TI 2021-080/III
Tinbergen Institute Discussion Paper

Moments, Shocks and Spillovers in Markov Switching VAR Models

Dick van Dijk[1]
Erik Kole[1]

[1] Erasmus University Rotterdam

Moments, Shocks and Spillovers in Markov Switching VAR Models*

Erik Kole†
Dick van Dijk
Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

September 14, 2021

Abstract

To investigate how economies, financial markets or institutions can deal with stress, we nowadays often analyze the effects of shocks conditional on a recession or a bear market. MSVAR models are ideally suited for such analyses because they combine gradual movement with sudden switches. In this paper, we develop a comprehensive framework with methods to conduct these analyses. We first derive first and second moments conditional on only a set of regime probabilities. Next, we propose generalized impulse response functions of first and second moments to shocks originating from the regime process, the structural innovations and the variables themselves. By formulating the MSVAR as an extended linear non-Gaussian VAR for the combination of the regime process and the level and squares of the observable variables, all results are in closed-form, which eases a detailed investigation. We illustrate our methods with an application to stock and bond return predictability. Our results show how regime switching combined with predictor variables influences means, volatilities and (auto-)correlations. The impulse response functions show that the effect of shocks becomes highly nonlinear, and that 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 participants in the EEA-ESEM Virtual 2021 Conference, the Annual Conference of the IAAE 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 already in a bad state. 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 observable variables $\boldsymbol{y}_t$ in an MSVAR model 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 $\mathrm{E}[\boldsymbol{y}_{t+h}|S_{t-1}]$, $h \geq 0$. Whereas the conditional expectations $\mathrm{E}[\boldsymbol{y}_{t+h}|S_{t-1}, \boldsymbol{y}_{t-1}]$ are relatively straightforward to derive, the conditional expectations that do not condition on $\boldsymbol{y}_{t-1}$, are more complicated, as the substitution of the VAR recursion requires expectations of the form $\mathrm{E}[\boldsymbol{y}_{t-k}|S_{t-1}]$, $k > 1$. We derive these expectations, 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 comprehensive and complete analysis of moments, shocks and spillovers in MSVAR models on a VAR model whose parameters switch according to a latent homogeneous first order Markov chain with a fixed number of regimes, as in Bianchi (2016). The central part of our methods is formed by the joint specification of the processes followed by the level of the 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.

As our first theoretical result 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 $\mathrm{E}[\boldsymbol{y}_{t+h}|S_{t-1}]$ and (co)variances $\mathrm{Var}[\boldsymbol{y}_{t+h}|S_{t-1}]$ for $h \geq 0$, but also autocovariances $\mathrm{Cov}[\boldsymbol{y}_{t+h}, \boldsymbol{y}_{t+k}|S_{t-1}]$ for $h, k \geq 0$. Here we extend Timmermann (2000), who considers univariate processes with a slightly different specification than the processes we consider. We also extend Bianchi (2016) who derives moments conditional on a specific regime prevailing forever, $\mathrm{E}[\boldsymbol{y}_{t+h}|S_{t-1} = S_{t-2} = \ldots]$, which essentially ignores any current or past regime switching, though he does allow for future regime switching.

Second, we present a complete framework with analytical expressions for impulse response analysis (IRA) for both the level and the squared process. We allow for different sources of shocks in line with the generalized impulse response analysis of Koop et al. (1996), that is, in the structural innovations, the regime process, and the observable variables. Because the model has Markov switching features, the effect of shocks become time and size dependent. The analytical expressions for the impulse response functions that we find, make it straightforward to analyze the responses to shocks, and how they are affected by 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 forecast (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 forecast (co)variances are not affected by shocks at all, and in standard Markov Switching models, 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 show how our framework for impulse response analysis can be used to construct a Generalized Forecast Error Variance Decomposition (GFEVD) for different horizons as in Lütkepohl (2005); Pesaran and Shin (1998). Because the effect 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.

In the empirical part of our paper, we use our theoretical results to analyze the risk-return trade-off of 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, 2008; Guidolin, 2011). Moreover, Guidolin and Hyde (2012a,b, 2014) show that simple VAR-models with different number of predictors and lags cannot beat simple Markov switching models. Our model with 2 regimes and a VAR(1) component is supported by Guidolin and Ono (2006) who shows 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 values, which are high for stocks but low for bonds in the low volatility regime, and reversed in the high volatility regime. We also show that the implied means that results under the assumption that a particular regime has prevailed forever differ considerably and lead to misleading implications. 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 volatilities of stock and bond returns at different horizons and depend on the regime distribution at $t-1$. Our results for the expected returns show three channels via which shocks propagate in this system: a direct channel that follows from the VAR-part of the model, an indirect channel via the contemporaneous correlation of the variables with propagation via the VAR-part, and a channel that follows from the updating of the forecast regime probabilities based on the shock with propagation via the Markov chain of the model. Because the parameters that govern the first two channels depend on the regime, and the updating in the third channel is nonlinear, the total effect is nonlinear. Moreover, the channels can reinforce or counteract each other. We show how these aspects depend on the model parameters, the regime distribution and the size of the shock.

Our analysis of the impact of shocks on the volatilities show that here we can discern four channels via which the shocks can propagate. First, a shock in one variable at time t lowers the forecast variance for other variables at t. Second, the updating of the probability for the different regimes may change the likelihood of the low and high volatility regimes. However, part of the variance stems from the possibility of a regime switch itself, which can also change. Finally, because the propagation of the shock differs per regime, the path that is implied by the shock is different for each regime, and this further contributes to the variance. This last channel depends positively on the size of the shock, and can be compared to the effect that shocks have in GARCH-models. Also here the effects are highly nonlinear, and the channels can reinforce or counteract each other. These effects on the variances stand in stark contrast with standard VAR models where these effects are completely absent, or with standard Markov switching models, where only the first three channels are present, and effects die out quickly.

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 forecast distribution of the variables for any point in the future, taking a specific current regime distribution as starting point. 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 theoretical results are closely related to Krolzig (2006), who was among the first to analyze MSVAR models, and Bianchi (2016) who also analyzes moments of MSVAR models. Krolzig (2006) focuses on expectations conditional on both past observations and the regime distribution. Bianchi (2016) extends the focus to first and second moments that are conditional on past observations and the regime distribution. We contribute to this set of results by deriving the first and second moments that only take the regime distribution at a particular point in time as given. Timmermann (2000) derives similar results for univariate processes. Though it is appealing and easy to calculate these moments with the assumption that a particular regime has prevailed infinitely long and will prevail in the future, this approach ignores the very regime switching that the model is designed to capture, and, as we show, can lead to quite different implications.

Our results for the impulse response analysis also extend Krolzig (2006) and Bianchi (2016). Krolzig (2006) investigates the responses of expected values to shocks to the regime process, or the structural innovations for a given regime path. Bianchi (2016) conducts an impulse response analysis for first and second moments under the assumption that a regime has prevailed infinitely long, and is known at the time of the shock. Our approach is more general and shows 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. We also extend Karamé (2010, 2012, 2015) who includes the effect of regime switching by simulations, as we show how the IRA can be done completely in closed form.

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. This finding shows the relevance and the added value of MSVAR models for analyses of financial markets.

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. Section 5 concludes.

2 MSVAR models and their moments

We consider a vector of n variables $\boldsymbol{y}_t$ that follow a VAR model of order 1, whose parameters are subject to Markov 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 $n \times n$ lower-triangular 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 $\boldsymbol{\varepsilon}_t$ is given by $\mathrm{Var}[\boldsymbol{\varepsilon}_t|S_t] = \boldsymbol{\Sigma}_{S_t} = \boldsymbol{\Lambda}_{S_t}\boldsymbol{\Lambda}'_{S_t}$. We assume that the innovations are independent over time, $\mathrm{E}[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+l}] = \mathbf{O}$, for all $l \neq 0$.

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 more convenient for the results in this paper. We assume that the Markov chain is irreducible and ergodic. We assume that the processes S_t and $\boldsymbol{\varepsilon}_{t+l}$ are independent for all l. We use ξ_{it} to denote a generic probability for state $i = 1, \ldots, m$ to occur at time t, and collect these in the vector $\boldsymbol{\xi}$. $\bar{\xi}_i = \Pr[S_t = i]$ denotes the ergodic probability for state i. They satisfy $\boldsymbol{P}\bar{\boldsymbol{\xi}} = \bar{\boldsymbol{\xi}}$. 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 eq. (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 eq. (1) to different independent Markov processes as in Hubrich and Tetlow (2015).

