41. This is a large number

---

[10]Results using the CRSP market return are very similar and available upon request.
[11]Results using dividends reinvested at the risk-free rate are very similar and available upon request.

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}	0.057	0.414	0.005	0.139
	$[0.035, 0.072]$	$[0.310, 0.461]$	$[-0.010, 0.012]$	$[0.098, 0.158]$
pd_{t+1}	3.524	-0.421	0.906	-0.171
	$[3.402, 4.087]$	$[-0.730, -0.279]$	$[0.861, 0.928]$	$[-0.295, -0.112]$
r_{t+1}	0.090	0.018	-0.115	-0.026
	$[0.053, 0.104]$	$[-0.270, 0.152]$	$[-0.157, -0.096]$	$[-0.142, 0.030]$
		$\boldsymbol{\Sigma}$ (corr/std)		
		u^d_{t+1}	u^{pd}_{t+1}	u^r_{t+1}
u^d_{t+1}		0.054		
		$[0.050, 0.059]$		
u^{pd}_{t+1}		-0.316	0.166	
		$[-0.422, -0.209]$	$[0.152, 0.180]$	
u^r_{t+1}		0.023	0.939	0.153
		$[-0.094, 0.139]$	$[0.926, 0.953]$	$[0.139, 0.166]$

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 68th 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 variance-covariance matrix.

in economic terms. For comparison, annual US real GDP in the same period displays an autocorrelation coefficient of 0.14. Almost all of the posterior distribution is in the positive region, with the 5th percentile at 0.31. Dividend growth is clearly persistent, even after controlling for the price-dividend ratio and returns; hence, the common approach of dropping dividend growth from the VAR is not justified and will have important economic consequences as we will see when analyzing the IRFs. 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. Consistent with CS restriction (5), lagged dividends

negatively predict the subsequent price-dividend ratio; i.e., $\phi_{pd,d} < 0$. Therefore, even though dividend growth is not a useful predictor of one-period-ahead returns, it affects expected returns over longer horizons. In line with Cochrane's (2008) conclusions, we also find that the price-dividend ratio is highly persistent, and that $\phi_{r,pd} < 0$, whereas $\phi_{d,pd}$ is approximately zero. Interestingly, we find that $\phi_{d,r} > 0$ and significant (with a posterior mean value of 0.14), meaning there is some predictability from returns to dividends. There is, however, little evidence of serial correlation 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). The combination of $\phi_{d,r} > 0$ and $\phi_{r,r} = 0$ strengthens the channel of dividend momentum, as a dividend growth shock will lead to higher returns, which in turn feed back into dividend growth.

Panel (a) of Figure 2 reproduces the IRFs in Panel (a) of Figure 1 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 by means of 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 of 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 2 displays the results. The price-dividend ratio shock looks

Figure 2: Impulse Response Functions Under Flat Priors

(a) 3-variable VAR
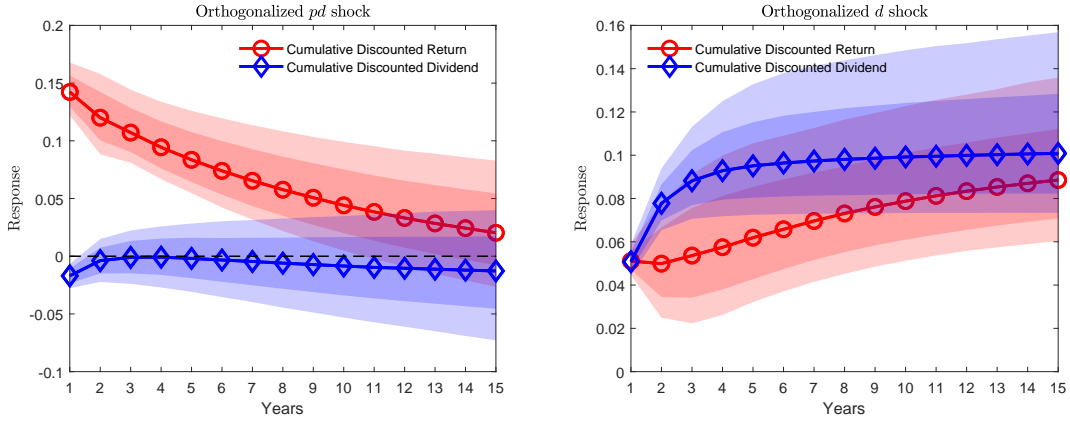
(b) 2-variable VAR omitting dividend growth
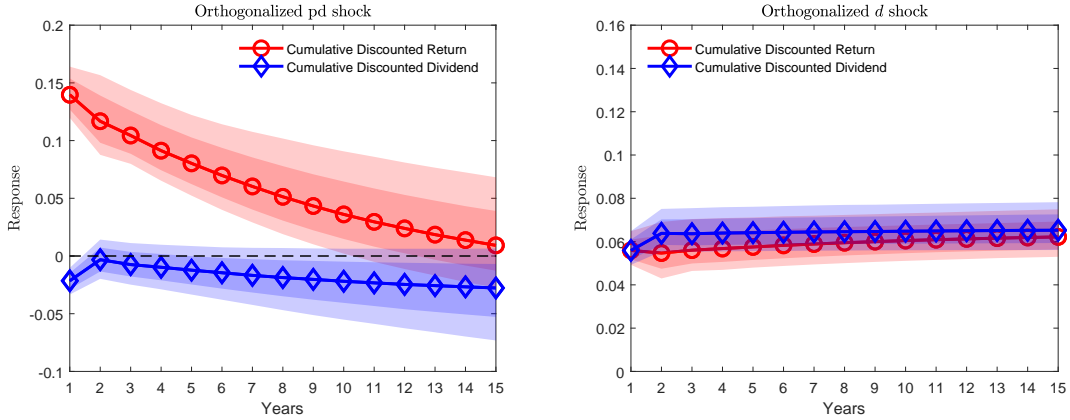
Note: The solid lines represent the median posterior response. The darker shadow area represents the 68th posterior credible intervals, while the lighter shadow area represents the 95th posterior credible intervals.

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 will not find dividend momentum and leads to similar results as in Panel (b) of Figure 1. This result highlights the conclusion of Section 3: dropping dividend growth effectively imposes $\phi_{d,d} = 0$ and arbitrarily rules out dividend momentum.

Table 2 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

25

Table 2: SHOCK CONTRIBUTION TO NCF_{t+1} AND NDR_{t+1} UNDER FLAT PRIORS

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

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

both NCF_{t+1} and NDR_{t+1}, and their correlation is meaningfully larger.

5.3   Drawbacks of Flat Priors and Informative Priors

It is known that inference under flat priors in VARs is problematic. OLS estimates of VAR parameters are plagued with finite-sample bias that may seriously distort inference when the model contains variables that are highly persistent (see, e.g., Bekaert, Hodrick, and Marshall, 1997).

It is well documented that high degrees of a priori return predictability implies poor out-of-sample performance, a problem that will worsen with additional predictors and lags. From an economic point of view, parameter combinations that imply very high degrees of return predictability should be a priori implausible (see 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). As Figure 3 shows, a flat prior over the VAR coefficients implies a prior distribution over the one-period-ahead R^2 of returns that is heavily concentrated around high values: both the 68th and the 90th percentiles are above 99 percent.

The high persistence of the price-dividend ratio is also concerning. The presence of

Figure 3: One-Period-Ahead Return Equation R-squared
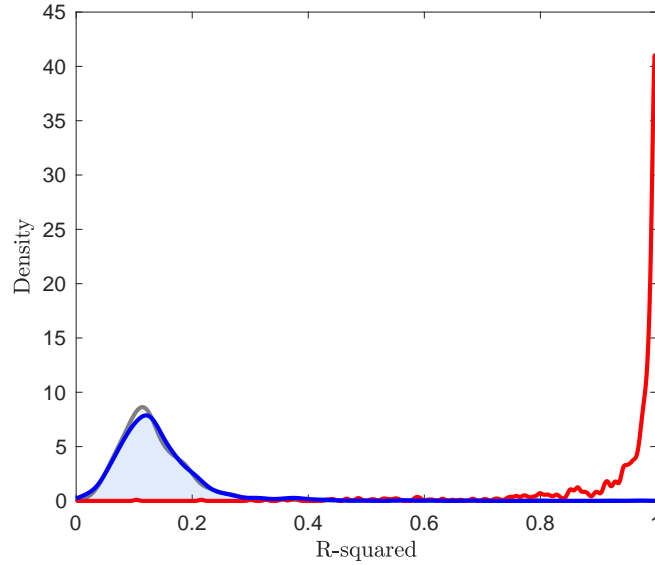
Note: Red represents the prior, grey the likelihood, and blue the posterior. The R^2 for the flat prior is computed by simulation.

downward bias in OLS estimates of autoregressive parameters when these are close to unit root has been known since Hurwicz (1950). 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 are biased away from zero. So not only $\phi_{pd,pd}$ is 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 that are below one for $\phi_{pd,pd}$, and therefore on high amounts of predictability. The bias and excess predictability problems will also worsen as additional persistent predictors are added.[12]

Finally, the flat priors combined with the standard use of the conditional likelihood also

---

[12]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 3.

imply that the initial values of the data are far away from their unconditional mean. This can be seen in Table 1, where the posterior mean of the unconditional mean of the price-dividend ratio is estimated to be much higher than the initial value in the data; in particular the posterior mean equals 3.5 (with 95 percent of the posterior above 3.4) while the initial data is 2.8. This implies an implausibly good forecasting power of initial conditions and determinist components (see, e.g., Sims and Uhlig, 1991; Sims, 2000; Jarocinski and Marcet, 2015; Giannone, Lenza, and Primiceri, 2019). In practice, VAR deterministic components over-fit the low-frequency variation in the data, a problem that again gets worse as lags or additional variables are included in the system.

To solve these issues one can consider informative priors. We will use a class of priors for VARs originally proposed by Doan, Litterman, and Sims (1984), commonly known in the macroeconometrics literature as "Minnesota" priors, to handle downward bias and excess predictability. In particular:

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

where in the simplest case $\boldsymbol{\Omega} = \lambda^2 (diag\left(\left[\underline{\sigma}_d^2, \underline{\sigma}_{pd}^2, \underline{\sigma}_r^2\right]\right))^{-1}$, 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.[13] As desired, the Minnesota prior pushes $\phi_{pd,pd}$ toward one, and both $\phi_{r,pd}$ and $\phi_{d,d}$ toward zero. For $\boldsymbol{\Phi}_0$, the Minnesota prior is usually specified as flat.

We combine the Minnesota prior with the Single Unit Root prior proposed by Sims (1993) and Sims and Zha (1998) to address the problem of the excessive explanatory power of initial conditions and deterministic components. A scalar hyperparameter θ controls the tightness of the prior. For the prior mean nominal return, $\underline{\mu}_r$ we chose a value of 10.5 percent, consistent with a 4 percent risk-free rate and a 6.5 percent equity risk premium, and for

---

[13]We follow the common practice of setting $\underline{\sigma}_i^2$ to the residual variance of an AR(1) model. In particular, $\underline{\sigma}_d^2 = 0.0034$, $\underline{\sigma}_{pd}^2 = 0.0284$, and $\underline{\sigma}_r^2 = 0.0254$

the prior 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 prior mean 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 prior hyperparameters $\lambda = 0.17$ and $\theta = 0.05$ are chosen to maximize the value of the marginal likelihood, as proposed by Giannone, Lenza, and Primiceri (2019). Finally, the prior for Σ is set to $p(\Sigma) \sim \mathcal{IW}(diag([\underline{\sigma}_1^2, \ldots, \underline{\sigma}_n^2]), n + 2)$.

In line with the suggestion of Wachter and Warusawitharana (2009), this informative prior lowers the return equation R^2 at the one-period horizon. The prior median for the one-period-ahead R^2 is 12 percent for the stock return equation, with a 68th percentile of 15 percent and a 90th percentile of 20%.

5.4 Informative Priors and the CS Restrictions

A problem with general prior distributions such as the one above is that they do not satisfy the CS restrictions. One could write a prior that is centered around the CS restrictions, but even if the prior mean satisfies the CS restrictions, the prior puts positive probability on parameter combinations that violate them. Although the likelihood may closely satisfy the CS restrictions, the more informative the prior is, the more likely it is that the posterior distribution violates the restrictions.

Figure 4 illustrates this point for the informative prior described above. Each column of Figure 4 corresponds to one of the CS restrictions (4)-(9). The first row plots draws of the likelihood, the second row plots draws from the prior, and the third row plots draws from the posterior. For instance, in the first panel of Figure 4, draws of c^r are plotted against the value of $\rho c^{pd} + c^d$ of the same likelihood draw. If the CS restrictions were to be satisfied, all points should be aligned along the 45-degree line.

As we can see from the first row, the likelihood respects the CS restrictions very closely.

