TI 2021-080/III
Tinbergen Institute Discussion Paper

Moments, Shocks and Spillovers in Markov-Switching VAR Models

Revision: January 2023

Erik Kole[1]
Dick van Dijk[1]

1 Erasmus University Rotterdam and Tinbergen Institute

Moments, Shocks and Spillovers in Markov-switching VAR Models*

Erik Kole†[1,2] and Dick van Dijk[1,2,3]

[1]Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
[2]Tinbergen Institute
[3]Erasmus Insititute of Management

December, 2022

Abstract

To investigate how economies, financial markets or institutions can deal with stress, we often analyze the effects of shocks conditional on being in a recession or a bear market. MSVAR models are perfectly suited for such analyses because they combine gradual movements with sudden regime switches. In this paper, we develop a comprehensive methodology to conduct these analyses. We derive first and second moments conditional only on the regime distribution and propose impulse response functions for both moments. By formulating the MSVAR as an extended linear non-Gaussian VAR, all results are available in closed-form. We illustrate our methods with an application to stock and bond return predictability. We show how forecasts of means, volatilities and (auto-)correlations depend on the regimes. The effect of shocks becomes highly nonlinear, and they propagate via different channels. During bear markets, shocks have stronger effects on means and volatilities and die out more slowly.

Keywords: Markov-switching VAR, moments, impulse response analysis, bull and bear markets
JEL classification: C32, C58, G01, G17

*We thank the Editor Serena Ng, the Associate Editor, two anonymous referees, Sílvia Gonçalves, Rutger-Jan Lange, Bram van Os, Richard Paap, Chen Zhou as well as participants in the NBER-NSF Time Series Conference (2022, Boston), SoFiE Annual Meeting (2022, Cambridge), EEA-ESEM conference (2021, virtual), the Annual Conference of the IAAE (2021, Rotterdam) and in seminars at Erasmus University Rotterdam and Pompeu Fabra University for helpful comments and suggestions.

†Corresponding author. Address: Burg. Oudlaan 50, Room ET-31, P.O. Box 1738, 3000DR Rotterdam, The Netherlands, Tel. +31 10 408 12 58. E-mail addresses `kole@ese.eur.nl`, `djvandijk@ese.eur.nl`

1 Introduction

Analyses of the economic vulnerabilities of countries, financial institutions or investments nowadays often take the form of investigating how they are affected by shocks, conditional on the economy or financial system being in a bad state. Also in the analysis of fiscal or monetary policy shocks, many authors show the relevance of conditioning on the state of the economy.[1] Markov-switching VAR (MSVAR) models are a promising class of models to conduct these analyses with. MSVAR models combine the gradual movement of economic and financial variables with sudden switches between a typically small number of regimes. For example, Hubrich and Tetlow (2015) propose an MSVAR model to capture the effect of financial crises on macro variables.

In this paper, we develop a comprehensive toolkit with new methods to conduct exactly these kinds of analyses with MSVAR models. In particular, we address the question how shocks affect the mean and variance of the observable variables $\boldsymbol{y}_t$ from the point in time t onward, given that the latent discrete regime process S_{t-1} is in a specific regime or has a given distribution at time $t-1$. To answer this question we need conditional expectations of the form $\mathrm{E}[y_{t+h}^j | S_{t-1}]$, $h \geq 0$, $j = 1, 2$. In words, we want to forecast the (squared) observable variables at time $t+h$ conditional on, for example, the stock market being bearish or the economy being recessionary at $t-1$, but without any past information about the observable variables. When the interest is in the effect of shocks in a particular state, one may typically wish to abstract from conditioning on specific past observations. Whereas the conditional expectations that use such information, that is $\mathrm{E}[y_{t+h}^j | S_{t-1}, \boldsymbol{y}_{t-1}]$, are relatively straightforward to derive, the conditional expectations that do not condition on $\boldsymbol{y}_{t-1}$ are more complicated. Applying the law of iterated expectations yields $\mathrm{E}[y_{t+h}^j | S_{t-1}] = \mathrm{E}\big[\mathrm{E}[y_{t+h}^j | S_{t-1}, \boldsymbol{y}_{t-1}]\big| S_{t-1}\big]$, which, due to the VAR structure, leads to a recursion over all paths for $\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots$ conditional on S_{t-1}. We derive these expectations in closed form, and show how to use them to compare the dynamics of an economy or financial market in different regimes, and to set up a formal impulse response analysis in the same way as for standard VAR models.

We base our analysis of moments, shocks and spillovers in MSVAR models on a VAR(1) model whose parameters switch according to a latent homogeneous first-order Markov chain with a fixed number of regimes, as in Bianchi (2016). The key part of our methods is formed by the joint specification of the processes followed by the level of the observable variables and their squares, and the latent state process. We show that this extended vector of variables follows a linear VAR(1) model with non-Gaussian innovations. Though non-Gaussian, this extended VAR model is Markovian, which is the driving force of our results. Despite the simple model specification, the resulting framework can be easily extended to higher-order VAR processes and more involved Markov chains.

We briefly summarize our four new theoretical results. First, we derive expressions for the expectations of this extended VAR process for different horizons conditional on a specific regime or regime distribution at time t. They allow us to calculate expectations $\mathrm{E}[\boldsymbol{y}_{t+h} | S_{t-1}]$, (co)variances

[1] See for example Tenreyro and Thwaites (2016) and the discussion in Ramey and Zubairy (2018).

$\text{Var}[\boldsymbol{y}_{t+h}|S_{t-1}]$ for $h \geq 0$, and autocovariances $\text{Cov}[\boldsymbol{y}_{t+h}, \boldsymbol{y}_{t+k}|S_{t-1}]$ for $h, k \geq 0$. Here we extend Timmermann (2000), who considers univariate processes with a slightly different specification than ours. We also extend Bianchi (2016), who derives moments conditional on a specific regime having prevailed for a long time, that is $\text{E}[\boldsymbol{y}_{t+h}|S_{t-1} = S_{t-2} = \ldots]$. His approach conditions on a single path of the regime process without any past regime switching and can therefore use the standard VAR expressions for unconditional moments, whereas ours accounts for all possible paths leading up to S_{t-1}.

Second, we present analytical expressions for the generalized impulse response functions (GI) for both the level of the observable variables and their squares. In line with Koop et al. (1996), we allow for different specifications of shocks as well as different sets of conditioning information, that is in- or excluding past observations. Because the model has Markov-switching features, the effects of shocks become time- and size-dependent. Our analytical expressions make it straightforward to analyze these responses and their dependence on the model parameters, the regime distribution or parameter uncertainty.

Third, we define the variance impulse response function to analyze the effect that shocks have on the future (co)variances. We show how they can be derived from the extended VAR specification that includes the squared process for the observable variables. In a standard VAR model, the (co)variances are entirely unaffected by shocks, and in Markov-switching models without VAR components, the effects at horizons $h > 0$ run completely via the updating of the forecast regime probabilities. To the contrary, in MSVAR models the interaction of the VAR and Markov-switching features makes the effect of shocks on future (co)variances larger and highly nonlinear.

As our final theoretical result we construct a Generalized Forecast Error Variance Decomposition (GFEVD) as in Lütkepohl (2005); Pesaran and Shin (1998). Because the effects of shocks depend on their sizes and the regime distribution, the GFEVD shows the same dependence. This result implies that spillover indexes in the style of Diebold and Yılmaz (2009, 2012, 2014) become time-varying as well, with large shocks having potentially different effects than small ones.

We demonstrate the empirical relevance of our theoretical results by analyzing the risk-return trade-off for stocks and bonds with the T-Bill rate and the dividend-to-price ratio as predictors. We base this part on an MSVAR model with one lag and two regimes. Though simple, this model accommodates both return predictability and the presence of regimes. Many authors have documented the importance of return predictability for long-term portfolio allocation (see Campbell and Viceira, 1999; Campbell et al., 2003; Barberis, 2000, amongst others). Also the presence and implications of regime switching have been well documented (see Ang and Bekaert, 2002; Guidolin and Timmermann, 2006a,b, 2007; Guidolin, 2011). Our model with 2 regimes and a VAR(1) component is supported by Guidolin and Ono (2006) who show that it outperforms models with only VAR or Markov-switching features.

In line with these earlier papers, we find a low- and a high-volatility regime. We use our theoretical results to determine the implied expected returns, which are high (low) for stocks but low (high) for bonds conditional on the low (high) volatility regime prevailing at t. However, conditioning on the low- or high-volatility regime having prevailed forever produces considerably different and potentially misleading means. For stocks, the high-volatility regime would falsely imply a higher mean than the low-volatility regime. Our analysis also shows how the predictability varies over the regimes.

We then investigate how shocks to the different variables at time t impact the expectations and (co)variances of stock and bond returns at different horizons and depend on the regime distribution at $t-1$. We show how the transmission of shocks is affected by the VAR and the Markov-switching parts of the model. The combined effects are highly non-linear, and can reinforce or dampen each other. We also find that shocks are more persistent than in the corresponding VAR or MS models.

We conclude that MSVAR models are a useful tool for investigating economic and financial processes under stress. Our proposed methods can be used to characterize the conditional distribution of the observable variables for any point in the future, taking only a specific current regime distribution as given. Similarly, they can also be used to investigate how the variables respond to shocks. Our empirical analysis shows that the combination of low and high-volatility regimes with return predictability leads to rich and interesting dynamics. Next to the correlation between stock and bond returns being higher starting from the high-volatility regime, we also find that predictability is stronger. Consequently, shocks die out more slowly, and have stronger and more prolonged effects on both the expectation and volatilities of returns.

Our results for the moments and impulse response functions extend Krolzig (2006) and Bianchi (2016). Krolzig (2006) analyzes expectations conditional on both past observations and the regime distribution, and first-order impulse responses to the structural innovations for a given regime path. Bianchi (2016) extends the focus to first and second moments that are conditional on past observations and the regime distribution. He conducts an impulse response analysis for first and second moments under the assumption that a regime has prevailed for a long time to show how responses depend on the prevailing regime in situations where regimes can be persistent (e.g., monetary policy regimes). Our results are complementary and apply to situations where regimes are less persistent, and where shocks hit an endogenous variable, leading to a contemporaneous update of the forecast probabilities for the regimes. We systematically show how regime switching makes the response nonlinear, and how the responses depend on the sign and size of the shock, and the information set at the time before the shock hits. We also extend Karamé (2010, 2012, 2015) who includes the effect of regime switching by simulations, as we show how the IRFs can be derived completely in closed form. Our approach differs from the local projection approach of Jordà (2005, 2009), as the shocks can trigger a regime switch in our case, though it requires a specification of the regime process (see also Gonçalves et al., 2022).

Our empirical analysis contributes to the large literature about the effect of regime switching and return predictability on the risk and return characteristics of assets for different horizons. We show that MSVAR models inherit the well-known effects of standard models with only Markov-switching or only VAR components. Our analysis of the implications for the risk-return trade-off complements Campbell and Viceira (2005) who investigate the term-structure of risk and return for VAR models and Taamouti (2012) for Markov-switching models without a VAR component. Next to exhibiting the features of both models, their combination makes the (co)variances respond to shocks in a way akin to GARCH models, which neither of the contributing parts exhibit.

The empirical analysis also shows how the closed-form expressions for the GIs enhance our understanding of the transmission of shocks. In particular, they clarify how the transmission of shocks within a regime combines with their effect on the forecast regime distribution. The insights of our imupulse response analysis of the risk-return trade-off carry over to the impact of shocks in present-value models of Campbell and Shiller (1988) as shown by Bianchi (2016, 2020) and macroeconomic models (see e.g. Hubrich and Tetlow, 2015; Hamilton, 2016). They are also relevant for analyses of (in)stability and contagion in the banking system (see e.g. Clerc et al., 2016; Cont and Schaanning, 2019) or the impact of shocks in the energy market (see Hou and Nguyen, 2018). These fields of application testify to the relevance and usefulness of the comprehensive framework we propose.

The remainder of this article is structured as follows. In Section 2 we introduce the general formulation of MSVAR models, and derive their moments. In Section 3 we propose a general framework for first- and second-order impulse response analysis. In Section 4 we apply these methods to study the risk and return characteristics of stocks and bonds. We conclude in section 5.

2 The MSVAR model and its moments

We consider a vector of n variables $\boldsymbol{y}_t$ that follow a VAR model of order 1, whose parameters are subject to regime switching. The switching results from a latent Markov chain S_t that can be in one out of m regimes, numbered 1 to m. We formulate the model as

$$\boldsymbol{y}_t = \boldsymbol{c}_{S_t} + \boldsymbol{\Phi}_{S_t}\boldsymbol{y}_{t-1} + \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{I}_n\right), \tag{1}$$

where $\boldsymbol{c}_{S_t}$ is an n-vector containing the regime-specific intercepts, $\boldsymbol{\Phi}_{S_t}$ is an $n \times n$ matrix with the regime-specific autoregressive coefficients, $\boldsymbol{\Lambda}_{S_t}$ is a regime-specific $n \times n$ matrix, and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. The n-vector $\boldsymbol{\varepsilon}_t$ contains the structural innovations. Conditional on the regime, the variance of the innovation in $\boldsymbol{y}_t$ is given by $\mathrm{Var}[\boldsymbol{y}_t|S_t, \boldsymbol{y}_{t-1}] = \boldsymbol{\Sigma}_{S_t} = \boldsymbol{\Lambda}_{S_t}\boldsymbol{\Lambda}'_{S_t}$. We assume that the structural innovations are independent over time, $\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+l}] = \mathbf{O}$, for all $l \neq 0$.

To prevent label switching of the regimes, we order the regimes ascendingly based on the variance of the innovation in y_{1t}, $\boldsymbol{\Sigma}_{11,i} < \boldsymbol{\Sigma}_{11,i+1}$ for $i = 1, 2, \ldots, m-1$. For identification of the shocks, we impose that $\boldsymbol{\Lambda}_i$ is lower triangular with $\boldsymbol{\Lambda}_{11,i} > 0$ for $i = 1, 2, \ldots, m$. When shocks are restricted to one particular variable, only that shock needs to be identified, and our Cholesky identification scheme can be relaxed (see for example Gonçalves et al., 2022). Our results for moments in this section also apply if the identification scheme is relaxed, and, more generally, to structural VARs with Markov-switching.

The regime process S_t follows a first-order Markov chain with transition matrix $\boldsymbol{P}$, where

$$p_{ij} = \Pr[S_t = i|S_{t-1} = j]. \tag{2}$$

Formulating the transition matrix such that the columns sum to one is convenient for the results in this paper. We assume that the Markov chain is irreducible and ergodic, and that the processes S_t and $\boldsymbol{\varepsilon}_{t+l}$

are independent for all l. We use $\xi_{it} = \Pr[S_t = i|\mathcal{I}_t]$ to denote the probability for state $i = 1, \ldots, m$ to occur at time t conditional on an information set $\mathcal{I}_t$, and collect these in the vector $\boldsymbol{\xi}_t$. $\bar{\boldsymbol{\xi}}$ with typical element $\bar{\xi}_i = \Pr[S_t = i]$ denotes the vector of ergodic probabilities satisfying $\boldsymbol{P}\bar{\boldsymbol{\xi}} = \bar{\boldsymbol{\xi}}$. The initial regime probabilities $\boldsymbol{\xi}_0$ can be part of the model specification or equated to the ergodic probabilities.

It is straightforward to introduce more lags in Equation (1). By writing the resulting higher-order VAR in its companion form our results for the VAR(1)-case can still be used. In a similar way, the regime process S_t can be extended to a higher-order Markov process, which can then be written as a first-order Markov chain over a larger set of states. It is also possible to adapt the coefficients in Equation (1) to different independent Markov chains. In many practical applications, researchers introduce separate chains for the (co)variances next to the intercept and VAR-parameters, see, for example, Hubrich and Tetlow (2015) and Bianchi (2016, 2020). These independent chains can be written as a single Markov chain consisting of all state combinations with appropriate restrictions on the transition matrix as shown by, for example, Hamilton and Lin (1996).

Models of this form have been studied extensively by Krolzig (2013), who introduces the shorthand notation MSIAH(m)-VAR(l) for a VAR of order l with Markov-switching in the Intercept, Autoregressive coefficients and Heteroskedasticity, driven by a first-order Markov chain with m states. The model in Equation (1) is hence an MSIAH(m)-VAR(1) model. Bianchi (2016) also studies the MSIAH(m)-VAR(1) model, and his approach to deriving moments is the starting point for our analysis. Timmermann (2000) focuses on a slightly different type of Markov-Switching model, in which the regime-specific mean $\mathrm{E}[\boldsymbol{y}_t|S_t]$ is part of the model specification instead of the regime-specific intercept in Equation (1). We derive the expression for the regime-specific mean in Section 2.4. Moreover, he focuses on the single-variable case.

The core of our methodology consists of an extended VAR formulation that encompasses the level and the quadratic processes of the observable variables, and the regime process. We show that it takes the form of a linear VAR model with non-Gaussian innovations. We derive explicit expressions for the VAR part as well as the innovations. Whereas results for the VAR part have been derived before in Krolzig (2006) and Bianchi (2016), the innovation part is new and necessary for the generalized impulse response analysis that we propose. To give a complete framework for analyzing the moments of MSVAR models, we restate some of the results of these earlier papers.

2.1 Extended VAR formulation for the level process

We rewrite the model given by Equations (1) and (2) to make explicit that the Markov chain implies a selection of the VAR coefficients from a larger but fixed set of coefficients. To do so, we define the random m-vector $\boldsymbol{s}_t$ with $s_{it} = I(S_t = i)$, where I denotes the indicator function. Hence, the ith element of $\boldsymbol{s}_t$ equals 1 if $S_t = i$ and zero otherwise. Consequently, we can formulate the Markov chain as a linear VAR(1) model (see Hamilton, 1994)

$$\boldsymbol{s}_t = \boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{u}_t, \tag{3}$$

where $\boldsymbol{u}_t$ is a martingale difference sequence (MDS). Its conditional variance is equal to

$$\mathrm{E}[\boldsymbol{u}_t\boldsymbol{u}_t'|\boldsymbol{s}_{t-1}] = \mathrm{diag}(\boldsymbol{P}\boldsymbol{s}_{t-1}) - \boldsymbol{P}\boldsymbol{s}_{t-1}\boldsymbol{s}_{t-1}'\boldsymbol{P}',$$

where the diag operator produces a diagonal matrix with the input vector on the diagonal.

We also define the bdiag operator as in Timmermann (2000) and Bianchi (2016), which for a set of $n \times l$ matrices $\boldsymbol{A}_i$, $i = 1, \ldots, m$ produces the block-diagonal $mn \times ml$-matrix

$$\boldsymbol{A} = \mathrm{bdiag}_{i=1}^{m}(\boldsymbol{A}_i) = \mathrm{bdiag}(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_m) = \begin{pmatrix} \boldsymbol{A}_1 & \mathbf{O}_{n\times l} & \cdots & \mathbf{O}_{n\times l} \\ \mathbf{O}_{n\times l} & \boldsymbol{A}_2 & \cdots & \mathbf{O}_{n\times l} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{n\times l} & \mathbf{O}_{n\times l} & \cdots & \boldsymbol{A}_m \end{pmatrix}. \tag{4}$$

The matrix $\boldsymbol{A}_i$ can be retrieved from $\boldsymbol{A}$ as

$$\boldsymbol{A}_i = (\boldsymbol{e}_i' \otimes \boldsymbol{I}_n)\boldsymbol{A}(\boldsymbol{e}_i \otimes \boldsymbol{I}_l), \tag{5}$$

where $\boldsymbol{e}_i$ is the unit vector of length m with a 1 at position i and zero elsewhere. Since $\boldsymbol{s}_t = \boldsymbol{e}_i$ if $S_t = i$, we can use this relation to write, for example, $\boldsymbol{c}_{S_t} = (\boldsymbol{s}_t \otimes \boldsymbol{I}_n)'\boldsymbol{C}\boldsymbol{s}_t$, where $\boldsymbol{C} = \mathrm{bdiag}_{i=1}^{m}(\boldsymbol{c}_i)$ which makes the Markov-Switching of the coefficients in Equation (1) explicit.

Next, we define the random vectors $\boldsymbol{y}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{y}_t$, which combines the latent state process S_t with the observable process $\boldsymbol{y}_t$, and $\tilde{\boldsymbol{y}}_t = \left(\boldsymbol{y}_t^{*\prime}, \boldsymbol{s}_t'\right)'$ by stacking $\boldsymbol{y}_t^*$ and $\boldsymbol{s}_t$. Then we can prove the following proposition. (All proofs are provided in the Supplementary Material).

Proposition 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Define $\boldsymbol{y}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{y}_t$ and $\tilde{\boldsymbol{y}}_t = \left(\boldsymbol{y}_t^{*\prime}, \boldsymbol{s}_t'\right)'$. Then $\boldsymbol{y}_t^*$ follows the process*

$$\boldsymbol{y}_t^* = \boldsymbol{C}\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\varepsilon}_t^*, \tag{6}$$

with $\boldsymbol{C} = \mathrm{bdiag}_{i=1}^{m}(\boldsymbol{c}_i)$, $\boldsymbol{\Phi} = \mathrm{bdiag}_{i=1}^{m}(\boldsymbol{\Phi}_i)$, and

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\Lambda}(\boldsymbol{P} \otimes \boldsymbol{I}_n)(\boldsymbol{s}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Lambda}(\boldsymbol{u}_t \otimes \boldsymbol{\varepsilon}_t),$$

with $\boldsymbol{\Lambda} = \mathrm{bdiag}_{i=1}^{m}(\boldsymbol{\Lambda}_i)$, and $\boldsymbol{u}_t$ as defined in Equation (3). $\tilde{\boldsymbol{y}}_t$ follows the process

$$\tilde{\boldsymbol{y}}_t = \begin{pmatrix} \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t, \tag{7}$$

with

$$\tilde{\boldsymbol{\Phi}} = \begin{pmatrix} \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & \boldsymbol{C}\boldsymbol{P} \\ \mathbf{O}_{m\times nm} & \boldsymbol{P} \end{pmatrix}.$$

and $\tilde{\boldsymbol{\varepsilon}}_t = (\boldsymbol{\varepsilon}_t^{*\prime}, \boldsymbol{u}_t')'$. Moreover, $\boldsymbol{\varepsilon}_t^*$ and $\tilde{\boldsymbol{\varepsilon}}_t$ are MDS.

Part of this proposition is due to Krolzig (2006), though he does not derive the expression for $\boldsymbol{\varepsilon}_t^*$. It shows that $\tilde{\boldsymbol{y}}_t$ follows a first-order linear VAR model with non-Gaussian innovations. Even though the innovations are non-Gaussian, the representation is quite convenient for analysing the properties of MSVAR models, because it is Markovian. For example, we can still write $\tilde{\boldsymbol{y}}_{t+h}$ with $h > 0$ as the sum of $\tilde{\boldsymbol{y}}_t$ and the innovations between t and $t+h$,

$$\tilde{\boldsymbol{y}}_{t+h}|\tilde{\boldsymbol{y}}_t = \sum_{j=0}^{h-1} \tilde{\boldsymbol{\Phi}}^j \tilde{\boldsymbol{\varepsilon}}_{t+h-j} + \tilde{\boldsymbol{\Phi}}^h \tilde{\boldsymbol{y}}_t.$$

This expression is useful for the calculation of expectations and impulse response analysis. For the latter we will also use the explicit expression for $\boldsymbol{\varepsilon}_t^*$.

The variable $\tilde{\boldsymbol{y}}_t$ contains all information of $\boldsymbol{y}_t$ and S_t. We can obtain $\boldsymbol{y}_t$ from $\tilde{\boldsymbol{y}}_t$ by summing the appropriate elements,

$$\boldsymbol{y}_t = \widetilde{\boldsymbol{G}}_{\boldsymbol{y}} \tilde{\boldsymbol{y}}_t, \tag{8}$$

where $\widetilde{\boldsymbol{G}}_{\boldsymbol{y}} = \left(\boldsymbol{G}_{\boldsymbol{y}},\ \mathbf{O}_{n\times m}\right)$ has dimensions $n \times (n+1)m$ and $\boldsymbol{G}_{\boldsymbol{y}} = \boldsymbol{\imath}_m' \otimes \boldsymbol{I}_n$, with $\boldsymbol{\imath}_m$ being an m-vector filled with ones. It also follows that $\boldsymbol{y}_t = \boldsymbol{G}_{\boldsymbol{y}} \boldsymbol{y}_t^*$. We obtain $\boldsymbol{s}_t$ by selecting the last m elements of $\tilde{\boldsymbol{y}}_t$, which we can write as

$$\boldsymbol{s}_t = \widetilde{\boldsymbol{G}}_{\boldsymbol{s}} \tilde{\boldsymbol{y}}_t, \tag{9}$$

with $\widetilde{\boldsymbol{G}}_{\boldsymbol{s}} = \left(\mathbf{O}_{m\times nm},\ \boldsymbol{I}_m\right)$. We also define $\widetilde{\boldsymbol{G}}_{y_i} = \boldsymbol{e}_i' \widetilde{\boldsymbol{G}}_{\boldsymbol{y}}$, $\boldsymbol{G}_{y_i} = \boldsymbol{e}_i' \boldsymbol{G}_{\boldsymbol{y}}$, and $\widetilde{\boldsymbol{G}}_{s_j} = \boldsymbol{e}_j' \widetilde{\boldsymbol{G}}_{\boldsymbol{s}}$ to obtain a particular variable y_i or regime s_j, where $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ denote the appropriate unit vectors.

Based on the Markovian property in Equation (7), we directly find the expectation of $\tilde{\boldsymbol{y}}_{t+h}$ for $h \geq 0$ conditional on $\boldsymbol{y}_t$ and state probabilities $\boldsymbol{\xi}_t$ as[2]

$$\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix}. \tag{10}$$

Using Equation (8) the conditional expectation of $\boldsymbol{y}_{t+h}$ follows as $\mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \widetilde{\boldsymbol{G}}_{\boldsymbol{y}} \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t]$. This result corresponds with eq. (5) in Bianchi (2016).[3] We can use the relation $\boldsymbol{y}_t = \boldsymbol{G}_{\boldsymbol{y}} \boldsymbol{y}_t^*$ to calculate the first moments from the recursion

$$\mathrm{E}[\boldsymbol{y}_{t+h}^*|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\, \mathrm{E}[\boldsymbol{y}_{t+h-1}^*|\boldsymbol{y}_t, \boldsymbol{\xi}_t] + \boldsymbol{C}\boldsymbol{P}^h \boldsymbol{\xi}_t \tag{11}$$

with initial condition $\mathrm{E}[\boldsymbol{y}_t^*|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t$, which is useful when a series of expectations is required.

The matrix $\tilde{\boldsymbol{\Phi}}$ is not convergent (ie., $\lim_{h\to\infty} \tilde{\boldsymbol{\Phi}}^h \neq \mathbf{O}$), because it has at least one eigenvalue equal to one. The eigenvalues of $\tilde{\boldsymbol{\Phi}}$ are given by the eigenvalues of $\boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)$ and $\boldsymbol{P}$, due to the particular block structure of $\tilde{\boldsymbol{\Phi}}$. Because the columns of $\boldsymbol{P}$ sum to one, $\boldsymbol{P}$ has at least one eigenvalue equal to one (see Hamilton, 1994, Ch. 22), and hence so has $\tilde{\boldsymbol{\Phi}}$. We discuss the conditions for covariance-stationarity of $\boldsymbol{y}_t$ in Section 2.3.

[2] For clarity, we explicitly include $\boldsymbol{\xi}_t$ as conditioning information even though it need not be the realization of a random variable. This notation enables us to distinguish cases in which information about the probability distribution of the regimes captured by $\boldsymbol{\xi}_t$ is or is not combined with the realization $\boldsymbol{y}_t$.

[3] Bianchi (2016) does not define the VAR of Equation (7) but starts from its expectation.

2.2 Extended VAR formulation for the level and quadratic process

To determine second moments and the effect of shocks on (co)variances and correlation, we need to analyze the quadratic process $\boldsymbol{z}_t = \boldsymbol{y}_t \otimes \boldsymbol{y}_t$. As before, we also define the random vectors $\boldsymbol{z}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{z}_t$, which now combines the latent state process S_t with the quadratic process $\boldsymbol{z}_t$, and $\tilde{\boldsymbol{z}}_t = \left(\boldsymbol{z}_t^{*\prime}, \boldsymbol{y}_t^{*\prime}, \boldsymbol{s}_t^{\prime}\right)^{\prime}$ by stacking $\boldsymbol{z}_t^*$, $\boldsymbol{y}_t^*$ and $\boldsymbol{s}_t$. We can then prove the following proposition.