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 Markov chain of order m. The model in eq. (1) is hence an MSIAH(m)-VAR(1) model. Bianchi (2016) also studies the MSIAH(m)-VAR(1) model, and his approach to deriving conditional and unconditional 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 eq. (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 state space 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 linear VAR with non-Gaussian innovations. We derive explicit expressions for the VAR part as well as the innovations. Whereas result 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 The state space formulation for the level process

We rewrite the model given by eqs. (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 (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

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

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 $\boldsymbol{s}_t$. As inTimmermann (2000) and Bianchi (2016), we define the bdiag operator, which produces for a series of $n \times l$ matrices $\boldsymbol{A}_i$, $i = 1, \ldots, m$, the block-diagonal $mn \times ml$-matrix

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

Then we can prove the following proposition.

Proposition 1. *Let* $\boldsymbol{y}_t$ *follow the MS-VAR process as specified in eqs.* (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*

$$\boldsymbol{y}_t^* = \boldsymbol{C}\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\varepsilon}_t^*, \tag{5}$$

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 eq.* (3). $\tilde{\boldsymbol{y}}_t$ *follows the process*

$$\tilde{\boldsymbol{y}}_t = \begin{pmatrix} \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t, \tag{6}$$

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.*

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 with non-Gaussian innovations. Even though the innovations are non-Gaussian, the representation is quite useful in analysing the properties of MSVAR models, because it is Markovian. For example, we can write $\tilde{\boldsymbol{y}}_{t+h}$ as the sum of $\tilde{\boldsymbol{y}}_t$ and the innovations between t and $t+h$,

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

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,

$$\boldsymbol{y}_t = \tilde{\boldsymbol{G}}_{\boldsymbol{y}}\tilde{\boldsymbol{y}}_t, \tag{7}$$

where $\tilde{\boldsymbol{G}}_{\boldsymbol{y}} = \left(\boldsymbol{G}_{\boldsymbol{y}},\ \mathbf{O}_{n \times m}\right)$ has dimensions $n \times (n+1)m$ and $\boldsymbol{G}_{\boldsymbol{y}} = \boldsymbol{\imath}_m' \otimes \boldsymbol{I}_n$. 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

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

with $\tilde{\boldsymbol{G}}_{\boldsymbol{s}} = \left(\mathbf{O}_{m \times nm},\ \boldsymbol{I}_m\right)$. We also define $\tilde{\boldsymbol{G}}_{y_i} = \boldsymbol{e}_i'\tilde{\boldsymbol{G}}_{\boldsymbol{y}}$, $\boldsymbol{G}_{y_i} = \boldsymbol{e}_i'\boldsymbol{G}_{\boldsymbol{y}}$, and $\tilde{\boldsymbol{G}}_{s_j} = \boldsymbol{e}_j'\tilde{\boldsymbol{G}}_{\boldsymbol{s}}$ 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 eq. (6), 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

$$\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix}. \tag{9}$$

Using eq. (7) the conditional expectation of $\boldsymbol{y}_{t+h}$ follows as $\mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{G}}_{\boldsymbol{y}} \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t]$. This result corresponds with eq. (5) in Bianchi (2016).[1] We can use the relation $\boldsymbol{y}_t = \boldsymbol{G}_{\boldsymbol{y}} \boldsymbol{y}_t^*$ to calculate the first moments from the recursion

$$\mathrm{E}[\boldsymbol{y}^*_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n) \mathrm{E}[\boldsymbol{y}^*_{t+h-1}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] + \boldsymbol{C} \mathrm{E}[\boldsymbol{s}_{t+h}|\boldsymbol{\xi}_t] \tag{10}$$

with initial condition $\mathrm{E}[y^*_t|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t$, 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 $\tilde{\boldsymbol{\Phi}}$.

2.2 The state space formulation for the level and the 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 $\tilde{\boldsymbol{z}}_t = \left(\boldsymbol{z}^{*\prime}_t, \boldsymbol{y}^{*\prime}_t, \boldsymbol{s}'_t\right)'$ 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 eqs.* (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 $\tilde{\boldsymbol{z}}_t = \left(\boldsymbol{z}^{*\prime}_t, \boldsymbol{y}^{*\prime}_t, \boldsymbol{s}'_t\right)'$, with $\boldsymbol{y}^*_t$ as defined in proposition 1. Then $\boldsymbol{z}_t$ follows the process*

$$\boldsymbol{z}_t = \boldsymbol{\gamma}_{S_t} + \boldsymbol{\omega}_{S_t} + \boldsymbol{\Psi}_{S_t} \boldsymbol{y}_{t-1} + \boldsymbol{\Upsilon}_{S_t} \boldsymbol{z}_{t-1} + \boldsymbol{\zeta}_t, \tag{11}$$

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

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

[1]Bianchi (2016) does not define the VAR of eq. (6) but starts from its expectation.

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

$$\boldsymbol{z}_t^* = (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\boldsymbol{z}_{t-1}^* + \boldsymbol{\zeta}_t^*, \tag{12}$$

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

$$\tilde{\boldsymbol{z}}_t = \begin{pmatrix} \boldsymbol{z}_t^* \\ \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Upsilon}}\tilde{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{\zeta}}_t, \tag{13}$$

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. Morover, $\boldsymbol{\zeta}_t$, $\boldsymbol{\zeta}_t^*$ and $\tilde{\boldsymbol{\zeta}}_t$ are MDS.

This proposition shows that $\tilde{\boldsymbol{z}}_t$, so the combination of the state variable, measurement variables, and their squares, also follows a first-order linear VAR 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,

$$\boldsymbol{a}_t = \tilde{\boldsymbol{H}}_{\boldsymbol{a}}\tilde{\boldsymbol{z}}_t, \qquad \text{for } \boldsymbol{a} = \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{s}, \tag{14}$$

where $\tilde{\boldsymbol{H}}_{\boldsymbol{z}} = \left(\boldsymbol{H}_{\boldsymbol{z}}, \mathbf{O}_{n^2 \times m(n+1)}\right)$ with $\boldsymbol{H}_{\boldsymbol{z}} = \boldsymbol{\iota}_m^\prime \otimes \boldsymbol{I}_{n^2}$, $\tilde{\boldsymbol{H}}_{\boldsymbol{y}} = (\mathbf{O}_{n \times mn^2}, \tilde{\boldsymbol{G}}_{\boldsymbol{y}}) = (\mathbf{O}_{n \times mn^2}, \boldsymbol{G}_{\boldsymbol{y}}, \mathbf{O}_{n \times m})$, and $\tilde{\boldsymbol{H}}_{\boldsymbol{s}} = (\mathbf{O}_{m \times mn^2}, \tilde{\boldsymbol{G}}_{\boldsymbol{s}}) = (\mathbf{O}_{m \times mn(n+1)}, \boldsymbol{I}_m)$. 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 $h > 0$ conditional on $\boldsymbol{y}_t$ and state probabilities $\boldsymbol{\xi}_t$ follows directly as

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h} | \boldsymbol{y}_t, \boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix}. \tag{15}$$

We can 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

$$\begin{aligned}
\operatorname{vec}\left(\operatorname{Var}[\boldsymbol{y}_{t+h}]\right) &= \mathrm{E}[\boldsymbol{y}_{t+h} \otimes \boldsymbol{y}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] - \mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] \otimes \mathrm{E}[\boldsymbol{y}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] \\
&= \tilde{\boldsymbol{H}}_{\boldsymbol{z}} E[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] - \tilde{\boldsymbol{H}}_{\boldsymbol{y}} E[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] \otimes \tilde{\boldsymbol{H}}_{\boldsymbol{y}} E[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t]. \qquad (16)
\end{aligned}$$

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

$$\begin{aligned}
\mathrm{E}[\boldsymbol{z}^*_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] =& \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\, \mathrm{E}[\boldsymbol{z}^*_{t-h-1}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] + \\
& \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\, \mathrm{E}[\boldsymbol{y}^*_{t-h-1}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] + (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\, \mathrm{E}[\boldsymbol{s}_{t+h}|\boldsymbol{\xi}_t], \qquad (17)
\end{aligned}$$

in combination with eq. (10), and initial conditions $\mathrm{E}[\boldsymbol{z}^*_{t-h-1}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \otimes \boldsymbol{y}_t$, $\mathrm{E}[y^*_t|\boldsymbol{y}_t, \boldsymbol{\xi}_t] = \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t$.