Figure 4: CS Restrictions for $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ under informative priors
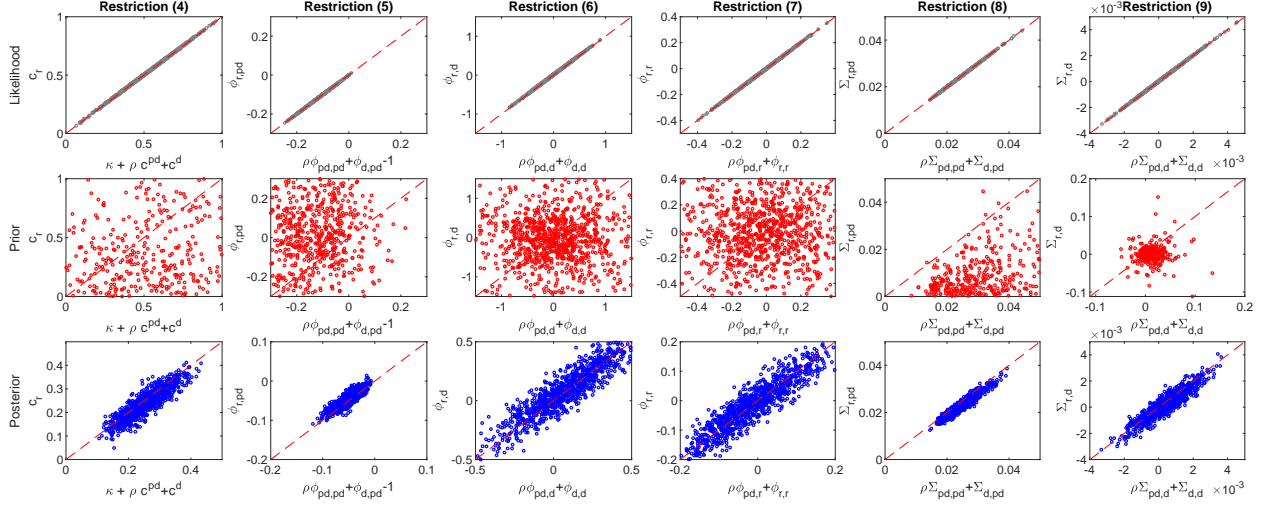
Note: Red represents the prior, grey the likelihood, and blue the posterior.

However, since the prior draws are not restricted to satisfing the CS restrictions, it is not surprising that the resulting posterior draws do not satisfy them either. In the next section we derive a method to draw from any posterior conditional on the CS restrictions on $\mathbf{\Phi}$ and $\mathbf{\Sigma}$.

6   Bayesian Estimation and the CS Restrictions

In this section we present a general methodology for drawing independently from any posterior distribution of a macro-finance VAR conditional on the CS restrictions on $\mathbf{\Phi}$ and $\mathbf{\Sigma}$. We will write our algorithm as independently drawing from the conjugate family of NIW posterior distributions conditional on the CS restrictions. Conjugacy and independent drawing are particularly useful in the Bayesian paradigm, as they open the possibility of estimating models with a large number of variables and lags. However, our techniques are not limited to the NIW family and can be applied to any prior. So far, we have considered a setting in which the VAR contains only one asset (stock returns) and one associated CS identity. One

could consider a model with k assets, such as the one of Campbell, Chan, and Viceira (2003), in which case there would be k associated CS identities. For the sake of generality we write our algorithms below to allow for this possibility.

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

6.1 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 $\mathbf{R}_{\boldsymbol{\Phi}} \boldsymbol{\alpha} = \mathbf{r}_{\boldsymbol{\Phi}}$. Thus, to draw from the restricted normal posterior distribution of $\boldsymbol{\alpha}$ conditional on $\boldsymbol{\Sigma}$ we should draw from $\mathcal{N}\left(\tilde{\boldsymbol{\alpha}}, \widetilde{\mathbf{V}}\right)$ where $\tilde{\boldsymbol{\alpha}} = \overline{\boldsymbol{\alpha}} - \mathbf{F}\left(\mathbf{R}_{\boldsymbol{\Phi}} \overline{\boldsymbol{\alpha}} - \mathbf{r}_{\boldsymbol{\Phi}}\right)$, $\widetilde{\mathbf{V}} = \left(\mathbf{I}_{nK} - \mathbf{F}\mathbf{R}_{\boldsymbol{\Phi}}\right)\left(\boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)\left(\mathbf{I}_{nK} - \mathbf{F}\mathbf{R}_{\boldsymbol{\Phi}}\right)'$, and $\mathbf{F} = \left(\boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)\mathbf{R}_{\boldsymbol{\Phi}}'\left(\mathbf{R}_{\boldsymbol{\Phi}}\left(\boldsymbol{\Sigma} \otimes \overline{\mathbf{V}}\right)\mathbf{R}_{\boldsymbol{\Phi}}'\right)^{-1}$. If $\mathcal{U}_{\boldsymbol{\Phi}}$ 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 $\mathcal{U}_{\boldsymbol{\Phi}}$. As mentioned in Section 2, the stationarity restriction (10) requires that the system is ultimately stationary, even if the price-dividend ratio is highly persistent as argued by Cochrane (2008). We implement this restriction by discarding draws of the posterior that do not satisfy the stationary restriction. This truncates and re-normalizes the restricted normal posterior.

6.2 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, will use an importance sampler to accomplish our objective.

The CS restrictions on the $n \times 1$ vector of innovations, $\mathbf{u}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$, can be represented as k linear restrictions and $n - k$ orthogonality restrictions. The k linear restrictions are:

$$\mathbf{L}\mathbf{u}_t = \boldsymbol{\eta}_t \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}\right) \tag{22}$$

where $\mathbf{L}$ is a given $k \times n$ matrix. Equation (22) 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, $k = 1$ and the CS restriction on the innovations represented by Equation (3) can be mapped into Equation (22) by specifying $\mathbf{L} = [-1, -\rho, 1]$.

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

$$\mathbb{E}\left(\boldsymbol{\Xi}\mathbf{u}_t\boldsymbol{\eta}_t'\right) = \mathbf{0}_{(n-k)\times k} \tag{23}$$

where $\boldsymbol{\Xi}$ is a given $(n - k) \times n$ selection matrix. For the case considered in Section 2 we have that $\boldsymbol{\Xi} = [\mathbf{I}_2, \mathbf{0}_{2\times 1}]$. Putting Equations (22) and (23) together, we obtain the result that the CS restrictions on $\boldsymbol{\Sigma}$ can be represented as:

$$\boldsymbol{\Xi}\mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right)\mathbf{L}' = \boldsymbol{\Xi}\boldsymbol{\Sigma}\mathbf{L}' = \mathbf{0}_{(n-k)\times k}. \tag{24}$$

Vectorizing Equation (24) implies the following linear restrictions:

$$\mathbf{R}_\Sigma \operatorname{vec}(\boldsymbol{\Sigma}) = \mathbf{0}_{(n-k)k \times 1} \tag{25}$$

where $\mathbf{R}_\Sigma = \mathbf{L} \otimes \boldsymbol{\Xi}$. Equation (25) appropriately imposes the CS restrictions on $\boldsymbol{\Sigma}$. For instance, for the case considered in Section 2, these are represented by Equations (8) and (9).

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

$$\mathbf{W} = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}', \tag{26}$$

where:

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

Using the above mapping, one can show that the CS restrictions on $\boldsymbol{\Sigma}$ hold if and only if $\mathbf{W}$ is block diagonal. To see this, notice that on the one hand, the mapping implies that $\mathbf{W}_{12} = \boldsymbol{\Xi}\boldsymbol{\Sigma}\mathbf{L}'$. Hence, if Equation (25) holds, it is the case that $\mathbf{W}_{12} = \mathbf{0}_{(n-k)\times k}$. On the other hand, the inverse mapping implies that $\mathbf{R}_\Sigma \operatorname{vec}(\boldsymbol{\Sigma}) = \mathbf{R}_\Sigma (\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}) \operatorname{vec}(\mathbf{W})$. Then one can show that:

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

and $\mathbf{R}_\Sigma (\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}) \operatorname{vec}(\mathbf{W}) = \operatorname{vec}(\mathbf{W}_{12})$. Thus, if $\mathbf{W}_{12} = \mathbf{0}_{(n-k)\times k}$, it is the case that Equation (25) holds.

It is the result above that is key to being able to make independent draws from the set of all $\boldsymbol{\Sigma}$ satisfying the CS restrictions using the following algorithm.

Algorithm 1. *The following makes independent draws from a distribution over $\boldsymbol{\Sigma}$ conditional*

33

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

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 $\mathrm{vec}(\mathbf{\Sigma})$, 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 for the importance sampler that we derive to draw from the desired distribution over $\mathbf{\Sigma}$. We will call $\pi_{(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}})}(\mathbf{\Sigma}) = \mathcal{IW}_{(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}})}(\mathbf{W}_{11})\,\mathcal{IW}_{(\mathbf{S}_{\mathbf{\Omega}}, \nu_{\mathbf{\Omega}})}(\mathbf{\Omega})$, where $\mathbf{W}_{11} = \mathbf{\Xi}\mathbf{\Sigma}\mathbf{\Xi}'$ and $\mathbf{\Omega} = \mathbf{L}\mathbf{\Sigma}\mathbf{L}'$, 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 $\mathbf{S}_{\mathbf{W}_{11}} = \mathbf{\Xi}\mathbf{S}\mathbf{\Xi}'$ and $\mathbf{S}_{\mathbf{\Omega}} = \mathbf{L}\mathbf{S}\mathbf{L}'$, while one could choose $\nu_{\mathbf{W}_{11}} = \nu_{\mathbf{\Omega}} = \nu$.

6.2.1   No Approximation Error

It is relatively easy to adapt Algorithm 1 for the case where there is no approximation error. This is the case when $\mathbf{\Omega} = \mathbf{0}_{k\times k}$ in Equation (22). 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, considering the fact that the variance-covariance

matrix is symmetric, leads to $\frac{(k+1)k}{2}$ additional restrictions:

$$\widetilde{\mathbf{R}}_\Sigma \operatorname{vec}(\mathbf{\Sigma}) = \mathbf{0}_{\frac{(k+1)k}{2}} \tag{27}$$

where $\widetilde{\mathbf{R}}_\Sigma = \mathbf{D}_k^+ (\mathbf{L} \otimes \mathbf{L})$ 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 (26), one can show that the CS restrictions on $\mathbf{\Sigma}$ hold if and only if $\mathbf{W}$ has the following form:

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

when the CS restrictions on $\mathbf{\Sigma}$ hold. To see this, notice that on the one hand, the mapping implies that $\mathbf{W}_{12} = \mathbf{\Xi\Sigma L'}$ and $\mathbf{W}_{22} = \mathbf{L\Sigma L'}$. Hence, if Equations (25) and (27) 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 $\left[\mathbf{R}_\Sigma', \widetilde{\mathbf{R}}_\Sigma'\right]' \operatorname{vec}(\mathbf{\Sigma}) = \left[\mathbf{R}_\Sigma', \widetilde{\mathbf{R}}_\Sigma'\right]' (\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}) \operatorname{vec}(\mathbf{W})$. It is easy to show that:

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

therefore $\left[\mathbf{R}_\Sigma', \widetilde{\mathbf{R}}_\Sigma'\right]' (\mathbf{H}^{-1} \otimes \mathbf{H}^{-1}) \operatorname{vec}(\mathbf{W}) = \left[\operatorname{vec}(\mathbf{W}_{12})', \operatorname{vech}(\mathbf{W}_{22})'\right]'$. 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 (25) and (27) 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 $\mathbf{\Omega} = \mathbf{0}_{k\times k}$. In this case, we will call $\pi_{(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}})}(\mathbf{\Sigma}) = \mathcal{IW}_{(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}})}(\mathbf{W}_{11})$, where $\mathbf{W}_{11} = \mathbf{\Xi\Sigma\Xi'}$ 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 $\mathbf{S}_{\mathbf{W}_{11}} = \mathbf{\Xi S\Xi'}$, while one could choose $\nu_{\mathbf{W}_{11}} = \nu$.

6.2.2 An Importance Sampler

Since our objective is to independently draw from the $\mathcal{IW}(\mathbf{S}, \nu)$ conditional on the CS restrictions on $\mathbf{\Sigma}$, the aforementioned results 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 $\pi\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\Omega}, \nu_{\Omega}\right)$.*

2. *Set its importance weight to*

$$\frac{\mathcal{IW}_{(\mathbf{S}, \nu)}(\mathbf{\Sigma})}{\pi_{\left(\mathbf{S}_{\mathbf{W}_{11}}, \nu_{\mathbf{W}_{11}}, \mathbf{S}_{\Omega}, \nu_{\Omega}\right)}(\mathbf{\Sigma})}.$$

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 just 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 just 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.

6.3 Drawing Any Posterior

Results in Sections 6.1 and 6.2 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 $\boldsymbol{\Phi}$ and $\mathbf{\Sigma}$.*

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

36

2. Use $\mathcal{N}\left(\tilde{\boldsymbol{\alpha}}, \widetilde{\mathbf{V}}\right)$ to 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.

As mentioned several times already, Algorithm 3 can be modified to independently draw from any desired posterior distribution conditional on the CS restrictions on $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$. When that is the case, one will need to add an importance sampling step where the restricted NIW posterior of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ is the proposal and the desired posterior distribution conditional on the CS restrictions on $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ is the target.

7   In-sample Results

In this section we present the results using the restricted NIW prior of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ in the 3-variable model described in Section 2. We parameterize the restricted NIW prior as in Section 5.3.[14] We call such a prior the restricted informative prior. Along the same lines, we call the implied restricted NIW posterior of $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ the restricted informative posterior. After analyzing the restricted posterior, we will focus on return predictability, dividend momentum, and 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.

7.1   The Restricted Posterior

Figure 5 reports the restricted informative prior, the likelihood, and the restricted informative posterior for $\boldsymbol{\Phi_1}$. The first thing to notice about the posterior distributions is that they

---

[14]We use the same values for λ and θ as in Section 5.3. Ideally one would want to calculate the marginal likelihood for the restricted model but an analytical solution is not available. We leave this for future research.