Proposition 2. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Define $\boldsymbol{z}_t = \boldsymbol{y}_t \otimes \boldsymbol{y}_t$, $\boldsymbol{z}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{z}_t$ and $\tilde{\boldsymbol{z}}_t = \left(\boldsymbol{z}_t^{*\prime}, \boldsymbol{y}_t^{*\prime}, \boldsymbol{s}_t^{\prime}\right)^{\prime}$, with $\boldsymbol{y}_t^*$ as defined in Proposition 1. Then $\boldsymbol{z}_t$ follows the process*

$$\boldsymbol{z}_t = \boldsymbol{\gamma}_{S_t} + \boldsymbol{\omega}_{S_t} + \boldsymbol{\Psi}_{S_t}\boldsymbol{y}_{t-1} + \boldsymbol{\Upsilon}_{S_t}\boldsymbol{z}_{t-1} + \boldsymbol{\zeta}_t, \tag{12}$$

where $\boldsymbol{\gamma}_{S_t} = \boldsymbol{c}_{S_t} \otimes \boldsymbol{c}_{S_t}$, $\boldsymbol{\omega}_{S_t} = \operatorname{vec}(\boldsymbol{\Sigma}_{S_t})$, $\boldsymbol{\Psi}_{S_t} = \boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{c}_{S_t} + \boldsymbol{c}_{S_t} \otimes \boldsymbol{\Phi}_{S_t}$, $\boldsymbol{\Upsilon}_{S_t} = \boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{\Phi}_{S_t}$, and

$$\begin{aligned}\boldsymbol{\zeta}_t =& (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{c}_{S_t} + \boldsymbol{c}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})\boldsymbol{\varepsilon}_t + (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{\Phi}_{S_t})(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{y}_{t-1}) + \\ & (\boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})(\boldsymbol{y}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t - \operatorname{vec}(\boldsymbol{I}_n)).\end{aligned}$$

$\boldsymbol{z}_t^*$ follows the process

$$\boldsymbol{z}_t^* = (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\boldsymbol{z}_{t-1}^* + \boldsymbol{\zeta}_t^*, \tag{13}$$

with $\boldsymbol{\Gamma} = \operatorname{bdiag}_{i=1}^m(\gamma_i)$, $\boldsymbol{\Omega} = \operatorname{bdiag}_{i=1}^m(\boldsymbol{\omega}_i)$, $\boldsymbol{\Psi} = \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Psi}_i)$, and $\boldsymbol{\Upsilon} = \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Upsilon}_i)$, and

$$\begin{aligned}\boldsymbol{\zeta}_t^* =& (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t \otimes \boldsymbol{z}_{t-1}) + \\ & \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i \otimes \boldsymbol{c}_i + \boldsymbol{c}_i \otimes \boldsymbol{\Lambda}_i)(\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t) + \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Phi}_i)(\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t \otimes \boldsymbol{y}_{t-1}) + \\ & \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i \otimes \boldsymbol{\Lambda}_i)(\boldsymbol{s}_t \otimes \boldsymbol{y}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Lambda}_i)(\boldsymbol{s}_t \otimes (\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t - \operatorname{vec}(\boldsymbol{I}_n))).\end{aligned}$$

$\tilde{\boldsymbol{z}}_t$ follows the process

$$\tilde{\boldsymbol{z}}_t = \begin{pmatrix} \boldsymbol{z}_t^* \\ \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Upsilon}}\tilde{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{\zeta}}_t, \tag{14}$$

with

$$\tilde{\boldsymbol{\Upsilon}} = \begin{pmatrix} \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2}) & \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P} \\ \mathbf{O} & \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & \boldsymbol{C}\boldsymbol{P} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{P} \end{pmatrix},$$

and $\tilde{\boldsymbol{\zeta}}_t = (\boldsymbol{\zeta}_t^{*\prime}, \boldsymbol{\varepsilon}_t^{*\prime}, \boldsymbol{u}_t^{\prime})^{\prime}$, with $\boldsymbol{\varepsilon}_t^*$ as defined in Proposition 1. Moreover, $\boldsymbol{\zeta}_t$, $\boldsymbol{\zeta}_t^*$ and $\tilde{\boldsymbol{\zeta}}_t$ are MDS.

This proposition shows that $\tilde{\boldsymbol{z}}_t$, so the combination of the state variable, observable variables, and their squares, also follows a first-order linear VAR model with non-Gaussian innovations. Its Markovian property makes it again useful to analyze both first and second moments of MSVAR models. $\tilde{\boldsymbol{\Upsilon}}$ is also non-convergent because of the presence of $\boldsymbol{P}$.

The variable $\tilde{\boldsymbol{z}}_t$ contains all information of $\boldsymbol{z}_t$, $\boldsymbol{y}_t$ and S_t, which we can obtain by summations of the form,

$$\boldsymbol{a}_t = \tilde{\boldsymbol{H}}_{\boldsymbol{a}} \tilde{\boldsymbol{z}}_t, \qquad \text{for } \boldsymbol{a} = \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{s}, \tag{15}$$

where $\tilde{\boldsymbol{H}}_{\boldsymbol{z}} = \left(\boldsymbol{H}_{\boldsymbol{z}}, \mathbf{O}_{n^2\times m(n+1)}\right)$ with $\boldsymbol{H}_{\boldsymbol{z}} = \boldsymbol{\iota}_m' \otimes \boldsymbol{I}_{n^2}$, $\tilde{\boldsymbol{H}}_{\boldsymbol{y}} = (\mathbf{O}_{n\times mn^2}, \widetilde{\boldsymbol{G}}_{\boldsymbol{y}}) = (\mathbf{O}_{n\times mn^2}, \boldsymbol{G}_{\boldsymbol{y}}, \mathbf{O}_{n\times m})$, and $\tilde{\boldsymbol{H}}_{\boldsymbol{s}} = (\mathbf{O}_{m\times mn^2}, \widetilde{\boldsymbol{G}}_{\boldsymbol{s}}) = (\mathbf{O}_{m\times mn(n+1)}, \boldsymbol{I}_m)$. Also, $\boldsymbol{z}_t = \boldsymbol{H}_{\boldsymbol{z}} \boldsymbol{z}_t^*$. It means that we can analyze both second and first moments and shocks to it by analyzing $\tilde{\boldsymbol{z}}_t$, and we do not need to additionally analyze $\tilde{\boldsymbol{y}}_t$.

The conditional expectation of $\tilde{\boldsymbol{z}}_{t+h}$ for $h \geq 0$ conditional on $\boldsymbol{y}_t$ and state probabilities $\boldsymbol{\xi}_t$ follows directly as

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix}. \tag{16}$$

We extract second moments from this result as $\mathrm{E}[\boldsymbol{y}_{t+h} \otimes \boldsymbol{y}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \mathrm{E}[\boldsymbol{z}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}} E[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t]$, which corresponds with eq. (9) in Bianchi (2016), and first moments as $\mathrm{E}[\boldsymbol{y}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{y}} E[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t]$. The vectorized variance matrix then follows as

$$\begin{aligned} \operatorname{vec}\left(\mathrm{Var}[\boldsymbol{y}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t]\right) &= \mathrm{E}[\boldsymbol{y}_{t+h} \otimes \boldsymbol{y}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] - \mathrm{E}[\boldsymbol{y}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] \otimes \mathrm{E}[\boldsymbol{y}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] \\ &= \tilde{\boldsymbol{H}}_{\boldsymbol{z}} E[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] - \tilde{\boldsymbol{H}}_{\boldsymbol{y}} E[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] \otimes \tilde{\boldsymbol{H}}_{\boldsymbol{y}} E[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t]. \end{aligned} \tag{17}$$

When a series of expectations is required, the relation $\boldsymbol{z}_t = \boldsymbol{H}_{\boldsymbol{z}} \boldsymbol{z}_t^*$ can be used with the recursion

$$\begin{aligned} \mathrm{E}[\boldsymbol{z}^*_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] =& \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\, \mathrm{E}[\boldsymbol{z}^*_{t+h-1} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] + \\ & \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\, \mathrm{E}[\boldsymbol{y}^*_{t+h-1} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] + (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\, \mathrm{E}[\boldsymbol{s}_{t+h} | \boldsymbol{\xi}_t], \end{aligned} \tag{18}$$

in combination with Equation (11), and initial conditions $\mathrm{E}[\boldsymbol{z}^*_t | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \otimes \boldsymbol{y}_t$, $\mathrm{E}[\boldsymbol{y}^*_t | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t$.

We use a similar approach to determine autocovariances and covariances for different leads or lags. We define the lead processes of order $k \geq 0$, $\boldsymbol{z}_{t+k,t} = \boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t$, and combine the lead and level process with the latent state process, $\boldsymbol{z}^*_{t+k,t} = \boldsymbol{s}_{t+k} \otimes \boldsymbol{z}_{t+k,t}$ and $\boldsymbol{y}^*_{t+k,t} = \boldsymbol{s}_{t+k} \otimes \boldsymbol{y}_t$. Because $\boldsymbol{z}_{t+k,t} = \mathrm{vec}(\boldsymbol{y}_t \boldsymbol{y}'_{t+1})$, it gives rise to $\mathrm{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+1})$. The process for $\boldsymbol{z}_{t,t+k}$ can be derived by premultiplying $\boldsymbol{z}_{t+k,t}$ by the appropriate vectorized transpose matrix. We use them in the following proposition.

Proposition 3. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Define $\boldsymbol{z}_{t+k,t} = \boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t$, $\boldsymbol{z}^*_{t+k,t} = \boldsymbol{s}_{t+k} \otimes \boldsymbol{z}_{t+k,t}$ and $\boldsymbol{y}^*_{t+k,t} = \boldsymbol{s}_{t+k} \otimes \boldsymbol{y}_t$ for $h \geq 0$. The process defined by $\tilde{\boldsymbol{z}}_{t+k,t} = \left(\boldsymbol{z}^{*\prime}_{t+k,t}, \boldsymbol{y}^{*\prime}_{t+k,t}, \boldsymbol{s}'_{t+k}\right)'$ follows*

$$\tilde{\boldsymbol{z}}_{t+k,t} = \begin{pmatrix} \tilde{\boldsymbol{\Phi}} \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P} \end{pmatrix} \tilde{\boldsymbol{z}}_{t+k-1,t} + \begin{pmatrix} \tilde{\boldsymbol{\varepsilon}}_{t+k} \otimes \boldsymbol{y}_t \\ \boldsymbol{u}_{t+k} \end{pmatrix}, \quad k \geq 1 \tag{19}$$

with $\tilde{\boldsymbol{\Phi}}$ as in Proposition 1 and the second term a MDS.

This proposition shows that $\tilde{\boldsymbol{z}}_{t+k,t}$ also follows a first-order linear VAR model with non-Gaussian innovations. When $k = 0$, $\boldsymbol{y}^*_{t,t} = \boldsymbol{y}^*_t$, $\boldsymbol{z}_{t,t} = \boldsymbol{z}_t$ and $\boldsymbol{z}^*_{t,t} = \boldsymbol{z}^*_t$ result as defined in the previous propositions. The process $\boldsymbol{z}_{t+k,t}$ can be obtained by $\boldsymbol{z}_{t+k,t} = \tilde{\boldsymbol{H}}_{\boldsymbol{z}}\tilde{\boldsymbol{z}}_{t+k,t}$. The conditional expectation of $\tilde{\boldsymbol{z}}_{t+h+k,t+h}$ for $h > 0$ conditional on $\boldsymbol{y}_t$ and state probabilities $\boldsymbol{\xi}_t$ follows as

$$
\begin{aligned}
\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h+k,t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] &= \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] \\
&= \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix}.
\end{aligned} \tag{20}
$$

We can use the recursion in Equation (19) and the structure of $\tilde{\boldsymbol{\Phi}}$ to write the conditional expectation of $\boldsymbol{z}^*_{t+h+k,t+h}$ recursively as

$$
\mathrm{E}[\boldsymbol{z}^*_{t+h+k,t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] = (\boldsymbol{\Phi}\otimes\boldsymbol{I}_n)(\boldsymbol{P}\otimes\boldsymbol{I}_{n^2})\,\mathrm{E}[\boldsymbol{z}^*_{t+h+k-1,t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] + (\boldsymbol{C}\boldsymbol{P}^k\otimes\boldsymbol{I}_n)\,\mathrm{E}[\boldsymbol{y}^*_{t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t], \tag{21}
$$

where we use that $\boldsymbol{\Phi}(\boldsymbol{P}\otimes\boldsymbol{I}_n)\otimes\boldsymbol{I}_n = (\boldsymbol{\Phi}\otimes\boldsymbol{I}_n)(\boldsymbol{P}\otimes\boldsymbol{I}_n\otimes\boldsymbol{I}_n) = (\boldsymbol{\Phi}\otimes\boldsymbol{I}_n)(\boldsymbol{P}\otimes\boldsymbol{I}_{n^2})$, $\mathrm{E}[\boldsymbol{y}^*_{t+h+k-1,t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] = \mathrm{E}\big[\mathrm{E}[\boldsymbol{s}_{t+h+k-1}\otimes\boldsymbol{y}_{t+h}|\boldsymbol{y}^*_{t+h}]\big|\boldsymbol{y}_t,\boldsymbol{\xi}_t\big] = (\boldsymbol{P}^{k-1}\otimes\boldsymbol{I}_n)\,\mathrm{E}[\boldsymbol{y}^*_{t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t]$ and $\mathrm{E}[\boldsymbol{y}^*_{t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t]$ follows from Proposition 1. This result corresponds with proposition 3 in Bianchi (2016). When a Markov-switching model without a VAR component is considered (i.e. an MSIH, or MSI model), the first terms vanishes because $\boldsymbol{\Phi} = \mathbf{O}$, and the autocovariance structure is completely driven by the transition matrix.

We conclude this section by noting that along the lines of these propositions, it is possible to derive n^{th} order moments by recursively defining the n-times repeated Kronecker product $\boldsymbol{y}_t^{\otimes n} = \boldsymbol{y}_t\otimes\boldsymbol{y}_t^{\otimes(n-1)}$ with $\boldsymbol{y}_t^{\otimes 1} = \boldsymbol{y}_t$. The combination of the n processes $\boldsymbol{s}_t\otimes\boldsymbol{y}^{\otimes n}, \boldsymbol{s}_t\otimes\boldsymbol{y}^{\otimes(n-1)}, \ldots, \boldsymbol{y}^*_t$ and $\boldsymbol{s}_t$ can be shown to follow a first-order linear VAR model with non-Gaussian innovations. In particular, combining the results for third- and fourth-order moments with the first-and second-order moments can give insights into the skewness and kurtosis implied by the model. We leave such an analysis for future work.

2.3 Stationarity and unconditional moments

Stationarity of (V)ARMA processes with Markov switching has been investigated before by Yang (2000); Francq and Zakoïan (2001); Zhang and Stine (2001); Stelzer (2009) and Bianchi (2016). Based on their results, the MS-VAR process as specified in Equations (1) and (2) is second-order stationary if and only if the spectral radius of the matrix $\boldsymbol{\Upsilon}(\boldsymbol{P}\otimes\boldsymbol{I}_{n^2})$ is smaller than 1, where $\boldsymbol{\Upsilon}$ is defined in Proposition 2. When this condition is satisfied, the first and second moment exist, and follow as

$$
\bar{\boldsymbol{y}} = \lim_{t\to\infty}\mathrm{E}[\boldsymbol{y}_t|\boldsymbol{y}_0,\boldsymbol{\xi}_0] = \boldsymbol{G}_{\boldsymbol{y}}\bar{\boldsymbol{y}}^*, \tag{22}
$$

$$
\bar{\boldsymbol{z}} = \lim_{t\to\infty}\mathrm{E}[\boldsymbol{z}_t|\boldsymbol{y}_0,\boldsymbol{\xi}_0] = \boldsymbol{H}_{\boldsymbol{z}}\bar{\boldsymbol{z}}^*, \tag{23}
$$

where

$$\bar{\boldsymbol{y}}^* = \lim_{t\to\infty} \mathrm{E}[\boldsymbol{y}_t^*|\boldsymbol{y}_0, \boldsymbol{\xi}_0] = \left(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{P}\otimes\boldsymbol{I}_n)\right)^{-1}\boldsymbol{C}\bar{\boldsymbol{\xi}}. \tag{24}$$

$$\bar{\boldsymbol{z}}^* = \lim_{t\to\infty} \mathrm{E}[\boldsymbol{z}_t^*|\boldsymbol{y}_0, \boldsymbol{\xi}_0] = \left(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{P}\otimes\boldsymbol{I}_{n^2})\right)^{-1}\left((\boldsymbol{\Gamma}+\boldsymbol{\Omega})\bar{\boldsymbol{\xi}} + \boldsymbol{\Psi}(\boldsymbol{P}\otimes\boldsymbol{I}_n)\bar{\boldsymbol{y}}^*\right), \tag{25}$$

and $\bar{\boldsymbol{\xi}}$ are the ergodic probabilities of the regime process. The unconditional expectation of $\boldsymbol{z}_{t+k,t}$ then also exists and is given by

$$\bar{z}_k = \lim_{t\to\infty} \mathrm{E}[\boldsymbol{z}_{t+k,t}|\boldsymbol{y}_0, \boldsymbol{\xi}_0] = \boldsymbol{H}_{\boldsymbol{z}}\bar{\boldsymbol{z}}_k^*, \tag{26}$$

where $\bar{\boldsymbol{z}}_k^* = \lim_{t\to\infty} \mathrm{E}[\boldsymbol{z}_{t+k,t}^*|\boldsymbol{y}_0, \boldsymbol{\xi}_0]$ follows together with $\bar{\boldsymbol{y}}_k^* = \lim_{t\to\infty} \mathrm{E}[\boldsymbol{y}_{t+k,t}^*|\boldsymbol{y}_0, \boldsymbol{\xi}_0]$ from Equation (20) as

$$\begin{pmatrix}\bar{\boldsymbol{z}}_k^* \\ \bar{\boldsymbol{y}}_k^*\end{pmatrix} = \left(\tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n\right)\begin{pmatrix}\bar{\boldsymbol{z}}^* \\ \bar{\boldsymbol{y}}^*\end{pmatrix}. \tag{27}$$

Cavicchioli (2017a,b) derives conditions for and expressions of unconditional third and fourth order moments.

The condition for second-order stationarity depends on the multiplication of $\boldsymbol{\Upsilon}$ and $\boldsymbol{P}\otimes\boldsymbol{I}_{n^2}$ and is necessary and sufficient. It is not possible to reduce this condition to a set of regime-specific conditions. Moreover, Francq and Zakoïan (2001) show that within-regime stationarity (i.e., the spectral radius of each matrix $\boldsymbol{\Phi}_i$, $i = 1, \ldots, m$ is smaller than 1), is neither a necessary nor a sufficient condition, which they illustrate in their Example 5. Consequently, regimes can exhibit local non-stationary behavior, for example bubble formation, as long as they are not too persistent.

2.4 Regime-specific future moments

We now turn to first and second (future) moments when the current regime or regime probabilities are given. Specifically, we consider the situation where the prevailing regime or regime distribution is the only conditioning information, and the history of the regimes and past observable variables are not included. This framework is useful for analyses when one wants to assume that, say, the economy is a recession, or the stock market is bearish, without specifying particular values on key indicators, and without the assumption that the state process has been in that particular state indefinitely long. We conduct such analyses in the next section.

Mathematically, we need to calculate the expectation or variance of $\boldsymbol{y}_{t+h}$ for $h \geq 0$ given a particular regime S_t or regime distribution $\boldsymbol{\xi}_t$, but without any information about how the process arrived there. To derive these expectations, we define the "time-reversed" Markov chain, which describes the sequence of S_{t-1} given S_t. With the time-reversed Markov chain, we can determine the probabilities of the different sequences leading to S_t.

Definition 1. Let S_t be the irreducible, ergodic Markov chain defined in Equation (2). Then the corresponding time-reversed Markov chain is governed by the transition matrix $\boldsymbol{Q}$ with elements

$$q_{ij} = \Pr[S_{t-1} = i|S_t = j] = p_{ji}\frac{\bar{\xi}_i}{\bar{\xi}_j}, \tag{28}$$

where $\bar{\xi}_i = \Pr[S_t = i]$ denotes the ergodic probability for state i.

The expression for q_{ij} follows from the application of Bayes' rule. We can also write $\boldsymbol{Q} = \operatorname{diag}(\bar{\boldsymbol{\xi}})\boldsymbol{P}' \operatorname{diag}(\bar{\boldsymbol{\xi}})^{-1}$, which shows that the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}'$ are similar, and hence have the same characteristic equation. We use this definition in the following propositions.

Proposition 4. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and assume that it is second-order stationary. Let $\boldsymbol{\mu}_j = \mathrm{E}[\boldsymbol{y}_t|S_t = j]$, and stack these conditional expectations in the $mn \times 1$ vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \dots \boldsymbol{\mu}_m')'$. Then*

$$\boldsymbol{\mu} = \big(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\big)^{-1}\boldsymbol{c}, \tag{29}$$

where $\boldsymbol{c} = \big(\boldsymbol{c}_1', \dots, \boldsymbol{c}_m'\big)'$, $\boldsymbol{\Phi} = \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i)$, and $\boldsymbol{Q}$ is the transition matrix of the time-reversed Markov chain of the proces S_t. The expectation of $\boldsymbol{y}_t$ conditional on the state distribution $\boldsymbol{\xi}_t$ follows as

$$\mathrm{E}[\boldsymbol{y}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}, \tag{30}$$

The expectation of $\boldsymbol{y}_{t+h}$ for $h \geq 0$ conditional on the current state distribution $\boldsymbol{\xi}_t$ follows as $\mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{\xi}_t] = \widetilde{\boldsymbol{G}}_{\boldsymbol{y}} \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\xi}_t]$ with

$$\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} (\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{31}$$

with $\tilde{\boldsymbol{y}}_{t+h}$ and $\tilde{\boldsymbol{\Phi}}$ defined in Proposition 1 and $\widetilde{\boldsymbol{G}}_{\boldsymbol{y}}$ as in (8).

Proposition 5. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and assume that it is second-order stationary. Let $\boldsymbol{z}_t = \boldsymbol{y}_t \otimes \boldsymbol{y}_t$, $\boldsymbol{\mu}_j = \mathrm{E}[\boldsymbol{y}_t|S_t = j]$, $\boldsymbol{\nu}_j = \mathrm{E}[\boldsymbol{z}_t|S_t = j]$ with stacked versions $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \dots \boldsymbol{\mu}_m')'$ and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \dots \boldsymbol{\nu}_m')'$. Then*

$$\boldsymbol{\nu} = \big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}\big(\gamma + \boldsymbol{\omega} + \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}\big), \tag{32}$$

where $\boldsymbol{\gamma} = (\gamma_1', \dots, \gamma_m')'$ with $\boldsymbol{\gamma}_j = \boldsymbol{c}_j \otimes \boldsymbol{c}_j$, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1', \dots, \boldsymbol{\omega}_m')'$ with $\boldsymbol{\omega}_j = \operatorname{vec}(\boldsymbol{\Sigma}_j)$, $\boldsymbol{\Upsilon} = \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j)$, $\boldsymbol{\Psi} = \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{c}_j + \boldsymbol{c}_j \otimes \boldsymbol{\Phi}_j)$, $\boldsymbol{Q}$ is the transition matrix of the time-reversed Markov chain of the proces S_t, and $\boldsymbol{\mu}$ is given in Proposition 4. The expectation of $\boldsymbol{z}_t$ conditional on the state distribution $\boldsymbol{\xi}_t$ follows as

$$\mathrm{E}[\boldsymbol{z}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}. \tag{33}$$

The expectation of $\boldsymbol{z}_{t+h}$ for $h \geq 0$ conditional on the current state distribution $\boldsymbol{\xi}_t$ follows as $\mathrm{E}[\boldsymbol{z}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t]$ with

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} (\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu} \\ (\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{34}$$

with $\tilde{\boldsymbol{z}}_{t+h}$ and $\tilde{\boldsymbol{\Upsilon}}$ defined in Proposition 2.

These propositions complement the results of Bianchi (2016). He derives conditional steady states of the form $\lim_{\tau\to\infty} \mathrm{E}[\boldsymbol{y}_{t+h}|S_t = \cdots = S_{t-\tau} = j]$ and $\lim_{\tau\to\infty} \mathrm{E}[\boldsymbol{z}_{t+h}|S_t = \cdots = S_{t-\tau} = j]$, which can be interpreted as a situation in which regime j has already prevailed for a long time. Such a situation occurs when regimes are strongly persistent and regime switches are infrequent. Our results give the expectations conditional on information about S_t, but without any knowledge or assumptions about its past, and without past observations. They are relevant when regimes are less persistent and switches occur more frequently.

The propositions are also related to Timmermann (2000), who considers state-dependent autoregressive dynamics in his section 4 of the form

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{\Phi}_{S_{t-1}}(\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_{S_{t-1}}) + \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t,$$

though limited to the one-dimensional case. The main differences with our results are that the dynamics in the AR-part of the model depend on the current state in our model (that is, $\boldsymbol{\Phi}_{S_t}$ in Equation (1)), whereas they depend on the prior state in his, and that the autoregressive part is driven by past values $\boldsymbol{y}_{t-1}$ in our specification, and by deviations of past values from state-specific means $\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_{S_{t-1}}$ in his. As a consequence, the expectations $\boldsymbol{\mu}_j = \mathrm{E}[\boldsymbol{y}_t|S_t = j]$ are part of the model specification. The expressions for the second moments and their derivation are similar in spirit to our propositions and proofs.

Combining the results of Proposition 3 with Proposition 5 completes the necessary building blocks to analyze the regime-specific autocovariances, for which we need the expectation of (future values of) the process $\boldsymbol{z}_{t+k,t} = \boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t$ conditional on the state process.

Proposition 6. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and assume that it is second-order stationary. Define $\boldsymbol{z}_{t+k,t} = \boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t$. Then the expectation of $\boldsymbol{z}_{t+h+k,t+h}$ for $h \geq 0$ conditional on the current state distribution $\boldsymbol{\xi}_t$ follows as*

$$\mathrm{E}[\boldsymbol{z}_{t+h+k,t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}} \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t], \tag{35}$$

with $\boldsymbol{\Phi} = \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i)$ and $\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t]$ as in Proposition 5.

The expression in this proposition concerns the general case $h \geq 0$, and simplifies when $h = 0$. In line with the results based on Proposition 3, we can calculate $\mathrm{E}[\boldsymbol{z}_{t+k,t}|\boldsymbol{\xi}_t] = \boldsymbol{H}_{\boldsymbol{z}}\, \mathrm{E}[\boldsymbol{z}^*_{t+k,t}|\boldsymbol{\xi}_t]$ with the recursion

$$\mathrm{E}[\boldsymbol{z}^*_{t+k,t}|\boldsymbol{\xi}_t] = (\boldsymbol{\Phi} \otimes \boldsymbol{I}_n)(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\, \mathrm{E}[\boldsymbol{z}^*_{t+k-1,t}|\boldsymbol{\xi}_t] + (\boldsymbol{C}\boldsymbol{P}^k \otimes \boldsymbol{I}_n)\, \mathrm{E}[\boldsymbol{y}^*_t|\boldsymbol{\xi}_t], \tag{36}$$

with $\mathrm{E}[\boldsymbol{z}^*_{t,t}|\boldsymbol{\xi}_t] = \mathrm{E}[\boldsymbol{z}^*_t|\boldsymbol{\xi}_t] = (\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}$ and $\mathrm{E}[\boldsymbol{y}^*_t|\boldsymbol{\xi}_t] = (\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n)\boldsymbol{\mu}$.

The combination of Proposition 4 and Proposition 6 gives the autocovariances as

$$\mathrm{Cov}[\boldsymbol{y}_{t+h}, \boldsymbol{y}_{t+h+k}|\boldsymbol{\xi}_t] = \mathrm{E}[\boldsymbol{y}_{t+h+k}\boldsymbol{y}'_{t+h}|\boldsymbol{\xi}_t] - \mathrm{E}[\boldsymbol{y}_{t+h+k}|\boldsymbol{\xi}_t]\, \mathrm{E}[\boldsymbol{y}'_{t+h}|\boldsymbol{\xi}_t]. \tag{37}$$

The autocorrelations can be found by scaling them by the appropriate variances based on the conditional expectations of the squared processes in Proposition 5.

The propositions of this section enable us to derive the autocovariance structure of MSIH-models as a special case.

Corollary 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2) with $\Phi_i = \mathbf{O}$ for $i = 1, \ldots, m$, and assume that it is second-order stationary. Then*

$$\operatorname{vec}\bigl(\operatorname{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t]\bigr) = (\boldsymbol{\iota}_m' \otimes \boldsymbol{I}_{n^2})\bigl(\boldsymbol{C}\boldsymbol{P}^k\bigl(\operatorname{diag}(\boldsymbol{\xi}_t) - \boldsymbol{\xi}_t\boldsymbol{\xi}_t'\bigr) \otimes \boldsymbol{I}_n\bigr)\boldsymbol{\mu}. \tag{38}$$

This corollary shows that even when Markov-switching models do not have a (V)AR-component in their specification, the switching component leads to non-zero autocovariance. When $\boldsymbol{\xi}_t = \boldsymbol{e}_i$, that is, regime i occurs with probability 1, $\boldsymbol{\xi}_t\boldsymbol{\xi}_t' = \boldsymbol{e}_i\boldsymbol{e}_i' = \operatorname{diag}(\boldsymbol{e}_i)$, and hence $\operatorname{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t = \boldsymbol{e}_i] = \mathbf{O}$.

3 Impulse response analysis

The results of the previous section enable an analysis of the static properties of MSVAR models. To further understand what the implications of these models are for the behavior of the variables, we now turn to their dynamic properties. Therefore we present a framework to analyze how shocks propagate in MSVAR models. We use the extended VAR formulation of the previous section to derive the first-order as well as the second-order impulse response functions.

3.1 The Generalized Impulse Response Function

The VAR-representation in Proposition 1 fits naturally with the definition of the Generalized Impulse Response Function (GI) of Koop et al. (1996),

$$GI_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h\, \mathrm{E}[\tilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}], \tag{39}$$

where h gives the horizon, and $\mathcal{I}_{t-1}$ denotes the information set at time $t-1$. We specify the GI for the extended process $\tilde{\boldsymbol{y}}$, from which we can easily derive the GI for $\boldsymbol{y}$ and $\boldsymbol{s}$, using

$$GI_{\boldsymbol{a}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \widetilde{\boldsymbol{G}}_{\boldsymbol{a}}\, GI_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}), \quad \text{for } \boldsymbol{a} = \boldsymbol{y}, \boldsymbol{s}, \tag{40}$$

with $\widetilde{\boldsymbol{G}}_{\boldsymbol{a}}$ as in Equations (8) and (9).

As pointed out by Koop et al. (1996) and Pesaran and Shin (1998), this setting is particularly useful when shocks are endogenous. In our empirical application, we analyze how shocks to asset returns or state variables at t affect asset returns h periods ahead, conditional on the information set $\mathcal{I}_{t-1}$. In the baseline setting, no information about shocks is available, so they are integrated out. When one is interested in the effect of a exogenous shock, e.g., a particular policy decision, a baseline setting where the shock is absent as in Gallant et al. (1993) may be more relevant.