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 $\boldsymbol{z}_{t+k,t} = \operatorname{vec}(\boldsymbol{y}_t \boldsymbol{y}_{t+1})$, it gives rise to $\operatorname{Cov}(\boldsymbol{y}_t, \boldsymbol{y}_{t+1})$. 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 eqs.* (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*

$$\tilde{\boldsymbol{z}}_{t+k,t} = \begin{pmatrix} \tilde{\boldsymbol{\Phi}} \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P} \end{pmatrix} \tilde{\boldsymbol{z}}_{t+k-1,t} + \begin{pmatrix} \tilde{\boldsymbol{\varepsilon}}_{t+k} \otimes \boldsymbol{y}_t \\ \boldsymbol{u}_{t+k} \end{pmatrix}, \quad k \geq 1 \qquad (18)$$

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 with non-Gaussian innovations. When $k = 0$, $\boldsymbol{y}^*_{t,t} = \boldsymbol{y}^*_t$, $\boldsymbol{z}_{t,t} = \boldsymbol{z}_t$ and 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

$$\begin{aligned}
\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h+k,t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] &= \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{y}_t, \boldsymbol{\xi}_t] \\
&= \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \otimes \boldsymbol{y}_t \\ \boldsymbol{\xi}_t \end{pmatrix}. \qquad (19)
\end{aligned}$$

We can use the recursion in eq. (18) and the structure of $\tilde{\boldsymbol{\Phi}}$ to write the conditional expectation of $\boldsymbol{z}^*_{t+h+k,t+h}$ recursively as

$$\mathrm{E}[\boldsymbol{z}^*_{t+h+k,t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] = (\boldsymbol{\Phi}\otimes\boldsymbol{I}_n)(\boldsymbol{P}\otimes\boldsymbol{I}_{n^2})\,\mathrm{E}[\boldsymbol{z}^*_{t+h+k-1,t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t] + (\boldsymbol{C}\boldsymbol{P}^k\otimes\boldsymbol{I}_n)\,\mathrm{E}[\boldsymbol{y}^*_{t+h}|\boldsymbol{y}_t,\boldsymbol{\xi}_t], \tag{20}$$

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.

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 eqs. (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

$$\bar{\boldsymbol{y}} = \lim_{t\to\infty}\mathrm{E}[\boldsymbol{y}_t|\boldsymbol{y}_0,\boldsymbol{\xi}_0] = \boldsymbol{G}_{\boldsymbol{y}}\bar{\boldsymbol{y}}^*, \tag{21}$$

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

where

$$\bar{\boldsymbol{y}}^* = \lim_{t\to\infty}\mathrm{E}[\boldsymbol{y}^*_t|\boldsymbol{y}_0,\boldsymbol{\xi}_0] = \big(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{P}\otimes\boldsymbol{I}_n)\big)^{-1}\boldsymbol{C}\bar{\boldsymbol{\xi}}. \tag{23}$$

$$\bar{\boldsymbol{z}}^* = \lim_{t\to\infty}\mathrm{E}[\boldsymbol{z}^*_t|\boldsymbol{y}_0,\boldsymbol{\xi}_0] = \big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{P}\otimes\boldsymbol{I}_{n^2})\big)^{-1}\big((\boldsymbol{\Gamma}+\boldsymbol{\Omega})\bar{\boldsymbol{\xi}} + \boldsymbol{\Psi}(\boldsymbol{P}\otimes\boldsymbol{I}_n)\bar{\boldsymbol{y}}^*\big), \tag{24}$$

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

$$\bar{\boldsymbol{z}}_k = \lim_{t\to\infty}\mathrm{E}[\boldsymbol{z}_{t+k,t}|\boldsymbol{y}_0,\boldsymbol{\xi}_0] = \boldsymbol{H}_{\boldsymbol{z}}\bar{\boldsymbol{z}}^*_k, \tag{25}$$

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 eq. (19) as

$$\begin{pmatrix}\bar{\boldsymbol{z}}^*_k\\ \bar{\boldsymbol{y}}^*_k\end{pmatrix} = \big(\tilde{\boldsymbol{\Phi}}^k\otimes\boldsymbol{I}_n\big)\begin{pmatrix}\bar{\boldsymbol{z}}^*\\ \bar{\boldsymbol{y}}^*\end{pmatrix}. \tag{26}$$

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

2.4 Regime-specific moments

We now turn to first and second moments when the regime or regime probabilities are given. These expectations correspond with situations where the current regime or regime distribution is given, but no information about the history of the regimes is available. In other words, we calculate the expectation or variance of $\boldsymbol{y}_t$ 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 need to define the "time-reversed" Markov chain.

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

$$q_{ij} = \Pr[S_{t-1} = i | S_t = j] = p_{ji}\frac{\bar{\xi}_i}{\bar{\xi}_j}, \tag{27}$$

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 lemma and propositions.

Lemma 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (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.*

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. We use this property in the proofs of the next propositions.

Proposition 4. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (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*

$$\boldsymbol{\mu} = \left(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\right)^{-1}\boldsymbol{c}, \tag{28}$$

where $\boldsymbol{c} = \left(\boldsymbol{c}_1', \dots, \boldsymbol{c}_m'\right)'$, $\boldsymbol{\Phi} = \operatorname{bdiag}_{i=1}^m(\boldsymbol{\Phi}_i)$, 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

$$\mathrm{E}[\boldsymbol{y}_t | \boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}, \tag{29}$$

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

$$\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{30}$$

with $\tilde{\boldsymbol{y}}_{t+h}$ *and* $\tilde{\boldsymbol{\Phi}}$ *defined in proposition 1 and* $\tilde{\boldsymbol{G}}_{\boldsymbol{y}}$ *as in* (7)*.*

Proposition 5. *Let* $\boldsymbol{y}_t$ *follow the MS-VAR process as specified in eqs.* (1) *and* (2)*, and assume that 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* $\boldsymbol{\mu} = \big(\boldsymbol{\mu}_1', \dots \boldsymbol{\mu}_m'\big)'$ *and* $\boldsymbol{\nu} = \big(\boldsymbol{\nu}_1', \dots \boldsymbol{\nu}_m'\big)'$*. Then*

$$\boldsymbol{\nu} = \big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}\big(\boldsymbol{\gamma} + \boldsymbol{\omega} + \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}\big), \tag{31}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \dots, \boldsymbol{\gamma}_m')'$ *with* $\boldsymbol{\gamma}_j = \boldsymbol{c}_j \otimes \boldsymbol{c}_j$*,* $\boldsymbol{\omega} = (\boldsymbol{\omega}_1', \dots, \boldsymbol{\omega}_m')'$ *with* $\boldsymbol{\omega}_j = \mathrm{vec}(\boldsymbol{\Sigma}_j)$*,* $\boldsymbol{\Upsilon} = \mathrm{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j)$*,* $\boldsymbol{\Psi} = \mathrm{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{c}_j + \boldsymbol{c}_j \otimes \boldsymbol{\Phi}_j)$*,* $\boldsymbol{Q}$ *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*

$$\mathrm{E}[\boldsymbol{z}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}. \tag{32}$$

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*

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_{n^2}\big)\boldsymbol{\nu} \\ \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{33}$$

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 $\lim_{\tau\to\infty} \mathrm{E}[\boldsymbol{y}_t|S_t = \cdots = S_{t-\tau} = j]$ and $\lim_{\tau\to\infty} \mathrm{E}[\boldsymbol{z}_t||S_t = \cdots = S_{t-\tau} = j]$, which essentially correspond with a situation without any regime switching. Our results give the expectation conditional on knowing the current state, but without any knowledge about the past, and without past observations. They can be used 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 propose a framework for such analyses in the next section.