Figure 5: BAYESIAN RESTRICTED PRIOR AND POSTERIOR OF $\mathbf{\Phi}_1$
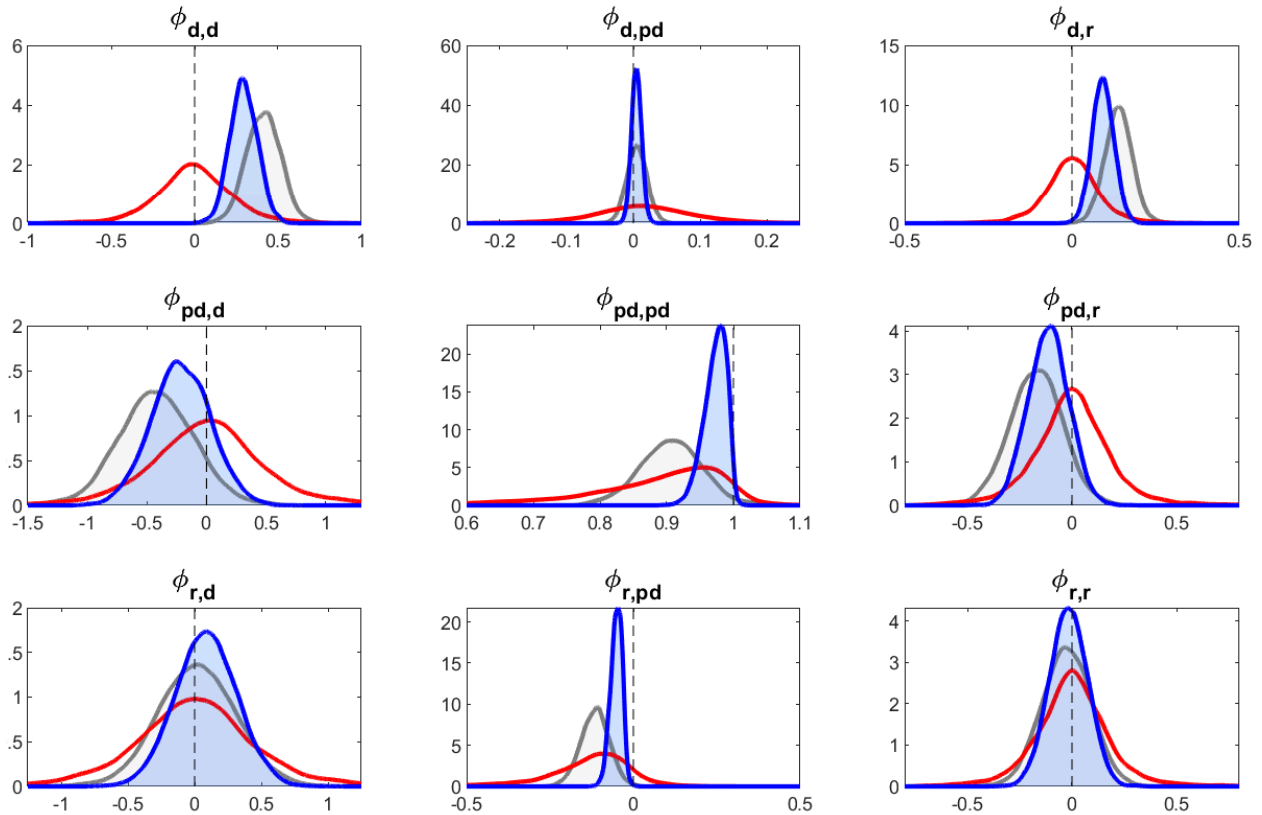
Note: Red represents the prior, grey the likelihood, and blue the posterior.

are generally more concentrated than both the prior and the likelihood, meaning that the posterior incorporates information from both. Looking at the first column, we observe that the coefficient $\phi_{d,d}$ is attenuated, from 0.42 in the likelihood to 0.3 in the restricted posterior, but the entirety of the posterior distribution is still in the positive region. The remaining two coefficients, $\phi_{pd,d}$ and $\phi_{r,d}$, move somewhat toward zero and gain precision relative to the likelihood. In the second column we observe important differences between likelihood and posterior. The first is an increase in $\phi_{pd,pd}$, which becomes very concentrated around a mode of 0.98. The stationarity restriction implies that this coefficient is highly skewed a feature that is visible already from the prior density, while it is absent from the likelihood,

Figure 6: BAYESIAN PRIOR AND POSTERIOR OF $\boldsymbol{\mu}$
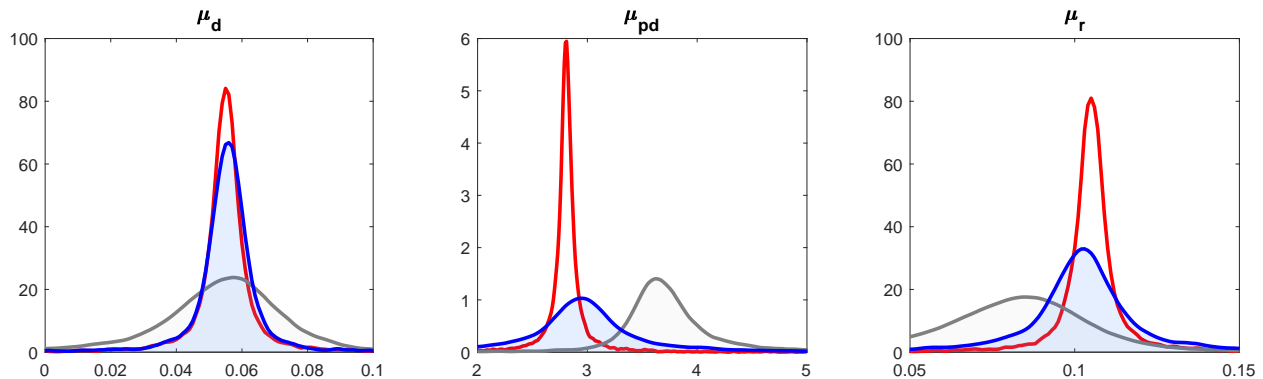
Note: Red represents the prior, grey the likelihood, and blue the posterior.

which does not discard non-stationary roots. The stationarity restriction, coupled with the CS restriction (9), implies that price-dividend ratios predict a priori either dividend growth or returns, or both. In the posterior, the coefficient $\phi_{d,pd}$, which captures dividend growth predictability coming from the price-dividend ratio, is strongly concentrated around zero. Hence, in line with Cochrane's (2008) famous conclusion, the price-dividend ratio shows no ability to predict dividend growth. The persistence of the price-dividend ratio, taken together with the near zero posterior mean for $\phi_{d,pd}$ and the CS restriction (9), implies that the posterior of $\phi_{r,pd}$ must be concentrated around negative values. As seen in the figure by comparing priors and posteriors, a more persistent price-dividend ratio means slower mean reversion after price-dividend ratio shocks and less predictable returns. Finally, the third column displays posteriors that are again more precise and tilted toward zero with respect to the likelihood. In any case, the coefficient $\phi_{d,r}$, which captures the predictability of dividends using lagged returns and, as discussed above, amplifies the dividend momentum channel, remains significantly positive despite the shrinkage encoded in the prior.

Figure 6 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,

Figure 7: BAYESIAN RESTRICTED PRIOR AND POSTERIOR OF Σ
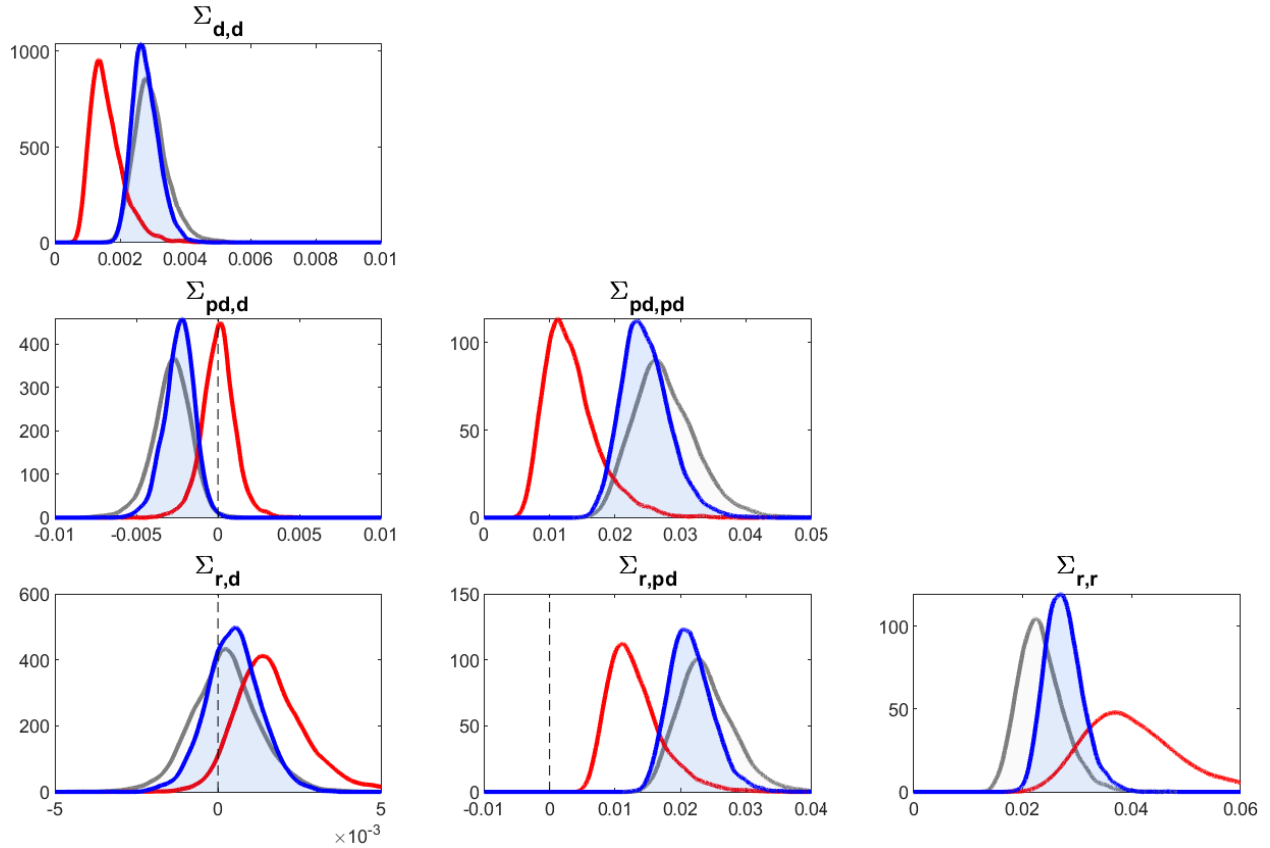
Note: Red represents the prior, grey the likelihood, and blue the posterior.

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 that is close to non-stationarity. For μ_r, the posterior is consistent with a markedly higher and less uncertain unconditional equity return in nominal terms.

Figure 7 reports the distribution of the variance-covariance elements. The CS restrictions push the prior variance-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 that 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 8: RETURN EQUATION R-SQUARED

(a) One-Period-Ahead R^2

(b) R^2 as a function of horizon
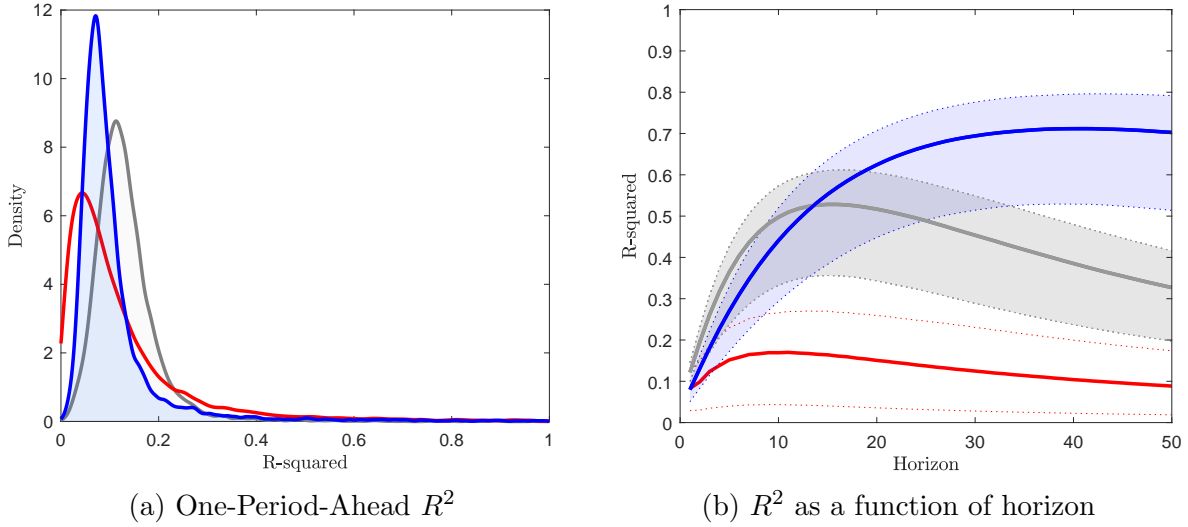
Note: Red represents the prior, grey the likelihood, and blue the posterior. In Panel (b) the solid line is the median value, while the shadow area represents the 68th posterior credible intervals.