Our approach allows for three specifications of the information set. In the first one, it contains both the most recent observation $\boldsymbol{y}_{t-1}$ and a given set of probabilities for each regime $\boldsymbol{\xi}_{t-1}$. In the second one, a series of past observations $\mathcal{Y}_{t-1} = \{\boldsymbol{y}_\tau\}_{\tau=0}^{t-1}$ is present, and we calculate $\boldsymbol{\xi}_{t-1} = \mathrm{E}[\boldsymbol{s}_{t-1}|\mathcal{Y}_{t-1}, \boldsymbol{\xi}_0]$ or $\boldsymbol{\xi}_{t-1} = \mathrm{E}[\boldsymbol{s}_{t-1}|\mathcal{Y}_{t-1}]$ if $\boldsymbol{\xi}_0$ is not specified. In the third one, only the set of probabilities for each regime $\boldsymbol{\xi}_{t-1}$ is given, and we use $\mathrm{E}[\boldsymbol{y}_{t-1}|\boldsymbol{\xi}_{t-1}]$ for $\boldsymbol{y}_{t-1}$. To simplify our notation, we assume in this section that the information set is given in terms of $\boldsymbol{y}_{t-1}$ and $\boldsymbol{\xi}_{t-1}$, though the one may actually be calculated as the expectation of (a series of) the other.

In the generalized impulse response analysis typically one or a few of the shocks are specified. In an MS-VAR model, there are three ways to specify the shocks. Following directly from the model, shocks can be specified in terms of the regime innovations $\boldsymbol{u}_t$ and the structural innovations ε_{it}.[4] They can also be specified with respect to the observed variables, which we denote by $\eta_{it} = y_{it} - \mathrm{E}[y_{it}|\mathcal{I}_{t-1}]$. As a consequence, we define three GIs depending on the type of shocks.

Proposition 7. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{y}}_t$ be defined by Equation (39) and the results in Proposition 1. Let the vector $\boldsymbol{y}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\mathrm{E}[\boldsymbol{y}_{t-1}|\mathcal{I}_{t-1}]$. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\mathrm{E}[\boldsymbol{s}_{t-1}|\mathcal{I}_{t-1}]$. Let the matrices $\boldsymbol{C}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}$, and $\tilde{\boldsymbol{\Phi}}$ be defined as in Proposition 1. When the shock originates from the regime process, the corresponding GI satisfies*

$$GI_{\tilde{\boldsymbol{y}}}^{\boldsymbol{u}}(h, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \emptyset, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \tag{41}$$

When the shock is specified in terms of the structural innovation ε_{it}, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{y}}}^{\varepsilon_i}(h, \varepsilon_{it}, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, \emptyset, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}. \tag{42}$$

When the shock is specified as $\eta_{it} = y_{it} - \mathrm{E}[y_{it}|\mathcal{I}_{t-1}]$, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{y}}}^{y_i}(h, \eta_{it}, \mathcal{I}_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|y_{it}, \mathcal{I}_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, \mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] \end{pmatrix}. \tag{43}$$

The second conditional expectation satisfies $\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] = \mathrm{E}[\boldsymbol{s}_t|y_{it}, \mathcal{I}_{t-1}] - \mathrm{E}[\boldsymbol{s}_t|\mathcal{I}_{t-1}]$ with

$$\begin{aligned} \mathrm{E}[\boldsymbol{s}_t|\mathcal{I}_{t-1}] &= \boldsymbol{P}\boldsymbol{\xi}_{t-1}, \\ \mathrm{E}[\boldsymbol{s}_t|y_{it}, \mathcal{I}_{t-1}] &= \frac{1}{\boldsymbol{f}'\boldsymbol{P}\boldsymbol{\xi}_{t-1}} \boldsymbol{f} \odot \boldsymbol{P}\boldsymbol{\xi}_{t-1}, \end{aligned}$$

where $\boldsymbol{f}$ is a vector of size m whose element j is equal to the pdf of the marginal distribution of y_{it} under regime j, $y_{it}|S_t = j \sim \mathrm{N}(\boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j\boldsymbol{y}_{t-1}), \boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i)$. The first conditional expectation satisfies

$$\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, \mathcal{I}_{t-1}] = \boldsymbol{C}\,\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] + \boldsymbol{\Phi}(\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Lambda}\,\mathrm{E}[\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t|y_{it}, \mathcal{I}_{t-1}],$$

[4] The vector $\boldsymbol{u}_t$ cannot be chosen freely but should satisfy the restriction that $\boldsymbol{P}\boldsymbol{\xi}_{t-1} + \boldsymbol{u}_t$ is in the unit simplex. A necessary condition is that $\sum_{j=1}^m u_i = 0$.

with last term

$$\mathrm{E}[\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t | y_{it}, \mathcal{I}_{t-1}] = \begin{pmatrix} \mathrm{E}\big[s_{1t}\big|y_{it}, \mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = 1, \mathcal{I}_{t-1}\big] \\ \vdots \\ \mathrm{E}\big[s_{mt}\big|y_{it}, \mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = m, \mathcal{I}_{t-1}\big] \end{pmatrix},$$

and

$$\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = j, \mathcal{I}_{t-1}\big] = \boldsymbol{\Lambda}_j^{-1}\left(\frac{y_{it} - \boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j \boldsymbol{y}_{t-1})}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j \boldsymbol{e}_i}\boldsymbol{\Sigma}_j \boldsymbol{e}_i\right).$$

The function $GI_{\tilde{\boldsymbol{y}}}^{\boldsymbol{u}}$ shows the effect of shocks to the regime process. It can be used to assess the consequences of a switch to a particular regime. If the switch is to a particular regime j, ie. $\boldsymbol{s}_t = \boldsymbol{e}_j$, we can substitute $\boldsymbol{u}_t = \boldsymbol{e}_j - \boldsymbol{P}\boldsymbol{\xi}_{t-1}$. The GI for this type of shock is also presented by Krolzig (2006). When the regimes only affect the covariance matrices via $\boldsymbol{\Lambda}_{S_t}$, shocks to the regime process have effects that are comparable to the volatility shocks in Gorodnichenko and Ng (2017).

The function $GI_{\tilde{\boldsymbol{y}}}^{\varepsilon_i}$ is in line with the traditional framework of Sims (1980). In our setting, shocks in ε_{it} translate to different contemporaneous effects on $\boldsymbol{y}_t$ in the different regimes depending on the regime-specific matrices $\boldsymbol{\Lambda}_{S_t}$. Specifying shocks that are structurally identified has become popular in macroeconomic analyses. Identification can stem from a structural model such as the Markov-Switching DSGE model in Bianchi (2016), or by imposing additional restrictions in a structural VAR as in Karamé (2015). As in Bianchi (2016), the extended VAR specification enables us to express $GI_{\tilde{\boldsymbol{y}}}^{\varepsilon_i}$ in closed form in Equation (42), whereas Karamé (2015) reverts to simulations to include the possibility of future regime switches.

The function $GI_{\tilde{\boldsymbol{y}}}^{y_i}$ shows that shocks specified with respect to a particular variable y_i generate a contemporaneous response in two ways. The term $\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}]$ captures the effect of the shock on the inference about the regime process. Though $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are independent, $\boldsymbol{y}_t$ and $\boldsymbol{u}_t$ are not, and hence $\mathrm{E}\big[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}\big]$ is not necessarily zero. The effect on the inferred regime has a direct effect on the expectation of the regime specific innovations $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, \mathcal{I}_{t-1}]$. However, shocks in y_{it} may be correlated with shocks in the other variables, also depending on the regime, and this effect is captured by the term $\mathrm{E}[\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t|y_{it}, \mathcal{I}_{t-1}]$. The proposition shows how to derive the effects for the different regimes.

Specifying shocks in terms of the observable variables $\boldsymbol{y}_t$ has been proposed by Pesaran and Shin (1998). In this specification shocks can be correlated and are not structurally identified contrary to the first two GIs. This specification is more common in settings outside macro such as finance, for example to determine the connectedness of (financial) variables in a network and spillovers between them based on the Generalized Forecast Error Variance Decomposition. Diebold and Yılmaz (2014) specifically advocate this approach because of its "comparatively agnostic data-based spirit" (p. 121). Other examples are Diebold and Yılmaz (2015); Barigozzi et al. (2021); Miao et al. (2022) and Chen and Schienle (2022). We discuss these variance decompositions in Section 3.3.

Our set of GI functions complement the impulse response analyses proposed by other authors for Markov-switching VAR models. The differences relate to the information about the regimes at the time of the shock up to the horizon h and the specification of the shock. In our setup, a shock can be specified in three ways, occurs at time t and information about the regimes pertains to $t-1$, hence the regime in which the shock occurs and the further regime path are not known. To the contrary, in Ehrmann et al. (2003) shocks can only occur in the structural innovations, the regime at the time of the shock is known, and assumed to prevail up to the forecast horizon h. Karamé (2010, 2012) relaxes this last assumption, and allows for regime switching after the time of the shock, whereas Karamé (2015) also removes the assumption that the regime at the time of the shock is known. Krolzig (2006) assumes that the regime at the time of the shock is known, but specifies the shock in one variable while assuming that the other variables are not shocked. The first-order GI in Bianchi (2016) is in line with Karamé (2010), though with the additional assumption that the regime process has spent "a significant amount of time" in a particular regime. The effect that shocks have on the inference of the regime process is also present in Bianchi (2016, 2020)'s application of MSVAR models to the present value decomposition of Campbell and Shiller (1988). In this decomposition, changes in excess returns can be related to discount rate or cashflow news. In the MSVAR setting, both shocks may propagate differently in a particular regime, but they also lead to an update of the regime distribution.

3.2 Second-order responses

Because of its nonlinear nature, shocks in a Markov-switching model also affect higher-order moments, whereas these are unaffected in the setting of a linear VAR model. A shock may signal an increase or a decrease of the future variance of the system, depending on the information set as well as the sign and the size of the shock. The framework we have developed so far allows for a straightforward analysis of these effects.

The VAR-representation for the extended squared process in Proposition 2 gives rise to

$$GI_{\tilde{\boldsymbol{z}}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Upsilon}}^h \tilde{\boldsymbol{\zeta}}, \tag{44}$$

which extends Equation (39). It is important to note that $\tilde{\boldsymbol{\zeta}}_t$ results as a combination of the innovations in the state probabilities $\boldsymbol{u}_t$ and the structural innovations $\boldsymbol{\varepsilon}_t$. Because $\tilde{\boldsymbol{z}}_t$ contains all information regarding the level, squared and state processes $\boldsymbol{z}_t$, $\boldsymbol{y}_t$, and S_t, we can derive their GIs from $GI_{\tilde{\boldsymbol{z}}}$ by

$$GI_{\boldsymbol{a}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{H}}_{\boldsymbol{a}} GI_{\tilde{\boldsymbol{z}}}(h, \tilde{\boldsymbol{\zeta}}_t, \mathcal{I}_{t-1}), \quad \text{for } \boldsymbol{a} = \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{s}, \tag{45}$$

as in Equation (15). As an extension of Proposition 7 for the first-order impulse responses, we define separate GIs for $\tilde{z}$ based on the origin of shocks.

Proposition 8. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{z}}_t$ be defined by Equation (44) and the results in Proposition 2. Let the vector $\boldsymbol{y}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\mathrm{E}[\boldsymbol{y}_{t-1}|\mathcal{I}_{t-1}]$. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\mathrm{E}[\boldsymbol{s}_{t-1}|\mathcal{I}_{t-1}]$. Let the matrices $\boldsymbol{C}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}$, and $\tilde{\boldsymbol{\Phi}}$*

be defined as in Proposition 1, and $\boldsymbol{\Gamma}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Upsilon}$ and $\tilde{\boldsymbol{\Upsilon}}$ as in Proposition 2. When the shock originates from the regime process, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{z}}}^{\boldsymbol{u}}(h, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \emptyset, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t \otimes \boldsymbol{z}_{t-1}) \\ \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \quad (46)$$

When the shock is specified in terms of the innovation ε_{it}, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{z}}}^{\varepsilon_i}(h, \varepsilon_{it}, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \varepsilon_{it}, \emptyset, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\zeta}_t^* | \varepsilon_{it}, \mathcal{I}_{t-1}] \\ \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}, \quad (47)$$

with

$$\begin{aligned}\mathrm{E}[\boldsymbol{\zeta}_t^* | \varepsilon_{it}, \mathcal{I}_{t-1}] = & \varepsilon_{it} \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Lambda}_j \otimes \boldsymbol{c}_j + \boldsymbol{c}_j \otimes \boldsymbol{\Lambda}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) + \\ & \varepsilon_{it} \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Lambda}_j \otimes \boldsymbol{\Phi}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i \otimes \boldsymbol{y}_{t-1}) + \\ & \varepsilon_{it} \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Lambda}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{y}_{t-1} \otimes \boldsymbol{e}_i) + \\ & (\varepsilon_{it}^2 - 1) \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Lambda}_j \otimes \boldsymbol{\Lambda}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i \otimes \boldsymbol{e}_i).\end{aligned}$$

When the shock is specified as $\eta_{it} = y_{it} - \mathrm{E}[y_{it}|\mathcal{I}_{t-1}]$, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{z}}}^{y_i}(h, \eta_{it}, \mathcal{I}_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|y_{it}, \mathcal{I}_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{z}_t^*|y_{it}, \mathcal{I}_{t-1}] - \mathrm{E}[\boldsymbol{z}_t^*|\mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, \mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] \end{pmatrix}, \quad (48)$$

where the last two conditional expectations have been defined in Proposition 7, and

$$\mathrm{E}[\boldsymbol{z}_t^*|y_{it}, \mathcal{I}_{t-1}] = \begin{pmatrix} \mathrm{E}\big[s_{1t}|y_{it}, \mathcal{I}_{t-1}\big]\, \mathrm{E}\big[\boldsymbol{z}_t|y_{it}, S_t = 1\big] \\ \vdots \\ \mathrm{E}\big[s_{mt}|y_{it}, \mathcal{I}_{t-1}\big]\, \mathrm{E}\big[\boldsymbol{z}_t|y_{it}, S_t = m\big] \end{pmatrix}$$

and

$$\begin{aligned}\mathrm{E}[\boldsymbol{z}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}] = & \operatorname{vec}\left(\mathrm{Var}[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}]\right) - \\ & \mathrm{E}[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}] \otimes \mathrm{E}[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}].\end{aligned}$$

with

$$\begin{aligned}\mathrm{E}\big[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}\big] &= \mathrm{E}\big[\boldsymbol{y}_t|S_t = j, \mathcal{I}_{t-1}\big] + \frac{y_{it} - \boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j\boldsymbol{y}_{t-1})}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i}\boldsymbol{\Sigma}_j\boldsymbol{e}_i, \\ \mathrm{Var}[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}] &= \boldsymbol{\Sigma}_j - \frac{1}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i}\boldsymbol{\Sigma}_j\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Sigma}_j.\end{aligned}$$

Whereas many authors have studied first-order GIs, Bianchi (2016) is the first to address second-order responses. Similar to his first-order GI, he assumes that the shock occurs in a structural innovation, that the regime at the time of the shock is known, and that the process has not encountered a regime switch before the shock. In our proposition, we extend his analysis to the case where the regime at the time of the shock is unknown, and where the shock can occur in the regime process or an observable variable. We do not make the assumption that the process has been in a particular regime for an infinitely long time.

The results for the squared process $\boldsymbol{z}_t$ are of course mostly interesting to determine how shocks influence the (co)variance or correlation of the different variables. Therefore, we introduce the variance impulse response function (VI)

$$VI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \mathrm{Var}[\boldsymbol{y}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}] - \mathrm{Var}[\boldsymbol{y}_{t+h}|\mathcal{I}_{t-1}] \tag{49}$$

It is related to $GI_{\tilde{\boldsymbol{z}}}$ by

$$\begin{aligned}
&\mathrm{vec}\, VI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) \\
&\quad = \mathrm{E}[\boldsymbol{z}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}] - \mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}] \otimes \mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}] - \\
&\qquad (\mathrm{E}[\boldsymbol{z}_{t+h}|\mathcal{I}_{t-1}] - \mathrm{E}[\boldsymbol{y}_{t+h}|\mathcal{I}_{t-1}] \otimes \mathrm{E}[\boldsymbol{y}_{t+h}|\mathcal{I}_{t-1}]) \\
&\quad = GI_{\boldsymbol{z}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) - GI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) \otimes GI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) - \\
&\qquad \mathrm{E}[\boldsymbol{y}_{t+h}|\mathcal{I}_{t-1}] \otimes GI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) - GI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, \mathcal{I}_{t-1}) \otimes \mathrm{E}[\boldsymbol{y}_{t+h}|\mathcal{I}_{t-1}].
\end{aligned}$$

Variance or volatility impulse responses show up for any model of heteroskedasticity. Hafner and Herwartz (2006) introduce the concept in relation to multivariate GARCH models.

3.3 Conditional Variance Decompositions

We can analyze the dynamics of an MS-VAR model by decomposing the variance of the forecast error (see Lütkepohl, 2005; Pesaran and Shin, 1998), and transforming the result into spillover indexes developed by Diebold and Yılmaz (2009, 2012). However, in contrast to linear Gaussian VAR models, regime switching models produce decompositions and indexes that are both time-varying and depend on the size of the shocks. The GIs in Equations (41) to (43) vary over time because of their dependence on $\mathcal{I}_{t-1}$. When shocks are specified with respect to $\boldsymbol{y}_t$ as in Equation (43), their effect on this expectation is nonlinear and asymmetric. Small or positive shocks typically lead to a different update on the prevailing regime than large or negative shocks.

As a starting point, we propose to quantify the shocks y_{it} as a proportion δ of the square root of the one-step-ahead forecast variance

$$v_{it} = \mathrm{var}[y_{it}|\mathcal{I}_{t-1}] = \mathrm{E}[y_{it}^2|\mathcal{I}_{t-1}] - \mathrm{E}[y_{it}|\mathcal{I}_{t-1}]^2, \tag{50}$$

where $\mathrm{E}[y_{it}^2|\mathcal{I}_{t-1}]$ and $\mathrm{E}[y_{it}|\mathcal{I}_{t-1}]$ follow from Equation (16). This leads to the standardized generalized impulse response function

$$\boldsymbol{\psi}_{\boldsymbol{a}}^{y_i}(h, \delta, \mathcal{I}_{t-1}) = \widetilde{\boldsymbol{G}}_{\boldsymbol{a}} GI_{\tilde{\boldsymbol{y}}}^{y_i}(h, \eta_{it} = \delta\sqrt{v_{it}}, \mathcal{I}_{t-1}), \quad \boldsymbol{a} = \boldsymbol{y}, \boldsymbol{s} \tag{51}$$

which gives the effect of a shock of δ standard deviations to y_{it} on $\boldsymbol{y}_{t+h}$ or $\boldsymbol{s}_{t+h}$. Following Lanne and Nyberg (2016), we define the generalized forecast error variance decomposition as the proportion of the total of impulse responses of variable y_j or regime s_j which is accounted for by the GI of variable y_i conditional on $\mathcal{I}_{t-1}$ and innovation size δ standard deviations as

$$\theta^{y_i}_{a_j}(h,\delta,\mathcal{I}_{t-1}) = \frac{\sum_{l=0}^{h}(\boldsymbol{e}_j'\boldsymbol{\psi}^{y_i}_{\boldsymbol{a}}(l,\delta,\mathcal{I}_{t-1}))^2}{\sum_{l=0}^{h}\sum_{k=1}^{n}(\boldsymbol{e}_j'\boldsymbol{\psi}^{y_k}_{\boldsymbol{a}}(l,\delta,\mathcal{I}_{t-1}))^2}, \quad \boldsymbol{a}=\boldsymbol{y},\boldsymbol{s}. \tag{52}$$

Lanne and Nyberg (2016) propose this definition as an alternative to Pesaran and Shin (1998) who use $\delta\,\mathrm{var}[y_{it+h}|\mathcal{I}_{t-1}]$ in the denominator, to ensure that $\sum_{i=1}^{n}\theta^{y_i}_{a_j}(h,\delta,\mathcal{I}_{t-1})=1$.

4 Application to Investments

In this section, we use our theoretical results to analyze the risk-return trade-off for stocks and bonds. MSVAR models are well suited for such analyses, as they comprise the insights from two strands of literature. First, the literature on return predictability, for example in the seminal papers by Campbell and Viceira (1999); Campbell et al. (2003) and Barberis (2000), shows that state variables such as the dividend-to-price ratio and the short rate have persistent effects on the expected returns and (co)variances. This predictability is typically captured by a VAR(1) model. Second, the presence of regime-switching in asset returns and their pronounced implications for investments have been widely documented.[5] MSVAR models accommodate both features. In their analysis of these models for US stock and bond returns, Guidolin and Ono (2006) show that an MSIH(4)-VAR(1) model works best, and yields better predictions than simpler VAR or MSIH models (see also the evidence in Guidolin and Hyde, 2012, 2014).

In line with these findings we analyze the risk and returns of US stocks and bonds with the dividend-to-price ratio and the short rate as predictor variables. We investigate means, (co)variances and the impact of shocks for the different regimes. Our base model is an MSIAH(2)-VAR(1) model, so a VAR model of order 1, with intercepts, autoregressive parameters and (co)variances that switch between two regimes. In Appendix C we report results for the nested MSIH(2) and VAR(1) models. We keep the model specification simple, because we do not aim for the best fitting model.

4.1 Data and Estimation Results

Our data consist of monthly observations taken from CRSP. As stock returns we take returns including dividends on the S&P500. We calculate the log dividend-to-price ratio at month t as the log of the sum of the dividends over months $t-11$ through t minus the log of the level of the index at the end of month t. Bond returns are calculated from the CRSP Fixed-Term Index of 10-year Treasury bonds. As the short rate, we take the yield on 1-month Treasury Bills from the CRSP Risk-Free Rates Files. The sample period runs from January 1952 to December 2018 (804 observations).

[5] Notably Ang and Bekaert (2002), Guidolin and Timmermann (2006a,b, 2007). See Guidolin (2011) for a survey.

[Table 1 about here.]

The summary statistics in Table 1 show that both stocks and bonds yield on average positive excess returns, with excess returns on equity being four times higher on average. This is nicely in line with the excess return volatilities, which is twice as high for equity. All variables exhibit non-zero skewness, indicating deviations from normality that a standard linear VAR model cannot capture. The kurtosis of stock and bond returns also point at deviations from normality. Panel b shows that the T-bill rate and the D/P ratio are highly persistent, and may even exhibit a unit root. Correlations are overall close to zero, with the exception of the T-bill rate and the D/P ratio. The autocorrelations and cross-correlations paint a rich picture that Markov-switching models without a VAR term may find difficult to emulate. In particular, we find larger correlations between stocks and bonds at leads and lags than contemporaneously. As expected, we see a small but positive correlation between stock returns and the lagged D/P ratio. The lagged T-bill rate is negatively correlated with both stock and bond returns.

Maximum-likelihood parameter estimates based on the Expectation-Maximization algorithm of Dempster et al. (1977) (see also Hamilton, 1990) are reported in Table 2. They show low- and high-volatility regimes that are persistent. The VAR-parts of the model show predictability between stock and bond returns, and from the T-bill rate and D/P ratio to stock returns. These features are comparable to the estimation results for the MSIH(2) and VAR(1) models in Tables C.1 and C.2 in Appendix C. However, predictability varies across the regimes. In the low-volatility regime, stocks and bonds exhibit negative autocorrelation, whereas it is positive in the high-volatility regime. The effect of the T-bill rate on stock returns is significant in the low, but not in the high volaltility regime. To the contrary, the effect of the D/P ratio is concentrated in the high-volatility regime. Next, the persistence of the T-Bill rate and D/P ratio is very strong in the low-volatility regime. We also find coefficients for which the difference between the regimes is not very large compared to their standard errors, for example the effect of the lagged bond returns and T-Bill rate on stock returns.[6] Together, these differences can produce substantial differences in the return distributions and predictability in the different regimes. We analyze their implications in the next subsections.

[Table 2 about here.]

We plot the regime probabilities in Figure 1 based on the smoother of Kim (1994). The red line indicate the smoothed probability for the low-volatility regime. The gray bars show when this probability is below 1/2, in other words when the high-volatility regime applies. The figure shows the well known alternation of these periods. Notable periods with high volatility occur in the early '80s and between 2008-2010 during the great recession. Periods with low volatility tend to last longer, as also indicated by the transition probabilities in Table 2. They imply an average duration of 8.9 (4.4)

[6] A comparison of the likelihood functions and information criteria for the stocks and bonds part of the models shows that the improvement offered by allowing switches for all VAR-coefficients is limited relative to the additional number of parameters. See Table C.3 for the results.

months for the low (high) volatility regime. These results are comparable to Kole and Dijk (2017), but imply less persistence than for the macro series in Bianchi (2016) or Karamé (2015).

[Figure 1 about here.]

4.2 Regime-specific moments

We use the results of Section 2.4 to get more insights in the behavior of the variables conditional on a specific regime. We present the means, volatilities, and correlations of each variable conditional on the prevailing regime being 1 or 2 in Table 3, and turn to the implications for the autocovariance structure in Table 4. We also calculate these moments under the much stronger assumption that the process has been in the a specific regime forever. This assumption simplifies the calculations considerably to the standard results for VAR models and circumvents the need of Propositions 4–6. However, ruling out any past regime switches is unrealistic, and may be too strong if regimes are short-lived. The comparison of both results shows the consequences of this assumption. For completeness we also include the unconditional moments. We calculate 90%-confidence intervals by simulating from the distribution of the estimates, for which we assume it has converged to a multivariate normal distribution. Because the moments result from highly non-linear transformations of the parameters, we report confidence intervals instead of standard errors.

Table 3 shows that the effect of regime switching is mostly confined to stock and bonds, even though the parameter estimates in Table 2 for the T-Bill rate and D/P ratio vary considerably over the regimes. Regime 1 clearly shows a high mean for stocks, a low mean for bonds and low volatilities for both, whereas regime 2 shows the opposite pattern. The difference between the mean for stocks in the two regimes is large, 0.75% versus -0.67% per month, but the confidence intervals around both moments are wide, reflecting the sometimes large standard errors in Table 2. Still, based on the simulations of the estimated parameters we reject equality of the regime-specific means in favor of the first one being higher with a p-value of 0.046. The difference between the means of the bond returns is smaller (0.01% versus 0.37%) with also wide confidence intervals, and here we cannot reject equality (p-value of 0.131). The difference between the volatilities in both regimes is large and more precise for both asset returns, in line with the common result that asset return regimes largely reflect differences in volatilities (cf. Kole and Dijk, 2017).[7] The correlations show small variations over the different regimes. Stocks and bonds are more strongly correlated in regime 2, indicating less diversification benefits. However, the simulations indicate that the evidence for the regime-2 correlation being larger is marginal at most (p-value of 0.113).

[Table 3 about here.]

[7] A richer model specification with separate Markov chains for the intercept and VAR-coefficients on the one hand and (co)variances on the other hand as in Hubrich and Tetlow (2015) and Bianchi (2016, 2020) may provide further insights, but we leave this for future research.

To illustrate the differences between moments conditional on regime j prevailing at time t or having prevailed infinitely long (so $S_t = j$ versus $S_t = S_{t-1} = \ldots = j$), we report these latter moments with the suffix "inf." after the regime. For the means, these differences are huge and would lead to very different implications. If the first regime has prevailed infinitely long, the resulting means for stocks and bonds would be 0.23% and 0.02%; for regime two we find 0.26% and 0.28%. Interpreting the regimes based on these calculations, so ignoring the effect of Markov switching, would make the first regime seem less attractive, and the second regime more. We would also conclude that the state variables show regime switching. Surprisingly, the differences between the volatilities are much smaller. For the correlations they are again larger. Though the moments are easy to calculate under the "infinitely long" assumption, the assumption is not realistic and the resulting moments do not reflect a particular aspect of the data. This also implies that the different values for the means under this assumption do not have a clear interpretation.

The first-order autocorrelations in Table 4 show quite some differences between the regimes. Autocorrelation of stocks is negative in regime 1, but positive in 2. We reject equality in favor of the second being larger with a p-value of 0.010. Perhaps more consequential are the differences in magnitude. The auto- and cross-correlation coefficients between stocks and bonds are considerably larger in regime 2. Together, they make regime 2 riskier than regime 1, on top of the higher volatilities. The correlation between the D/P ratio at t and stock returns at $t+1$ increases from 0.022 in regime 1 to 0.102 in regime 2, indicating that predictability of stock returns is stronger in regime 2. Even though the confidence intervals are relatively wide and partly overlap, we reject equality with a p-value of 0.049. The autocorrelations of the T-bill rate and the D/P ratio are a bit lower in regime 2. The unconditional autocorrelations lie between the regime-specific autocorrelations, and are also close to the empirical ones in Table 1 and the ones implied by the VAR(1) model in Table C.5. This last table shows that the MSIH(2) cannot match the empirical autocorrelation, and as we show in Corollary 1, the regime-conditional lead-lag correlations of any plain MSIH model are zero.

[Table 4 about here.]

The rows with suffix "inf." indicate that ignoring regime switches leads again to quite different results. The autocorrelations of stocks and bonds become more extreme. For example, if regime 1 (2) prevailed forever, the autocorrelation of stocks is -0.101 (0.157) compared to -0.060 (0.112) when regimes can switch. The correlations between the state variables at t and stock and bond returns at $t+1$ in the lower-left part of the table tend to be closer to zero, indicating that the implied predictive effect of the state variables would be smaller when regime switching is ignored. Similar as in Table 3, the differences in the behavior of the state variables between the two regimes seem larger when regimes are assumed to have prevailed forever.

We conclude that both Markov-switching and predictability have profound implications for the risk-return trade-off. Our results show the presence of a bull regime where average returns are high for stocks but low for bonds, volatilities and correlations are low, and predictability is weak. We also find a bear regime where average returns are low for stocks but high for bonds, volatilities and correlations are higher, and predictability is stronger. The combination of Markov-switching with a VAR model indicates that the bear regime is riskier than the bull regime, and the difference is larger than indicated by the results for the MSIH(2) and VAR(1) models in Appendix C. So investors should pay close attention to the detrimental risk-return trade-off in the bear market regime, as well as the different time-series dynamics. We also show that the assumption that a particular regime has prevailed forever gives quite different moments. These differences can of course lead to quite different decisions or implications.