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 eq. (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 eqs.* (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*

$$\mathrm{E}[\boldsymbol{z}_{t+h+k,t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}} \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t], \tag{34}$$

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

$$\mathrm{E}[\boldsymbol{z}^*_{t+k,t}|\boldsymbol{\xi}_t] = (\boldsymbol{\Phi} \otimes \boldsymbol{I}_n)(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\,\mathrm{E}[\boldsymbol{z}^*_{t+k-1,t}|\boldsymbol{\xi}_t] + (\boldsymbol{C}\boldsymbol{P}^k \otimes \boldsymbol{I}_n)\,\mathrm{E}[\boldsymbol{y}^*_t|\boldsymbol{\xi}_t], \tag{35}$$

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

$$\mathrm{Cov}[\boldsymbol{y}_{t+h}, \boldsymbol{y}_{t+h+k}|\boldsymbol{\xi}_t] = \mathrm{E}[\boldsymbol{y}_{t+h+k}\boldsymbol{y}'_{t+h}|\boldsymbol{\xi}_t] - \mathrm{E}[\boldsymbol{y}_{t+h+k}|\boldsymbol{\xi}_t]\,\mathrm{E}[\boldsymbol{y}'_{t+h}|\boldsymbol{\xi}_t]. \tag{36}$$

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 eqs.* (1) *and* (2) *with* $\Phi_i = \mathbf{O}$ *for* $i = 1, \ldots, m$, *and assume that it is second-order stationary. Then*

$$\operatorname{vec}\big(\operatorname{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t]\big) = (\boldsymbol{\iota}_m' \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k\big(\operatorname{diag}(\boldsymbol{\xi}_t) - \boldsymbol{\xi}_t\boldsymbol{\xi}_t'\big) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}. \tag{37}$$

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 $\operatorname{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{e}_i] = \mathbf{O}$.

3 Impulse responses 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 state space 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),

$$GI_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|I_{t-1}] = \tilde{\boldsymbol{\Phi}}^h\tilde{\boldsymbol{\varepsilon}}_t, \tag{38}$$

where h gives the horizon, and I_t denotes the information set at time t. We have specified the GI for the extended process $\tilde{\boldsymbol{y}}$, from which we can easily derive the GI for $\boldsymbol{y}$ and $\boldsymbol{s}$, using

$$GI_{\boldsymbol{a}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}) = \tilde{\boldsymbol{G}}_{\boldsymbol{a}} GI_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}), \quad \text{for } \boldsymbol{a} = \boldsymbol{y}, \boldsymbol{s}, \tag{39}$$

with $\tilde{\boldsymbol{G}}_{\boldsymbol{a}}$ as in eqs. (7) and (8).

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 innovations are specified. In an MS-VAR model, there are three types of innovations that can be jointly or separately specified. Following directly from the model specification are regime innovations $\boldsymbol{u}_t$ and innovations ε_{it}.[2] They can also be specified with respect to the observed variables, which we denote by $\eta_{it} = y_{it} - \mathrm{E}[y_{it}|I_{t-1}]$. As a consequence, we define three GI functions depending on the type of innovation.

Proposition 7. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (1) *and* (2)*, and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{y}}_t$ be defined by eq.* (38) *and the results in proposition 1. Let the vector $\boldsymbol{y}_{t-1}$ be part of I_{t-1} or calculated as $\mathrm{E}[\boldsymbol{y}_{t-1}|I_{t-1}]$. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of I_{t-1} or calculated as $\mathrm{E}[\boldsymbol{s}_{t-1}|I_{t-1}]$. 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*

$$GI^{\boldsymbol{u}}_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{u}_t, I_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \emptyset, \boldsymbol{u}_t, I_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \tag{40}$$

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

$$GI^{\varepsilon_i}_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, I_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, \emptyset, I_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}. \tag{41}$$

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

$$GI^{y_i}_{\tilde{\boldsymbol{y}}}(h, \eta_{it}, I_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|y_{it}, I_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|I_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\varepsilon}^*_t|y_{it}, I_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|y_{it}, I_{t-1}] \end{pmatrix}. \tag{42}$$

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

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

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

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

[2] 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

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

and

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

The impulse response 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. $\boldsymbol{s}_t = \boldsymbol{e}_j$, with $\boldsymbol{e}_j$ the standard basis vector for dimension j, we can substitute $\boldsymbol{u}_t = \boldsymbol{e}_j - \boldsymbol{P}\boldsymbol{\xi}_{t-1}$.

The impulse response function $GI_{\tilde{\boldsymbol{y}}}^{\varepsilon_i}$ is in line with the traditional framework of Sims (1980). However, shocks in ε_{it} are not very interesting because they translate to different shocks in the different regimes depending on the regime-specific matrices $\boldsymbol{\Lambda}_{S_t}$. Only when the regime at t is known with certainty, one can interpret and compare the resulting GIRF for shocks of ε_i standard deviations. Bianchi (2016) uses this particular setting for his impulse response analysis of a system of macroeconomic variables.

The impulse response function $GI_{\tilde{\boldsymbol{y}}}^{y_i}$ shows that shocks in a particular variable y_i generate a contemporaneous response in two ways. The term $\mathrm{E}[\boldsymbol{u}_t|y_{it}, I_{t-1}]$ captures the effect of the shock on the inference of the regime process. Though $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{u}_t$ are independent, $\boldsymbol{y}_t$ and $\boldsymbol{u}_t$ are not, and hence $\mathrm{E}\big[\boldsymbol{u}_t|y_{it}, I_{t-1}\big]$ is not necessarily zero. The effect on the inferred regime has a direct effect on the expectation of the regime specific innovations $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, I_{t-1}]$. 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 $\mathrm{E}[\boldsymbol{s}_t \otimes \boldsymbol{\varepsilon}_t|y_{it}, I_{t-1}]$. The proposition shows how to derive the effects for the different regimes.

Our set of GI functions complement the impulse response analyses (IRA) proposed by other authors. The differences with our IRA relate to the information about the regimes at the time of the shock, and the nature of the shock. In our IRA, the shock occurs at time t, can have three sources, and information about the regimes pertains to $t-1$. If the shock occurs in the structural innovation ε_{it} or the observable y_{it}, the regime in which the shock occurs is hence not known. To the contrast, 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. Krolzig (2006) also 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. Karamé (2015) specifies shocks similarly to Krolzig (2006), but does not assume that the regime at the time of the shock is known. The first order IRA 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, i.e., no regime switches have occured until the time of the shock. We argue that our setting is a realistic one, because (i) shocks can signal a regime switch, and (ii) we do not require further assumptions on the regime process, only its distribution at $t-1$.

Our approach is closest to Karamé (2015), but we consider both shocks to the structural innovations, and to one observable variable y_{it} without the assumption that the other variables are not shocked. Another important difference is that we can express the GI in closed form due to the extended VAR specification in proposition 1, whereas he reverts to simulations. The GIs in Krolzig (2006) and Bianchi (2016) are also in closed form.

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 a the setting of a linear VAR. 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 analyses of these effects.

The VAR-respresentation for the extended squared process in proposition 2 gives rise to

$$GI_{\tilde{\boldsymbol{z}}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|I_{t-1}] = \tilde{\boldsymbol{\Upsilon}}^h \tilde{\boldsymbol{\zeta}}, \tag{43}$$

which extends eq. (38). 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

$$GI_{\boldsymbol{a}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}) = \tilde{\boldsymbol{H}}_{\boldsymbol{a}} GI_{\tilde{\boldsymbol{z}}}(h, \tilde{\boldsymbol{\zeta}}_t, I_{t-1}), \quad \text{for } \boldsymbol{a} = \boldsymbol{z}, \boldsymbol{y}, \boldsymbol{s}, \tag{44}$$

as in eq. (14). 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 eqs.* (1) *and* (2)*, and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{z}}_t$ be defined by eq.* (43) *and the results in proposition 2. Let the vector $\boldsymbol{y}_{t-1}$ be part of I_{t-1} or calculated as* $\mathrm{E}[\boldsymbol{y}_{t-1}|I_{t-1}]$*. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of I_{t-1} or calculated as* $\mathrm{E}[\boldsymbol{s}_{t-1}|I_{t-1}]$*. 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

$$GI_{\tilde{\boldsymbol{z}}}^{\boldsymbol{u}}(h, \boldsymbol{u}_t, I_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \emptyset, \boldsymbol{u}_t, I_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t \otimes \boldsymbol{z}_{t-1}) \\ \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \tag{45}$$