7.2 Return Predictability

Figure 8 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. Clearly, 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 68th interval associated with the restricted informative prior, likelihood, and restricted informative posterior for the R^2 of the return equation as

Figure 9: One-Period-Ahead Expected Return
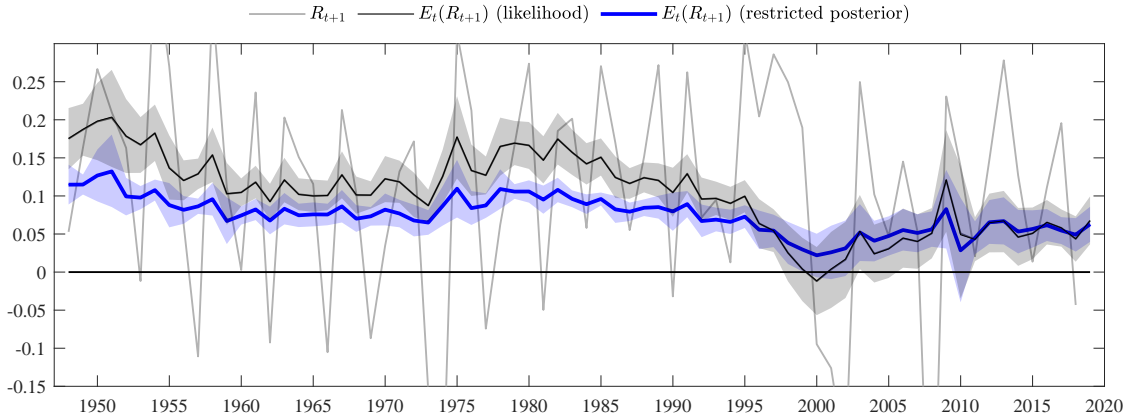
Note: The solid line is the median value while, the shadow area represents the 68th posterior credible intervals.

a function of the horizon.[15] 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 the longer horizons. This is seen from an R^2 that increases 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, Richardson, and Whitelaw (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 9 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 8. The restricted informative posterior also shows less uncertainty. Notably, the median of the restricted informative posterior estimate never becomes negative,

---

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

a desirable property as emphasized by Campbell and Thompson (2008). Nevertheless, there is still a meaningful amount of time variation in expected returns in the restricted informative posterior.

7.3   Dividend Momentum

Figure 10 plots the IRFs for cumulative discounted returns and dividend growth for the price-dividend ratio and dividend growth shocks implied by the restricted informative posterior. We identify the shocks using the Cholesky approach described above. Clearly, there is dividend momentum after a dividend growth shock. Compared with Panel (a) of Figure 2 where we use flat priors, the IRFs are flatter and display much narrower bands; the restricted informative prior sharpens the inference toward less predictability. Mean reversion after price-dividend ratio shocks is still present, but occurs at a much slower rate once the restricted informative prior pushes up the persistence of the price-dividend ratio. Similarly, the dividend momentum effect, reflected in the upward sloping IRFs of both dividend growth and returns after a dividend growth shock is, still present, although again attenuated by the restricted informative prior.

Finally, we analyze the Campbell and Ammer (1993) decomposition of return innovations into news about cash flows and discount rates. Table 3 analyzes how both price-dividend ratio and dividend growth shocks contribute to NDR_{t+1} and NCF_{t+1} using the restricted informative posterior. As expected, when dividend momentum is present, the table shows that NDR_{t+1} and NCF_{t+1} are correlated, and most of the correlation comes from the dividend momentum associated with dividend growth shocks. In line with the results in Section 4, the contribution of dividend growth shocks to the variances and correlation increases with respect to the results in the 2-variable VAR shown in Table 2.

Figure 10: Impulse Response Functions
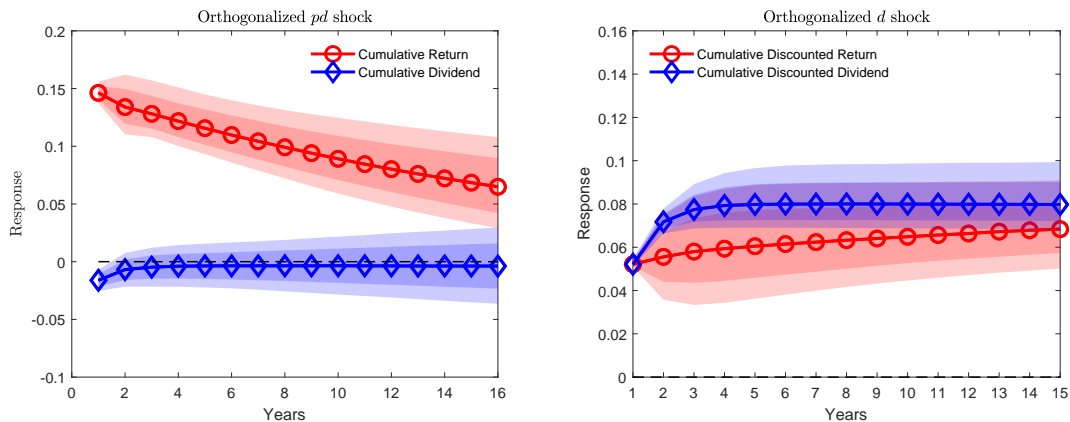
Note: The solid lines represent the median posterior response. The darker shadow area represents the 68th posterior credible intervals, while the lighter shadow area represents the 95th posterior credible intervals.

Table 3: Shock Contribution's to NCF_{t+1} and NDR_{t+1}

	Total	u_{t+1}^{pd}	u_{t+1}^{d}
$Var(NDR_{t+1})$	0.023	96.6%	3.3%
	[0.014, 0.036]	[93.1%, 98.5%]	[1.5%, 6.9%]
$Var(NCF_{t+1})$	0.007	8.2%	91.6%
	[0.006, 0.011]	[0.8%, 30.6%]	[69.2%, 99.0%]
$Corr(NDR_{t+1}, NCF_{t+1})$	20.7%	1.8%	17.1%
	[−18.8%, 57.6%]	[−37.6%, 40.9%]	[11.2%, 24.5%]

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

7.4 Variance of Long-Horizon Return

As discussed in Section 4.1, the ratio between $\mathtt{Var}_T(r_{T,T+k})$ and $k\mathtt{Var}_T(r_{T,T+1})$ is crucial to determining how risky stocks are for the the long-horizon investor. Figure 11 reports the ratio as a function of the horizon k and evaluates the contribution of dividend momentum to its shape. 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. We eliminate dividend momentum by canceling any effect of dividend growth

44

Figure 11: VARIANCE RATIO
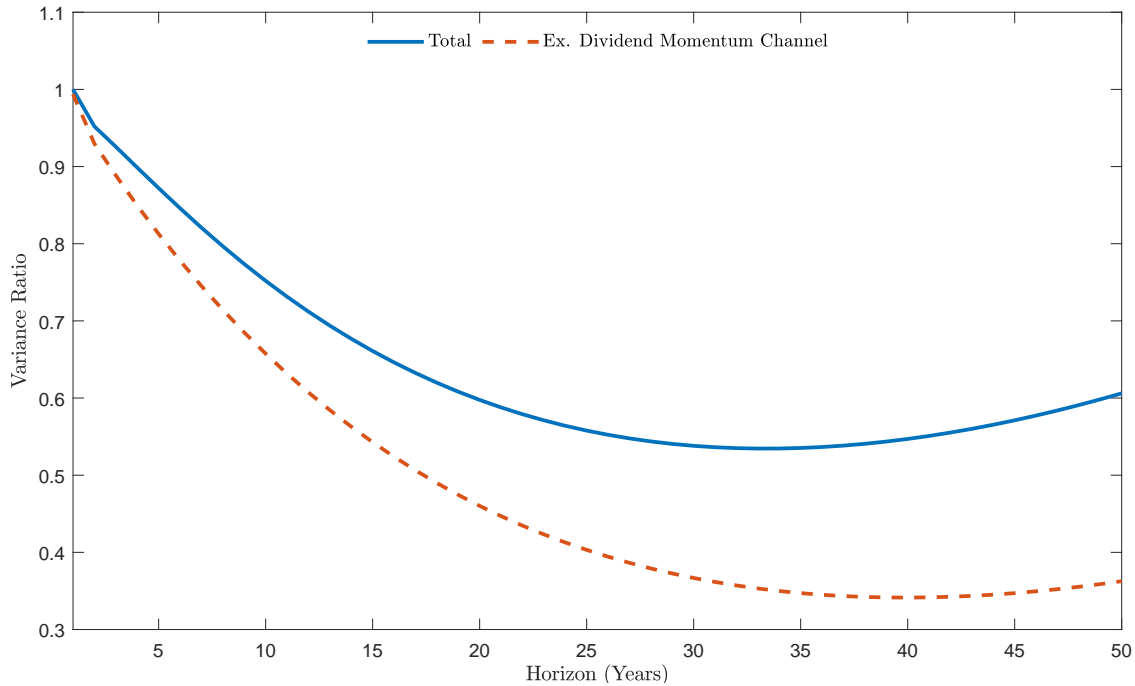
Note: The blue solid line reflects the ratio when both mean reversion and dividend momentum are present. The orange dashed line reflects the ratio when dividend momentum is excluded.

shocks after impact. See Appendix D for details. As the reader can see, dividend momentum increases this ratio for all of the reported horizons.

The presence of dividend momentum implies that dividend growth shocks increase both current returns and future expected returns. Therefore, these shocks increase the variance of long-run returns, $\text{Var}_T(r_{T,T+k})$, 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. In the absence of dividend momentum, the variance ratio would be roughly one-third smaller. Thus, this analysis confirms the results of the simplified model considered in Section 4.1; the presence of dividend momentum generates a negative hedging demand component that partially offsets the traditional positive hedging demand arising

45

from the mean reversion in returns (see Koijen, Rodríguez, and Sbuelz, 2009). As we will see in the next section, this negative hedging demand is empirically important and affects the portfolio choice of long-horizon investors trying to time the market.

8   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. This is even more important for the long-run investor (see, e.g., Campbell and Viceira, 1999, 2002). 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. To do that we need to consider a more general model that includes the risk-free rate. In Appendix F we describe this more general model and the CS restrictions associated with it and how the results in Section 3 still hold.

We compare the performance of an investor who uses our restricted informative prior with that if four other investors who use different priors. Two of the investors use the 4-variable VAR described above. Of those, one will use flat priors and another informative priors without imposing the CS restrictions. Appendix G describes in detail the flat and the informative prior parameterizations used in this section. A third investor follows the common approach in the literature and drops dividend growth from the VAR; she runs a 3-variable VAR with the risk-free rate, the price-dividend ratio, and the excess stock return using flat priors. Finally, a naive investor uses the historical average returns, computed every year as new data become available. Each year, the five investors estimate the parameters of their models and produce forecasts using only information available at each point in time in an expanding window.

Table 4: Out-of-Sample Return Equation R-squared

| | Flat priors | | Informative Priors | |
	Omitting Divs.	Including Divs.	Unrestricted	Restricted
$h = 1$	-11%	-13%	-1%	5%
$h = 2$	-8%	-10%	0%	14%
$h = 3$	-27%	-30%	-4%	22%
$h = 4$	-41%	-40%	-9%	27%
$h = 5$	-48%	-45%	-20%	29%
$h = 6$	-89%	-82%	-46%	27%
$h = 7$	-141%	-129%	-76%	22%
$h = 8$	-186%	-170%	-106%	19%
$h = 9$	-246%	-223%	-137%	17%
$h = 10$	-355%	-319%	-189%	11%

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

As a first step, Table 4 evaluates the out-of-sample forecasts, looking at the cumulative excess return $\sum_{i=1}^{h}(r_{t+h} - r_{t+h}^{f})$ at horizons $h = 1, \ldots, 10$ made by the different investors. 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 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. Interestingly, the addition of dividend growth to the VAR with flat priors leads to a moderate improvement at horizons greater than or equal to 4 years. Both 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. Nevertheless, only the investor using the restricted informative prior beats the naive investor out-of-sample. The gains peak at around 5 years, but remain economically significant up to 10 years ahead.

We now discuss optimal long-horizon allocations. The investors maximize the expected utility of the terminal value of wealth under Constant Relative Risk Aversion (CRRA) preferences.[16] Using different forecasts, each investor choose the allocation between the

---

[16]This problem resembles the one of target-date funds, whose dramatic increase in importance in the last

risk-free asset and the stock return for one year. To compute the allocations, we use the solution derived by Jurek and Viceira (2011) where the investors re-balance every year, taking into account how many years are left until retirement. The investors start with a planning horizon of 45 years starting at the end of 1973 and can invest in equity and a risk-free cash instrument and we will use the posterior mean at each period. We consider a risk aversion coefficient of $\gamma = 5$.

Figure 12 displays the results. Panel (a) displays the profile of wealth accumulation for each of the five investors. To facilitate comparisons, each curve is adjusted by its ex-post volatility. Panel (b) in turn presents the history of the weights for the risky asset. Two points 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 of the 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. Both of the investors using informative priors have more moderate weight profiles, but the restricted informative prior leads to weights that are leveraged only occasionally and by marginal amounts, and never take short positions. These two desirable properties lead to the superior performance shown 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 of the investors using informative priors, since it does not take advantage of the moderate degree of return predictability estimated in the data.

Figure 13 shows which part of the risk-adjusted wealth gains of the investor using restricted informative priors with respect to the investor who uses informative priors but does not impose the restrictions is due to the imposition of the CS restrictions on (1) $\boldsymbol{\Phi}$ and (2) $\boldsymbol{\Sigma}$, respectively. Of the almost 9 percentage point gain with respect to using informative priors but not imposing the restrictions, 6 percentage points come from the CS restrictions on $\boldsymbol{\Phi}$ while 3 come from imposing the CS restrictions on $\boldsymbol{\Sigma}$.