4.3 Impulse Response Analysis

We continue our analysis by investigating how shocks affect the risk-return trade-off at different horizons. As we showed in the derivation of the GIs, shocks in Markov-switching models have a nonlinear effect where both the sign and the magnitude of the shock matter, contrary to the VAR(1) model where the effect of shocks is symmetric in its sign and proportional to its magnitude. We concentrate on the impulse responses of stocks and bonds, because these are the variables that investors are most interested in.

We start with the generalized impulse responses of the first order $GI_{\boldsymbol{y}}^{y_i}(h, \eta_{it}, \mathcal{I}_{t-1})$, defined in Proposition 7, to determine the effect of shocks on the expected returns, and then consider the responses of the second order $GI_{\boldsymbol{z}}^{y_i}(h, \eta_{it}, \mathcal{I}_{t-1})$, defined in Proposition 8, for the variances. We assume that the information set $\mathcal{I}_{t-1}$ only contains the regime process at time $t-1$, which can be in regime 1 or 2. No information about $\boldsymbol{y}_\tau$, $\tau < t$ is contained in $\mathcal{I}_{t-1}$. We calculate the GIs for $h = 0, 1, 6$.

4.3.1 First Order Impulse Responses

Figure 2 shows the GIs that result from a shock to stock returns at t. Recall from Proposition 7 that the GI for $\boldsymbol{y}$ reads,

$$GI_{\boldsymbol{y}}^{y_i}(h, \eta_{it}, \mathcal{I}_{t-1}) = \boldsymbol{G}_{\boldsymbol{y}} \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\varepsilon}_t^* | y_{it}, \mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t | y_{it}, \mathcal{I}_{t-1}] \end{pmatrix},$$

meaning that a shock has contemporaneous effects on the probability forecasts for the regimes through $\mathrm{E}[\boldsymbol{u}_t | y_{it}, \mathcal{I}_{t-1}]$ and on the other variables through $\mathrm{E}[\boldsymbol{\varepsilon}_t^* | y_{it}, \mathcal{I}_{t-1}]$. We show the first effect for the probability of regime 1 in panel a, and the second effect for expected bond returns in panel b. The shocks are then propagated through $\tilde{\boldsymbol{\Phi}}^h$. We show these effects on both expected stock and bond returns for horizons $h = 1$ and $h = 6$ months in panels c–f. In all figures the red (blue) lines correspond with low-volatility regime 1 (high-volatility regime 2) at $t-1$. The dashed lines indicate the bounds of 90% confidence intervals, obtained with simulation as discussed before.

[Figure 2 about here.]

Panel a shows the probability forecast for regime 1 that has been updated with the information of the shock (solid lines) as a function of the shock size (in %). As a reference we include the probability forecast without the shock (dotted horizontal lines). For either regime prevailing at $t-1$, we see that small shocks increase the likelihood of regime 1, whereas large shocks make regime 1 less likely. The difference between the solid and dotted lines corresponds with the effect of the shock $\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}]$. The effect of small shocks is largest when regime 2 prevails at $t-1$, whereas large shocks have the largest effect when regime 1 prevails. Very large shocks lead to a concentration of the probability forecast on regime 2 independent of their sign or the regime at $t-1$, but small shocks do never lead to a complete concentration on regime 1. The top of the solid lines reflects the differences in means and volatilities reported in Table 3. The confidence intervals show that the effect of positive shocks is a bit more uncertain than of negative shocks.

The contemporaneous effect of a shock to stock returns on expected bond returns is shown in panel b. In a VAR model this effect would be a linear function of the shock, but because of the presence of regimes it becomes non-linear. When shocks are small and regime 1 prevails at $t-1$, the process is likely to remain in regime 1, which exhibits a small positive correlation between stock and bond returns. Consequently, expected bond returns do not react strongly, as indicated by the flat part of the red line. Larger shocks signal a potential switch to regime 2 with higher expected bond returns and a stronger correlation with stock returns. This explains why the red line is steeper for shocks between -5 and -10% and between 7 and 12%. When shocks become even larger, the slope reflects mainly the larger correlation in regime 2, as the probability forecast is already concentrated on this regime. The blue line, which corresponds with regime 2 at $t-1$ shows less pronounced non-linearity, because the effect of shocks on the regime forecast are smaller as shown in panel a. However, the confidence intervals show quite some overlap, which means that the effects in the different regimes may be hard to distinguish.

Panel c shows the effect on the expected stock return at $t+1$. Particularly when regime 1 prevails at $t-1$, this is a highly non-linear function of the shock size. For small shocks, regime 1 is likely to remain, and the effect is overall negative, reflecting the net impact of the negative autocorrelation of stock returns (coefficient of -0.107 in Table 2) and the cross effects of the other variables. Larger shocks make regime 2 more likely, and because of the different VAR coefficients for the stock equation in regime 2, most notably the positive AR coefficient of 0.142, the slope gradually changes from negative to positive. When regime 2 prevails at $t-1$, the GI is again less non-linear. The confidence intervals do not overlap to a large degree when shocks are small, so in this case, we can distinguish the effects in the different regimes. When shocks are large, the second regime dominates the effect independent of the regime at $t-1$, and consequently, the blue and red lines are close and their confidence intervals overlap.

The GI of bond returns at $t+1$ in panel d shows that the overall positive relation at t changes to a negative one. This reflects the negative VAR coefficients for stock returns in the bond equation in both regimes, but also the propagation of the contemporaneous effect on the other variables. The overall effect does not differ much between regime 1 and 2 at $t-1$ when shocks are small. For larger shocks, both positive and negative, the effect is shifted upward when regime 1 prevails at $t-1$. Large shocks make regime 2 more likely, in which expected bond returns are higher, as shown in Table 3.

Panels e and f show that the effect of shocks reverts and diminishes but remains noticeable after 6 months. The slow mean-reversion of the T/Bill rate and particularly the D/P ratio explains this result, as discussed by, for example, Campbell et al. (2003). As in the other panels, the non-linearity of the effect is most pronounced when regime 1 prevails at $t-1$. The red and blue lines are quite close in panel (e) which means that the regime at $t-1$ does not matter much. In panel f, they are further apart, but here the confidence intervals are quite wide and overlapping.

The GIs that result from shocks to bonds returns at time t are shown in Figure 3. The different subfigures lead to the same conclusions as for shocks to stock returns in three respects. First, the impact on expected stock and bond returns is to a larger (smaller) extent non-linear when regime 1 (2) prevails at $t-1$. Second, the effect of shocks diminishes but remains present at $t+6$. Third, confidence intervals become wider for larger horizons.

[Figure 3 about here.]

We also observe some interesting differences, though. First, the contemporaneous effect of shocks to bonds returns on expected stock returns in Figure 3b is more asymmetric than the one from stocks to bonds in Figure 2b, in particular for regime 1 at $t-1$. The asymmetry follows from two effects that dampen each other for positive shocks, but amplify for negative ones. First, stock and bond returns are hardly correlated in regime 1, and positively correlated in regime 2. Second, larger shocks make regime 2 more likely, in which expected stock returns are lower. So, when bond returns encounter a small but positive shock, regime 1 with correlation close to zero becomes more likely, which leads to a small positive effect on expected stock returns. When shocks become larger, regime 2 becomes more likely, but the effect of the larger positive correlation is offset by the lower mean in regime 2. For negative shocks, the large positive correlation and lower mean work in the same direction.

Second, the slope of the GI in Figure 3c does not become negative when regime 1 prevails at $t-1$ and shocks are small as in Figure 2c. Instead, the effect of autocorrelation is flat and close to zero for small shocks, in line with the small coefficients in the bond equation for regime 1 in Table 2. Larger shocks make regime 2 more likely, which exhibits a higher mean for bond returns, and larger coefficients in the bond equation, leading to an overall positive effect reflected in the slope of both the blue and the red lines, and an upward shift in the red line compared to the blue one.

Third, the GIs for $t+6$ show that the slow mean-reversion of the state variables has a smaller impact on the transmission of shocks to bond returns. The GIs show that the transmission mostly follows from the updating of the regime forecast, as indicated by the (inverted) U-shape in panel e (f). In Figures 2e and 2f, the U-shape is tilted, indicating the combined effect of the Markov-switching and the VAR-part.

We show the figures with the GIs when shocks originate in the T-Bill rate and the D/P ratio in Appendix C.3 and summarize the main insights here. Shocks to the T-Bill rate affect expected stock and bond returns mostly through their update of the regime forecasts for two reasons. First, the difference between the volatilities of the innovations in the two regime is larger than for stocks and bonds, which means that small shocks have a stronger effect on the probability of regime 1. Second, the regime-specific cross-correlations with stock and bond returns are mostly close to zero (see Tables 3 and 4). Since the T-Bill rate is strongly persistent, shocks die out more slowly. Shocks to the D/P ratio are by construction strongly negatively correlated with shocks to stock returns (-0.952 and -0.920 in regimes 1 and 2), and consequently their GIs are the mirror images of those for shocks to stocks returns in Figure 2.

To determine which source of shocks is most important, and how it depends on the regime and the sign and size of the shock we decompose the forecast error variance as in Section 3.3. We express shocks as a multiple δ of the forecast standard deviation of the different variables for time t, and plot the results in Figure 4 for either regime 1 (solid) or 2 (dashed) prevailing at $t-1$. We focus on the horizon $h = 1$ because this horizon combines both the contemporaneous and VAR(1) transmission of the shocks. Panel a shows that slightly over 50% of forecast error variances of stock returns originates in shocks to stock returns themselves, and over 40% comes from the D/P ratio. Of the remaining 5-10% most can be attributed to bonds when shocks are negative or the T-Bill rate when shocks are positive. For shock sizes close to zero, the lines for stock returns and the D/P ratio show a pronounced trough. This results from the fact that knowing that there is no shock leads to a strong reduction of the contribution to the forecast error variance at $t+1$ when shocks originate in stock returns or the D/P ratio, but much less so when they come from bond returns or the T/Bill rate. The regimes at $t-1$ lead to small variations, with the D/P ratio being more important for positive and small negative shocks in regime 1 compared to regime 2, whereas bond returns show the opposite pattern.

[Figure 4 about here.]

Figure 4b shows that bond returns are more isolated from the other variables, as 90-95% of its forecast error variance can be attributed to own shocks. This holds in particular for small shocks with regime 1 at $t-1$. The remainder is about equally divided between shocks originating in stock returns and the D/P ratio. The impact of the T-Bill rate is surprisingly small.

For investors, this means that the estimates in Table 2, which correspond with the average effect at $h = 0$, do not present the complete picture. Because of the VAR nature of the model, shocks die out more slowly, and the Markov-switching component makes their effect nonlinear. Smaller shocks to stock returns are expected to be reversed in the next period, but not so for large shocks, or any shocks in the high-volatility regime. Shocks to bond returns have a small contemporaneous effect of the same direction on stock returns, but their effect dies out slowly, reducing the diversification opportunities. We also conclude that the high-volatility regime is riskier but also a bit easier to predict and understand, because the effect of regime switching is less. The consequences regarding expected returns of bonds are a bit more positive than for stocks. Shocks have a less persistent effect on bond returns, and more importantly, when shocks originate from stock returns they revert. The contemporaneous effect of these shocks has the same direction as the shock, but it becomes opposite in the next period, indicating some protection.

4.3.2 Second Order Impulse Responses

Having investigated the effect of shocks on the return part of the risk-return trade-off, we now turn to the second-order impulse responses for the risk part. With 2 regimes the law of total variance makes it easy to discern four different channels through which shocks affect the variance of $\boldsymbol{y}_{t+h}$,

$$\mathrm{Var}[\boldsymbol{y}_{t+h}|y_{it},\mathcal{I}_{t-1}] = \xi_{1,t+h}\boldsymbol{\Sigma}_{1,t+h}+\xi_{2,t+h}\boldsymbol{\Sigma}_{2,t+h}+\xi_{1,t+h}\xi_{2,t+h}(\boldsymbol{\mu}_{1,t+h}-\boldsymbol{\mu}_{2,t+h})(\boldsymbol{\mu}_{1,t+h}-\boldsymbol{\mu}_{2,t+h})', \quad (53)$$

where $\xi_{j,t+h} = \Pr[S_{t+h} = j|y_{it},\mathcal{I}_{t-1}]$, $\boldsymbol{\mu}_{j,t+1} = \mathrm{E}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it},\mathcal{I}_{t-1}]$, and $\boldsymbol{\Sigma}_{j,t+h} = \mathrm{Var}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it},\mathcal{I}_{t-1}]$ for $j = 1, 2$. The first channel runs via the effect that the shock in y_{it} has on the regime-specific variances at $t + h$. If $h = 0$, the contemporaneous correlation of the variables leads to a decrease in the variance of the other variables. If $h > 0$, $\boldsymbol{\Sigma}_{j,t+h}$ is equal to the variance matrix of regime j. Next and as before, shocks lead to an updating of the regime forecasts $\xi_{i,t+h}$. On top of this effect, the variability of the regime process also contributes to the total variance via the term $\xi_{1,t+h}\xi_{2,t+h} = \xi_{1,t+h}(1-\xi_{1,t+h})$. The final channel stems from the difference between the autoregressive components in the different regimes, $(\boldsymbol{\mu}_{1,t+h}-\boldsymbol{\mu}_{2,t+h})$. We investigate how much the different channels contribute, and how the effects depend on the size of the shock and the regime at $t - 1$.

[Figure 5 about here.]

Figure 5 shows how shocks to stock returns affect the volatilities of stock and bond returns at different horizons, with the contemporaneous effect on bond volatility in panel a. In both regimes a shock to stock returns at t reduces the contemporaneous bond volatility, but it also leads to an update of the regime forecast as shown in Figure 2a. When shocks are small, both effects work in the same direction, which means that the forecast volatilities go down compared to the forecast without a shock (the dotted lines), and that the drop is largest when the high-volatility regime (the blue line) prevails at $t - 1$. For large shocks, the regime update dominates, so even though the conditional variance in either regime drops, the increased likelihood of the second regime leads to an increase of the forecast volatility. Obviously, the increase is largest when starting in the low-volatility regime. The confidence intervals indicate a reasonable degree of precision, though they also imply that effects may be zero when regime 2 prevails at $t - 1$ or when regime 1 prevails and shocks are small.

Panels b and c show the effect moving one period forward for both bond and stock volatilities. The picture for bond volatility resembles the effect at t pretty much, though with a smaller magnitude, as shocks die out. The picture for stock volatility shows richer dynamics. As for bond volatility, stock volatility is forecast to go down when shocks are small. If shocks are large, forecast volatility goes up, but it does not simply converge to the volatility of regime 2. Instead, the volatility remains increasing in the size of the shock, which steams from the second part in Equation (53). This part accounts for the difference in the expected stock returns at $t+1$, which is largest when shocks are negative. This is a big difference with both MSIH models and VAR models where the effect on the volatility is bounded or absent. The confidence intervals indicate that the effect of small shocks may be difficult to discern from zero, but not so for large shocks.

The effect of shocks on the volatilities remains present at longer horizons as indicated by panels d and e for a horizon of 6 months, though the magnitude diminishes. At this horizon, the effect of positive shocks to stock returns at t is increasing in size on both stock and bond volatility but not as pronounced as for a one-month horizon. Parameter uncertainty accumulates and leads to relatively large confidence intervals.

We show the effects of shocks to the other variables in Appendix C.3. The figures are generally comparable to Figure 5. When shocks originate from bond returns, both bond and stock volatilities show a clear increasing pattern for increasing shock size, independent of the size of the shock, and in line with the strong effects in Figures 3c and 3d. Shocks to the T-Bill rate do not have strong VAR effects, so the regime updating channel dominates here. The large negative correlation of shocks to stock returns and the D/P ratio means that the shocks to the latter have the mirrored effect of shocks to the former.

We finish this part with an analysis of the effects that shocks have on the forecast correlation of stock and bond returns at different horizons. Figure 6 shows that also the effects that shocks have on correlation are highly non-linear, and are also quite different depending on their source. Shocks to stock returns lead to small decreases in correlation when shocks are small and regime 1 prevails at $t-1$. For larger shocks (below -7% or above 4%), the forecast correlation quickly increases towards the forecast correlation from regime 2 at $t-1$, and goes up substantially beyond it when shocks are above 9%. When shocks become more negative, below -15%, the opposite happens, and the correlation starts to decrease. When regime 2 prevails at $t-1$, positive as well as small negative shocks lead to increases, but also here large negative shocks lead to decreases. Confidence intervals are wide, though.

[Figure 6 about here.]

Shocks to bond returns have a more symmetric effect as shown in panel b. Independent of the regime at $t-1$, small shocks lead to a decrease in the forecast correlation, whereas large shocks lead to an increase. Also here, the increase continues well beyond the correlation that is forecast from the high-volatility regime. The differences between panels a and b show that the combination of Markov-switching and VAR features lead to intricate dynamics. Panels c and d show that the effect of shocks remains present for longer horizons, though again the magnitude has diminished. Shocks to stock returns die out more slowly because of their effect on the D/P ratio which is very persistent. Also here, the accumulation of parameter uncertainty means that the difference should be interpreted with care.

Summarizing, our analyses show that shocks can lead to an increase or a reduction of risk. However, the increases are more substantial than the decreases, and the decreases only occur for a limited domain of small shocks. Moreover, the effects of shocks on the second moments are prolonged. These results stand in stark contrast with the implications of VAR models where shocks do not affect second moments at all. They are also stronger than what we observe for simpler models with only Markov switching, indicating that the combination of VAR and Markov-switching properties amplify the effect of shocks.

Combining the results for risk with those for the return part of the previous subsection shows that our framework presents a unified way to analyze the effects of shocks. Shocks can have positive or negative effects on the expected returns, depending on their sign and size. The effects are nonlinear, but less so in the high-volatility regime. Shocks can have positive or negative effects on the volatility and correlation of stocks and bond returns. In the high-volatility regime, shocks easily lead to a further and prolonged increase of volatility forecasts. Shocks can hence lead to a deterioration of the risk-return trade-off by decreasing expected returns and increasing risk and reducing diversification opportunities.

5 Conclusion

We propose a unified framework that enables the calculation of moments and the analysis of impulse responses for MSVAR models in a setting where only the distribution of the regimes at a particular point in time is given. As the key ingredient of our framework we show that the processes of the level of the observable variables, their squares and the latent regime indicator can be combined such that they form a linear VAR(1) model with innovations that form a non-Gaussian martingale difference sequence. Whereas the VAR(1)-part has been shown before by Bianchi (2016); Krolzig (2006), we explicitly derive and use the MDS-part.

We show how to use this extended VAR(1) formulation to derive future first and second moments conditional on the regime distribution at one particular point in time t only, so without any assumption on the values of the observable variables or the regime distribution up to t. In this derivation, we use the time-reversed version of the regime process, and show that the same stationarity conditions apply as for the original MSVAR. We then derive closed-form expressions for impulse responses in the framework of Koop et al. (1996). Shocks affect forecasts of both the first and second moment in this model, and hence we propose the variance impulse response function next to the traditional generalized impulse response function for the first moment. We also show how to construct the generalized forecast error variance decomposition.

We apply our methods for an analysis of the risk-return trade-off of investments in stocks and bonds, where we include predictability by the T-Bill rate and dividend-to-price ratio, and switching between low- and high-volatility regimes. We use our theoretical results to characterize the regimes. Consistent with the stylized facts, we find a bull regime with a high (low) mean for returns on stocks (bonds), and low volatility for both, and a bear regime where these features are reversed. Next to that, we show that the predictability also varies across the regimes, with stronger effects in the bear regime. The impulse response analysis for first and second moments shows that the effect of shocks is asymmetric, nonlinear, and regime-dependent. We use our framework to discern the different channels via which the shocks propagate, which further helps understanding the shocks and linking them to the different features of the model. While the effect of shocks on the expected returns can be positive or negative in both regimes, shocks tend to have an upward effect on risk by increasing volatilities and correlations, in particular when they are large or occur in the high-volatility regime.

Our theoretical and empirical results are useful for modern risk management. The tools we propose can be directly used to investigate what happens when a shock hits in a bad state. In our empirical application this setting corresponds with a shock hitting during a bear market, but of course a similar analysis can be done for financial institutions during a crisis regime, or countries during a recession. Because all results are available in closed form, the tools are easy to use and do not require simulations or numerical approximations.

References

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, 15(4):1137–1187.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55(1):225–264.

Barigozzi, M., Hallin, M., Soccorsi, S., and von Sachs, R. (2021). Time-varying general dynamic factor models and the measurement of financial connectedness. *Journal of Econometrics*, 222(1):324–343.

Bianchi, F. (2016). Methods for measuring expectations and uncertainty in Markov-switching models. *Journal of Econometrics*, 190(1):79–99.

Bianchi, F. (2020). The great depression and the great recession: A view from financial markets. *Journal of Monetary Economics*, 114:240–261.

Campbell, J. Y., Chan, Y. L., and Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics*, 67(1):41–80.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.

Campbell, J. Y. and Viceira, L. M. (1999). Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics*, 114(2):433–495.

Campbell, J. Y. and Viceira, L. M. (2005). The term-structure of the risk-return trade-off. *Financial Analysts Journal*, 61(1):34–44.

Cavicchioli, M. (2017a). Higher order moments of Markov switching VARMA models. *Econometric Theory*, 33(6):1502–1515.

Cavicchioli, M. (2017b). Third and fourth moments of vector autoregressions with regime switching. *Communications in Statistics - Theory and Methods*, 46(9):4181–4194.

Chen, S. and Schienle, M. (2022). Large spillover networks of nonstationary systems. *Journal of Business & Economic Statistics*, pages 1–15.

Clerc, L., Giovannini, A., Langfield, S., Peltonen, T., Portes, R., and Scheicher, M. (2016). Indirect contagion: the policy problem. ESRB occasional paper series no. 9, European Systemic Risk Board (ESRB), European System of Financial Supervision.

Cont, R. and Schaanning, E. (2019). Monitoring indirect contagion. *Journal of Banking & Finance*, 104:85–102.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–22.

Diebold, F. X. and Yılmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

Diebold, F. X. and Yılmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

Diebold, F. X. and Yılmaz, K. (2015). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press.

Ehrmann, M., Ellison, M., and Valla, N. (2003). Regime-dependent impulse response functions in a Markov-switching vector autoregression model. *Economics Letters*, 78(3):295–299.

Francq, C. and Zakoïan, J.-M. (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics*, 102(2):339–364.

Gallant, A. R., Rossi, P. E., and Tauchen, G. (1993). Nonlinear dynamic structures. *Econometrica*, 61(4):871.

Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2022). When do state-dependent local projections work? Working paper, McGill University.

Gorodnichenko, Y. and Ng, S. (2017). Level and volatility factors in macroeconomic data. *Journal of Monetary Economics*, 91:52–68.

Guidolin, M. (2011). Markov switching models in empirical finance. In Drukker, D. M., editor, *Missing Data Methods: Time-Series Methods and Applications*, volume 27 of *Advances in Econometrics*, pages 1–86. Emerald Group Publishing Limited.

Guidolin, M. and Hyde, S. (2012). Can VAR models capture regime shifts in asset returns? A long-horizon strategic asset allocation perspective. *Journal of Banking & Finance*, 36(3):695–716.

Guidolin, M. and Hyde, S. (2014). Linear predictability vs. bull and bear market models in strategic asset allocation decisions: evidence from UK data. *Quantitative Finance*, 14(12):2135–2153.

Guidolin, M. and Ono, S. (2006). Are the dynamic linkages between the macroeconomy and asset prices time-varying? *Journal of Economics & Business*, 58(5-6):480–518.

Guidolin, M. and Timmermann, A. (2006a). An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. *Journal of Applied Econometrics*, 21(1):1–22.

Guidolin, M. and Timmermann, A. (2006b). Term structure of risk under alternative econometric specifications. *Journal of Econometrics*, 131:285–308.

Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economics Dynamics & Control*, 31(11):3503–3544.

Hafner, C. M. and Herwartz, H. (2006). Volatility impulse responses for multivariate GARCH models: An exchange rate illustration. *Journal of International Money and Finance*, 25(5):719–740.

Hamilton, J. (2016). Macroeconomic regimes and regime shifts. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, chapter 3, pages 163–201. Elsevier.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hamilton, J. D. and Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11:573–593.

Hou, C. and Nguyen, B. H. (2018). Understanding the US natural gas market: A Markov-switching VAR approach. *Energy Economics*, 75:42–53.

Hubrich, K. and Tetlow, R. J. (2015). Financial stress and economic dynamics: the transmission of crises. *Journal of Monetary Economics*, 70:100–115.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95(1):161–182.

Jordà, Ò. (2009). Simultaneous confidence regions for impulse responses. *Review of Economics and Statistics*, 91(3):629–647.

Karamé, F. (2010). Impulse–response functions in Markov-switching structural vector autoregressions: A step further. *Economics Letters*, 106(3):162–165.

Karamé, F. (2012). An algorithm for generalized impulse-response functions in Markov-switching structural VAR. *Economics Letters*, 117(1):230–234.

Karamé, F. (2015). Asymmetries and Markov-switching structural VAR. *Journal of Economic Dynamics and Control*, 53:85–102.

Kim, C.-J. (1994). Dynamic linear models with Markov-Switching. *Journal of Econometrics*, 60(1):1–22.

Kole, E. and Dijk, D. (2017). How to identify and forecast bull and bear markets? *Journal of Applied Econometrics*, 32(1):120–139.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147.

Krolzig, H.-M. (2006). Impulse response analysis in Markov-switching vector autoregressive models. Working paper, University of Kent, UK.

Krolzig, H.-M. (2013). *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*, volume 454 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag Berlin Heidelberg.

Lanne, M. and Nyberg, H. (2016). Generalized forecast error variance decomposition for linear and nonlinear multivariate models. *Oxford Bulletin of Economics and Statistics*, 78(4):595–603.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer-Verlag GmbH.

Miao, K., Phillips, P. C., and Su, L. (2022). High-dimensional VARs with common factors. *Journal of Econometrics*.

Pesaran, H. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics Letters*, 58(1):17–29.

Ramey, V. A. and Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from US historical data. *Journal of Political Economy*, 126(2):850–901.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Stelzer, R. (2009). On Markov-switching ARMA processes – stationarity, existence of moments, and geometric ergodicity. *Econometric Theory*, 25(1):43–62.

Taamouti, A. (2012). Moments of multivariate regime switching with application to risk-return trade-off. *Journal of Empirical Finance*, 19(2):292–308.

Tenreyro, S. and Thwaites, G. (2016). Pushing on a string: US monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics*, 8(4):43–74.

Timmermann, A. (2000). Moments of Markov-switching models. *Journal of Econometrics*, 96(1):75–111.

Yang, M. (2000). Some properties of vector autoregressive processes with Markov-switching coefficients. *Econometric Theory*, 16(1):23–43.

Zhang, J. and Stine, R. A. (2001). Autocovariance structure of Markov regime switching models and model selection. *Journal of Time Series Analysis*, 22(1):107–124.

Table 1: Summary statistics

(a) Marginal moments

	Mean	Volatility	Skewness	Kurtosis
Stock returns	0.51	4.17	-0.67	5.48
Bond returns	0.13	2.07	0.20	4.56
T-bill rate	0.34	0.25	0.88	4.10
D/P ratio	-3.53	0.40	-0.29	2.30

(b) contemporaneous and lag-1 correlations

	Stock returns	Bond returns	T-bill rate	D/P ratio	Stock returns	Bond returns	T-bill rate	D/P ratio
Stock returns	1	0.085	-0.097	-0.044	0.052	-0.150	-0.095	-0.051
Bond returns		1	-0.009	-0.015	0.114	0.061	-0.065	-0.025
T-bill rate			1	0.433	-0.077	-0.003	0.983	0.437
D/P ratio				1	0.059	-0.005	0.430	0.994

This table show the means, volatilities (both in %), skewness and kurtosis, and correlations of excess log stock and bond returns, the log T-bill rate and the log d/p ratio. The left four columns of panel b show the contemporaneous correlations. The right four columns show the lag-1 correlations with the lagged variables on the rows. The sample period runs from January 1952 to December 2018 (804 observations).

Table 2: Parameter estimates MSIAH(2)-VAR(1) Model

	Regime 1				Regime 2			
	Stock returns	Bond returns	T-bill rate	D/P ratio	Stock returns	Bond returns	T-bill rate	D/P ratio
Intercept	3.380	0.077	0.006	−0.026	10.541	0.242	0.036	−0.091
	(1.403)	(0.877)	(0.008)	(0.015)	(1.036)	(2.282)	(0.065)	(0.019)
VAR(1) matrix								
Stock returns	−0.107	−0.049	−0.0005	0.0012	0.142	−0.109	0.0018	−0.0015
	(0.043)	(0.020)	(0.0002)	(0.0005)	(0.070)	(0.037)	(0.0008)	(0.0009)
Bond returns	0.205	−0.003	−0.0016	−0.0022	0.265	0.140	−0.012	−0.0021
	(0.084)	(0.041)	(0.0005)	(0.0009)	(0.161)	(0.070)	(0.002)	(0.0018)
T-bill rate	−1.653	−0.296	0.993	0.005	−1.879	−0.272	0.967	0.014
	(0.806)	(0.393)	(0.005)	(0.009)	(1.343)	(0.747)	(0.018)	(0.014)
D/P ratio	0.535	−0.022	0.0005	0.995	2.955	−0.045	0.006	0.974
	(0.370)	(0.224)	(0.0023)	(0.004)	(0.356)	(0.604)	(0.018)	(0.005)
Volatilities and correlations								
Stock returns	3.168				5.317			
	(0.104)				(0.257)			
Bond returns	0.020	1.552			0.159	2.761		
	(0.045)	(0.051)			(0.060)	(0.138)		
T-bill rate	−0.067	−0.010	0.018		−0.116	0.020	0.069	
	(0.051)	(0.053)	(0.001)		(0.080)	(0.068)	(0.002)	
D/P ratio	−0.952	−0.035	0.049	0.034	−0.920	−0.159	0.136	0.058
	(0.121)	(0.046)	(0.051)	(0.001)	(0.160)	(0.066)	(0.085)	(0.003)

Transition and Initial Probabilities

	Regime 1	Regime 2	Initial
Regime 1	0.887	0.228	1
	(0.019)	(0.034)	–
Regime 2	0.112	0.772	0
	(0.019)	(0.034)	–

This table shows the parameter estimates for a Markov-switching VAR(1) model with switches in the means, autoregressive coefficients and variances between two regimes. The parameters have been estimated by the Expectation Maximization algorithm. Standard errors calculated by the outer product of the gradient are in parentheses. The sample period runs from January 1952 to December 2018 (804 observations). The VAR(1) matrices reports the lagged variables on the rows. We report estimated volatilities on the diagonals in the third panel, and estimated correlations in the lower triangle. We do not report a standard error for the parameter $\boldsymbol{\zeta} = \Pr[\boldsymbol{s}_1]$, because it is at the boundary of its domain.