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

$$GI_{\tilde{\boldsymbol{z}}}^{\varepsilon_i}(h, \varepsilon_{it}, I_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \varepsilon_{it}, \emptyset, I_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\zeta}_t^* | \varepsilon_{it}, I_{t-1}] \\ \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}, \tag{46}$$

with

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

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

$$GI_{\tilde{\boldsymbol{z}}}^{y_i}(h, \eta_{it}, I_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h} | y_{it}, I_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h} | I_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{z}_t^* | y_{it}, I_{t-1}] - \mathrm{E}[\boldsymbol{z}_t^* | I_{t-1}] \\ \mathrm{E}[\boldsymbol{\varepsilon}_t^* | y_{it}, I_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t | y_{it}, I_{t-1}] \end{pmatrix}, \tag{47}$$

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

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

and

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

with

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

$$\operatorname{Var}[\boldsymbol{y}_t | y_{it}, S_t = j, I_{t-1}] = \boldsymbol{\Sigma}_j - \frac{1}{\boldsymbol{e}_i' \boldsymbol{\Sigma}_j \boldsymbol{e}_i} \boldsymbol{\Sigma}_j \boldsymbol{e}_i \boldsymbol{e}_i' \boldsymbol{\Sigma}_j.$$

Whereas many authors have studied the analysis of first order impulse responses, Bianchi (2016) is the first to address second order IRA. Similar to his first order IRA, 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

$$VI_{\boldsymbol{y}}(h, \boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}) = \mathrm{Var}[\boldsymbol{y}_{t+h}|\boldsymbol{\varepsilon}_t, \boldsymbol{u}_t, I_{t-1}] - \mathrm{Var}[\boldsymbol{y}_{t+h}|I_{t-1}] \tag{48}$$

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

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

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 eqs. (40) to (42) vary over time because of their dependence on I_{t-1}. When shocks are defined directly with respect to $\boldsymbol{y}_t$ as in eq. (42), their effect on this expectation is nonlinear and asymmetric. Small or positive shocks typically lead to a different update on the prevailing regime than big 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

$$v_{it} = \mathrm{var}[y_{it}|I_{t-1}] = \mathrm{E}[y_{it}^2|I_{t-1}] - \mathrm{E}[y_{it}|I_{t-1}]^2, \tag{49}$$

where $\mathrm{E}[y_{it}^2|I_{t-1}]$ and $\mathrm{E}[y_{it}|I_{t-1}]$ follow from eq. (15). This leads to the standardized generalized impulse response function

$$\boldsymbol{\psi}_{\boldsymbol{a}}^{y_i}(h,\delta,I_{t-1}) = \tilde{\boldsymbol{G}}_{\boldsymbol{a}}\, GI_{\tilde{\boldsymbol{y}}}^{y_i}(h,\eta_{it}=\delta\sqrt{v_{it}},I_{t-1}), \quad \boldsymbol{a}=\boldsymbol{y},\boldsymbol{s} \tag{50}$$

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 I_t and innovation size δ standard deviations as

$$\theta_{a_j}^{y_i}(h,\delta,I_{t-1}) = \frac{\sum_{l=0}^{h}(\boldsymbol{e}_j'\boldsymbol{\psi}_{\boldsymbol{a}}^{y_i}(l,\delta,I_{t-1}))^2}{\sum_{l=0}^{h}\sum_{k=1}^{n}(\boldsymbol{e}_j'\boldsymbol{\psi}_{\boldsymbol{a}}^{y_k}(l,\delta,I_{t-1}))^2}, \quad \boldsymbol{a}=\boldsymbol{y},\boldsymbol{s}. \tag{51}$$

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

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.[3] MSVAR models accommodate both features. In their analysis of these models for US stock and bond returns, Guidolin and Ono (2006) show than an MSIH(4)-VAR(1) model works best, and yields better predictions than simpler VAR or MSIH models.

Based on these findings in the literature, we analyze the risk and returns of US stocks and bond returns 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. We compare its implications to those of restricted models to determine the impact of predictability by predictor variables

[3]Notably Ang and Bekaert (2002), Guidolin and Timmermann (2006a,b, 2007, 2008). See Guidolin (2011) for a survey.

and regime switching. To be precise, the comparison with a simple Markov switching model for stocks and bonds with two regimes (an MSIH(2) model) shows the impact of predictability, whereas the comparison with a simple VAR(1) for the returns and predictor variables shows the effect of regime switching. We keep the model specifications simple, because we do not aim for the best fitting model.

4.1 Data and estimation results

We base our analysis on monthly observations taken from CRSP. As stock returns we take value-weighted 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 the past 12 months minus the log of the level of the index at the end of month t. As bonds returns we take the monthly return series from the CRSP Fixed-Term Index of 10-year Treasury bonds. As the short rate, we take the yield on 1-month Treasury Bills as supplied by the CRSP Risk-Free Rates Files. The data set runs from January 1952 to December 2018 (804 observations).

[Table 1 about here.]

The summary statistics in table 1 show that both stocks and bonds yield on average positive excess returns, with equity volatility being twice as high as bond volatility. All variables exhibit non-zero skewness, indicating deviations from normality that a standard linear VAR 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 lead and lag than contemporaneous correlations between stocks and bonds. 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.

We report the parameters estimates of the MSIH(2) Model in table 2. The estimation uses the Expectation-Maximation algorithm of Dempster et al. (1977), see also Hamilton (1990). 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 2 about here.]

The estimation results for the VAR(1) model are in table 3. 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. The combination of their predictive effect with their strong persistence leads to the long-term effects on means and (co)variances presented by Campbell and Viceira (2005). Correlations are generally low, except between stock returns and D/P ratio, which is by construction.

[Table 3 about here.]

[Table 4 about here.]

The parameter estimates for the MSIAH(2)-VAR(1) model in table 4 show that the features of the two restricted models remain present. As for the MSIH(2) model, we see low and high volatility regimes that are persistent. Conform the results for the VAR(1) model, we observe predictability between stock and bond returns, and from the T-bill rate and D/P ratio to stock returns. However, predictability varies over 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 vol regime, but not in the high vol 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 are very strong in the low volatility regime. 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.

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 5, and turn to the implications for the autocovariance structure in table 6. We also calculate these moments under the much stronger assumption that the process has been in the current 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 both assumptions. For completeness we also include the unconditional moments.

Table 5 shows that the effect of regime switching is mostly confined to stock and bonds, even though the parameter estimates in table 4 for the T-Bill rate and D/P ratio vary considerably over the regimes. As for the MSIH(2) model, the MSIAH(2)-VAR(1) model yields a regime with a high mean for stocks and a low mean for bonds and low volatilities for both, and one with the opposite pattern. Whereas the differences between the regime-specific volatilities implied by both models are minor, the differences between the means for stocks are more substantial, at 1.00% vs. 0.75% in regime 1, and -0.48% vs. -0.67% in regime 2. Consequently, the unconditional mean is also substantially lower. Of course, the differences in these moments may be related to different identification of bull and bear markets, as the MSIAH(2)-VAR(1) model has a larger dimension.[4] The inclusion of the predictor variables has a negative effect on the steady state mean of stocks, as well as its regime-specific means. The correlations also vary over the different regimes. Stocks and bonds are more strongly correlated in regime 2, indicating less diversification benefits. Correlations with the predictor variables are generally a bit lower. As an exception, the correlation between the T-Bill rate and D/P ratio is large at 0.455 and shows no switching. The unconditional correlations implied by the different models are comparable, and generally lie between the regime-specific correlations.

[Table 5 about here.]

To illustrate the differences between moments conditional on the current regime and moments conditional on the current regime having prevailed infinitely long, we also report these latter moments with the label "MSIAH(2)-VAR(1) inf." 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 are 0.22% 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

[4]The correlation between the smoothed probabilities for the bull regimes in the two models is equal to 0.64, indicating that the models do not identify the regimes identically.

larger. Though the moments are easy to calculate under the "infinitely long" assumption, the assumption is not realistic and consequently, the resulting moments do not reflect a particular aspect of the data. That also means that the different values for the means we calculate under this assumption do not have a clear meaning or interpretation.