---

decade has been documented by Parker, Schoar, and Sun (2020).

Figure 12: ASSET ALLOCATION RESULTS: LONG-TERM STRATEGY

(a) Risk-Adjusted Log Wealth
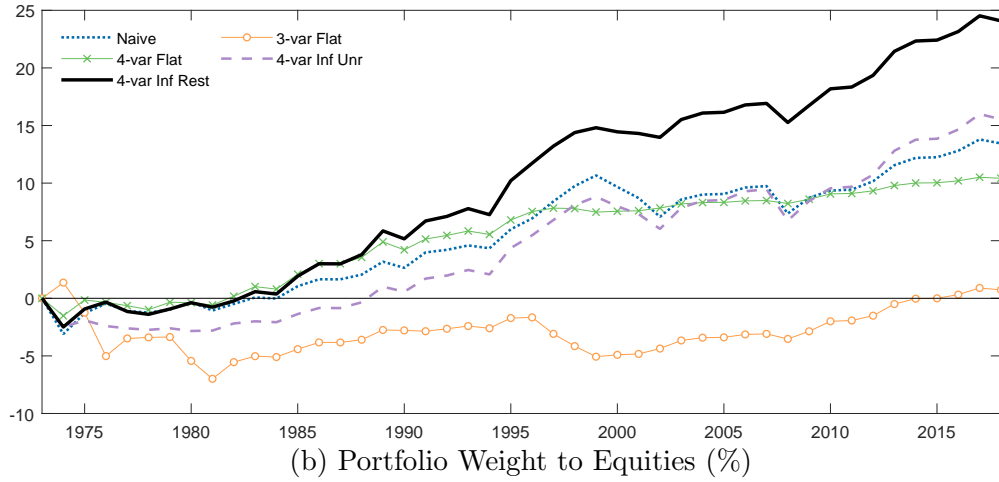
(b) Portfolio Weight to Equities (%)
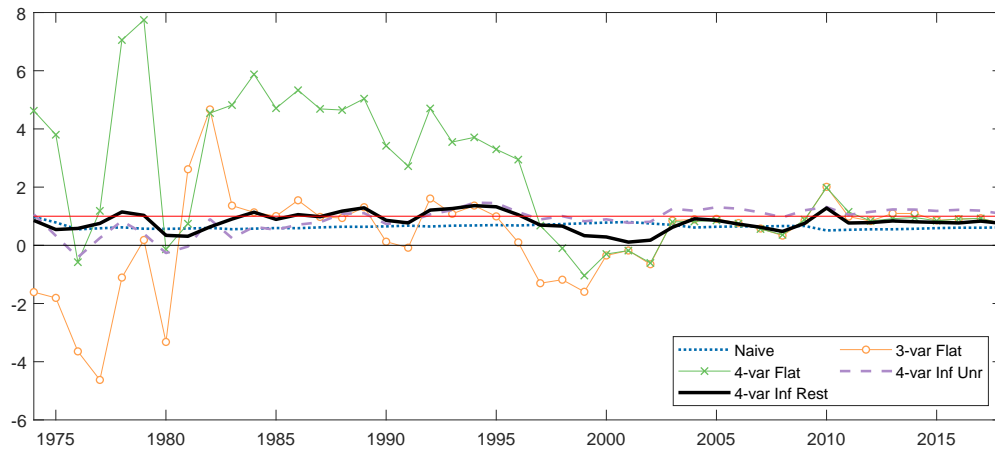
Why does the model with restricted informative priors lead to less leverage and less aggressive timing? Figure 14 plots the mean allocations as a function of investment horizon, together with bands representing the expected standard deviation of the allocations using the posterior mean using data until 2018. 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

Figure 13: RISK-ADJUSTED LOG WEALTH BREAKDOWN
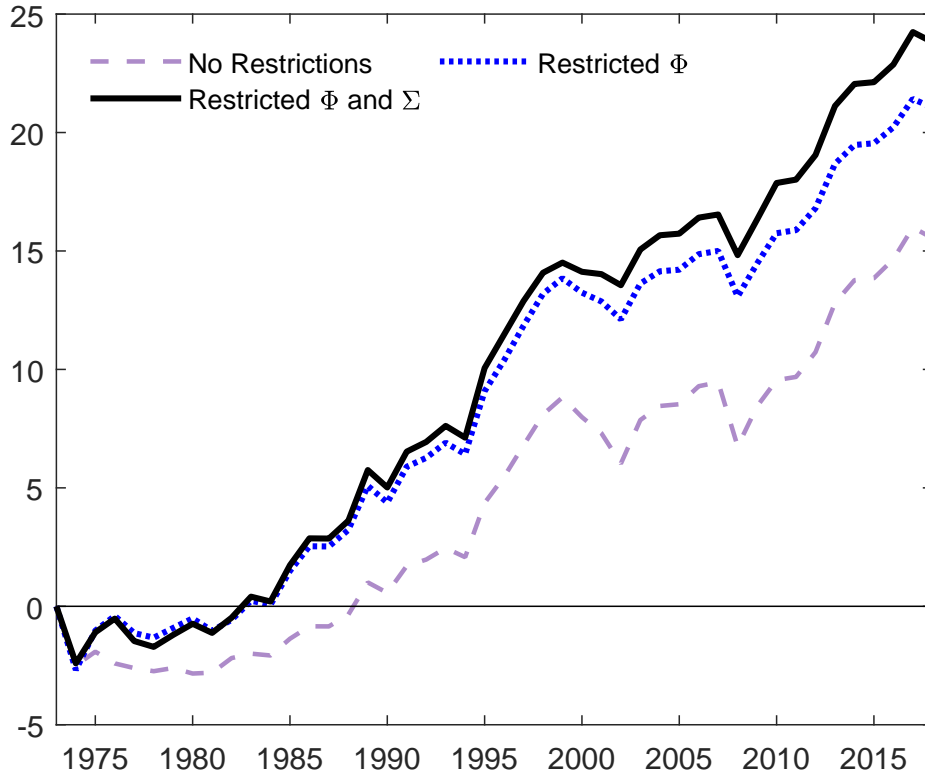
opportunities.

The first model considered is the one with flat priors excluding dividend growth. This model 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 a great deal of timing around the average position (wide bands). 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. 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

Figure 14: Steady-State Allocation to Stocks under Different Models
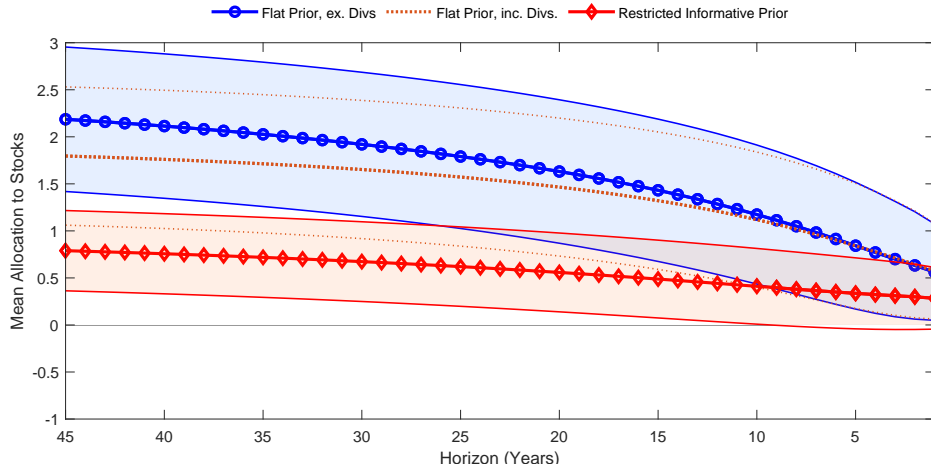
allocation by tilting the model parameters toward values consistent with a lower degree of return predictability. This leads to positions that are close to 100 percent stocks at the beginning of the planning horizon, and close to 50 percent when the investor nears retirement, and positions that feature 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 not consistent with the quantitative results of portfolio choice problems, but we show that the use of priors that encode shrinkage of return predictability, together with 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.

8.1 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, Lenza, and Primiceri (2019), using only data up to 1973. The previous results highlight that imposing informative priors delivers

Figure 15: Sharpe Ratio as a function of Hyperparameters
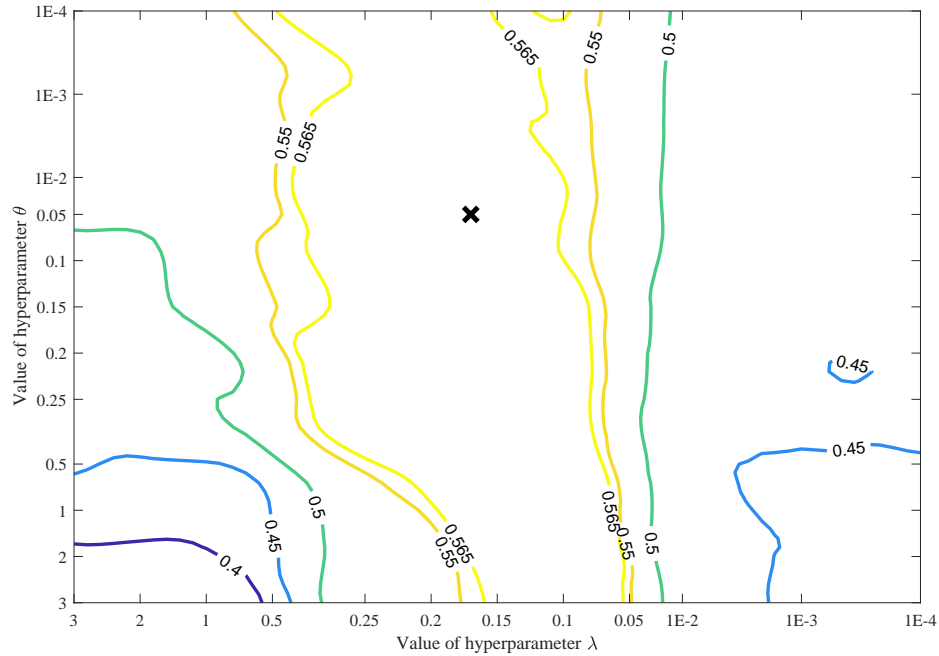
Note: The hyperparameter values used throughout this section are marked with a cross in the figure.

clear gains with respect to any alternative prior when measured in terms of risk-adjusted wealth at the end of the 45 years. Figure 15 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 45-year exercise 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, Lenza, and Primiceri (2019) succeeds in selecting hyperparameters that are close to the ones 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.[17] Of course, if the parameter is set 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.

8.2 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 16 replicates Figure 12 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 of the investors using flat priors take large short positions that lead to large draw-downs, particularly in the last half of the 1990s. Both of the investors using informative priors have more moderate weight profiles, but the restricted informative prior leads to weights that are short only occasionally and by marginal amounts, and never require leverage. These two desirable properties lead to superior performance. Figure 17 replicates Figure 15 for the case of myopic investors. As before, the figure shows that in order 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 also. 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.
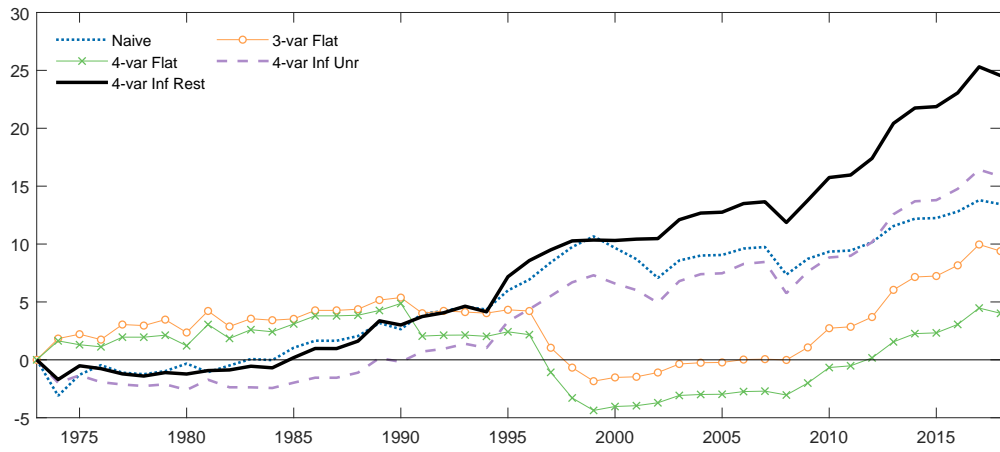
---