Table 3: Regime-conditional and unconditional first and second moments

Means (% per month)

	Stocks		Bonds		T-Bill rate		D/P ratio	
Regime 1	0.75	(0.22, 1.72)	0.01	(-0.24, 0.32)	0.40	(-0.43, 0.74)	−3.58	(-4.59, -3.18)
Regime 1 inf.	0.23	(-1.88, 1.52)	0.02	(-0.69, 1.15)	0.49	(-0.55, 2.24)	−4.34	(-8.14, -2.91)
Regime 2	−0.67	(-1.60, 0.43)	0.37	(-0.13, 1.37)	0.40	(-0.50, 0.74)	−3.54	(-4.48, -3.14)
Regime 2 inf.	0.26	(-0.34, 1.03)	0.28	(-0.39, 1.35)	0.44	(-0.48, 0.91)	−3.24	(-4.09, -2.69)
Unconditional	0.28	(-0.02, 0.89)	0.13	(-0.12, 0.62)	0.40	(-0.47, 0.74)	−3.57	(-4.56, -3.17)

Volatilities (% per month)

	Stocks		Bonds		T-Bill rate		D/P ratio	
Regime 1	3.23	(3.07, 3.49)	1.56	(1.49, 1.67)	0.28	(0.21, 0.62)	0.34	(0.27, 0.78)
Regime 1 inf.	3.21	(3.06, 3.41)	1.56	(1.48, 1.67)	0.18	(0.11, 0.53)	0.34	(0.23, 1.09)
Regime 2	5.49	(5.13, 6.00)	2.83	(2.64, 3.14)	0.29	(0.22, 0.63)	0.33	(0.26, 0.75)
Regime 2 inf.	5.50	(5.11, 5.96)	2.84	(2.66, 3.15)	0.32	(0.23, 0.71)	0.32	(0.24, 0.70)
Unconditional	4.18	(4.01, 4.57)	2.08	(1.99, 2.30)	0.28	(0.21, 0.63)	0.34	(0.27, 0.77)

Correlations

	Stocks, Bonds		Stocks, T-Bill		Stocks, D/P ratio	
Regime 1	0.037	(-0.039, 0.117)	−0.114	(-0.303, -0.014)	−0.097	(-0.234, -0.030)
Regime 1 inf.	0.032	(-0.041, 0.113)	−0.073	(-0.222, -0.022)	−0.065	(-0.151, 0.037)
Regime 2	0.140	(0.030, 0.245)	−0.063	(-0.213, 0.101)	−0.035	(-0.119, 0.145)
Regime 2 inf.	0.138	(0.035, 0.235)	−0.069	(-0.204, 0.037)	−0.066	(-0.131, 0.027)
Unconditional	0.084	(-0.0002, 0.161)	−0.085	(-0.174, -0.029)	−0.074	(-0.122, -0.027)

	Bonds, T-Bill		Bonds, D/P ratio		T-Bill, D/P ratio	
Regime 1	−0.043	(-0.189, 0.074)	−0.024	(-0.139, 0.066)	0.455	(-0.067, 0.835)
Regime 1 inf.	−0.028	(-0.173, 0.053)	−0.015	(-0.148, 0.084)	0.351	(-0.507, 0.832)
Regime 2	−0.032	(-0.231, 0.079)	−0.036	(-0.216, 0.076)	0.455	(-0.050, 0.838)
Regime 2 inf.	−0.036	(-0.262, 0.092)	−0.037	(-0.213, 0.084)	0.510	(-0.063, 0.877)
Unconditional	−0.037	(-0.171, 0.034)	−0.024	(-0.136, 0.044)	0.454	(-0.068, 0.830)

This table gives the regime-conditional and unconditional means, volatilities and correlations for the MSIAH(2)-VAR(1) model with parameter values reported in Table 2, and 90% confidence intervals in parentheses. The regime-conditional means $\mathrm{E}[\boldsymbol{y}_t|S_t]$ follow from Proposition 4, whereas the regime conditional volatilities and correlations are calculated from $\mathrm{Var}[\boldsymbol{y}_t|S_t]$ based on Proposition 5. The rows with suffix "inf." report the moments conditional on the given regime having prevailed infinitely long, i.e. $\lim_{\tau\to\infty} \mathrm{E}[\boldsymbol{y}_t|S_t = \cdots = S_{t-\tau} = j]$, and $\lim_{\tau\to\infty} \mathrm{Var}[\boldsymbol{y}_t|S_t = \cdots = S_{t-\tau} = j]$. The unconditional moments follow from Equations (22) to (25). The confidence intervals are constructed by simulation with 10,000 draws of the parameters assuming a multivariate normal distribution for the estimates. The variance matrix of the estimates is calculated by the outer product of the gradient. Parameter draws that do not meet the stationarity requirement are discarded.

Table 4: Auto- and cross-correlations at (lead) order 1

	Stocks, Stocks		Stocks, Bonds		Stocks, T-Bill		Stocks, D/P ratio	
Regime 1	−0.06	(-0.11, 0.03)	−0.10	(-0.15, -0.03)	−0.12	(-0.30, -0.02)	−0.09	(-0.23, -0.02)
Regime 1 inf.	−0.10	(-0.16, 0.00)	−0.10	(-0.16, -0.01)	−0.08	(-0.22, -0.03)	−0.05	(-0.14, 0.04)
Regime 2	0.11	(0.03, 0.21)	−0.18	(-0.27, -0.07)	−0.05	(-0.20, 0.11)	−0.05	(-0.14, 0.13)
Regime 2 inf.	0.16	(0.06, 0.27)	−0.19	(-0.29, -0.06)	−0.05	(-0.18, 0.05)	−0.09	(-0.16, 0.01)
Unconditional	0.05	(0.00, 0.16)	−0.15	(-0.22, -0.07)	−0.08	(-0.17, -0.02)	−0.08	(-0.13, -0.03)

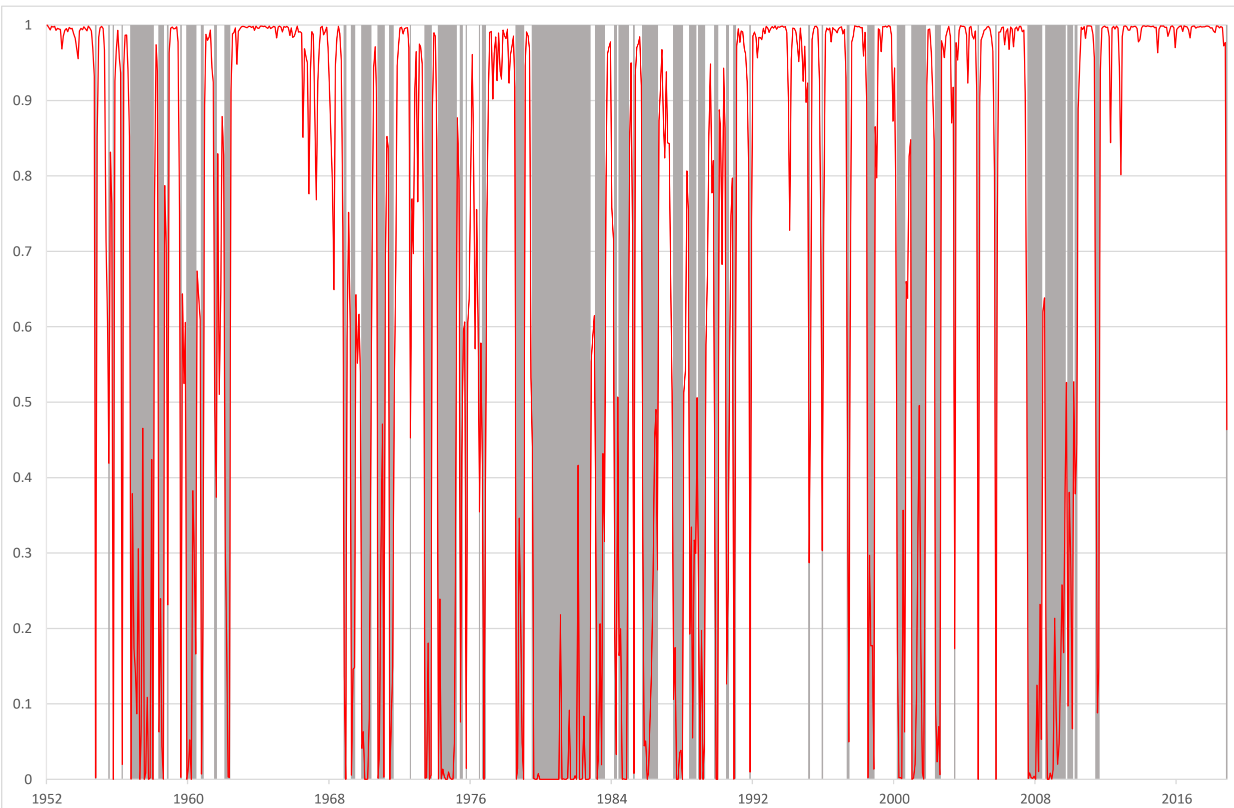
	Bonds, Stocks		Bonds, Bonds		Bonds, T-Bill		Bonds, D/P ratio	
Regime 1	0.09	(0.04, 0.15)	0.01	(-0.04, 0.08)	−0.06	(-0.20, 0.06)	−0.03	(-0.15, 0.06)
Regime 1 inf.	0.10	(0.03, 0.18)	0.00	(-0.07, 0.09)	−0.04	(-0.18, 0.04)	−0.02	(-0.15, 0.08)
Regime 2	0.15	(0.04, 0.27)	0.09	(0.00, 0.21)	−0.12	(-0.29, 0.00)	−0.06	(-0.23, 0.06)
Regime 2 inf.	0.15	(0.02, 0.29)	0.11	(0.00, 0.27)	−0.13	(-0.32, -0.02)	−0.06	(-0.23, 0.06)
Unconditional	0.12	(0.03, 0.20)	0.06	(0.01, 0.18)	−0.09	(-0.20, -0.02)	−0.04	(-0.15, 0.03)

	T-Bill, Stocks		T-Bill, Bonds		T-Bill, T-Bill		T-Bill, D/P ratio	
Regime 1	−0.09	(-0.22, -0.02)	−0.04	(-0.17, 0.05)	0.994	(0.990, 0.999)	0.46	(-0.07, 0.84)
Regime 1 inf.	−0.07	(-0.22, -0.01)	−0.03	(-0.18, 0.05)	0.995	(0.987, 0.999)	0.35	(-0.51, 0.83)
Regime 2	−0.04	(-0.17, 0.07)	−0.02	(-0.20, 0.08)	0.972	(0.954, 0.994)	0.46	(-0.05, 0.84)
Regime 2 inf.	−0.04	(-0.18, 0.08)	−0.02	(-0.26, 0.10)	0.971	(0.945, 0.994)	0.51	(-0.05, 0.88)
Unconditional	−0.07	(-0.16, -0.02)	−0.03	(-0.17, 0.04)	0.987	(0.977, 0.997)	0.45	(-0.07, 0.83)

	D/P ratio, Stocks		D/P ratio, Bonds		D/P ratio, T-Bill		D/P ratio, D/P ratio	
Regime 1	0.02	(-0.11, 0.09)	−0.02	(-0.12, 0.05)	0.45	(-0.07, 0.84)	0.994	(0.990, 0.999)
Regime 1 inf.	0.03	(-0.09, 0.10)	−0.01	(-0.14, 0.09)	0.35	(-0.51, 0.83)	0.995	(0.988, 1.000)
Regime 2	0.10	(0.01, 0.18)	−0.02	(-0.19, 0.08)	0.45	(-0.06, 0.84)	0.987	(0.979, 0.997)
Regime 2 inf.	0.10	(0.01, 0.17)	−0.01	(-0.19, 0.11)	0.50	(-0.07, 0.87)	0.983	(0.970, 0.996)
Unconditional	0.05	(-0.04, 0.10)	−0.01	(-0.13, 0.05)	0.45	(-0.06, 0.83)	0.991	(0.986, 0.998)

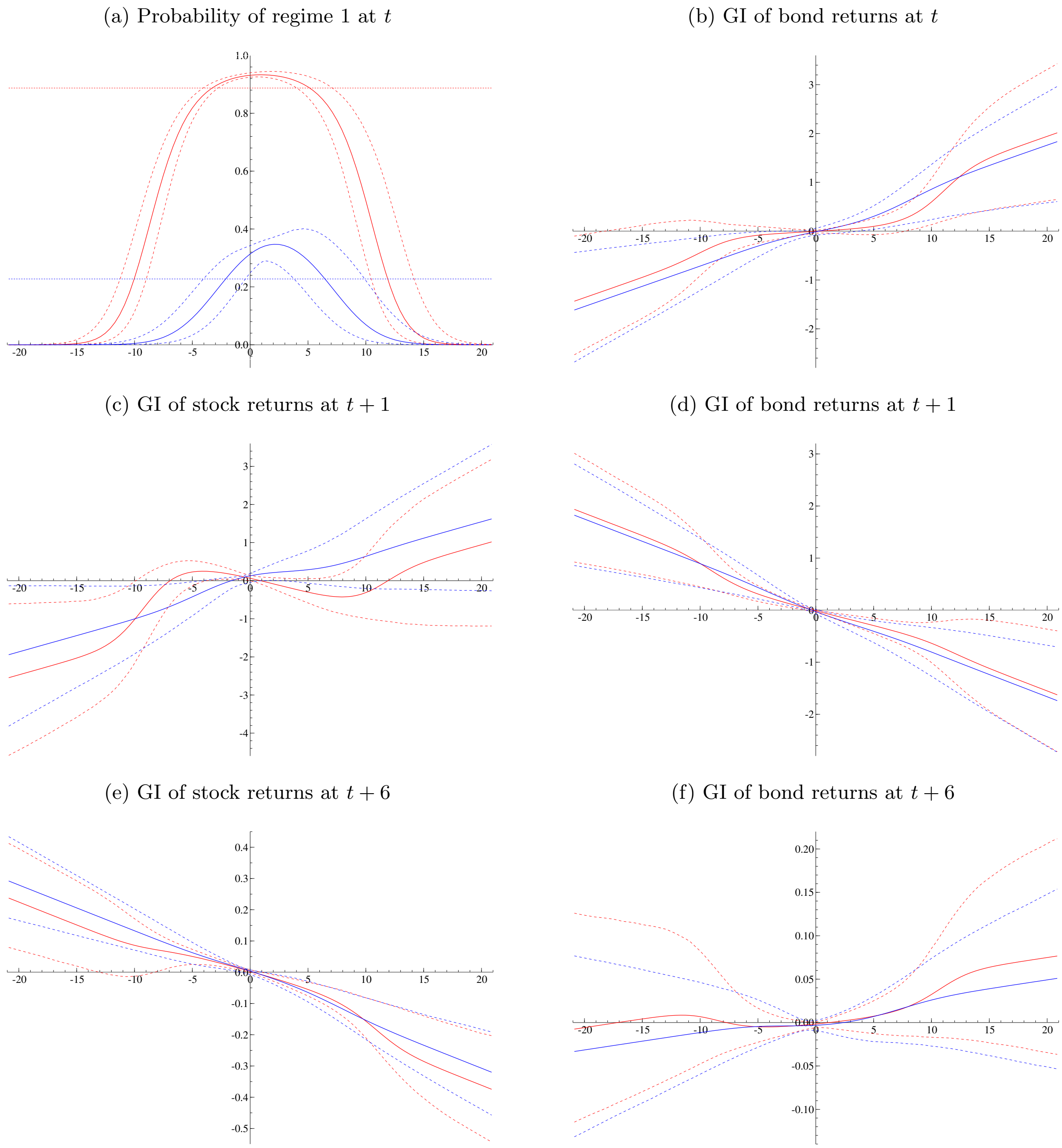
This table gives regime-conditional and unconditional auto- and cross-correlations with lead 1 for the MSIAH(2)-VAR(1) model with parameter values reported in Table 2 and 90% confidence intervals in parentheses. The heading of each block indicates the variable at t followed by the variable at $t+1$. The regime-conditional correlations are calculated from $\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+1}|S_t]$ based on Equation (37), and the conditional volatilities based on Proposition 5. The rows with suffix "inf." report the autocorrelations conditional on the given regime having prevailed infinitely long, calculated from $\lim_{\tau\to\infty} \mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+1}|S_{t+1} = S_t = \cdots = S_{t-\tau} = j]$, and the corresponding volatilities in Table 3. The unconditional correlations are calculated from $\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+1}]$ based on Equations (26) and (27), and the unconditional volatilities in Table 3. The confidence intervals are constructed as in Table 3.

Figure 1: Regime probabilities



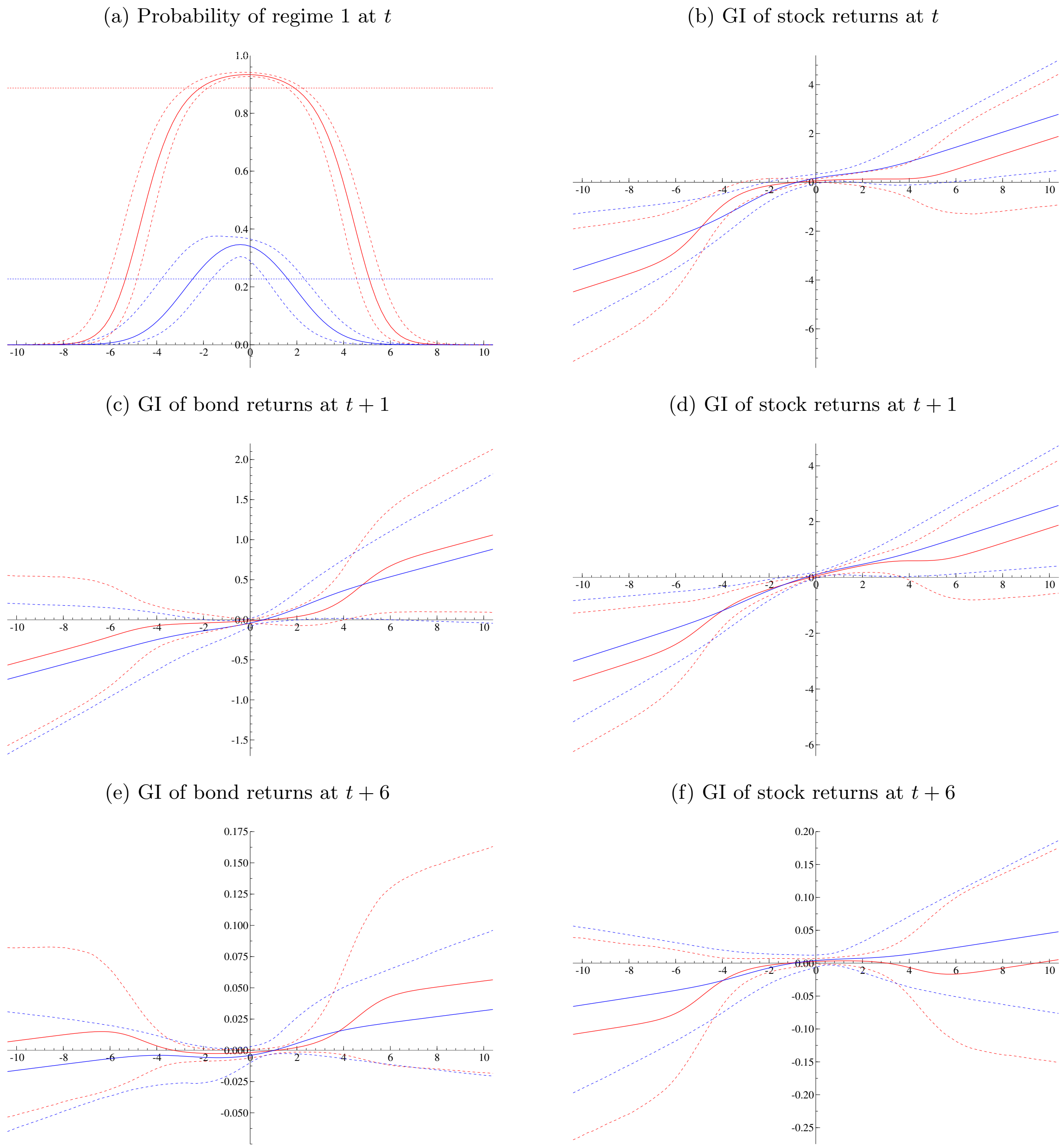
This figures shows the regime probabilities based on the smoother of Kim (1994) and the estimates of the MSIAH(2)-VAR(1) in Table 2. The solid red line gives the probability of regime 1. The gray bars indicate the periods for which the probability of regime 2 exceeds 1/2.

Figure 2: The effects of a shock to stock returns at t



This figure shows the Generalized Impulse Response Functions for the probability forecasts for regime 1 at time t (panel a), for expected bond returns at t, $t+1$ and $t+6$ (panels b, d and f), and for expected stock returns at $t+1$ and $t+6$ (panels c and e) as a function of a shock to stock returns at time t. Both the horizontal axes and the vertical axes (panels b to f) show returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The GI are calculated based on Proposition 7. The straight dotted lines in panel (a) give the forecast probability of state 1 in absence of a shock. The difference between the solid and the dotted lines corresponds with the GI. In the other panels the GIs are plotted.

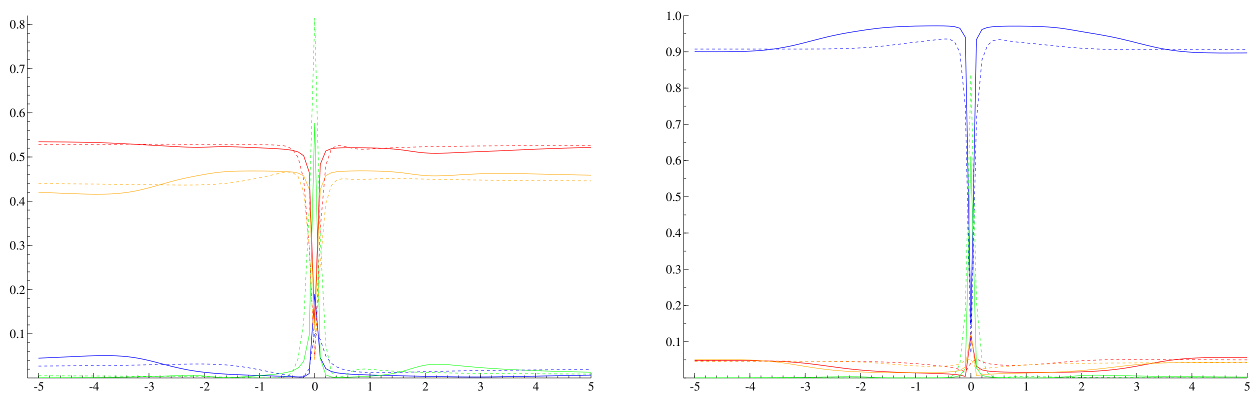
Figure 3: The effects of a shock to bond returns at t



This figure shows the Generalized Impulse Response Functions for the probability forecasts for regime 1 at time t (panel a), for expected stock returns at t, $t+1$ and $t+6$ (panels b, d and f), and for expected bond returns at $t+1$ and $t+6$ (panels c and e) as a function of a shock to bond returns at time t. Both the horizontal axes and the vertical axes (panels b to f) show returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The GI are calculated based on Proposition 7. The straight dotted lines in panel (a) give the forecast probability of state 1 in absence of a shock. The difference between the solid and the dotted lines corresponds with the GI. In the other panels the GIs are plotted.

Figure 4: The forecast error variance decomposition for stocks and bonds at $t+1$

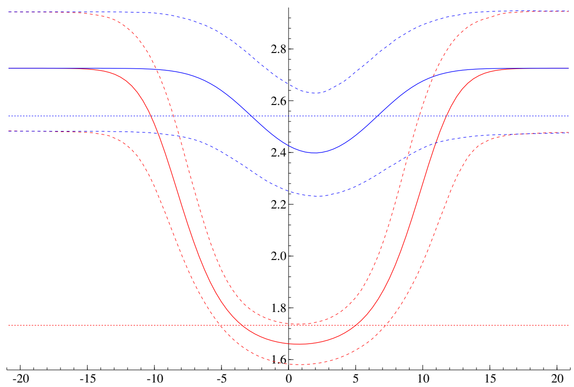
(a) Stocks (b) Bonds



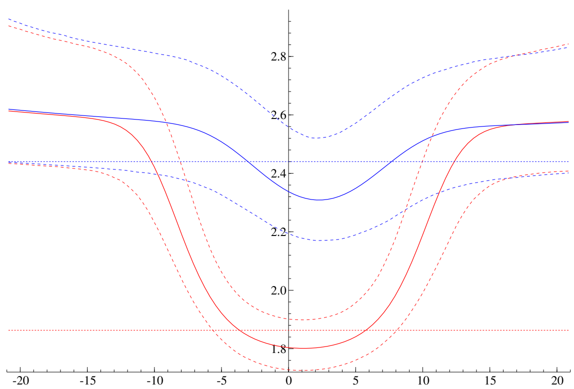
This figure shows the Generalized Forecast Error Variance Decomposition (GFEVD) for stock and bond returns at time $t+1$. The lines show the GFEVD as a function of shocks equal to δ times the time t forecast standard deviation of stock returns in red, bond returns in blue, the T-bill rate in green and the D/P ratio in orange based on Equation (52). The solid (dashed) lines corresponds with regime 1 (2) at $t-1$.

Figure 5: The effects of shocks to stock returns on the forecast volatility of stock and bond returns

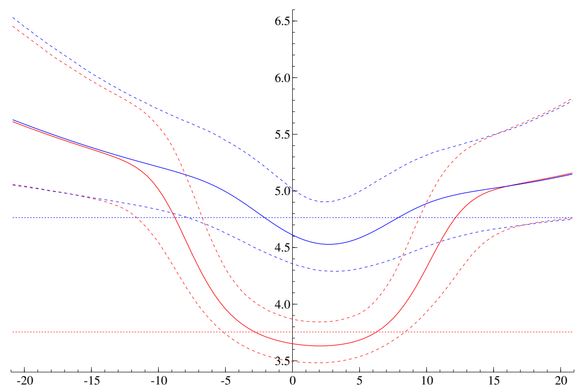
(a) Bond volatility at t



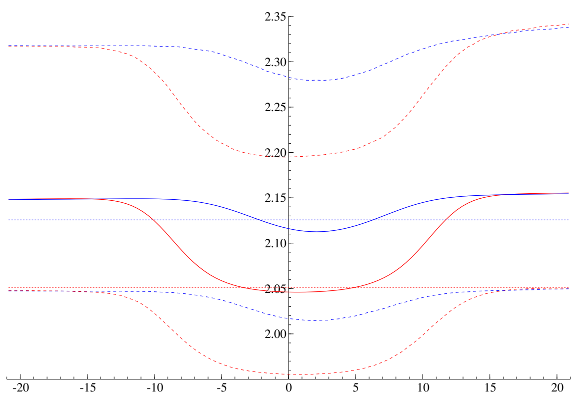
(b) Bond volatility at $t+1$



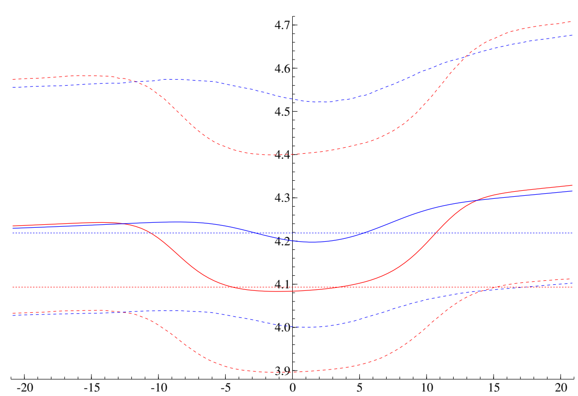
(c) Stock volatility at $t+1$



(d) Bond volatility at $t+6$

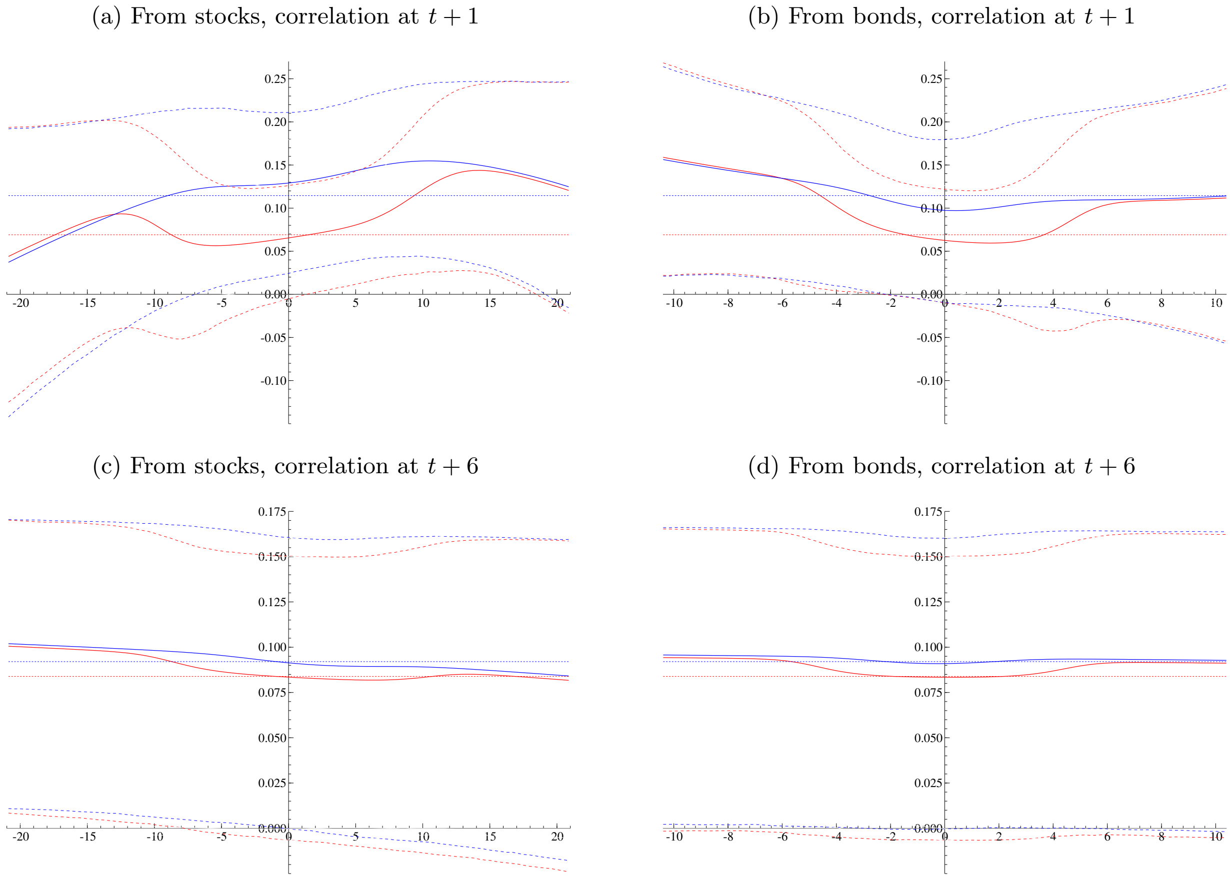


(e) Stock volatility at $t+6$



This figure shows the impulse responses for the forecast volatility of bond returns at time t, $t+1$ and $t+6$ (panels a, b and d) and stock returns at $t+1$ and $t+6$ (panels c and e) as a function of a shock to stock returns at time t based on Proposition 8. Both the horizontal and the vertical axes shows returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The straight dotted lines give the forecast volatility in absence of a shock. The difference between the solid and the dotted lines gives the volatility impulse response function.