The first order autocorrelation matrices in table 6 show quite some differences between the regimes. Stocks exhibit negative autocorrelation in regime 1, but positive autocorrelation in regime 2. 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, these increases make regime 2 more risky than regime 1, on top of its higher volatility of the assets. The correlation between stock returns at $t + 1$ and the D/P ratio at t increases from 0.022 in regime 1 to 0.102 in regime 2, indicating that predictability of stock returns in regime 2 is considerably stronger. The autocorrelations of the T-bill rate and the D/P ratio are a bit lower in regime 2.

[Table 6 about here.]

Comparing the regime-conditional results in panel (a) with the unconditional ones in panel (c) shows the impact of combining a VAR structure with Markov switching. Incorporating only Markov switching as in the MSIH(2) model leads to some unconditional autocorrelation, but the coefficients are much lower than those resulting from the VAR(1) and MSIAH(2)-VAR(1) models, and more importantly, they indicate weaker autocorrelation than what we find in the data in table 1. Also, as we show in corollary 1, the regime-conditional correlations of any plain MSIH model are zero. The unconditional autocorrelations of the VAR(1) and MSIAH(2)-VAR(1) models are comparable, and correspond well with the empirical autocorrelations. They lie again between the results for the two regimes in panel (a). But since the VAR(1) model does not exhibit different regimes, it means that it would miss the differences implied by the different regimes.

Panel (b) of table 6 reports the autocorrelation matrices calculated under the assumption that a regime prevails forever. They can be calculated by considering that specific regime as a VAR on its own. However, ignoring regime switches leads again to quite different results. The autocorrelations between stocks and bonds become more extreme. For example, if regime 1 (2) would prevail forever, the autocorrelation of stocks is -0.101 (0.158) compared to -0.060 (0.112) when regimes can switch. The autocorrelations between the state variables at t and stock and bond returns at $t + 1$ in the lower-left part of the matrices tend to be lower in panel (b), indication that the implied predictive effect of the state variables would be lower when regime switching is ignored. Similar as in table 5, the differences in the behavior of the state variables in the different regimes seem larger when regimes can prevail 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 differences with the implications from the simpler MSIH(2) or VAR(1) models are substantial for some key indicators. In particular the combination of Markov Switching with a VAR model indicates that the bearish regime is riskier than indicated by the simpler models. 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 one particular regime has prevailed forever gives quite different moments than the assumption that a regime prevailed only at t.

4.3 Impulse Response Analysis

We continue our analysis by investigating how shocks in the different variables affect the risk-return trade-off at different horizons. As we showed in the derivation of the GIRFs, shocks in Markov Switching models have a nonlinear effect where both the sign and the size of the shock matter, contrary to the VAR(1) model where the effect of shocks is linear in the size of the shock. Though the MSIAH(2)-VAR(1) model comprises four variables, we concentrate on the impulse responses of stocks and bonds, because these are the variables that investors are interested in. We consider the T-Bill rate and the D/P ratio as sources of shocks, but are less interested in these state variables themselves. To disentangle the effects of the Markov-Switching and the VAR components in the MSIAH(2)-VAR(1) model, we also conduct an analysis based on the simpler MSIH(2) model.

4.3.1 First Order Impulse Responses

Our theoretical results show that we can distinguish three channels through which shocks affect the levels. The first one is the direct channel that follows from the VAR-part of the model. A shock to one variable y_{it} influences all variables from $t+1$ onwards because of the VAR nature of the model. Next, we distinguish the indirect channel that shows up in the generalized framework of Koop et al. (1996) because of the contemporaneous correlation of the variables. A shock to the variable y_{it} affects the expectation of all variables in $\boldsymbol{y}_t$, which then also propagates because of the VAR structure. The third channel runs via the updating of the regime forecasts. A shock to the variable y_{it} also affects the regime distribution from t onwards because of the Markov switching part of the model. How much the different channels contribute depends on the model parameters, the size of the shock and the regime distribution at $t-1$. For our analysis we assume that the regime process at $t-1$ is in regime 1, 2 or in the steady state distribution, and we use proposition 7 to determine the Generalized Impulse Response Functions for $t+h$ with $h = 0, 1, \ldots, 12$. In the discussion below we focus on $h = 0, 1, 6$. We specify the shocks as δ standard deviations, which also reflect the regime distribution at $t-1$ as in eq. (49).

We start with the third channel, which we can identify by the GIRF for the regime process. Figure 1 shows how shocks in the different variables affect the forecast probability of regime 1 for different horizons and regime distributions at $t-1$. The regime process is in regime 1 in the top row, regime 2 in the middle row and the steady state in the bottom row. The straight dashed lines in the subfigures in the first three columns give the forecast probability for $t+h$ without the shock, whereas the solid lines give the effect of shocks. The difference between the solid and dashed lines gives the $GI_{\boldsymbol{s}}^{y_i}(h, \eta_{it}, I_{t-1})$ defined in section 3.1. For all horizons and regime distributions, small shocks are likely to signal that regime 1 prevails, whereas large shocks indicate regime 2. The curves for shocks to stock and bond returns and the D/P ratio resemble each other, because the differences between the volatilities of their innovations in the different regimes in table 4 are approximately equal to a factor 2. The location of their peaks is different, and reflect the differences between the mean forecasts. The curve for the T-Bill rate is more peaked and concentrated, because the difference between the regime volatilities is almost a factor 4. For each variable, there are two shock sizes that yield no update on the regime probabilities, i.e. $GI_{\boldsymbol{s}}^{y_i}(h, \eta_{it}, I_{t-1}) = \mathbf{0}$. And since $GI_{\boldsymbol{s}}^{y_i}(h, \eta_{it}, I_{t-1}) = \boldsymbol{P}^h \, \mathrm{E}[\boldsymbol{u}_t | y_{it}, I_{t-1}]$, these shock sizes are the same for each h.

[Figure 1 about here.]

Comparing the subfigures in the different rows shows that the effect of large shocks is larger when the low volatility regime prevails at $t-1$, and the effect of small shocks is larger when the high volatility regime prevails. This result is of course driven by how likely a shock signals a regime switch. Comparing the subfigures in the different columns shows that the shocks die out relatively quickly with only small effects remaining after six months.

The subfigures in the rightmost column show the decomposition of the forecast error variance. These proportions do not depend on the horizon h, because the numerator and denominator in eq. (51) change proportionally for different values of h. Because there are four different sources of shocks, that are equally informative when the shocks become infinitely large in magnitude, we observe an asymptote at 1/4 for all sources. The proportion for each variable converges faster to the asymptote when the high volatility regime prevails at $t-1$, because the forecast probability for the low volatility converges faster to zero. Because the curve for the T-Bill rate is more peaked in the impulse response columns, the proportion of the forecast error variance that can be attributed to it is generally largest, except for the small ranges of shock sizes where the T-Bill rate leads to only small updates in the forecast probability. Negative shocks are more informative stemming from stocks, and positive shocks from bonds.

Figure 2 shows the effect of shocks on stock returns. The leftmost subfigures show the combined effect of channels two and three. Expected stocks returns are affected by shocks to the other variables because they are correlated, and because shocks lead to an update of the regime forecast. When shocks originate from bond returns or the T-bill rate, their effect is nonlinear. Contemporaneously, shocks to the D/P ratio have an almost linear effect because the correlation with shocks to stock returns is large and about the same in both regimes.

[Figure 2 about here.]

The effect of shocks also depends on the sign of the shock. The blue line in fig. 2a shows that a large negative shock to bonds at t when the low volatility regime prevails at $t-1$ has a relatively large negative effect on stocks, because the positive correlation between stocks and bonds, and the inferred switch to the high volatility regime amplify each other. For positive shocks, the effects work in opposite direction and mitigate each other. Because the correlation between shocks to the T-bill rate and stock returns is negative, the green line reflects that positive shocks to the T-bill rate have a larger effect on expected stock returns than negative ones. Both lines in fig. 2d are closer to linear because the effect of switching is smaller in the high volatility regime. As can be expected, the results for the steady state at $t-1$ in the bottom row are in between the results for regime 1 and 2.