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

Figure 16: Asset Allocation Results: Myopic Strategy

(a) Risk-Adjusted Log Wealth

(b) Portfolio Weight to Equities (%)
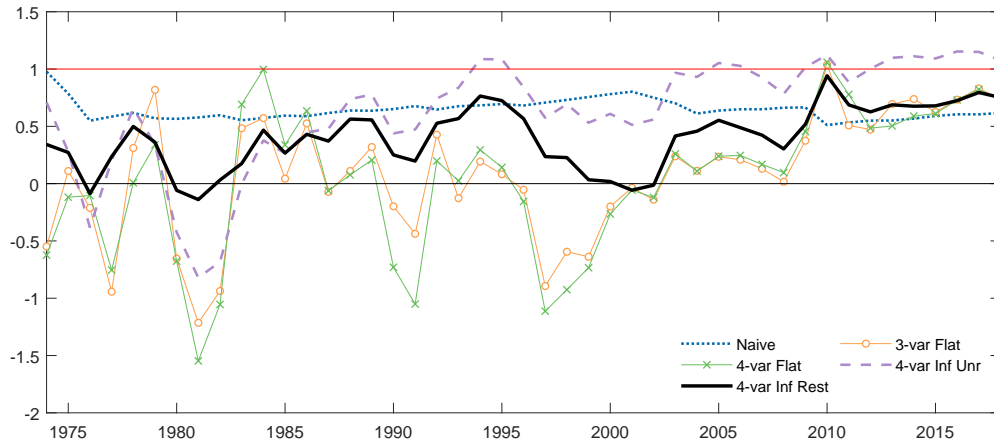
Figure 17: SHARPE RATIO AS A FUNCTION OF HYPERPARAMETERS
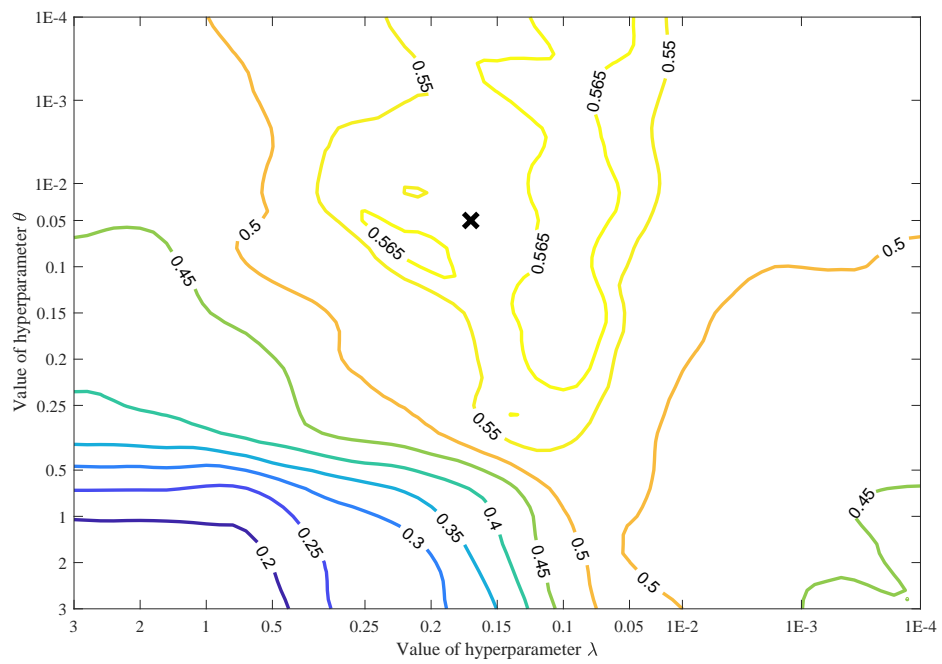
Note: The hyperparameter values used throughout this section are marked with a cross in the figure.

9   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. We highlight the importance of including dividend growth in a VAR together with the price-dividend ratio and returns: the common practice of omitting dividend growth is in general invalid, since it amounts to imposing the additional restriction that this variable is not persistent after controlling for lags of the remaining variables. Using postwar annual 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 the interpretation of cash flow and discount rate news and the optimal asset allocation of long-horizon investors. We conclude with some remarks on the implications of our methods for future research that we have not explored in the paper.

On the empirical front, there is a burgeoning literature that uses large data sets to search for evidence of predictability in aggregate stock returns (see, e.g., Kelly and Pruitt, 2013; Rapach and Zhou, 2020). Shrinkage or regularization of a potentially very 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 (2008): the price-dividend ratio and aggregate dividend growth are not *any* predictors: they are fundamentally linked to returns by the CS identity. Our approach opens the door to using large Bayesian VARs, which have been shown to be highly successful in forecasting macroeconomic and financial variables with hundreds of predictors (see Koop, 2013; Carriero, Kapetanios, and Marcellino, 2012; Carriero, Clark, and Marcellino, 2019), for the task of uncovering return predictability while respecting the cross-equation restrictions

implied by the CS type of identities.

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. In fact, while in the small system presented in Section 4, the condition $\phi_{d,d} > 0$ is necessary for dividend momentum to exist, this concept generalizes to larger systems where $\phi_{d,d} > 0$ may not be required. To see this, imagine that a variable is added to the system that forecasts dividend growth, driving $\phi_{d,d}$ toward zero. The logic of the CS identity means that any predictor of dividends, other than the price-dividend ratio itself, will either predict returns or predict the price-dividend ratio (with the opposite sign), therefore predicting returns with a lag. Thus, shocks to this predictor will also induce dividend momentum.

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 all those applications a highly persistent ratio or spread is the key long-run predictor of an asset 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.

We end the paper by noting that the presence of dividend momentum has 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, Huang, and Santos (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 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 the fact that such mechanisms are the largest driver of fluctuations in expected returns. Rather, it points to the presence of at least one shock

that generates pro-cyclical variation in expected returns, through channels that the literature has so far not explored.

References

ABADIR, K. M. AND J. MAGNUS (2005): *Matrix Algebra*, Cambridge University Press.

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., 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., 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.

———— (2017): *Financial Decisions and Markets: a Course in Asset Pricing*, Princeton University Press.

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.

CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2012): "Forecasting government bond yields with large Bayesian vector autoregressions," *Journal of Banking & Finance*, 36, 2026–2047.

CHEN, L., Z. DA, AND R. PRIESTLEY (2012): "Dividend Smoothing and Predictability," *Management Science*, 58, 1834–1853.

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

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

——— (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.

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

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.

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 pPerformance of Equity Premium Prediction," *The Review of Financial Studies*, 21, 1455–1508.

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.

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

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.

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.

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.

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.

——— (2000): "Using a Likelihood Perspective to Sharpen Econometric Discourse: Three Examples," *Journal of Econometrics*, 95, 443–462.

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.

Appendix A   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 Section 3 and Appendix F. Let us consider the following model with approximation error of $\mathbf{x}_t$:

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

Let $\mathbb{E}_t \left(\boldsymbol{\xi}_{t+1} \boldsymbol{\xi}'_{t+1} \right) = \boldsymbol{\Omega}_\xi$, $\mathbb{E}_t \left(\boldsymbol{\eta}_{t+1} \boldsymbol{\eta}'_{t+1} \right) = \boldsymbol{\Omega}_\eta$, and $\mathbb{E}_t \left(\boldsymbol{\xi}_{t+1} \boldsymbol{\eta}'_{t+1} \right) = \boldsymbol{\Omega}_{\xi\eta}$, where the last assumption reflects the fact that the approximation error $\boldsymbol{\eta}_{t+1}$ may be correlated with $\boldsymbol{\xi}_{t+1}$.[18]

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

$$\mathbf{x}_{t+1} = \mathbf{G}_1 \mathbf{x}_t + \mathbf{G}_2 \mathbf{x}_{t-1} + \mathbf{e}_{t+1} + \mathbf{D}_1 \mathbf{e}_t, \tag{A.2}$$

where $\mathbb{E}_t \left(\mathbf{e}_{t+1} \mathbf{e}'_{t+1} \right) = \boldsymbol{\Omega}_e$. We can now find $\mathbf{D}_1$ and $\boldsymbol{\Omega}_e$ as a function of $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{B}_1$, $\mathbf{M}_0$, $\boldsymbol{\Omega}_\xi$, $\boldsymbol{\Omega}_\eta$, and $\boldsymbol{\Omega}_{\xi\eta}$ so that the autocovariance function of Equation (A.2) matches the one of Equation (A.1).

Specifically, let the first and second autocovariance of the moving average component of Equation (A.1) be:

$$
\begin{aligned}
\boldsymbol{\Psi}_0 &= \boldsymbol{\Omega}_\xi + \mathbf{B}_1 \boldsymbol{\Omega}_\xi \mathbf{B}'_1 + \mathbf{M}_0 \boldsymbol{\Omega}_\eta \mathbf{M}'_0 + \mathbf{B}_1 \boldsymbol{\Omega}_{\xi\eta} \mathbf{M}'_0 + \mathbf{M}_0 \boldsymbol{\Omega}'_{\xi\eta} \mathbf{B}'_1 \text{ and} \\
\boldsymbol{\Psi}_1 &= \mathbf{B}_1 \boldsymbol{\Omega}_\xi + \mathbf{M}_0 \boldsymbol{\Omega}'_{\xi\eta} \mathbf{B}'_1,
\end{aligned}
\tag{A.3}
$$

---

[18]Equation (A.1) boils down to Equation (14) by defining $\boldsymbol{\eta}_{t+1} = \eta_{t+1}$, $\boldsymbol{\Omega}_\eta = \sigma^2_\eta$, $\mathbf{M}_0 = -\boldsymbol{\phi}_{12}$, and $\boldsymbol{\Omega}_{\xi\eta} = \begin{bmatrix} 0 & \sigma^2_\eta \end{bmatrix}'$. Notice that $\boldsymbol{\Omega}_{\xi\eta} = \begin{bmatrix} 0 & \sigma^2_\eta \end{bmatrix}'$ because we assume that η_{t+1} is only correlated with u^r_{t+1}.

then if Equation (A.2) is the VARMA(2,1) representation of $\mathbf{x}_t$, we have that:

$$\boldsymbol{\Psi}_0 \;\; = \;\; \boldsymbol{\Omega}_e + \mathbf{D}_1 \boldsymbol{\Omega}_e \mathbf{D}_1' \text{ and} \tag{A.4}$$

$$\boldsymbol{\Psi}_1 \;\; = \;\; \mathbf{D}_1 \boldsymbol{\Omega}_e. \tag{A.5}$$

To find $\mathbf{D}_1$ and $\boldsymbol{\Omega}_e$ note that from Equation (A.5) we have that:

$$\boldsymbol{\Omega}_e = \mathbf{D}_1^{-1} \boldsymbol{\Psi}_1.$$

Substituting the above expression into Equation (A.4) we get that:

$$\boldsymbol{\Psi}_0 = \mathbf{D}_1^{-1} \boldsymbol{\Omega}_1 + \mathbf{D}_1 \mathbf{D}_1^{-1} \boldsymbol{\Psi}_1 \mathbf{D}_1' = \mathbf{D}_1^{-1} \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_1 \mathbf{D}_1',$$

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

$$\mathbf{D}_1 \boldsymbol{\Psi}_1 \mathbf{D}_1' - \mathbf{D}_1 \boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 = \mathbf{0} \tag{A.6}$$

The solution to Equation (A.6) can be easily found numerically. Starting with a guess for $\mathbf{D}_1$ (e.g. $\mathbf{D}_1^0 = \mathbf{B}_1$), one can find the solution iterating over $\mathbf{D}_1^{k+1} = \boldsymbol{\Psi}_1 [\boldsymbol{\Psi}_0 - \boldsymbol{\Psi}_1 (\mathbf{D}_1^k)']^{-1}$.

Appendix B   Proof of Theorem 1

First, let us assume that $\phi_{r,d} = \phi_{pd,d} = 0$ (i.e. $\boldsymbol{\phi}_{12} = \mathbf{0}_{2\times1}$) holds. Then, from Equation (5) $\phi_{d,d} = 0$ which implies that $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$. Moreover, since $\boldsymbol{\phi}_{12} = \mathbf{0}_{2\times1}$, $\mathbf{G}_2 = \mathbf{0}_{2\times2}$ and $\mathbf{B}_1 = \mathbf{0}_{2\times2}$. If $\mathbf{B}_1 = \mathbf{0}_{2\times2}$ by Equation (A.3) we have that $\boldsymbol{\Psi}_1 = \mathbf{0}_{2\times2}$; hence, by Equation (A.5) we have that $\mathbf{D}_1 = \mathbf{0}_{2\times2}$.