Figure 6: The effects of shocks to stock or bond returns on the forecast correlation of stock and bond returns



This figure shows the impulse responses for the forecast correlation of stock and bond returns at time $t+1$ and $t+6$ as a function of shocks to stock (panels a and b) or bond returns (panels c and d) at time t based on Proposition 8. The horizontal axis shows returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The straight dotted lines give the forecast correlation in absence of a shock. The difference between the solid and the dotted lines gives the correlation impulse response function.

A Proofs for Section 2 (The MSVAR model and its moments)

The equation numbers in this section are copied from the main text with an appendix prefix.

Lemma 1. *Let $\boldsymbol{A}_i, i = 1, \ldots, m$ be a set of $n \times l$ matrices, and $\boldsymbol{A} = \text{bdiag}_{i=1}^m(\boldsymbol{A}_i)$. Then $\boldsymbol{e}_i \otimes \boldsymbol{A}_i = \boldsymbol{A}(\boldsymbol{e}_i \otimes \boldsymbol{I}_l)$.*

Proof. We can write $\boldsymbol{A}_i = (\boldsymbol{e}_i' \otimes \boldsymbol{I}_n)\boldsymbol{A}(\boldsymbol{e}_i \otimes \boldsymbol{I}_n)$. The multiplication $\boldsymbol{e}_i \otimes (\boldsymbol{e}_i' \otimes \boldsymbol{I}_n) = \boldsymbol{e}_i\boldsymbol{e}_i' \otimes \boldsymbol{I}_n$ gives a $mn \times mn$ block diagonal matrix whose ith block on the diagonal is equal to the identity matrix, so $\boldsymbol{e}_i \otimes (\boldsymbol{e}_i' \otimes \boldsymbol{I}_n)\boldsymbol{A}$ gives a $mn \times mn$ block diagonal matrix whose ith block on the diagonal is equal to $\boldsymbol{A}_i$. The post-multiplication $\boldsymbol{e}_i \otimes \boldsymbol{I}_l$ selects block-row i. Block-row i of $\boldsymbol{e}_i \otimes (\boldsymbol{e}_i' \otimes \boldsymbol{I}_n)\boldsymbol{A}$ and $\boldsymbol{A}$ are the same, as both consist of m blocks of size $n \times l$ with block i equal to $\boldsymbol{A}_i$ and all other blocks equal to the zero matrix. Hence the result of the multiplication is the same. □

Proposition 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Define $\boldsymbol{y}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{y}_t$ and $\tilde{\boldsymbol{y}}_t = \left(\boldsymbol{y}_t^{*\prime}, \boldsymbol{s}_t'\right)'$. Then $\boldsymbol{y}_t^*$ follows the process*

$$\boldsymbol{y}_t^* = \boldsymbol{C}\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\varepsilon}_t^*, \tag{A.6}$$

with $\boldsymbol{C} = \text{bdiag}_{i=1}^m(\boldsymbol{c}_i)$, $\boldsymbol{\Phi} = \text{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i)$, and

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\Lambda}(\boldsymbol{P} \otimes \boldsymbol{I}_n)(\boldsymbol{s}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Lambda}(\boldsymbol{u}_t \otimes \boldsymbol{\varepsilon}_t),$$

with $\boldsymbol{\Lambda} = \text{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i)$, and $\boldsymbol{u}_t$ as defined in Equation (3). $\tilde{\boldsymbol{y}}_t$ follows the process

$$\tilde{\boldsymbol{y}}_t = \begin{pmatrix} \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t, \tag{A.7}$$

with

$$\tilde{\boldsymbol{\Phi}} = \begin{pmatrix} \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & \boldsymbol{C}\boldsymbol{P} \\ \mathbf{O}_{m \times nm} & \boldsymbol{P} \end{pmatrix}.$$

and $\tilde{\boldsymbol{\varepsilon}}_t = (\boldsymbol{\varepsilon}_t^{*\prime}, \boldsymbol{u}_t')'$. Moreover, $\boldsymbol{\varepsilon}_t^*$ and $\tilde{\boldsymbol{\varepsilon}}_t$ are MDS.

Proof. From the definition of $\boldsymbol{y}_t^*$ follows

$$\boldsymbol{y}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{s}_t + \boldsymbol{\Phi}(\boldsymbol{s}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Lambda}(\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t),$$

where we use Lemma 1, and $(\boldsymbol{s}_t \otimes \boldsymbol{I}_n)\boldsymbol{b} = (\boldsymbol{s}_t \otimes \boldsymbol{I}_n\boldsymbol{b}) = \boldsymbol{s}_t \otimes \boldsymbol{b}$ for any vector $\boldsymbol{b}$ of size n. Substitution of Equation (3) for $\boldsymbol{s}_t$ yields

$$\begin{aligned} \boldsymbol{y}_t^* &= \boldsymbol{C}\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Phi}(\boldsymbol{P}\boldsymbol{s}_{t-1} \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\varepsilon}_t^* \\ &= \boldsymbol{C}\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\varepsilon}_t^*. \end{aligned}$$

and

$$\boldsymbol{\varepsilon}_t^* = \boldsymbol{\Lambda}(\boldsymbol{P} \otimes \boldsymbol{I}_n)(\boldsymbol{s}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Lambda}(\boldsymbol{u}_t \otimes \boldsymbol{\varepsilon}_t).$$

Equation (6) shows how $\boldsymbol{y}_t^*$ depends on past values. Stacking $\boldsymbol{y}_t^*$ and $\boldsymbol{s}_t$ in the vector $\tilde{\boldsymbol{y}}_t$ gives

$$\tilde{\boldsymbol{y}}_t = \begin{pmatrix} \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & \boldsymbol{C}\boldsymbol{P} \\ \mathbf{O} & \boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_{t-1}^* \\ \boldsymbol{s}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t^* \\ \boldsymbol{u}_t \end{pmatrix} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t,$$

with $\tilde{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{\varepsilon}}_t$ as defined in the proposition. Next,

$$\mathrm{E}[\boldsymbol{u}_t|\tilde{\boldsymbol{y}}_{t-1}] = \mathrm{E}[\boldsymbol{u}_t|\boldsymbol{s}_{t-1}] = \mathbf{0},$$

because conditional on $\boldsymbol{s}_{t-1}$, $\boldsymbol{s}_t$ is independent of $\boldsymbol{y}_{t-1}$. We also have $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|\tilde{\boldsymbol{y}}_{t-1}] = \mathbf{0}$, because $\boldsymbol{\varepsilon}_t$ is independent of S_t, S_{t-1} and hence $\boldsymbol{u}_t$, $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{u}_t|\tilde{\boldsymbol{y}}_{t-1}] = \mathbf{0}$. This establishes that $\tilde{\boldsymbol{\varepsilon}}_t$ is an MDS. □

Proposition 2. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Define $\boldsymbol{z}_t = \boldsymbol{y}_t \otimes \boldsymbol{y}_t$, $\boldsymbol{z}_t^* = \boldsymbol{s}_t \otimes \boldsymbol{z}_t$ and $\tilde{\boldsymbol{z}}_t = \left(\boldsymbol{z}_t^{*\prime}, \boldsymbol{y}_t^{*\prime}, \boldsymbol{s}_t'\right)'$, with $\boldsymbol{y}_t^*$ as defined in Proposition 1. Then $\boldsymbol{z}_t$ follows the process*

$$\boldsymbol{z}_t = \boldsymbol{\gamma}_{S_t} + \boldsymbol{\omega}_{S_t} + \boldsymbol{\Psi}_{S_t}\boldsymbol{y}_{t-1} + \boldsymbol{\Upsilon}_{S_t}\boldsymbol{z}_{t-1} + \boldsymbol{\zeta}_t, \tag{A.12}$$

where $\boldsymbol{\gamma}_{S_t} = \boldsymbol{c}_{S_t} \otimes \boldsymbol{c}_{S_t}$, $\boldsymbol{\omega}_{S_t} = \mathrm{vec}\,(\boldsymbol{\Sigma}_{S_t})$, $\boldsymbol{\Psi}_{S_t} = \boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{c}_{S_t} + \boldsymbol{c}_{S_t} \otimes \boldsymbol{\Phi}_{S_t}$, $\boldsymbol{\Upsilon}_{S_t} = \boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{\Phi}_{S_t}$, and

$$\begin{aligned}\boldsymbol{\zeta}_t =& (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{c}_{S_t} + \boldsymbol{c}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})\boldsymbol{\varepsilon}_t + (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{\Phi}_{S_t})(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{y}_{t-1}) + \\ & (\boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})(\boldsymbol{y}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t - \mathrm{vec}\,(\boldsymbol{I}_n)).\end{aligned}$$

$\boldsymbol{z}_t^*$ follows the process

$$\boldsymbol{z}_t^* = (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\boldsymbol{z}_{t-1}^* + \boldsymbol{\zeta}_t^*, \tag{A.13}$$

with $\boldsymbol{\Gamma} = \mathrm{bdiag}_{i=1}^m(\boldsymbol{\gamma}_i)$, $\boldsymbol{\Omega} = \mathrm{bdiag}_{i=1}^m(\boldsymbol{\omega}_i)$, $\boldsymbol{\Psi} = \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Psi}_i)$, and $\boldsymbol{\Upsilon} = \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Upsilon}_i)$, and

$$\begin{aligned}\boldsymbol{\zeta}_t^* =& (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t \otimes \boldsymbol{z}_{t-1}) + \\ & \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i \otimes \boldsymbol{c}_i + \boldsymbol{c}_i \otimes \boldsymbol{\Lambda}_i)(\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t) + \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Phi}_i)(\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t \otimes \boldsymbol{y}_{t-1}) + \\ & \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i \otimes \boldsymbol{\Lambda}_i)(\boldsymbol{s}_t \otimes \boldsymbol{y}_{t-1} \otimes \boldsymbol{\varepsilon}_t) + \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Lambda}_i \otimes \boldsymbol{\Lambda}_i)(\boldsymbol{s}_t \otimes (\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t - \mathrm{vec}\,(\boldsymbol{I}_n))).\end{aligned}$$

$\tilde{\boldsymbol{z}}_t$ follows the process

$$\tilde{\boldsymbol{z}}_t = \begin{pmatrix} \boldsymbol{z}_t^* \\ \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Upsilon}}\tilde{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{\zeta}}_t, \tag{A.14}$$

with

$$\tilde{\boldsymbol{\Upsilon}} = \begin{pmatrix} \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2}) & \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P} \\ \mathbf{O} & \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & \boldsymbol{C}\boldsymbol{P} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{P} \end{pmatrix},$$

and $\tilde{\boldsymbol{\zeta}}_t = (\boldsymbol{\zeta}_t^{*\prime}, \boldsymbol{\varepsilon}_t^{*\prime}, \boldsymbol{u}_t')'$, with $\boldsymbol{\varepsilon}_t^*$ as defined in Proposition 1. Moreover, $\boldsymbol{\zeta}_t$, $\boldsymbol{\zeta}_t^*$ and $\tilde{\boldsymbol{\zeta}}_t$ are MDS.

Proof. Substitution of Equation (1) in the definition of $\boldsymbol{z}_t$ yields

$$\boldsymbol{z}_t = \boldsymbol{y}_t \otimes \boldsymbol{y}_t = \left(\boldsymbol{c}_{S_t} + \boldsymbol{\Phi}_{S_t}\boldsymbol{y}_{t-1} + \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t\right) \otimes \left(\boldsymbol{c}_{S_t} + \boldsymbol{\Phi}_{S_t}\boldsymbol{y}_{t-1} + \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t\right),$$

and working out the multiplication gives Equation (12). We use that $\boldsymbol{c}_{S_t} \otimes \boldsymbol{\Phi}_{S_t}\boldsymbol{y}_{t-1} = (\boldsymbol{c}_{S_t} \otimes \boldsymbol{\Phi}_{S_t})\boldsymbol{y}_{t-1}$, $\boldsymbol{\Phi}_{S_t}\boldsymbol{y}_{t-1} \otimes \boldsymbol{c}_{S_t} = (\boldsymbol{\Phi}_{S_t} \otimes \boldsymbol{c}_{S_t})\boldsymbol{y}_{t-1}$ and $\boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t = (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t) = \operatorname{vec}(\boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'\boldsymbol{\Lambda}'_{S_t})$ and $\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t] = \boldsymbol{I}_n$ such that $\mathrm{E}[\boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t | S_t] = \operatorname{vec}(\boldsymbol{\Sigma}_{S_t}) = \boldsymbol{\omega}_{S_t}$. From the definition of $\boldsymbol{z}^*_t$ and Lemma 1 follows

$$\boldsymbol{z}^*_t = \boldsymbol{s}_t \otimes \boldsymbol{z}_t = (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{s}_t + \boldsymbol{\Psi}(\boldsymbol{s}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{s}_t \otimes \boldsymbol{z}_{t-1}) + \boldsymbol{s}_t \otimes \boldsymbol{\zeta}_t.$$

Substitution of Equation (3) yields

$$\begin{aligned}\boldsymbol{z}^*_t &= (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Psi}(\boldsymbol{P}\boldsymbol{s}_{t-1} \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{P}\boldsymbol{s}_{t-1} \otimes \boldsymbol{z}_{t-1}) + \boldsymbol{\zeta}^*_t \\ &(\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}^*_{t-1} + \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\boldsymbol{z}^*_{t-1} + \boldsymbol{\zeta}^*_t,\end{aligned}$$

which is Equation (13). Stacking $\boldsymbol{z}^*_t$, $\boldsymbol{y}^*_t$ and $\boldsymbol{s}_t$ in the vector $\tilde{\boldsymbol{z}}_t$ gives

$$\tilde{\boldsymbol{z}}_t = \begin{pmatrix} \boldsymbol{z}^*_t \\ \boldsymbol{y}^*_t \\ \boldsymbol{s}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2}) & \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P} \\ \mathbf{O} & \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) & \boldsymbol{C}\boldsymbol{P} \\ \mathbf{O} & \mathbf{O} & \boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{z}^*_{t-1} \\ \boldsymbol{y}^*_{t-1} \\ \boldsymbol{s}_{t-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}^*_t \\ \boldsymbol{\varepsilon}^*_t \\ \boldsymbol{u}_t \end{pmatrix} = \tilde{\boldsymbol{\Upsilon}}\tilde{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{\zeta}}_t,$$

with $\tilde{\boldsymbol{\Upsilon}}$ and $\tilde{\boldsymbol{\zeta}}_t$ as defined in the proposition. Next, $\mathrm{E}[\boldsymbol{\zeta}_t|\boldsymbol{y}_{t-1}] = \mathbf{0}$, because $\mathrm{E}[\boldsymbol{\varepsilon}_t|\boldsymbol{y}_{t-1}] = \mathbf{0}$, and $\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t|\boldsymbol{y}_{t-1}] = \boldsymbol{I}_n$. Also, $\mathrm{E}[\boldsymbol{\zeta}^*_t|\boldsymbol{y}_{t-1}, \boldsymbol{s}_{t-1}] = \mathbf{0}$ because $\boldsymbol{\varepsilon}_t$ is independent of S_t, S_{t-1} and hence $\boldsymbol{u}_t$, $\mathrm{E}[\boldsymbol{\varepsilon}_t] = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{u}_t|\tilde{\boldsymbol{y}}_{t-1}] = \mathbf{0}$. Because $\boldsymbol{\varepsilon}^*_t$ is also an MDS, it follows that $\tilde{\boldsymbol{\zeta}}_t$ is an MDS. □

Proposition 3. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Define $\boldsymbol{z}_{t+k,t} = \boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t$, $\boldsymbol{z}^*_{t+k,t} = \boldsymbol{s}_{t+k} \otimes \boldsymbol{z}_{t+k,t}$ and $\boldsymbol{y}^*_{t+k,t} = \boldsymbol{s}_{t+k} \otimes \boldsymbol{y}_t$ for $h \geq 0$. The process defined by $\tilde{\boldsymbol{z}}_{t+k,t} = \left(\boldsymbol{z}^*_{t+k,t}{}', \boldsymbol{y}^*_{t+k,t}{}', \boldsymbol{s}'_{t+k}\right)'$ follows*

$$\tilde{\boldsymbol{z}}_{t+k,t} = \begin{pmatrix} \tilde{\boldsymbol{\Phi}} \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P} \end{pmatrix} \tilde{\boldsymbol{z}}_{t+k-1,t} + \begin{pmatrix} \tilde{\boldsymbol{\varepsilon}}_{t+k} \otimes \boldsymbol{y}_t \\ \boldsymbol{u}_{t+k} \end{pmatrix}, \quad k \geq 1 \tag{A.19}$$

with $\tilde{\boldsymbol{\Phi}}$ as in Proposition 1 and the second term a MDS.

Proof. We use that

$$\begin{pmatrix} \boldsymbol{z}^*_{t+k,t} \\ \boldsymbol{y}^*_{t+k,t} \end{pmatrix} = \tilde{\boldsymbol{y}}_{t+k} \otimes \boldsymbol{y}_t = \left(\tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{y}}_{t+k-1} + \tilde{\boldsymbol{\varepsilon}}_{t+k}\right) \otimes \boldsymbol{y}_t = (\tilde{\boldsymbol{\Phi}} \otimes \boldsymbol{I}_n)(\tilde{\boldsymbol{y}}_{t+k-1} \otimes \boldsymbol{y}_t) + \tilde{\boldsymbol{\varepsilon}}_{t+k} \otimes \boldsymbol{y}_t,$$

where we have substitued Equation (7) in the second equality. Combining this result with $\boldsymbol{s}_{t+k} = \boldsymbol{P}\boldsymbol{s}_{t+k-1} + \boldsymbol{u}_{t+k}$ gives the result in Equation (19). The second term is an MDS, since $\tilde{\boldsymbol{\varepsilon}}_{t+k}$ and $\boldsymbol{u}_{t+k}$ are MDS, and consequently $\mathrm{E}[\tilde{\boldsymbol{\varepsilon}}_{t+k} \otimes \boldsymbol{y}_t|\boldsymbol{y}_{t+k-l}, S_{t+k-l}] = \mathbf{0}$ for all $k \geq 1, l \geq 1$. □

In the proofs that follow we use the following lemma.

Lemma 2. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2). Let $\boldsymbol{Q}$ be the transition matrix of the time-reversed Markov chain of the proces S_t, and let $\boldsymbol{\Upsilon} = \operatorname{bdiag}_{j=1}^{m}(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j)$. Then the matrices $\boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})$ and $\boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})$ are similar.*

Proof. Use $\boldsymbol{Q} = \operatorname{diag}(\bar{\boldsymbol{\xi}})\boldsymbol{P}' \operatorname{diag}(\bar{\boldsymbol{\xi}})^{-1}$ to find

$$\begin{aligned}\boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2}) &= \boldsymbol{\Upsilon}\Big(\big(\operatorname{diag}(\bar{\boldsymbol{\xi}})^{-1}\boldsymbol{P} \operatorname{diag}(\bar{\boldsymbol{\xi}})\big) \otimes \boldsymbol{I}_{n^2}\Big) \\ &= \boldsymbol{\Upsilon}\big(\operatorname{diag}(\bar{\boldsymbol{\xi}}) \otimes \boldsymbol{I}_{n^2}\big)^{-1}\big(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2}\big)\big(\operatorname{diag}(\bar{\boldsymbol{\xi}}) \otimes \boldsymbol{I}_{n^2}\big) \\ &= \big(\operatorname{diag}(\bar{\boldsymbol{\xi}}) \otimes \boldsymbol{I}_{n^2}\big)^{-1}\boldsymbol{\Upsilon}\big(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2}\big)\big(\operatorname{diag}(\bar{\boldsymbol{\xi}}) \otimes \boldsymbol{I}_{n^2}\big),\end{aligned}$$

which shows that $\boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})$ and $\boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})$ are similar. The last equality uses the fact that $\boldsymbol{\Upsilon}$ is a block-diagonal matrix, and $\big(\operatorname{diag}(\bar{\boldsymbol{\xi}}) \otimes \boldsymbol{I}_{n^2}\big)^{-1}$ is diagonal, which allows to switch the order of multiplication. □

This lemma implies that $\boldsymbol{y}_t$ combined with the original state process S_t governed by $\boldsymbol{P}$ or combined with the time-reversed process governed by $\boldsymbol{Q}$ have many properties in common. The matrices $\boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})$ and $\boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})$ have the same characteristic equation, because they are similar. As a consequence, they have the same eigenvalues.

Proposition 4. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and assume that it is second-order stationary. Let $\boldsymbol{\mu}_j = \mathrm{E}[\boldsymbol{y}_t | S_t = j]$, and stack these conditional expectations in the $mn \times 1$ vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots \boldsymbol{\mu}_m')'$. Then*

$$\boldsymbol{\mu} = \big(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\big)^{-1}\boldsymbol{c}, \tag{A.29}$$

where $\boldsymbol{c} = \big(\boldsymbol{c}_1', \ldots, \boldsymbol{c}_m'\big)'$, $\boldsymbol{\Phi} = \operatorname{bdiag}_{i=1}^{m}(\boldsymbol{\Phi}_i)$, and $\boldsymbol{Q}$ is the transition matrix of the time-reversed Markov chain of the proces S_t. The expectation of $\boldsymbol{y}_t$ conditional on the state distribution $\boldsymbol{\xi}_t$ follows as

$$\mathrm{E}[\boldsymbol{y}_t | \boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}, \tag{A.30}$$

The expectation of $\boldsymbol{y}_{t+h}$ for $h \geq 0$ conditional on the current state distribution $\boldsymbol{\xi}_t$ follows as $\mathrm{E}[\boldsymbol{y}_{t+h} | \boldsymbol{\xi}_t] = \widetilde{\boldsymbol{G}}_{\boldsymbol{y}} \,\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h} | \boldsymbol{\xi}_t]$ with

$$\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h} | \boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \big(\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{A.31}$$

with $\tilde{\boldsymbol{y}}_{t+h}$ and $\tilde{\boldsymbol{\Phi}}$ defined in Proposition 1 and $\widetilde{\boldsymbol{G}}_{\boldsymbol{y}}$ as in (8).

Proof. Start with

$$
\begin{aligned}
E[\boldsymbol{y}_t|S_t=j] &= \boldsymbol{\mu}_j = \boldsymbol{c}_j + \boldsymbol{\Phi}_j \operatorname{E}[y_{t-1}|S_t=j] \\
&= \boldsymbol{c}_j + \boldsymbol{\Phi}_j \sum_{i=1}^{m} \operatorname{E}[y_{t-1}|S_{t-1}=i] \Pr[S_{t-1}=i|S_t=j] \\
&= \boldsymbol{c}_j + \boldsymbol{\Phi}_j \sum_{i=1}^{m} \boldsymbol{\mu}_i q_{ij} \\
&= \boldsymbol{c}_j + \boldsymbol{\Phi}_j \begin{pmatrix} \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_m \end{pmatrix} \begin{pmatrix} q_{1j} \\ \vdots \\ q_{mj} \end{pmatrix} \\
&= \boldsymbol{c}_j + \boldsymbol{\Phi}_j \begin{pmatrix} \boldsymbol{\mu}_1 & \cdots & \boldsymbol{\mu}_m \end{pmatrix} \boldsymbol{Q}\boldsymbol{e}_j \\
&= \boldsymbol{c}_j + \boldsymbol{\Phi}_j (\boldsymbol{e}_j'\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}.
\end{aligned}
$$

We can write this system compactly as

$$
\boldsymbol{\mu} = \boldsymbol{c} + \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}.
$$

Under the assumption of stationarity, Lemma 2 implies that $\boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)$ is convergent, and the solution follows as

$$
\boldsymbol{\mu} = \big(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\big)^{-1}\boldsymbol{c}.
$$

By summing the appropriate elements with weights $\boldsymbol{\xi}_t$ we find $\operatorname{E}[\boldsymbol{y}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}$. To find the expression for $\operatorname{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\xi}_t]$, we first use Proposition 1 to establish

$$
\operatorname{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \operatorname{E}[\tilde{\boldsymbol{y}}_t|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \operatorname{E}[\boldsymbol{y}_t^*|\boldsymbol{\xi}_t] \\ \boldsymbol{\xi}_t \end{pmatrix},
$$

because $\operatorname{E}[\tilde{\boldsymbol{\varepsilon}}_{t+l}|\boldsymbol{\xi}_t] = \boldsymbol{0}$ for $l > 0$ and the definition of $\tilde{\boldsymbol{y}}_t$. Then,

$$
\operatorname{E}[\boldsymbol{y}_t^*|\boldsymbol{\xi}_t] = \begin{pmatrix} \xi_{1t} \operatorname{E}[\boldsymbol{y}_t|S_t=1] \\ \vdots \\ \xi_{mt} \operatorname{E}[\boldsymbol{y}_t|S_t=m] \end{pmatrix} = \big(\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}.
$$

□

Proposition 5. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and assume that it is second-order stationary. Let $\boldsymbol{z}_t = \boldsymbol{y}_t \otimes \boldsymbol{y}_t$, $\boldsymbol{\mu}_j = \operatorname{E}[\boldsymbol{y}_t|S_t=j]$, $\boldsymbol{\nu}_j = \operatorname{E}[\boldsymbol{z}_t|S_t=j]$ with stacked versions $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots \boldsymbol{\mu}_m')'$ and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \ldots \boldsymbol{\nu}_m')'$. Then*

$$
\boldsymbol{\nu} = \big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}\big(\gamma + \boldsymbol{\omega} + \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}\big), \tag{A.32}
$$

where $\boldsymbol{\gamma} = (\gamma_1', \ldots, \gamma_m')'$ with $\boldsymbol{\gamma}_j = \boldsymbol{c}_j \otimes \boldsymbol{c}_j$, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_m')'$ with $\boldsymbol{\omega}_j = \operatorname{vec}(\boldsymbol{\Sigma}_j)$, $\boldsymbol{\Upsilon} = \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j)$, $\boldsymbol{\Psi} = \operatorname{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{c}_j + \boldsymbol{c}_j \otimes \boldsymbol{\Phi}_j)$, $\boldsymbol{Q}$ is the transition matrix of the time-reversed Markov chain *of the proces S_t, and $\boldsymbol{\mu}$ is given in Proposition 4. The expectation of $\boldsymbol{z}_t$ conditional on the state distribution $\boldsymbol{\xi}_t$ follows as*

$$\mathrm{E}[\boldsymbol{z}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}. \tag{A.33}$$

The expectation of $\boldsymbol{z}_{t+h}$ for $h \geq 0$ conditional on the current state distribution $\boldsymbol{\xi}_t$ follows as $\mathrm{E}[\boldsymbol{z}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t]$ *with*

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_{n^2}\big)\boldsymbol{\nu} \\ \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{A.34}$$

with $\tilde{\boldsymbol{z}}_{t+h}$ and $\tilde{\boldsymbol{\Upsilon}}$ defined in Proposition 2.