The more substantial deviations for linearity show up for the expected returns at horizons of 1 and 6 months. Shocks have an effect at longer horizons because of the VAR parts in the different regimes, and the updates of the regime probability. These different channels can reinforce each other or work in opposite directions. For example, fig. 2b shows a positive relation between large shocks to the stocks return at t and its expectation at $t+1$, but a negative one for small shocks. This pattern is in line with the signs of the AR coefficients in table 4.

Shocks to the other variables affect expected stock returns via the three channels: the VAR-effects in the different regimes, the VAR-effect of the other variables combined with the contemporaneous effect of the shock, and the update of the regime probabilities. Based on eq. (42), the total effect from a shock to variable i equals $GI_{\boldsymbol{y}}^{y_i}(1, \eta_{it}, I_{t-1}) = \boldsymbol{G}_{\boldsymbol{y}}(\boldsymbol{\Phi}\,\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, I_{t-1}] + \boldsymbol{CP}\,\mathrm{E}[\boldsymbol{u}_t|y_{it}, I_{t-1}])$. The first effect is captured by the rows in $\boldsymbol{\Phi}$ that correspond with y_i, the second effect with the rows in $\boldsymbol{\Phi}$ that correspond with the stock returns combined with the contemporaneous effect in $\mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, I_{t-1}]$, whereas the third effect runs completely via the second term. With this knowledge we can explain why shocks to bond returns at t have a stronger effect on stock returns at $t+1$ than shocks to stock returns themselves. Shocks to stock returns translate almost one-on-one to oppositie shocks in the D/P ratio because of their strong correlation. In particular in regime 2, their effects on stock returns at $t+1$ are almost equal (see table 6), so they offset each other to a large extent. Shocks to bond returns also have a positive effect on stock returns at $t+1$, in particular in regime 2, but because they are not strongly correlated with the other innovations, they are not dampened. Also here, the curve shows the effect of regime switching by its nonlinearity. We further see that the effect of shocks to the T-Bill rate on stock returns at $t+1$ runs largely via the updating of the regime probability, and that the effect of shocks to the D/P ratio mirrors those of shocks to stock returns themselves because of their strong negative contemporaneous correlation.

The nonlinear effects of regime switching are less dominant in fig. 2e than in fig. 2b, in line with the smaller effects on the forecast probabilities in fig. 1. The results for the regime process being in the steady state in fig. 2h are in between those for the low and high volatility regime.

The subfigures for the longer horizon of 6 months in the right panels of fig. 2 show the familiar long-lasting effects of persistent state variables. Shocks to the T-bill rate and the D/P ratio die out slowly, though their magnitude is substantially lower than at t or $t+1$. The effect of shocks to stock returns also dissipate slowly because of the high negative correlation with innovations in the D/P ratio. The slow mean-reversion of the D/P-ratio also explains why the effects of shocks to stock returns and the D/P ratio switch signs. nonlinear effects remain present to the extent of shocks affecting the regime probabilities as in fig. 1.

For investors, this means that the estimates in table 4, which correspond with the average effect at $h = 0$, does not present the full 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 positive contemporaneous effect 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 bit easier to predict and understand, because the effect of regime switching is less here.

The generalized decompositions of the forecast error variances for stock returns in fig. 3 confirm these conclusions. By construction, the largest proportion at t comes from shocks to stock returns themselves, though the strong correlation between innovations in stock returns and the D/P ratio leads to a high proportion for the D/P ratio, too. At longer horizons, the proportions due to the other variables increase. In particular when the low volatility regime prevails at $t-1$, the proportions depend nonlinearly on the shock size. Negative shocks to the

D/P ratio are more consequential than positive shocks, and for bond returns shocks that are large in magnitude explain a larger proportion than small shocks. These effects are less obvious when the high volatility regime prevails because the effect of regime switching is smaller.

[Figure 3 about here.]

The responses of expected bond returns in fig. 4 also show a combination of direct and indirect VAR effects combined with the nonlinearity induced by the Markov switching part. The contemporaneous effects of shocks to the other variables in the left column are small and highly nonlinear because the probability update channel dominates. Because the VAR-effects are also not very strong, the nonlinear effect of this channel is still quite present at $t+1$, and because the VAR-effects fade out faster than the probability updating, the effects of shocks at $t+6$ are highly nonlinear (though small in magnitude). And since the effect of shocks on the forecast probabilities is larger when the low volatility regime prevails, the nonlinearity of this channel is strongest in the top row, and weakest in the middle row. The effect of shocks to stock returns (and to the D/P ratio because of their strong negative correlation) is larger than shocks to bond returns themselves, because a one-standard deviation shock to stock returns has a larger %-effect than the same shock to bond returns. Shocks to the T-Bill rate do have a strong effect on bond returns contemporaneously or at short horizons, but become more important at longer horizons as they die out very slowly.

[Figure 4 about here.]

The consequences for investors of this impulse response analysis for bond returns is a bit more positive than for stock returns. 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 is positive, but it becomes negative in the next period, indicating some protection.

The GFEVD in fig. 5 reflect the low correlations with and the weak VAR effect from the other variables. Typically more than 90% of the forecast variance in bond returns can be attributed to own shocks. The effect via the probability update channel causes the proportions attributed to other variables to increase when shocks are large in magnitude, in particular when the low volatility regime prevails at $t-1$.

[Figure 5 about here.]

4.3.2 Second Order Impulse Responses

Our analysis of the first order impulse responses shows how shocks effect the return part of the risk-return trade-off. We now turn to the second order impulse responses for the risk part. We use the law of total variance to discern the four different channels through which shocks affect the variance of $\boldsymbol{y}_{t+h}$,

$$\begin{aligned}\mathrm{Var}[\boldsymbol{y}_{t+h}|y_{it}, I_{t-1}] = &\sum_{j=1}^{m} \mathrm{Var}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it}, I_{t-1}] \Pr[S_{t+h} = j|y_{it}, I_{t-1}] + \\ &\sum_{j=1}^{m} \Pr[S_{t+h} = j|y_{it}, I_{t-1}](1 - \Pr[S_{t+h} = j|y_{it}, I_{t-1}]) \times \\ &\quad \mathrm{E}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it}, I_{t-1}] \, \mathrm{E}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it}, I_{t-1}]'.\end{aligned}$$

The first channel runs via the effect that the shock in y_{it} has on the regime-specific variances at $t + h$. When $h = 0$, the contemporaneous correlation of the variables leads to a decrease in the variance of the other variables. Next and as before, shocks lead to an updating of the regime forecasts $\Pr[S_{t+h} = j|y_{it}, I_{t-1}]$, and hence the forecast for the low or the high volatility regime. On top of this effect, the variability of the regime process also contributes to the total variance via the term $\Pr[S_{t+h} = j|y_{it}, I_{t-1}](1 - \Pr[S_{t+h} = j|y_{it}, I_{t-1}])$. The final channel stems from the difference between the autoregressive components in the different regimes. Shocks propagate differently in the different regimes, which is captured by the term $\mathrm{E}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it}, I_{t-1}] \, \mathrm{E}[\boldsymbol{y}_{t+h}|S_{t+h} = j, y_{it}, I_{t-1}]'$. We investigate again how much the different channels contribute, and how the effects depend on the size of the shock and the regime distribution at $t - 1$.

Figure 6 shows the effect of shocks on the forecast volatility of stock returns for different initial regimes and horizons. The figures in the left panels show the contemporaneous effect of shocks. In both regimes, the variance conditional on a shock is lower than the original variance, but the shock also leads to an updating of the regime forecast. The first effect dominates when shocks originate from the D/P ratio, because of its strong negative correlation with innovations in stocks returns. When shocks originate from bond returns or the T-Bill rate, the updating of the regime forecast and the increased variability of the regime process dominate and generally lead to an increase of the volatility. This third channel of increased variability explains why large shocks to bond returns and the T-Bill rate have approximately the same effect on the volatility, whereas the effect of small shocks differs, but not as much as in fig. 1. Large shocks indicate a switch to the high volatility regime with a high degree of certainty, whereas small shocks make a switch to the low volatility regime more likely without completely removing the uncertainty. This remaining uncertainty counteracts the effect of switching to the low volatility regime, and explains why the difference between the effects of bond returns and the T-Bill rate at $\delta = 0$ is smaller in fig. 6d than in fig. 1e.