Second, let us assume that $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$, $\mathbf{G}_2 = \mathbf{0}_{2\times2}$, and $\mathbf{D}_1 = \mathbf{0}_{2\times2}$. Since $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$, we have that $\phi_{d,d} = 0$. Because $\phi_{d,d} = 0$ and $\phi_{r,d} = \rho\phi_{pd,d} + \phi_{d,d}$, we have $\phi_{r,d} = \rho\phi_{pd,d}$

and $\phi_{12} = [\phi_{pd,d}, \phi_{r,d}]' = [\phi_{pd,d}, \rho\phi_{pd,d}]' = \phi_{pd,d}[1,\rho]'$ and hence $\mathbf{B}_1 = \phi_{pd,d}[1,\rho]'[-\rho,1]$ and $\mathbf{M}_0 = -\phi_{12} = -\phi_{pd,d}[1,\rho]'$. In this case $\mathbf{G}_2 = \mathbf{0}_{2\times 2}$ implies that $\phi_{21} = \mathbf{0}_{1\times 2}$. Since $\mathbf{D}_1 = \mathbf{0}_{2\times 2}$, we have that $\mathbf{\Psi}_1 = \mathbf{0}_{2\times 2}$. Hence $\mathbf{B}_1\mathbf{\Omega}_\xi + \mathbf{M}_0\mathbf{\Omega}'_{\xi\eta}\mathbf{B}'_1 = \mathbf{0}_{2\times 2}$. But, since $\phi_{d,d} = 0$, we have that:

$$\mathbf{B}_1\mathbf{\Omega}_\xi + \mathbf{M}_0\mathbf{\Omega}'_{\xi\eta}\mathbf{B}'_1 = \phi_{pd,d}\begin{pmatrix} -\rho\Sigma_{pd,pd} + \Sigma_{pd,r} - \phi_{pd,d}\sigma_\eta^2 & \Sigma_{r,r} - \rho\left(\Sigma_{pd,r} + \phi_{pd,d}\sigma_\eta^2\right) \\ \rho\left(-\rho\Sigma_{pd,pd} + \Sigma_{pd,r} - \phi_{pd,d}\sigma_\eta^2\right) & \rho\left(\Sigma_{r,r} - \rho\left(\Sigma_{pd,r} + \phi_{pd,d}\sigma_\eta^2\right)\right) \end{pmatrix},$$

where $\Sigma_{pd,pd} = \mathtt{Var}\left(u_{t+1}^{pd}\right)$, $\Sigma_{r,r} = \mathtt{Var}\left(u_{t+1}^r\right)$, and $\Sigma_{pd,r} = \mathtt{Cov}\left(u_{t+1}^{pd}, u_{t+1}^r\right)$. Taking into account the CS restrictions among the covariance terms (Equations (8) and (9)), the above matrix can be equal to $\mathbf{0}_{2\times 2}$ without $\phi_{pd,d} = 0$ if and only if $\phi_{pd,d} = \left(\Sigma_{d,d} + \rho\Sigma_{pd,d} + \sigma_\eta^2\right)/\rho\sigma_\eta^2 = \Sigma_{pd,d}/\sigma_\eta^2$, where $\Sigma_{d,d} = \mathtt{Var}\left(u_{t+1}^d\right)$ and $\Sigma_{pd,d} = \mathtt{Cov}\left(u_{t+1}^d, u_{t+1}^{pd}\right)$. But this is not possible because it would imply that $\Sigma_{d,d} = -\sigma_\eta^2$.

In the absence of approximation error, we have that:

$$\mathbf{B}_1\mathbf{\Omega}_\xi = \phi_{pd,d}\begin{pmatrix} -\rho\Sigma_{pd,pd} + \Sigma_{pd,r} & \Sigma_{r,r} - \rho\Sigma_{pd,r} \\ \rho\left(-\rho\Sigma_{pd,pd} + \Sigma_{pd,r}\right) & \rho\left(\Sigma_{r,r} - \rho\Sigma_{pd,r}\right) \end{pmatrix} = \phi_{pd,d}\begin{pmatrix} \Sigma_{pd,d} & \Sigma_{d,d} + \rho\Sigma_{pd,d} \\ \rho\Sigma_{pd,d} & \rho\left(\Sigma_{d,d} + \rho\Sigma_{pd,d}\right) \end{pmatrix},$$

where $\Sigma_{r,r} = \mathtt{Var}\left(u_{t+1}^r\right)$ and the last equality takes into account the restrictions arising from the CS restriction. Clearly unless $\Sigma_{d,d} = \Sigma_{pd,d} = 0$, the above matrix can never be equal to $\mathbf{0}_{2\times 2}$ without $\phi_{pd,d} = 0$. It should be clear that if $\phi_{12} \neq \mathbf{0}_{2\times 1}$ (i.e. $\Sigma_{d,d} = \Sigma_{pd,d} = 0$), $\mathbf{G}_2 = \mathbf{0}_{3\times 3}$ implies that $\phi_{21} = \mathbf{0}'_2$; hence, the dividend growth process is fixed to constant over time.

Appendix C   Variance of the Innovations in the VAR(1) Implied by VARMA(2,1)

The results in the appendix are general, so they can be used to obtain the variance of the innovations in the VAR(1) implied by VARMA(2,1) representations in both Section 3 and Appendix F.

We can use the definition of the innovations from the VAR(1) in Equation (17) to calculate $\boldsymbol{\Omega}_\varepsilon$. In particular, we have that:

$$
\begin{aligned}
\boldsymbol{\Omega}_\varepsilon \;=\; & \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)\boldsymbol{\Gamma}_0^{-1}\left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)' \\
& - \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \\
& -\mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)' + \mathbf{G}_2\boldsymbol{\Gamma}_0\mathbf{G}_2' + \boldsymbol{\Omega}_e \\
& -\mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)' + \mathbf{D}_1\boldsymbol{\Omega}_e\mathbf{D}_1' \\[4pt]
\;=\; & \mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' + \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' \\
& -\mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' - \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' \\
& -\mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{G}_2\boldsymbol{\Gamma}_0\mathbf{G}_2' + \boldsymbol{\Omega}_e \\
& -\mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\mathbf{D}_1'.
\end{aligned}
$$

We can simplify the above equation to:

$$
\begin{aligned}
\boldsymbol{\Omega}_\varepsilon \;=\; & -\mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{G}_2\boldsymbol{\Gamma}_0\mathbf{G}_2' + \boldsymbol{\Omega}_e \\
& -\mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_1\mathbf{G}_2' - \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\mathbf{D}_1' \\[4pt]
\;=\; & \boldsymbol{\Omega}_e + \mathbf{D}_1\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{G}_2\boldsymbol{\Gamma}_0\mathbf{G}_2' \\
& -\mathbf{G}_2\boldsymbol{\Gamma}_1'\boldsymbol{\Gamma}_0^{-1}\left(\boldsymbol{\Gamma}_1\mathbf{G}_2' + \boldsymbol{\Omega}_e\mathbf{D}_1'\right) - \mathbf{D}_1\boldsymbol{\Omega}_e\boldsymbol{\Gamma}_0^{-1}\left(\boldsymbol{\Gamma}_1\mathbf{G}_2' + \boldsymbol{\Omega}_e\mathbf{D}_1'\right) \\[4pt]
\;=\; & \boldsymbol{\Omega}_e + \mathbf{D}_1\boldsymbol{\Omega}_e\mathbf{D}_1' + \mathbf{G}_2\boldsymbol{\Gamma}_0\mathbf{G}_2' - \left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)\boldsymbol{\Gamma}_0^{-1}\left(\mathbf{G}_2\boldsymbol{\Gamma}_1' + \mathbf{D}_1\boldsymbol{\Omega}_e\right)'
\end{aligned}
$$

Appendix D   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 $r_{T,T+k} = \sum_{j=1}^{k} r_{T+1}$, the variance of long-horizon returns

can be decomposed as:

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

where $\boldsymbol{\Theta} = \{\boldsymbol{\Phi}, \boldsymbol{\Sigma}\}$ denotes the parameters in the VAR. The first term in Equation (D.1) corresponds to the expected variance of long-horizon returns, and since we have assumed a constant volatility in the VAR, $\text{Var}_T\left(r_{T,T+k}|\boldsymbol{\Theta}\right) = \text{Var}\left(r_{T,T+k}|\boldsymbol{\Theta}\right)$. 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:

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

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

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

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

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

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:

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

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:

$$\text{Var}\left(r_{T,T+k}|\boldsymbol{\Theta}\right) = \underbrace{k\mathbf{s}_r \boldsymbol{\Sigma} \mathbf{s}_r'}_{\text{iid uncertainty}} + \underbrace{2\mathbf{s}_r \boldsymbol{\Phi}_1 \sum_{j=1}^{k-1} \left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1} \left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right) \boldsymbol{\Sigma} \mathbf{s}_r'}_{\text{covariance component}} + \tag{D.2}$$

$$+ \underbrace{\sum_{j=1}^{k-1} \left[\mathbf{s}_r \boldsymbol{\Phi}_1 \left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1} \left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right)\right] \boldsymbol{\Sigma} \left[\mathbf{s}_r \boldsymbol{\Phi}_1 \left(\mathbf{I}_n - \boldsymbol{\Phi}_1\right)^{-1} \left(\mathbf{I}_n - \boldsymbol{\Phi}_1^j\right)\right]'}_{\text{future expected return uncertainty}}$$

The expected variance of long-horizon returns is the expectation of Equation (D.2). The covariance component is generally labelled 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.

D.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 IRFs 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 (D.2) substituting $\boldsymbol{\Sigma}$ by $\mathbf{B}_{(:,i)} \mathbf{B}_{(:,i)}'$ 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 return uncertainty. In order to compute the variance of long-horizon returns without dividend momentum, we substitute $\mathbf{\Sigma}$ by $\mathbf{B}_{(:,1)}\mathbf{B}'_{(:,1)} + \mathbf{B}_{(:,3)}\mathbf{B}'_{(:,3)}$ in the covariance component and future expected return uncertainty terms of Equation (D.2), since the dividend growth shock is the second shock in our simplified model.

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:

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

For multiple-period returns, we have that:

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

Appendix F   Results for a More General Model

The basic macro-finance VAR of Sections 2 and 3 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 $\left(n_w + 4\right) \times 1$ and assume it follows a

VAR(1) structure:

$$
\underbrace{\begin{bmatrix} \Delta d_{t+1} \\ \mathbf{w}_{t+1} \\ r^f_{t+1} \\ pd_{t+1} \\ r_{t+1} - r^f_{t+1} \end{bmatrix}}_{\mathbf{y}_{t+1}} = \underbrace{\begin{bmatrix} c^d \\ \mathbf{c}^w \\ c^{rf} \\ c^{pd} \\ c^r \end{bmatrix}}_{\mathbf{\Phi_0}} + \underbrace{\begin{bmatrix} \phi_{d,d} & \boldsymbol{\phi}_{d,w} & \phi_{d,rf} & \phi_{d,pd} & \phi_{d,r} \\ \boldsymbol{\phi}_{w,d} & \boldsymbol{\phi}_{w,w} & \boldsymbol{\phi}_{w,rf} & \boldsymbol{\phi}_{w,pd} & \boldsymbol{\phi}_{w,r} \\ \phi_{rf,d} & \boldsymbol{\phi}_{rf,w} & \phi_{rf,rf} & \phi_{rf,pd} & \phi_{rf,r} \\ \phi_{pd,d} & \boldsymbol{\phi}_{pd,w} & \phi_{pd,rf} & \phi_{pd,pd} & \phi_{pd,r} \\ \phi_{r,d} & \boldsymbol{\phi}_{r,w} & \phi_{r,rf} & \phi_{r,pd} & \phi_{r,r} \end{bmatrix}}_{\mathbf{\Phi_1}} \underbrace{\begin{bmatrix} \Delta d_t \\ \mathbf{w}_t \\ r^f_t \\ pd_t \\ r_t - r^f_t \end{bmatrix}}_{\mathbf{y}_t} + \underbrace{\begin{bmatrix} u^d_{t+1} \\ \mathbf{u}^w_{t+1} \\ u^{rf}_{t+1} \\ u^{pd}_{t+1} \\ u^r_{t+1} \end{bmatrix}}_{\mathbf{u}_{t+1}}
\tag{F.1}
$$

where $\mathbf{u}_{t+1}$ is normal with mean zero and $\mathbb{E}_t\left(\mathbf{u}_{t+1}\mathbf{u}'_{t+1}\right) = \boldsymbol{\Sigma}$. As before, this system can be written compactly as $\mathbf{y}_{t+1} = \mathbf{\Phi}\mathbf{z}'_t + \mathbf{u}_{t+1}$ where $\mathbf{z}_t = \left[1, \mathbf{y}'_t\right]$ and $\mathbf{\Phi} = \left[\mathbf{\Phi_0}, \mathbf{\Phi_1}\right]$.

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:

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

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

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

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

$$
\begin{aligned}
c^r &= c^d - c^{rf} + \rho c^{pd} + \kappa, \\
\phi_{r,d} &= \phi_{d,d} - \phi_{rf,d} + \rho\phi_{pd,d}, \\
\boldsymbol{\phi}'_{r,w} &= \boldsymbol{\phi}'_{d,w} - \boldsymbol{\phi}'_{rf,w} + \rho\boldsymbol{\phi}'_{pd,w}, \\
\phi_{r,rf} &= \phi_{d,rf} - \phi_{rf,rf} + \rho\phi_{pd,rf}, \\
\phi_{r,pd} &= \phi_{d,pd} - \phi_{rf,pd} + \rho\phi_{pd,pd} - 1, \text{ and} \\
\phi_{r,r} &= \phi_{d,r} - \phi_{rf,r} + \rho\phi_{pd,r}
\end{aligned}
$$

72

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

$$\text{Cov}\left(u_{t+1}^d, u_{t+1}^r\right) = \rho\text{Cov}\left(u_{t+1}^d, u_{t+1}^{pd}\right) + \text{Var}\left(u_{t+1}^d\right) - \text{Cov}\left(u_{t+1}^d, u_{t+1}^{rf}\right),$$

$$\text{Cov}\left(\mathbf{u}_{t+1}^w, u_{t+1}^r\right) = \rho\text{Cov}\left(\mathbf{u}_{t+1}^w, u_{t+1}^{pd}\right) + \text{Cov}\left(\mathbf{u}_{t+1}^w, u_{t+1}^d\right) - \text{Cov}\left(\mathbf{u}_{t+1}^w, u_{t+1}^{rf}\right),$$

$$\text{Cov}\left(u_{t+1}^{rf}, u_{t+1}^r\right) = \rho\text{Cov}\left(u_{t+1}^{rf}, u_{t+1}^{pd}\right) + \text{Cov}\left(u_t^{rf}, u_{t+1}^d\right) - \text{Var}\left(u_{t+1}^{rf}\right), \text{ and}$$

$$\text{Cov}\left(u_{t+1}^{pd}, u_{t+1}^r\right) = \rho\text{Var}\left(u_{t+1}^{pd}\right) + \text{Cov}\left(u_{t+1}^{pd}, u_{t+1}^d\right) - \text{Cov}\left(u_{t+1}^{pd}, u_{t+1}^{rf}\right)$$

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

$$\text{Var}\left(u_{t+1}^r\right) = \text{Var}\left(u_{t+1}^d\right) + \text{Var}\left(u_{t+1}^{rf}\right) + \rho^2\text{Var}\left(u_{t+1}^{pd}\right) +$$
$$2\rho\text{Cov}\left(u_{t+1}^{pd}, u_{t+1}^d\right) - 2\rho\text{Cov}\left(u_{t+1}^{pd}, u_{t+1}^{rf}\right) - 2\text{Cov}\left(u_{t+1}^d, u_{t+1}^{rf}\right).$$

The stationarity restriction is now:

$$\mathbf{\Phi}_1 \in \left\{\mathbf{Z} \in \mathbb{R}^{(n_w+4)\times(n_w+4)} : \max\{\texttt{eig}\left(\mathbf{Z}\right)\} < 1\right\}.$$

In Appendix F, we show that Theorem 1 and Corollary 2 hold for the more general model for the case both with and without approximation error.