Proof. Use Equation (12) to find

$$\begin{aligned}
\mathrm{E}[\boldsymbol{z}_t|S_t = j] = \boldsymbol{\nu}_j &= \boldsymbol{\gamma}_j + \boldsymbol{\omega}_j + \boldsymbol{\Psi}_j \mathrm{E}[\boldsymbol{y}_{t-1}|S_t = j] + \boldsymbol{\Upsilon}_j \mathrm{E}[\boldsymbol{y}_{t-1}|S_t = j] \\
&= \boldsymbol{\gamma}_j + \boldsymbol{\omega}_j + \boldsymbol{\Psi}_j \sum_{i=1}^m \boldsymbol{\mu}_i q_{ij} + \boldsymbol{\Upsilon}_j \sum_{i=1}^m \boldsymbol{\nu}_i q_{ij} \\
&= \boldsymbol{\gamma}_j + \boldsymbol{\omega}_j + \boldsymbol{\Psi}_j(\boldsymbol{e}_j'\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu} + \boldsymbol{\Upsilon}_j(\boldsymbol{e}_j'\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}.
\end{aligned}$$

We can combine the above equality for $\boldsymbol{\nu}_j$ with the equality for $\boldsymbol{\mu}_j$ from the proof of Proposition 4 to form the system

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma} + \boldsymbol{\omega} \\ \boldsymbol{c} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2}) & \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n) \\ \mathbf{O} & \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n) \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\mu} \end{pmatrix}$$

Under the assumption of second-order stationarity, Lemma 2 implies that $\boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})$ and $\boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)$ are convergent, and the solution follows as

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\mu} \end{pmatrix} = \left(\boldsymbol{I}_{n(n+1)m} - \begin{pmatrix} \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2}) & \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n) \\ \mathbf{O} & \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n) \end{pmatrix}\right)^{-1} \begin{pmatrix} \boldsymbol{\gamma} + \boldsymbol{\omega} \\ \boldsymbol{c} \end{pmatrix}.$$

For block-triangular matrices $\boldsymbol{A}$ of the form

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \mathbf{O} & \boldsymbol{A}_3 \end{pmatrix},$$

where $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ have dimensions $m_1 \times m_1$, $m_1 \times m_2$ and $m_2 \times m_2$, and $\boldsymbol{A}_1$ and $\boldsymbol{A}_3$ are convergent, the inverse of $\big(\boldsymbol{I}_{m_1+m_2} - \boldsymbol{A}\big)$ exist and takes the form

$$(\boldsymbol{I}_{m_1+m_2} - \boldsymbol{A})^{-1} = \begin{pmatrix} \big(\boldsymbol{I}_{m_1} - \boldsymbol{A}_1\big)^{-1} & \big(\boldsymbol{I}_{m_1} - \boldsymbol{A}_1\big)^{-1}\boldsymbol{A}_2\big(\boldsymbol{I}_{m_2} - \boldsymbol{A}_3\big)^{-1} \\ \mathbf{O} & \big(\boldsymbol{I}_{m_2} - \boldsymbol{A}_3\big)^{-1} \end{pmatrix}.$$

Applying this result to the combined expression for $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ leads to the same expression for $\boldsymbol{\mu}$ as in Proposition 4, and for $\boldsymbol{\nu}$ we find

$$\begin{aligned}\boldsymbol{\nu} =&\big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}(\gamma + \boldsymbol{\omega})+\\ &\big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}\boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\big(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\big)^{-1}\boldsymbol{c}\\ =&\big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}(\gamma + \boldsymbol{\omega} + \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}).\end{aligned}$$

By summing the appropriate elements with weights $\boldsymbol{\xi}_t$ we find $\mathrm{E}[\boldsymbol{z}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}' \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}$. To find the expression for $\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t]$, we first use Proposition 2 to establish

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h\,\mathrm{E}[\tilde{\boldsymbol{z}}_t|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h\begin{pmatrix}\mathrm{E}[\boldsymbol{z}_t^*|\boldsymbol{\xi}_t]\\ \mathrm{E}[\boldsymbol{y}_t^*|\boldsymbol{\xi}_t]\\ \boldsymbol{\xi}_t\end{pmatrix},$$

because $\mathrm{E}[\tilde{\boldsymbol{\zeta}}_{t+l}|\boldsymbol{\xi}_t] = \mathbf{0}$ and $\mathrm{E}[\tilde{\boldsymbol{\varepsilon}}_{t+l}|\boldsymbol{\xi}_t] = \mathbf{0}$ for $l > 0$ and the definition of $\tilde{\boldsymbol{z}}_t$. Then,

$$\mathrm{E}[\boldsymbol{z}_t^*|\boldsymbol{\xi}_t] = \begin{pmatrix}\xi_{1t}\,\mathrm{E}[\boldsymbol{z}_t|S_t = 1]\\ \vdots\\ \xi_{mt}\,\mathrm{E}[\boldsymbol{z}_t|S_t = m]\end{pmatrix} = \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_{n^2}\big)\boldsymbol{\nu},$$

and $\mathrm{E}[\boldsymbol{y}_t^*|\boldsymbol{\xi}_t] = \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}$ as in the proof of Proposition 4. □

Proposition 6. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and assume that it is second-order stationary. Define $\boldsymbol{z}_{t+k,t} = \boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t$. Then the expectation of $\boldsymbol{z}_{t+h+k,t+h}$ for $h \geq 0$ conditional on the current state distribution $\boldsymbol{\xi}_t$ follows as*

$$\mathrm{E}[\boldsymbol{z}_{t+h+k,t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}}\begin{pmatrix}\tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O}\\ \mathbf{O} & \boldsymbol{P}^k\end{pmatrix}\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t], \tag{A.35}$$

with $\boldsymbol{\Phi} = \mathrm{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i)$ and $\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t]$ as in Proposition 5.

Proof. This result follows from

$$\begin{aligned}\mathrm{E}[\boldsymbol{z}_{t+h+k,t+h}|\boldsymbol{\xi}_t] &= \tilde{\boldsymbol{H}}_{\boldsymbol{z}}\,\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h+k,t+h}|\boldsymbol{\xi}_t]\\ &= \tilde{\boldsymbol{H}}_{\boldsymbol{z}}\,\mathrm{E}\big[\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h+k,t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t]\big|\boldsymbol{\xi}_t\big]\\ &= \tilde{\boldsymbol{H}}_{\boldsymbol{z}}\begin{pmatrix}\tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O}\\ \mathbf{O} & \boldsymbol{P}^k\end{pmatrix}\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t],\end{aligned}$$

where we substituted Equation (20) in the last equality. □

Corollary 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2) with $\Phi_i = \mathbf{O}$ for $i = 1, \ldots, m$, and assume that it is second-order stationary. Then*

$$\mathrm{vec}\big(\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t]\big) = (\boldsymbol{\iota}_m' \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k\big(\mathrm{diag}(\boldsymbol{\xi}_t) - \boldsymbol{\xi}_t\boldsymbol{\xi}_t'\big) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}. \tag{A.38}$$

Proof. The derivation of Equation (38) follows as

$$
\begin{aligned}
&\operatorname{vec}\big(\operatorname{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t]\big)\\
&= \operatorname{E}[\boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t|\boldsymbol{\xi}_t] - \operatorname{E}[\boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t] \otimes \operatorname{E}[\boldsymbol{y}_t|\boldsymbol{\xi}_t]\\
&= \boldsymbol{H}_{\boldsymbol{z}} \operatorname{E}[\boldsymbol{z}_{t+k,t}|\boldsymbol{\xi}_t] - \boldsymbol{G}_y \operatorname{E}[\boldsymbol{y}^*_{t+k}|\boldsymbol{\xi}_t] \otimes \boldsymbol{G}_y \operatorname{E}[\boldsymbol{y}^*_t|\boldsymbol{\xi}_t]\\
&= (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_{n^2})(\boldsymbol{C}\boldsymbol{P}^k \otimes \boldsymbol{I}_n)(\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n)\boldsymbol{\mu} - (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_n)\boldsymbol{C}\boldsymbol{P}^k\boldsymbol{\xi}_t \otimes (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_n)(\operatorname{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n)\boldsymbol{\mu}\\
&= (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k \operatorname{diag}(\boldsymbol{\xi}_t)\big) \otimes \boldsymbol{I}_n)\boldsymbol{\mu} - (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_n)\boldsymbol{C}\boldsymbol{P}^k\boldsymbol{\xi}_t \otimes (\boldsymbol{\xi}'_t \otimes \boldsymbol{I}_n)\boldsymbol{\mu}\\
&= (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k \operatorname{diag}(\boldsymbol{\xi}_t)\big) \otimes \boldsymbol{I}_n)\boldsymbol{\mu} - \big((\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_n)\boldsymbol{C}\boldsymbol{P}^k\boldsymbol{\xi}_t\boldsymbol{\xi}'_t \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}\\
&= (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k \operatorname{diag}(\boldsymbol{\xi}_t)\big) \otimes \boldsymbol{I}_n)\boldsymbol{\mu} - (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k\boldsymbol{\xi}_t\boldsymbol{\xi}'_t \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}\\
&= (\boldsymbol{\iota}'_m \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k\big(\operatorname{diag}(\boldsymbol{\xi}_t) - \boldsymbol{\xi}_t\boldsymbol{\xi}'_t\big) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}.
\end{aligned}
$$

□

B Proofs for Section 3 (Impulse response analysis)

The equation numbers in this section are copied from the main text with an appendix prefix.

Proposition 7. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{y}}_t$ be defined by Equation (39) and the results in Proposition 1. Let the vector $\boldsymbol{y}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\operatorname{E}[\boldsymbol{y}_{t-1}|\mathcal{I}_{t-1}]$. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\operatorname{E}[\boldsymbol{s}_{t-1}|\mathcal{I}_{t-1}]$. Let the matrices $\boldsymbol{C}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}$, and $\tilde{\boldsymbol{\Phi}}$ be defined as in Proposition 1. When the shock originates from the regime process, the corresponding GI satisfies*

$$
GI^{\boldsymbol{u}}_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \emptyset, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \tag{B.41}
$$

When the shock is specified in terms of the structural innovation ε_{it}, the corresponding GI satisfies

$$
GI^{\varepsilon_i}_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, \emptyset, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}. \tag{B.42}
$$

When the shock is specified as $\eta_{it} = y_{it} - \operatorname{E}[y_{it}|\mathcal{I}_{t-1}]$, the corresponding GI satisfies

$$
GI^{y_i}_{\tilde{\boldsymbol{y}}}(h, \eta_{it}, \mathcal{I}_{t-1}) = \operatorname{E}[\tilde{\boldsymbol{y}}_{t+h}|y_{it}, \mathcal{I}_{t-1}] - \operatorname{E}[\tilde{\boldsymbol{y}}_{t+h}|\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \operatorname{E}[\boldsymbol{\varepsilon}^*_t|y_{it}, \mathcal{I}_{t-1}] \\ \operatorname{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] \end{pmatrix}. \tag{B.43}
$$

The second conditional expectation satisfies $\operatorname{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] = \operatorname{E}[\boldsymbol{s}_t|y_{it}, \mathcal{I}_{t-1}] - \operatorname{E}[\boldsymbol{s}_t|\mathcal{I}_{t-1}]$ with

$$
\begin{aligned}
\operatorname{E}[\boldsymbol{s}_t|\mathcal{I}_{t-1}] &= \boldsymbol{P}\boldsymbol{\xi}_{t-1},\\
\operatorname{E}[\boldsymbol{s}_t|y_{it}, \mathcal{I}_{t-1}] &= \frac{1}{\boldsymbol{f}'\boldsymbol{P}\boldsymbol{\xi}_{t-1}}\boldsymbol{f} \odot \boldsymbol{P}\boldsymbol{\xi}_{t-1},
\end{aligned}
$$

where $\boldsymbol{f}$ is a vector of size m whose element j is equal to the pdf of the marginal distribution of y_{it} under regime j, $y_{it}|S_t = j \sim \mathrm{N}(\boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j \boldsymbol{y}_{t-1}), \boldsymbol{e}_i'\boldsymbol{\Sigma}_j \boldsymbol{e}_i)$. The first conditional expectation satisfies

$$\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, \mathcal{I}_{t-1}] = \boldsymbol{C}\,\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] + \boldsymbol{\Phi}(\mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Lambda}\,\mathrm{E}[\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t|y_{it}, \mathcal{I}_{t-1}],$$

with last term

$$\mathrm{E}[\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t|y_{it}, \mathcal{I}_{t-1}] = \begin{pmatrix} \mathrm{E}\big[s_{1t}\big|y_{it}, \mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = 1, \mathcal{I}_{t-1}\big] \\ \vdots \\ \mathrm{E}\big[s_{mt}\big|y_{it}, \mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = m, \mathcal{I}_{t-1}\big] \end{pmatrix},$$

and

$$\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = j, \mathcal{I}_{t-1}\big] = \boldsymbol{\Lambda}_j^{-1}\left(\frac{y_{it} - \boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j \boldsymbol{y}_{t-1})}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j \boldsymbol{e}_i}\boldsymbol{\Sigma}_j \boldsymbol{e}_i\right).$$

Proof. Equation (41) follows from

$$GI_{\tilde{\boldsymbol{y}}}(h, \emptyset, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Phi}}^h\,\mathrm{E}[\tilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{u}_t, \mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\varepsilon}_t^*|\boldsymbol{u}_t, \mathcal{I}_{t-1}] \\ \boldsymbol{u}_t \end{pmatrix}.$$

Based on the expression for $\boldsymbol{\varepsilon}_t^*$ in Proposition 1, and the assumptions that $\mathrm{E}[\boldsymbol{\varepsilon}_t|\mathcal{I}_t] = \boldsymbol{0}$ and that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are independent,

$$\mathrm{E}[\boldsymbol{\varepsilon}_t^*|\boldsymbol{u}_t, \mathcal{I}_{t-1}] = \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}).$$

Equation (42) follows from

$$GI_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, \emptyset, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Phi}}^h\,\mathrm{E}[\tilde{\boldsymbol{\varepsilon}}_t|\varepsilon_{it}, \mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\varepsilon}_t^*|\varepsilon_{it}, \mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|\varepsilon_{it}, \mathcal{I}_{t-1}] \end{pmatrix}.$$

In this case, the first expectation reduces to

$$\mathrm{E}[\boldsymbol{\varepsilon}_t^*|\varepsilon_{it}, \mathcal{I}_{t-1}] = \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P} \otimes \boldsymbol{I}_n)(\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) = \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i),$$

and $\mathrm{E}[\boldsymbol{u}_t|\varepsilon_{it}, \mathcal{I}_{t-1}] = \boldsymbol{0}$ because $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are independent.

In the derivations for Equation (43), $\mathrm{E}[\boldsymbol{s}_t|\mathcal{I}_{t-1}]$ follows from the definition of the regime process, and $\mathrm{E}[\boldsymbol{s}_t|y_{it}, \mathcal{I}_{t-1}]$ from the application of the Hamilton (1990)-filter, adjusted for observing only a specific y_{it}. In the expression for $\boldsymbol{\varepsilon}_t^*$ we do not separate $\boldsymbol{s}_{t-1}$ and $\boldsymbol{u}_t$ as in Proposition 1, but consider their joint effect $\boldsymbol{\Lambda}(\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t)$ in the last term. Because the vector $\boldsymbol{\varepsilon}_t^*$ contains the regime-specific shocks, we first establish for a given regime j

$$\begin{aligned}\mathrm{E}\big[s_{jt}\boldsymbol{\varepsilon}_t|y_{it}, \mathcal{I}_{t-1}\big] &= \mathrm{E}\Big[\mathrm{E}\big[s_{jt}\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = j, \mathcal{I}_{t-1}\big]\Big|y_{it}, \mathcal{I}_{t-1}\Big] \\ &= \mathrm{E}\big[s_{jt}\big|y_{it}, \mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = j\big],\end{aligned}$$

where the second equality follows from the independence of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{s}_t$. The second term satisfies

$$\begin{aligned}
\mathrm{E}\big[\boldsymbol{\varepsilon}_t\big|y_{it}, S_t = j, \mathcal{I}_{t-1}\big] &= \boldsymbol{\Lambda}_j^{-1}\big(\mathrm{E}\big[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}\big] - \mathrm{E}\big[\boldsymbol{y}_t|S_t = j, \mathcal{I}_{t-1}\big]\big) \\
&= \boldsymbol{\Lambda}_j^{-1}\left(\frac{y_{it} - \boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j\boldsymbol{y}_{t-1})}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i}\boldsymbol{\Sigma}_j\boldsymbol{e}_i\right).
\end{aligned}$$

To arrive at the second equality, we use the fact that $\boldsymbol{y}_t$ conditional on regime j follows a normal distribution, which implies

$$\mathrm{E}\big[\boldsymbol{y}_t|y_{it}, S_t = j, \mathcal{I}_{t-1}\big] = \mathrm{E}\big[\boldsymbol{y}_t|S_t = j, \mathcal{I}_{t-1}\big] + \frac{y_{it} - \boldsymbol{e}_i'(\boldsymbol{c}_j + \boldsymbol{\Phi}_j\boldsymbol{y}_{t-1})}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i}\boldsymbol{\Sigma}_j\boldsymbol{e}_i.$$

$\square$

Proposition 8. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in Equations (1) and (2), and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{z}}_t$ be defined by Equation (44) and the results in Proposition 2. Let the vector $\boldsymbol{y}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\mathrm{E}[\boldsymbol{y}_{t-1}|\mathcal{I}_{t-1}]$. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of $\mathcal{I}_{t-1}$ or calculated as $\mathrm{E}[\boldsymbol{s}_{t-1}|\mathcal{I}_{t-1}]$. Let the matrices $\boldsymbol{C}, \boldsymbol{\Phi}, \boldsymbol{\Lambda}$, and $\tilde{\boldsymbol{\Phi}}$ be defined as in Proposition 1, and $\boldsymbol{\Gamma}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Upsilon}$ and $\tilde{\boldsymbol{\Upsilon}}$ as in Proposition 2. When the shock originates from the regime process, the corresponding GI satisfies*

$$GI_{\tilde{\boldsymbol{z}}}^{\boldsymbol{u}}(h, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \emptyset, \boldsymbol{u}_t, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t \otimes \boldsymbol{z}_{t-1}) \\ \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \tag{B.46}$$

When the shock is specified in terms of the innovation ε_{it}, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{z}}}^{\varepsilon_i}(h, \varepsilon_{it}, \mathcal{I}_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \varepsilon_{it}, \emptyset, \mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\zeta}_t^*|\varepsilon_{it}, \mathcal{I}_{t-1}] \\ \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}, \tag{B.47}$$

with

$$\begin{aligned}
\mathrm{E}[\boldsymbol{\zeta}_t^*|\varepsilon_{it}, \mathcal{I}_{t-1}] =& \varepsilon_{it}\,\mathrm{bdiag}_{j=1}^m(\boldsymbol{\Lambda}_j \otimes \boldsymbol{c}_j + \boldsymbol{c}_j \otimes \boldsymbol{\Lambda}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) + \\
& \varepsilon_{it}\,\mathrm{bdiag}_{j=1}^m(\boldsymbol{\Lambda}_j \otimes \boldsymbol{\Phi}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i \otimes \boldsymbol{y}_{t-1}) + \\
& \varepsilon_{it}\,\mathrm{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Lambda}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{y}_{t-1} \otimes \boldsymbol{e}_i) + \\
& (\varepsilon_{it}^2 - 1)\,\mathrm{bdiag}_{j=1}^m(\boldsymbol{\Lambda}_j \otimes \boldsymbol{\Lambda}_j)(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i \otimes \boldsymbol{e}_i).
\end{aligned}$$

When the shock is specified as $\eta_{it} = y_{it} - \mathrm{E}[y_{it}|\mathcal{I}_{t-1}]$, the corresponding GI satisfies

$$GI_{\tilde{\boldsymbol{z}}}^{y_i}(h, \eta_{it}, \mathcal{I}_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|y_{it}, \mathcal{I}_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{z}_t^*|y_{it}, \mathcal{I}_{t-1}] - \mathrm{E}[\boldsymbol{z}_t^*|\mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, \mathcal{I}_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|y_{it}, \mathcal{I}_{t-1}] \end{pmatrix}, \tag{B.48}$$

where the last two conditional expectations have been defined in Proposition 7, and

$$\mathrm{E}[\boldsymbol{z}_t^*|y_{it},\mathcal{I}_{t-1}] = \begin{pmatrix} \mathrm{E}\big[s_{1t}|y_{it},\mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{z}_t|y_{it},S_t=1\big] \\ \vdots \\ \mathrm{E}\big[s_{mt}|y_{it},\mathcal{I}_{t-1}\big]\,\mathrm{E}\big[\boldsymbol{z}_t|y_{it},S_t=m\big] \end{pmatrix}$$

and

$$\begin{aligned}\mathrm{E}[\boldsymbol{z}_t|y_{it},S_t=j,\mathcal{I}_{t-1}] = {}&\mathrm{vec}\left(\mathrm{Var}[\boldsymbol{y}_t|y_{it},S_t=j,\mathcal{I}_{t-1}]\right) - \\ &\mathrm{E}[\boldsymbol{y}_t|y_{it},S_t=j,\mathcal{I}_{t-1}]\otimes\mathrm{E}[\boldsymbol{y}_t|y_{it},S_t=j,\mathcal{I}_{t-1}].\end{aligned}$$

with

$$\begin{aligned}\mathrm{E}\big[\boldsymbol{y}_t|y_{it},S_t=j,\mathcal{I}_{t-1}\big] &= \mathrm{E}\big[\boldsymbol{y}_t|S_t=j,\mathcal{I}_{t-1}\big] + \frac{y_{it}-\boldsymbol{e}_i'(\boldsymbol{c}_j+\boldsymbol{\Phi}_j\boldsymbol{y}_{t-1})}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i}\boldsymbol{\Sigma}_j\boldsymbol{e}_i,\\ \mathrm{Var}[\boldsymbol{y}_t|y_{it},S_t=j,\mathcal{I}_{t-1}] &= \boldsymbol{\Sigma}_j - \frac{1}{\boldsymbol{e}_i'\boldsymbol{\Sigma}_j\boldsymbol{e}_i}\boldsymbol{\Sigma}_j\boldsymbol{e}_i\boldsymbol{e}_i'\boldsymbol{\Sigma}_j.\end{aligned}$$

Proof. Equation (46) follows from

$$GI_{\tilde{\boldsymbol{z}}}(h,\emptyset,\boldsymbol{u}_t,\mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h\begin{pmatrix}\mathrm{E}[\boldsymbol{\zeta}_t^*|\boldsymbol{u}_t,\mathcal{I}_{t-1}]\\ \mathrm{E}[\boldsymbol{\varepsilon}_t^*|\boldsymbol{u}_t,\mathcal{I}_{t-1}]\\ \boldsymbol{u}_t\end{pmatrix},$$

with $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|\boldsymbol{u}_t,\mathcal{I}_{t-1}]$ as in Proposition 8. The conditional expectation for $\boldsymbol{\zeta}_t^*$ follows as

$$\mathrm{E}[\boldsymbol{\zeta}_t^*|\boldsymbol{u}_t,\mathcal{I}_{t-1}] = (\boldsymbol{\Gamma}+\boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t\otimes\boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t\otimes\boldsymbol{z}_{t-1}),$$

from Proposition 2, and the assumptions that $\mathrm{E}[\boldsymbol{\varepsilon}_t|\mathcal{I}_t]=\boldsymbol{0}$ and that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are independent.

Equation (47) follows from

$$GI_{\tilde{\boldsymbol{z}}}(h,\varepsilon_{it},\emptyset,\mathcal{I}_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h\,\mathrm{E}[\tilde{\boldsymbol{\zeta}}_t|\varepsilon_{it},\mathcal{I}_{t-1}] = \tilde{\boldsymbol{\Upsilon}}^h\begin{pmatrix}\mathrm{E}[\boldsymbol{\zeta}_t^*|\varepsilon_{it},\mathcal{I}_{t-1}]\\ \mathrm{E}[\boldsymbol{\varepsilon}_t^*|\varepsilon_{it},\mathcal{I}_{t-1}]\\ \mathrm{E}[\boldsymbol{u}_t,|\varepsilon_{it},\mathcal{I}_{t-1}]\end{pmatrix},$$

with $\mathrm{E}[\boldsymbol{u}_t,|\varepsilon_{it},\mathcal{I}_{t-1}]=\boldsymbol{0}$ and $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|\varepsilon_{it},\mathcal{I}_{t-1}]=\varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1}\otimes\boldsymbol{e}_i)$ as derived in Proposition 7. The conditional expectation follows from the expression for $\boldsymbol{\zeta}_t^*$ in Proposition 8.

The derivation of $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it},\mathcal{I}_{t-1}]$ and $\mathrm{E}[\boldsymbol{u}_t|y_{it},\mathcal{I}_{t-1}]$ is given in Proposition 7. The derivation of $\mathrm{E}[\boldsymbol{\zeta}_t^*|y_{it},\mathcal{I}_{t-1}]$ uses

$$\begin{aligned}\mathrm{E}[s_{jt}\boldsymbol{z}_t|y_{it},\mathcal{I}_{t-1}] &= \mathrm{E}\left[\mathrm{E}[s_{jt}\boldsymbol{z}_t|y_{it},S_t=j\mathcal{I}_{t-1}]|\,y_{it},\mathcal{I}_{t-1}\right]\\ &= \mathrm{E}[s_{jt}|y_{it},\mathcal{I}_{t-1}]\,\mathrm{E}[\boldsymbol{z}_t|y_{it},S_t=j\mathcal{I}_{t-1}].\end{aligned}$$

The derivation of $\mathrm{E}[\boldsymbol{z}_t|y_{it},S_t=j\mathcal{I}_{t-1}]$ uses the rules for the conditional distributions of the multivariate normal distribution. □

C Additional empirical results

C.1 Estimation results for restricted models

We report the parameters estimates of the MSIH(2) Model in Table C.1. The estimation period starts with February 1952 to account for the lag in the other models. The results clearly show low- and high-volatility regimes, whose volatilities differ by a factor two. Excess stock returns are large and positive (about 1% per month) in the low-volatility regime, and large and negative (about -0.5% per month) when volatility is high, so we can also classify the regimes as bullish and bearish. For bonds we see the opposite pattern with returns close to zero in the low-volatility regime and large positive returns in the high-volatility regime. These return differences indicate that investors see bonds as safe haven during high volatility periods. Correlations are low and do not differ much between the regimes. Both regimes are persistent as indicated by the high values for p_{11} and p_{22}.

[Table C.1 about here.]

The estimation results for the VAR(1) model are in Table C.2. The autoregressive coefficients show the typical results found in the literature about predictability and its effect on long-run asset allocation. We see weak but significant predictability of stock and bond returns by their past values. Both the T-bill rate and D/P ratio are strongly persistent, with autoregressive coefficients close to 1. The lagged T-bill rate (D/P ratio) has a strong negative (positive) effect on excess stock returns, but neither predict excess bond returns. Correlations are generally low, except between stock returns and D/P ratio, which is by construction.

[Table C.2 about here.]

[Table C.3 about here.]

C.2 Moments of restricted models

[Table C.4 about here.]

[Table C.5 about here.]

C.3 Additional impulse response analyses

[Figure C.1 about here.]

[Figure C.2 about here.]

[Figure C.3 about here.]

[Figure C.4 about here.]

[Figure C.5 about here.]

Table C.1: Parameter estimates MSIH(2) Model

	Regime 1		Regime 2	
	Stock returns	Bond returns	Stock returns	Bond returns
Intercept	0.999	−0.008	−0.480	0.405
	(0.149)	(0.071)	(0.397)	(0.184)
Volatilities and correlations				
Stock returns	3.084		5.633	
	(0.123)		(0.241)	
Bond returns	0.073	1.450	0.119	2.935
	(0.050)	(0.057)	(0.056)	(0.149)
Transition and Initial Probabilities				
	Regime 1	Regime 2	Initial	
Regime 1	0.944	0.111	1	
	(0.015)	(0.029)	–	
Regime 2	0.056	0.889	0	
	(0.015)	(0.029)	–	

This table shows the parameter estimates for a Markov-switching model with switches in the means and variances between two regimes. The parameters have been estimated by the Expectation Maximization algorithm. Standard errors calculated by the outer product of the gradient are in parentheses. The estimation period runs from February 1952 to December 2018 (803 observations). We report estimated volatilities on the diagonals in the middle panel, and estimated correlations in the lower triangle. We do not report a standard error for the parameter $\boldsymbol{\zeta} = \Pr[\boldsymbol{s}_1]$ because it is at the boundary of its domain.

Table C.2: Parameter estimates VAR(1) Model

	Stock returns	Bond returns	T-bill rate	D/P ratio
Intercept	5.381	0.126	0.017	−0.044
	(1.535)	(0.982)	(0.021)	(0.017)
VAR(1) matrix				
Stock returns	0.035	−0.079	0.0005	−0.0003
	(0.031)	(0.016)	(0.0003)	(0.0003)
Bond returns	0.223	0.075	−0.0069	−0.0020
	(0.068)	(0.030)	(0.0006)	(0.0008)
T-bill rate	−2.057	−0.135	0.981	0.012
	(0.698)	(0.327)	(0.006)	(0.007)
D/P ratio	1.195	−0.023	0.003	0.989
	(0.401)	(0.254)	(0.006)	(0.004)
Volatilities and correlations				
Stock returns	4.114			
	(0.085)			
Bond returns	0.086	2.050		
	(0.030)	(0.044)		
T-bill rate	−0.086	−0.009	0.044	
	(0.034)	(0.029)	(0.001)	
D/P ratio	−0.935	−0.093	0.091	0.044
	(0.072)	(0.031)	(0.036)	(0.001)

This table shows the parameter estimates for a VAR(1) model. The parameters have been estimated by maximum likelihood. Standard errors calculated by the outer product of the gradient are in parentheses. The estimation period runs from January 1952 to December 2018 (804 observations). The VAR(1) matrix reports the lagged variables on the rows. We report estimated volatilities on the diagonal in the lower panel, and estimated correlations in the lower triangle.

Table C.3: Comparison of the likelihoods and information criteria based on stocks and bonds

Model	Log Lik.	# parameters	AIC	BIC
MSIAH(2)-VAR(1)	-3,891.73	29	7,841.46	7,977.42
MSIH(2)	-3905.25	13	7,836.50	7,897.45
VAR(1)	-3982.90	13	7,991.80	8,052.75

This table reports the likelihood and information criteria AIC and BIC for each model. The values of the likelihood functions are calculated by evaluating the density functions only for stocks and bonds, based on the parameter values in Tables 2, C.1 and C.2. The number of parameters also reflect the ones for stocks and bonds only.