Using the results in Appendix A, it can be shown that if we drop dividend growth and run a VAR(1) on $\mathbf{x}_{t+1}' = \left[\mathbf{w}_{t+1}', r_{t+1}^f, pd_{t+1}, r_{t+1} - r_{t+1}^f\right]$ we can obtain the same VARMA(2,1) representation of $\mathbf{x}_{t+1}$:

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

where $\mathbf{G}_1 = \mathbf{\Phi}_{11} + \mathbf{\Phi}_{dd}$, $\mathbf{\Phi}_{dd} = \phi_{d,d}\mathbf{I}_{n+3}$, $\mathbf{G}_2 = \phi_{12}\phi_{21} - \mathbf{\Phi}_{dd}\mathbf{\Phi}_{11}$, $\mathbf{B}_1 = \phi_{12}\left[\mathbf{0}_{n_w}', 1, -\rho, 1\right] - \mathbf{\Phi}_{dd}$,

and $\mathbb{E}_t \left(\mathbf{e}_{t+1} \mathbf{e}_{t+1}' \right) = \mathbf{\Omega}_e$ and where

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

We have defined $\boldsymbol{\eta}_{t+1} = \eta_{t+1}$, $\mathbf{\Omega}_\eta = \sigma_\eta^2$, $\mathbf{M}_0 = -\boldsymbol{\phi}_{12}$, and $\mathbf{\Omega}_{\xi\eta} = \begin{bmatrix} \mathbf{0}_{n_w}' & 0 & 0 & \sigma_\eta^2 \end{bmatrix}'$.[19] Consider now specifying a VAR(1) for $\mathbf{x}_{t+1}$:

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

where $\mathbb{E}_t \left(\boldsymbol{\varepsilon}_{t+1} \boldsymbol{\varepsilon}_{t+1}' \right) = \mathbf{\Omega}_\varepsilon$. The matrix $\mathbf{A}_1$ follows the expression in Equation (16); thus the link between $\mathbf{A}_1$ and the parameters in the VAR(1) in Equation (F.1) highlighted in Section 3 also exists here. The next theorem replicates Theorem 1 in this more general set up.

Theorem F.1. *The VARMA(2,1) in Equation (F.3) 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* $\phi_{rf,d} = \phi_{pd,d} = \phi_{r,d} = 0$, *with* $\phi_{d,d} = 0$ *following from Equation* (F.3).

Proof. See Appendix H. □

Of course, Theorem F.1 implies that Corollary 1 also holds for the more general model described in this section. The innovations in the VAR(1) specified in Equation (F.4) also follow the expression in Equation (17). Results in Appendix C can be used to show that its variance can also be obtained using Equation (18). As expected, it is also the case that Corollary 2 holds here. As expected, the results in Section 3.1 when the approximation error is not present also hold in this more general case.

---

[19]As before, without loss of generality, we abstract from the constant term.

Appendix G Priors for Section 8

In this section we describe the prior parameterizations used in Section 8.

G.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.

G.2 The Informative Prior

The Minnesota priors in this section can be written:

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

where in the simplest case $\boldsymbol{\Omega} = \lambda^2(diag\left(\left[\underline{\sigma}_d^2, \underline{\sigma}_{rf}^2, \underline{\sigma}_{pd}^2, \underline{\sigma}_r^2\right]\right))^{-1}$, 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 $\boldsymbol{\Phi}_0$, the Minnesota prior is usually specified as flat.

We choose a value of $\theta = 0.05$, which controls the tightness of Single Unit Root prior. As before, we follow Giannone, Lenza, and Primiceri (2019) when choosing λ and θ. In particular, we use the starting sample finishing in 1973 to choose them. For mean of the excess return, $\underline{\mu}_r$ we chose a value of 6.5%; the mean of the risk-free rate $\underline{\mu}_{rf}$ is chosen to be 4%; and for the mean nominal dividend growth, $\underline{\mu}_d$ we chose a value of 5.5%, 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, Lenza, and Primiceri (2019).

Appendix H   Proof of Theorem F.1

First, let us assume that $\boldsymbol{\phi}_{w,d} = \mathbf{0}_n$ and $\phi_{r,d} = \phi_{pd,d} = 0$ holds. Then, $\phi_{d,d} = 0$ because of Equation (F.3). This implies that $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$ and $\boldsymbol{\phi}_{12} = \mathbf{0}_{n+3}$. Clearly this implies that $\mathbf{G}_2 = \mathbf{0}_{(n+3)\times(n+3)}$ and $\mathbf{B}_1 = \mathbf{0}_{(n+3)\times(n+3)}$. If $\mathbf{B}_1 = \mathbf{0}_{(n+3)\times(n+3)}$ by Equation (A.3) we have that $\boldsymbol{\Psi}_1 = \mathbf{0}_{(n+3)\times(n+3)}$; hence, by Equation (A.5) we have that $\mathbf{D}_1 = \mathbf{0}_{(n+3)\times(n+3)}$.

Second, let us assume that $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$, $\mathbf{G}_2 = \mathbf{0}_{(n+3)\times(n+3)}$, and $\mathbf{D}_1 = \mathbf{0}_{(n+3)\times(n+3)}$. Since $\mathbf{G}_1 = \boldsymbol{\Phi}_{11}$, we have that $\phi_{d,d} = 0$, because $\phi_{d,d} = 0$, $\phi_{r,d} = \rho\phi_{pd,d} - \phi_{rf,d}$ and $\boldsymbol{\phi}_{12} = \left[\boldsymbol{\phi}'_{w,d}, \phi_{rf,d}, \phi_{pd,d}, \rho\phi_{pd,d} - \phi_{rf,d}\right]'$.

Since $\mathbf{D}_1 = \mathbf{0}_{(n+3)\times(n+3)}$, we have that $\boldsymbol{\Psi}_1 = \mathbf{0}_{(n+3)\times(n+3)}$. Hence $\mathbf{B}_1\boldsymbol{\Omega}_\xi + \mathbf{M}_0\boldsymbol{\Omega}'_{\xi\eta}\mathbf{B}'_1 = \mathbf{0}_{(n+3)\times(n+3)}$. Now remembering that $\mathbf{M}_0 = -\boldsymbol{\phi}_{12}$ and, with $\phi_{d,d} = 0$, $\mathbf{B}_1 = \boldsymbol{\phi}_{12}\left[\mathbf{0}'_n, 1, -\rho, 1\right]$, we have that:

$$\boldsymbol{\Psi}_1 = \boldsymbol{\phi}_{12}\left(\left[\mathbf{0}'_n, 1, -\rho, 1\right]\boldsymbol{\Omega}_\xi - \boldsymbol{\Omega}'_{\xi\eta}\left[\mathbf{0}'_n, 1, -\rho, 1\right]'\boldsymbol{\phi}'_{12}\right). \tag{H.1}$$

It is therefore useful to partition the matrix $\boldsymbol{\Psi}_1$ as:

$$\boldsymbol{\Psi}_1 = \begin{bmatrix} \boldsymbol{\Psi}_1^{11} & \boldsymbol{\Psi}_1^{12} \\ \boldsymbol{\Psi}_1^{21} & \boldsymbol{\Psi}_1^{22} \end{bmatrix}$$

where $\boldsymbol{\Psi}_1^{22}$ is a 3×3 matrix of the variance of the residuals that are restricted by the CS restriction. Given the structure of $\boldsymbol{\Psi}_1$ in Equation (H.1), and defining $\mathbf{c} = \left[\phi_{rf,d}, \phi_{pd,d}, \rho\phi_{pd,d} - \phi_{rf,d}\right]'$, $\boldsymbol{\Omega}_\xi^{22}$ as the lower right 3×3 submatrix of the matrix $\boldsymbol{\Omega}_\xi$ and $\boldsymbol{\Omega}_{\xi\eta}^{12} = [0, 0, \sigma_\eta^2]'$, we have

that:

$$\Psi_1^{22} = \mathbf{c}\left\{[1, -\rho, 1]\,\Omega_\xi^{22} - [0, 0, \sigma_\eta^2]\,[1, -\rho, 1]'\,\mathbf{c}'\right\}$$

$$= \mathbf{c}\left(\begin{bmatrix} \Sigma_{rf,rf} - \rho\Sigma_{rf,pd} + \Sigma_{r,rf} \\ \Sigma_{rf,pd} - \rho\Sigma_{pd,pd} + \Sigma_{r,pd} \\ \Sigma_{r,rf} - \rho\Sigma_{r,pd} + \Sigma_{r,r} \end{bmatrix}' - \sigma_\eta^2 \mathbf{c}'\right)$$

$$= \mathbf{c}\left(\begin{bmatrix} \Sigma_{rf,d} \\ \Sigma_{pd,d} \\ \Sigma_{r,d} + \sigma_\eta^2 \end{bmatrix}' - \sigma_\eta^2 \mathbf{c}'\right)$$

$$= \mathbf{c}\begin{bmatrix} \Sigma_{rf,d} - \sigma_\eta^2\phi_{rf,d} \\ \Sigma_{pd,d} - \sigma_\eta^2\phi_{pd,d} \\ \Sigma_{r,d} + \sigma_\eta^2 - \sigma_\eta^2\left(\rho\phi_{pd,d} - \phi_{rf,d}\right) \end{bmatrix}'$$

for $\Psi_1^{22} = \mathbf{0}_{3\times3}$ with $\mathbf{c} \neq \mathbf{0}_{3\times1}$ (i.e., $\phi_{rf,d} \neq 0$ and $\phi_{pd,d} \neq 0$), one requires that $\Sigma_{rf,d} - \sigma_\eta^2\phi_{rf,d} = 0$, which implies that $\phi_{rf,d} = \Sigma_{rf,d}/\sigma_\eta^2$, $\Sigma_{pd,d} - \sigma_\eta^2\phi_{pd,d} = 0$, which implies that $\phi_{pd,d} = \Sigma_{pd,d}/\sigma_\eta^2$, and $\Sigma_{r,d} + \sigma_\eta^2 - \sigma_\eta^2\left(\rho\phi_{pd,d} - \phi_{rf,d}\right) = 0$ which implies that $\Sigma_{r,d} + \sigma_\eta^2 - \rho\Sigma_{pd,d} + \Sigma_{rf,d} = 0$. Since the CS restriction implies that $\Sigma_{r,d} = \Sigma_{d,d} - \Sigma_{rf,d} + \rho\Sigma_{p,pd}$, the last equality implies that $\Sigma_{d,d} = -\sigma_\eta^2$, which cannot be.

Now notice that:

$$\Psi_1^{12} = \phi_{w,d}\begin{bmatrix} \Sigma_{rf,d} - \sigma_\eta^2\phi_{rf,d} \\ \Sigma_{pd,d} - \sigma_\eta^2\phi_{pd,d} \\ \Sigma_{r,d} + \sigma_\eta^2 - \sigma_\eta^2\left(\rho\phi_{pd,d} - \phi_{rf,d}\right) \end{bmatrix}'$$

for $\Psi_1^{12} = \mathbf{0}_{3\times1}$, since $\Sigma_{r,d} + \sigma_\eta^2 - \rho\Sigma_{pd,d} + \Sigma_{rf,d} \neq 0$, we have that $\phi_{w,d} = \mathbf{0}_n$.

In the absence of approximation error, we have that:

$$\mathbf{\Psi}_1^{22} = \mathbf{c} \begin{bmatrix} \Sigma_{rf,d} \\ \Sigma_{pd,d} \\ \Sigma_{r,d} \end{bmatrix}'$$

Clearly unless $\Sigma_{rf,d} = \Sigma_{pd,d} = \Sigma_{r,d} = 0$ (which implies that $\Sigma_{d,d} = 0$ because of the CS restriction), the above matrix can be equal to $\mathbf{0}_{3\times3}$ with $\mathbf{c} \neq \mathbf{0}_{3\times1}$. The same argument can be used with $\mathbf{\Psi}_1^{12}$ to show that $\boldsymbol{\phi}_{w,d} = \mathbf{0}_n$. As before, it should be clear that if $\boldsymbol{\phi}_{12} \neq \mathbf{0}_{2\times1}$ (i.e. $\Sigma_{rf,d} = \Sigma_{pd,d} = \Sigma_{r,d} = \Sigma_{d,d} = 0$), $\mathbf{G}_2 = \mathbf{0}_{(n+3)\times(n+3)}$ implies that $\boldsymbol{\phi}_{21} = \mathbf{0}'_{n+3}$; hence, the dividend growth process is fixed to constant over time.