Table C.4: Regime-conditional and unconditional first and second moments of the MSIH(2) and VAR models

Means (% per month)

	Stocks		Bonds		T-Bill rate		D/P ratio	
MSIH2, regime 1	1.00	(0.75, 1.25)	−0.01	(-0.13, 0.11)				
MSIH2, regime 2	−0.48	(-1.15, 0.14)	0.40	(0.11, 0.72)				
MSIH2, uncond.	0.50	(0.23, 0.75)	0.13	(0.00, 0.26)				
VAR1, uncond.	0.40	(0.16, 0.89)	0.15	(-0.04, 0.42)	0.31	(-0.38, 0.54)	−3.68	(-4.97, -3.26)

Volatilities (% per month)

	Stocks		Bonds		T-Bill rate		D/P ratio	
MSIH2, regime 1	3.08	(2.87, 3.27)	1.45	(1.36, 1.54)				
MSIH2, regime 2	5.63	(5.21, 6.00)	2.94	(2.67, 3.16)				
MSIH2, uncond.	4.18	(3.97, 4.38)	2.08	(1.95, 2.20)				
VAR1, uncond.	4.17	(4.03, 4.32)	2.08	(2.02, 2.16)	0.25	(0.20, 0.45)	0.39	(0.27, 0.83)

Stock-Bonds correlation in the MSIH2 model

Regime 1	0.073	(-0.009, 0.154)
Regime 2	0.119	(0.028, 0.210)
Unconditional	0.085	(0.017, 0.149)

Unconditional correlations in the VAR(1) model

	Stocks		Bonds		T-Bill rate	
Bonds	0.085	(0.034, 0.133)				
T-Bill rate	−0.091	(-0.161, -0.037)	−0.011	(-0.101, 0.042)		
D/P ratio	−0.065	(-0.108, -0.021)	−0.014	(-0.105, 0.051)	0.512	(0.113, 0.870)

This table gives the regime-conditional and unconditional means, volatilities and correlations for the MSIH(2) and VAR(1) models with parameter values reported in Tables C.1 and C.2, and 90% confidence intervals in parentheses. The unconditional moments follow from Equations (22) to (25). The confidence intervals are constructed with 1,000 simulations of the parameters assuming a multivariate normal distribution for the estimates. The variance matrix of the estimates is calculated by the outer product of the gradient.

Table C.5: Unconditional auto- and cross-correlations at (lead) order 1 of the MSIH(2) and MSVAR(1) models

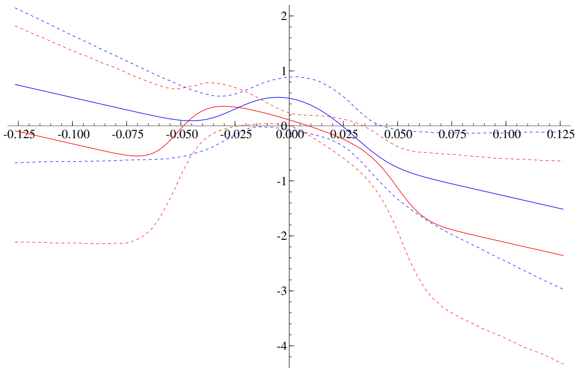
MSIH(2) model				
Stocks	0.023	(0.007, 0.050)	−0.013	(-0.027, -0.002)
Bonds	−0.013	(-0.027, -0.002)	0.007	(0.000, 0.024)

VAR(1) model	Stocks		Bonds		T-Bill rate		D/P ratio	
Stocks	0.049	(-0.001, 0.107)	−0.151	(-0.203, -0.096)	−0.087	(-0.159, -0.034)	−0.069	(-0.112, -0.025)
Bonds	0.114	(0.058, 0.170)	0.062	(0.017, 0.120)	−0.067	(-0.141, -0.013)	−0.025	(-0.111, 0.041)
T-Bill rate	−0.072	(-0.147, -0.017)	−0.005	(-0.095, 0.050)	0.983	(0.974, 0.995)	0.515	(0.121, 0.870)
D/P ratio	0.043	(-0.029, 0.084)	−0.003	(-0.100, 0.062)	0.507	(0.104, 0.868)	0.993	(0.987, 0.999)

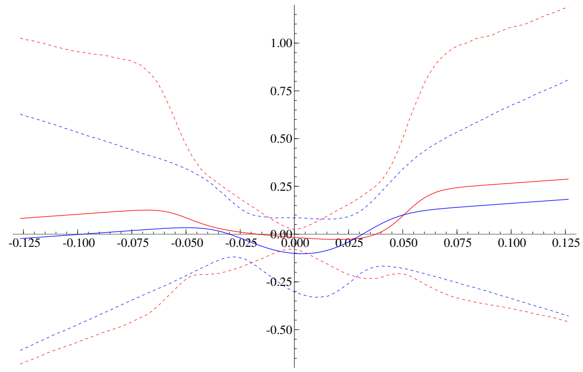
This table gives the unconditional auto- and cross-correlations with lead 1 for the MSIH(2) and VAR(1) models based on parameter values reported in Tables C.1 and C.2, and 90% confidence intervals in parentheses. The heading of each block indicates the variable at t and the rows indicate the variable at $t+1$. The unconditional correlations are calculated from $\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+1}]$ based on Equations (26) and (27), and the unconditional volatilities in Table 3. The confidence intervals are constructed as in Table 3.

Figure C.1: The effects of a shock to the T-Bill rate at t

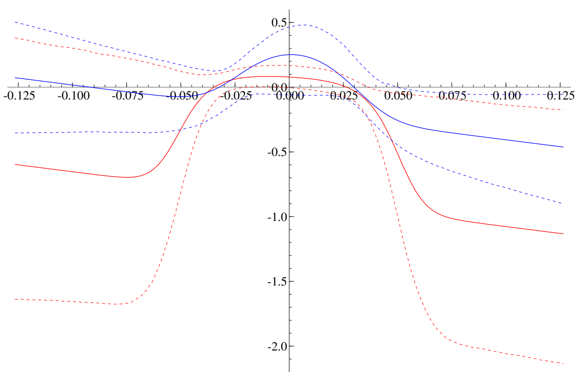
(a) GI of stock returns at t



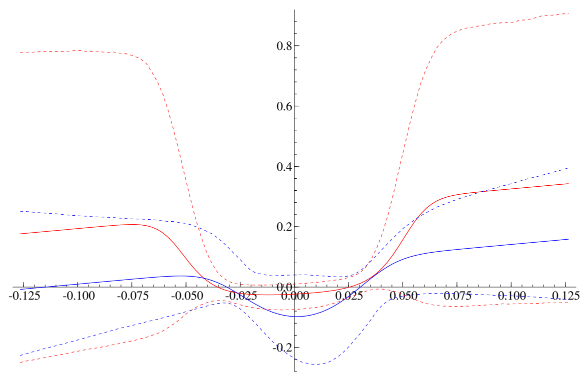
(b) GI of bond returns at t



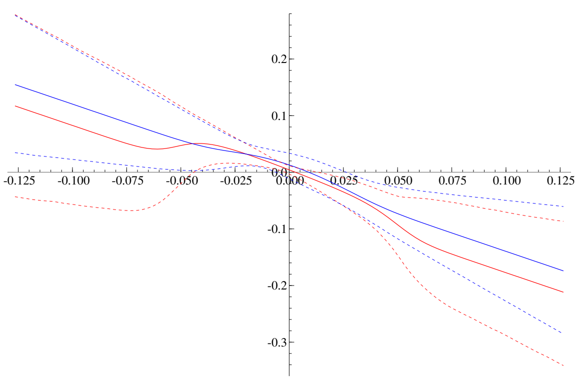
(c) GI of stock returns at $t+1$



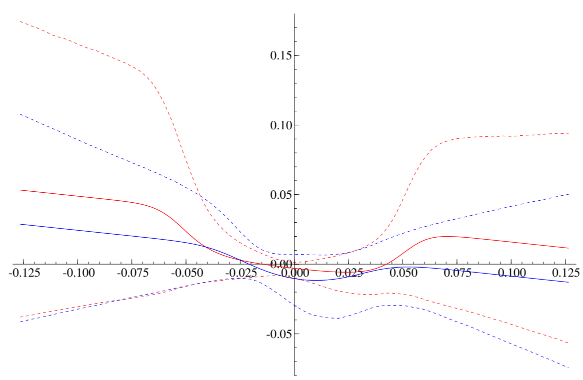
(d) GI of bond returns at $t+1$



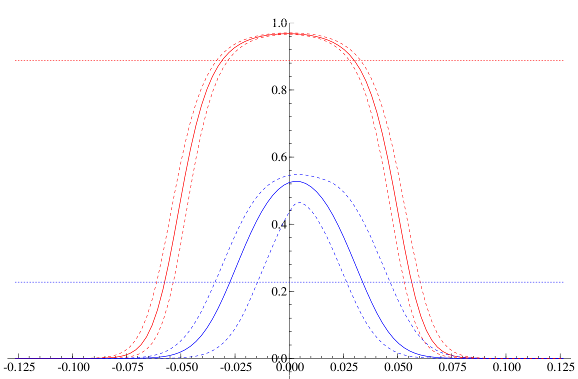
(e) GI of stock returns at $t+6$



(f) GI of bond returns at $t+6$



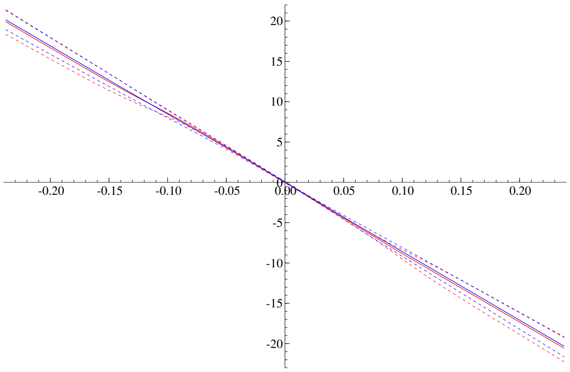
(g) Probability of regime 1 at t



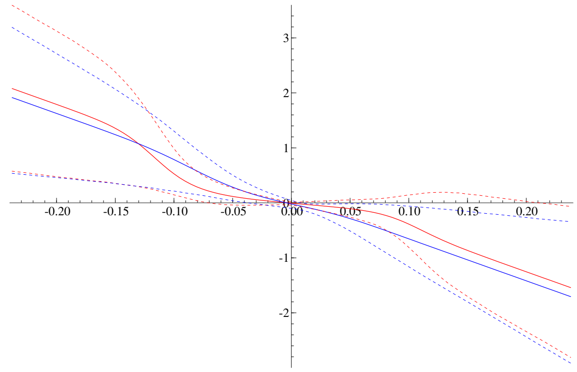
This figure shows the Generalized Impulse Response Functions for expected stock and bond returns at t, $t+1$ and $t+6$ (panels a to f), and for for the probability forecasts for regime 1 at time t (panel g) as a function of a shock to the T-Bill rate at time t. Both the horizontal axes and the vertical axes (panels a to f) show returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The GI are calculated based on Proposition 7. The straight dotted lines in panel g give the forecast probability of state 1 in absence of a shock. The difference between the solid and the dotted lines corresponds with the GI. In the other panels the GIs are plotted.

Figure C.2: The effects of a shock to the D/P ratio at t

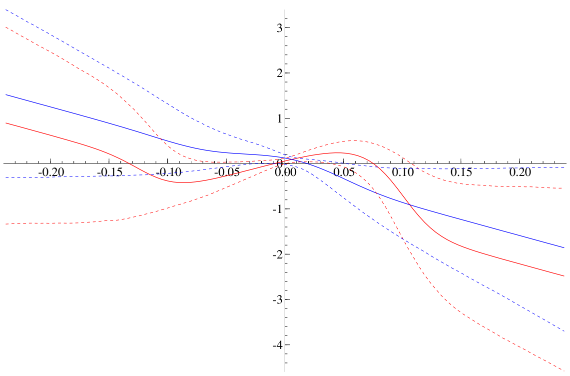
(a) GI of stock returns at t



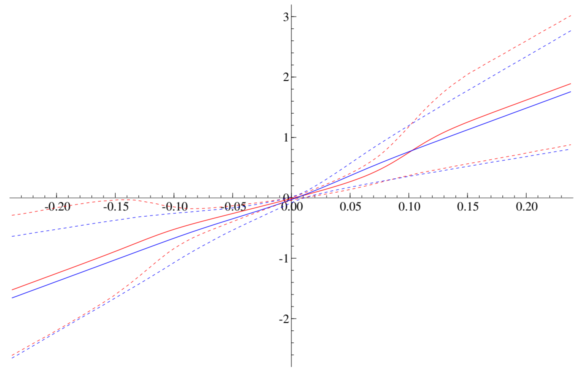
(b) GI of bond returns at t



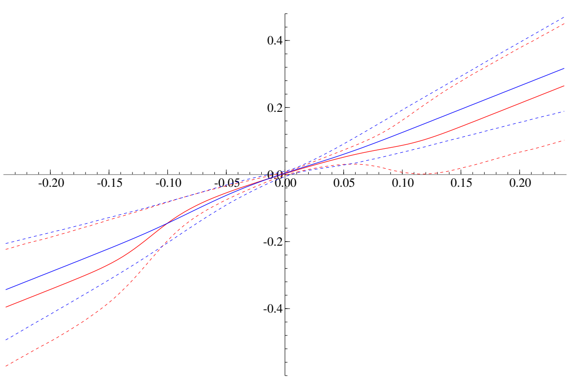
(c) GI of stock returns at $t+1$



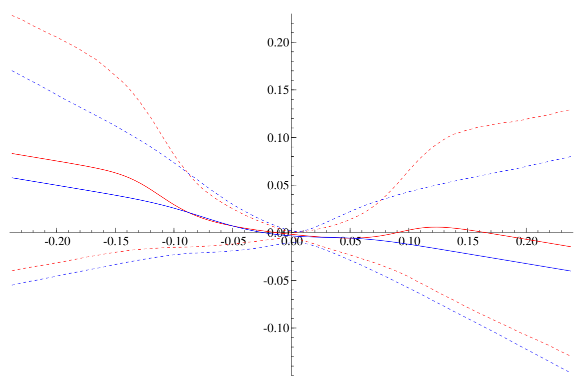
(d) GI of bond returns at $t+1$



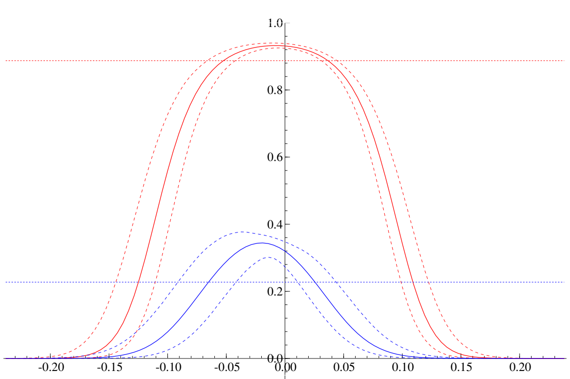
(e) GI of stock returns at $t+6$



(f) GI of bond returns at $t+6$



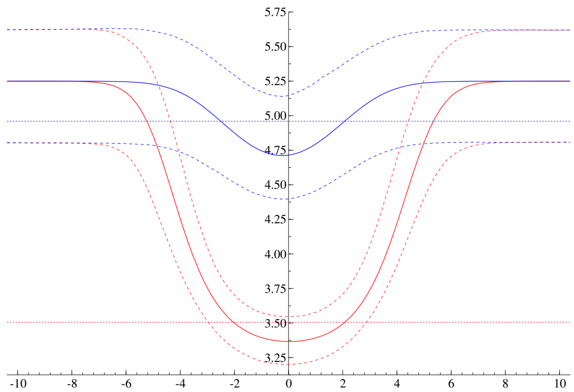
(g) Probability of regime 1 at t



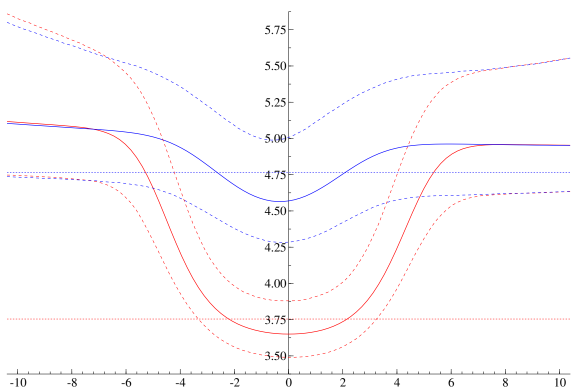
This figure shows the Generalized Impulse Response Functions for expected stock and bond returns at t, $t+1$ and $t+6$ (panels a to f), and for for the probability forecasts for regime 1 at time t (panel g) as a function of a shock to the D/P ratio at time t. The vertical axes (panels a to f) show returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The GI are calculated based on Proposition 7. The straight dotted lines in panel g give the forecast probability of state 1 in absence of a shock. The difference between the solid and the dotted lines corresponds with the GI. In the other panels the GIs are plotted.

Figure C.3: The effects of shocks to bond returns on the forecast volatility of stock and bond returns

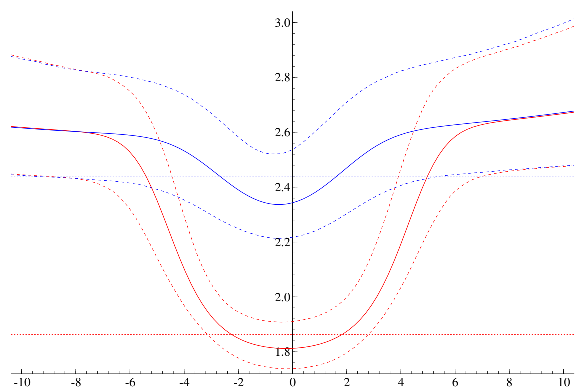
(a) Stock volatility at t



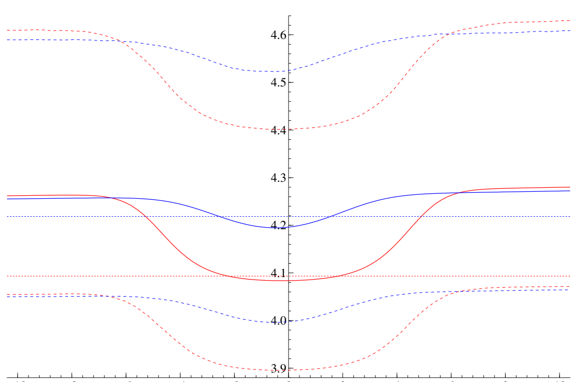
(b) Stock volatility at $t+1$



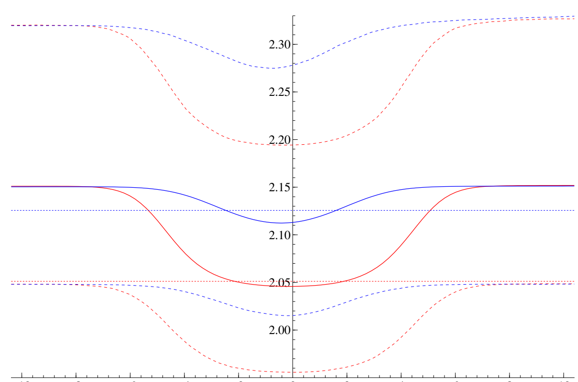
(c) Bond volatility at $t+1$



(d) Stock volatility at $t+6$



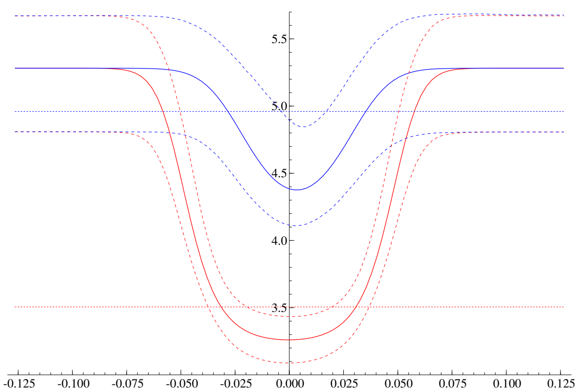
(e) Bond volatility at $t+6$



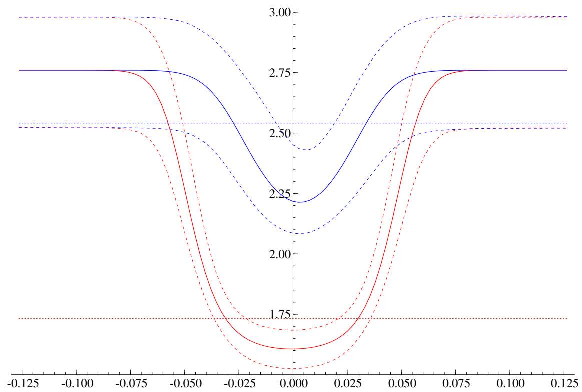
This figure shows the impulse responses for the forecast volatility of stock returns at time t, $t+1$ and $t+6$ (panels a, b and d) and bond returns at $t+1$ and $t+6$ (panels c and e) as a function of a shock to bond returns at time t based on Proposition 8. Both the horizontal and the vertical axes shows returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The straight dotted lines give the forecast volatility in absence of a shock. The difference between the solid and the dotted lines gives the volatility impulse response function.

Figure C.4: The effects of shocks to the T-Bill rate on the forecast volatility of stock and bond returns

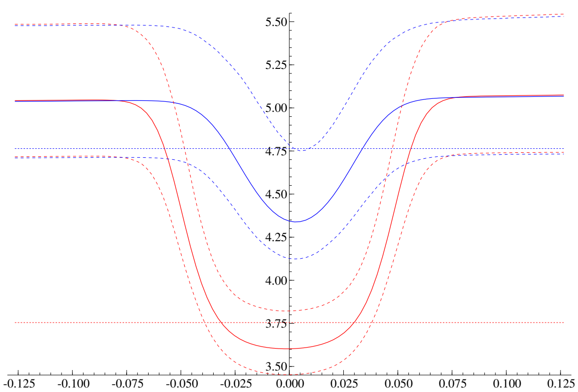
(a) Stock volatility at t



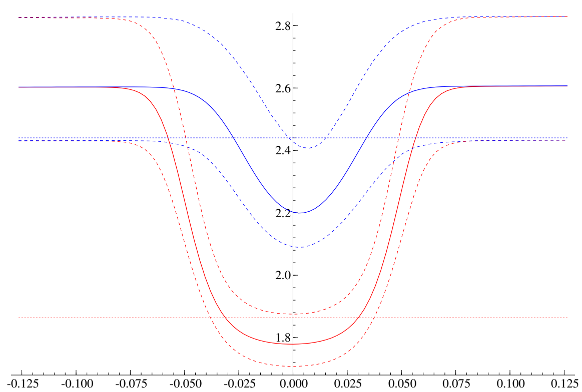
(b) Bond volatility at t



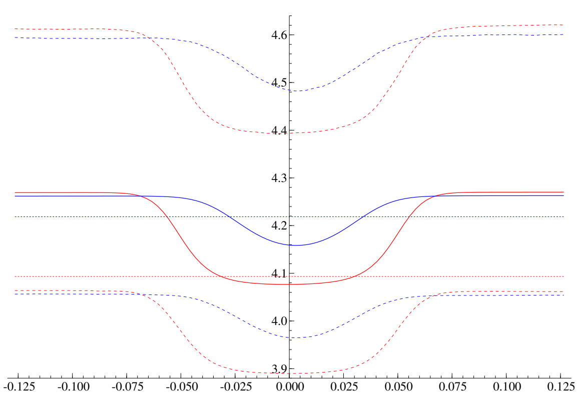
(c) Stock volatility at $t+1$



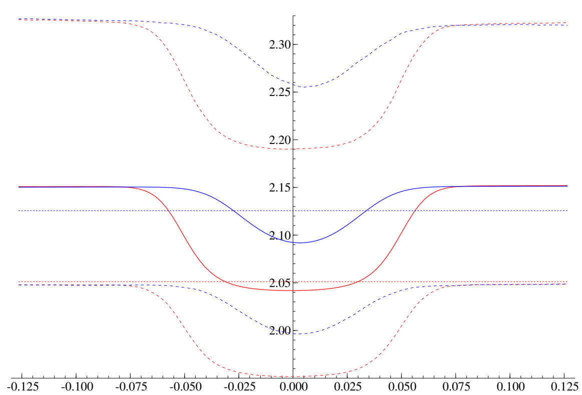
(d) Bond volatility at $t+1$



(e) Stock volatility at $t+6$



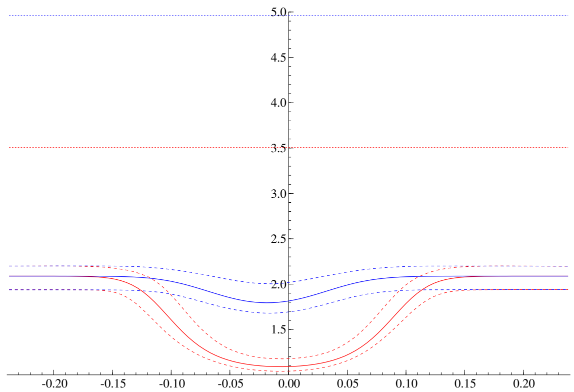
(f) Bond volatility at $t+6$



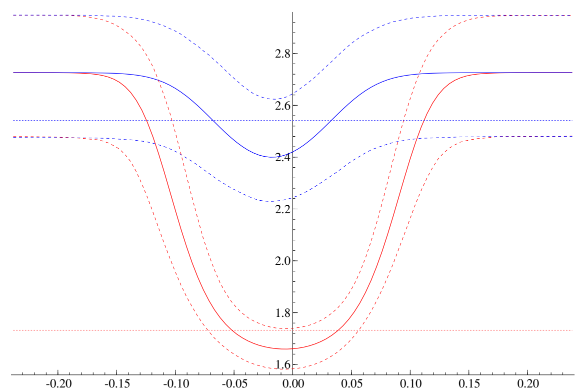
This figure shows the impulse responses for the forecast volatility of stock returns at time t, $t+1$ and $t+6$ (panels a, c and e) and bond returns at $t+1$ and $t+6$ (panels b, d and f) as a function of a shock to the T-Bill rate at time t based on Proposition 8. Both the horizontal and the vertical axes shows returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The straight dotted lines give the forecast volatility in absence of a shock. The difference between the solid and the dotted lines gives the volatility impulse response function.

Figure C.5: The effects of shocks to the D/P ratio on the forecast volatility of stock and bond returns

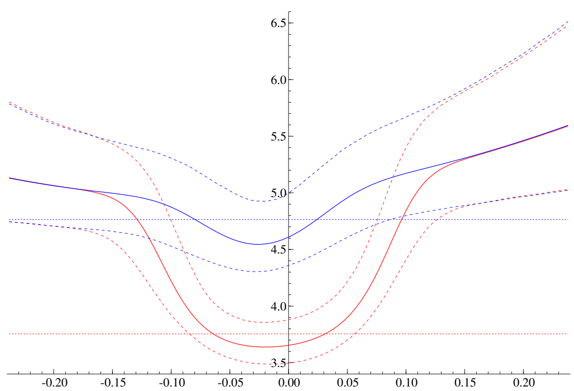
(a) Stock volatility at t



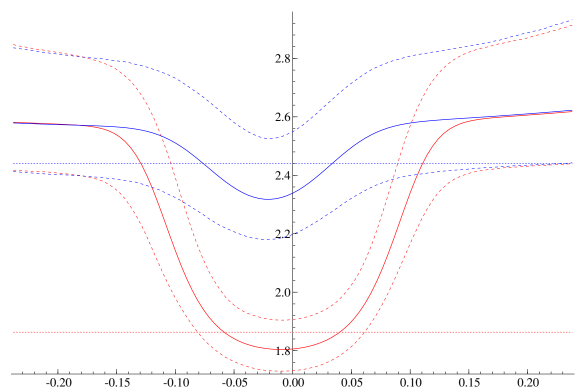
(b) Bond volatility at t



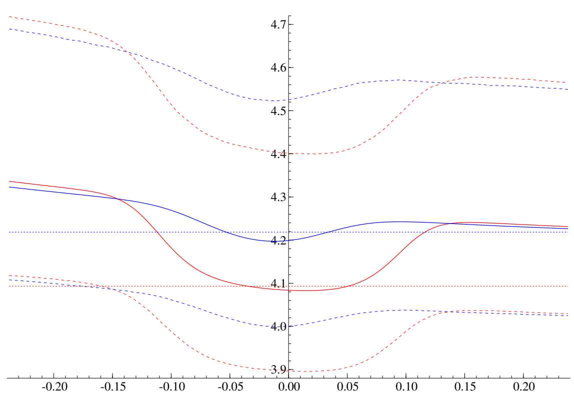
(c) Stock volatility at $t+1$



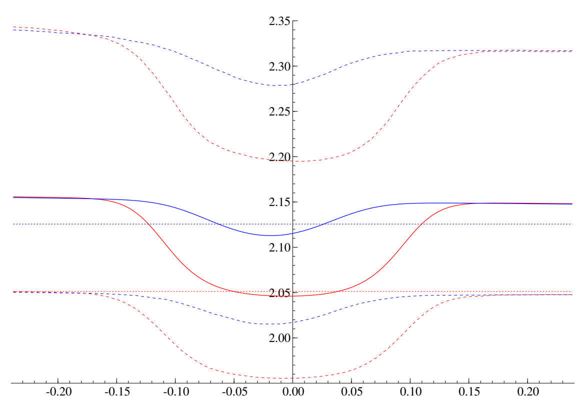
(d) Bond volatility at $t+1$



(e) Stock volatility at $t+6$



(f) Bond volatility at $t+6$



This figure shows the impulse responses for the forecast volatility of stock returns at time t, $t+1$ and $t+6$ (panels a, c and e) and bond returns at $t+1$ and $t+6$ (panels b, d and f) as a function of a shock to the D/P ratio at time t based on Proposition 8. Both the horizontal and the vertical axes shows returns in %. The solid red (blue) line shows the effect when regime 1 (2) prevails at $t-1$. The dashed lines show the bounds of the 90% confidence intervals (constructed as in Table 3). The straight dotted lines give the forecast volatility in absence of a shock. The difference between the solid and the dotted lines gives the volatility impulse response function.