[Figure 6 about here.]

The middle and right panels of fig. 6 for horizons of one and six months show that shocks mostly have an upward effect on the volatility, independent of the regime distribution at $t-1$. Only for small shocks do we see a small decrease in the forecast volatility. Moroever and contrary to their contemporaneous effect, the effect of large shocks increases when they become larger. This increase comes in via the fourth channel. To understand this effect, we have to realize that no matter the shock size at t, there is always uncertainty about the regimes at longer horizons. Consequently, the shock can propagate via either of the two regimes, and the difference between these two paths is increasing in the size of the shock, which in turn has an upward effect on the variance. This is a big difference with both MSIH models and VAR models where the effect on the volatility is bounded or constant. Because the shocks die out, the effects of shocks on the volatility at $t+6$ is much smaller, but the same channels as at $t+1$ remain present. For small shocks, the second and third channels we discerned at t operate in about the same way.

The effects of shock on the volatility of bond returns fig. 7 are less pronounced than for stock returns, though we see the presence of the same four channels. The correlations between innovations to bond returns and the other variables is weaker than for stock returns, which means that the first channel is weaker. The second channel depends on the difference between the volatilities in the regimes, which is about the same for stock and bond returns. The third channel depends on the Markov switching part, which is the same for all variables. Finally, the fourth channel depends on the predictability and the differences therein between the different regimes, and this aspect is again weaker for bond returns. As a consequence, we see an increasing effect of large negative and positive shocks on the volatility of bond returns at $t+1$, but it is much smaller than for stock returns.

[Figure 7 about here.]

Figures 8 and 9 provide a good view of the persistence of the effects on the volatilities for the different sources of the shocks and regime distributions at time $t-1$. The dashed gray lines show the forecast volatility in the absence of shocks, whereas the solid colored lines show the effect of shocks for sizes of -2, -1, 0, 1 and 2 standard deviations. All subfigures show that the effect of shocks is prolonged. They take longer to die out in volatilities than in returns, comparing the effects at t, $t+1$ and $t+6$ in figs. 2 and 4.[5] At $t+3$ about half of the effect at $t+1$ is still present, though at $t+12$, so after a year, most of the effect is gone.

The figures also show that the effect is asymmetric. The effect of shocks is increasing in the (absolute) size of the shock, but shocks of the same size that differ in sign affect volatilities differently, and also decay differently as can be concluded from the intersection of the lines. For example, fig. 8b shows that a shock to stock returns of -2 standard deviations at t has a smaller effect on the forecast stock volatility at $t+1$ than shocks of +1 and +2 standard deviations, but a larger effect from $t+3$ onwards. Because shocks to stock returns and to the D/P ratio are highly correlated, the effect of shocks to the D/P ratio on stock volatility is negative at t but positive at long horizons for larger shock sizes, as shows by figs. 8j to 8l.

[Figure 8 about here.]

[Figure 9 about here.]

Of course, shocks also affect the forecast correlations at different horizons. We plot the effects of shocks on the forecast stock-bond correlations over time for the different sources and regime distributions in fig. 10. The different subfigures show that shocks to stocks or bond returns themselves lead to intricate dynamics. Whereas the size of the effects decreases over time, they often change from positive to negative or vice versa. Shocks that originate from the state variables follow a more straightforward pattern.

[Figure 10 about here.]

Summarizing, our analyses show that shocks can lead to an increase or a decrease 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 just 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 of analysing 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 also have positive and negative effects on the volatility and correlation of stocks and bond returns. In the high

[5]We plot the effect of shocks on forecast stock and bond returns over time in figs. C.1 and C.2 in Appendix C.

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 decreasing diversification opportunities.

5 Conclusion

In this paper, 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 crucial step 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) 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 also derive and use the specification of the MDS-part.

As our main theoretical contribution, we show how to use this extensive VAR(1) formulation to derive its 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 the Generalized Forecast Error Variance Decomposition can be constructed.

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 bull and bear markets. 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 over 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 bear regime.

We conclude that our theoretical and empirical results are useful for modern risk management. The tools we propose can be used directly to investigate what happens when a shock hits in a bad situation. 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 without the need for 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.

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

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 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.

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.

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.

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. (2012a). 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. (2012b). Simple VARs cannot approximate Markov switching asset allocation decisions: An out-of-sample assessment. *Computational Statistics & Data Analysis*, 56(11):3546–3566.

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 and 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.

Guidolin, M. and Timmermann, A. (2008). International asset allocation under regime switching, skew and kurtosis preference. *Review of Financial Studies*, 21(2):889–935.

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. 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, Princeton, NJ, USA.

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

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.

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.

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

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.

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 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 3: 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 February 1952 to December 2018 (803 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 4: 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 model VAR 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 (803 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 5: **Regime-conditional and unconditional first and second moments**

	Means conditional on Regime 1				Volatilities conditional on Regime 1			
	Stocks	Bonds	T-Bill	D/P	Stocks	Bonds	T-Bill	D/P
MSIH(2)	1.00	-0.01			3.08	1.45		
MSIAH(2)-VAR(1)	0.75	0.01	0.40	-3.58	3.23	1.56	0.28	0.34
MSIAH(2)-VAR(1) inf.	0.22	0.02	0.49	-4.35	3.21	1.56	0.18	0.34
	Means conditional on Regime 2				Volatilities conditional on Regime 2			
MSIH(2)	-0.48	0.40			5.63	2.94		
MSIAH(2)-VAR(1)	-0.67	0.37	0.40	-3.54	5.49	2.83	0.29	0.33
MSIAH(2)-VAR(1) inf.	0.26	0.28	0.44	-3.24	5.50	2.84	0.32	0.32
	Unconditional means				Unconditional volatilities			
MSIH(2)	0.50	0.13			4.18	2.08		
VAR(1)	0.40	0.15	0.31	-3.68	4.17	2.08	0.25	0.39
MSIAH(2)-VAR(1)	0.28	0.13	0.40	-3.57	4.18	2.08	0.28	0.34

Regime-conditional correlations

	Regime 1						
	MSIH(2)	MSIAH(2)-VAR(1)			MSIAH(2)-VAR(1) inf.		
Bonds	0.073	0.037			0.032		
T-Bill		-0.114	-0.043		-0.073	-0.028	
D/P		-0.097	-0.024	0.455	-0.065	-0.015	0.351
	Regime 2						
	MSIH(2)	MSIAH(2)-VAR(1)			MSIAH(2)-VAR(1) inf.		
Bonds	0.119	0.140			0.138		
T-Bill		-0.063	-0.032		-0.069	-0.036	
D/P		-0.035	-0.036	0.455	-0.066	-0.037	0.510

Unconditional Correlations

	MSIH(2)	VAR(1)			MSIAH(2)-VAR(1)		
Bonds	0.085	0.085			0.084		
T-Bill		-0.091	-0.011		-0.085	-0.037	
D/P		-0.065	-0.014	0.512	-0.074	-0.024	0.454

This table gives the regime-conditional and unconditional means, volatilities and correlations for the MSIH(2), VAR(1) and MSIAH(2)-VAR(1) models with parameter values reported in Tables 2 to 4. The regime-conditional mean $\mathrm{E}[\boldsymbol{y}_t|S_t]$ 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 and blocks labeled "MSIAH(2)-VAR(1) inf." report the moments conditional on the given regime prevailing infinitely long in the past, 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 eqs. (21) to (24).

Table 6: Auto- and cross-correlations at (lead) order 1

(a) Conditional on current regimes

	Regime 1				Regime 2			
	Stocks	Bonds	T-Bill	D/P	Stocks	Bonds	T-Bill	D/P
Stocks	-0.060	-0.096	-0.117	-0.089	0.112	-0.181	-0.051	-0.052
Bonds	0.094	0.010	-0.059	-0.033	0.149	0.089	-0.120	-0.056
T-Bill	-0.092	-0.038	0.994	0.456	-0.041	-0.024	0.972	0.458
D/P	0.022	-0.016	0.454	0.994	0.102	-0.016	0.449	0.987

(b) Conditional on regimes prevailing forever

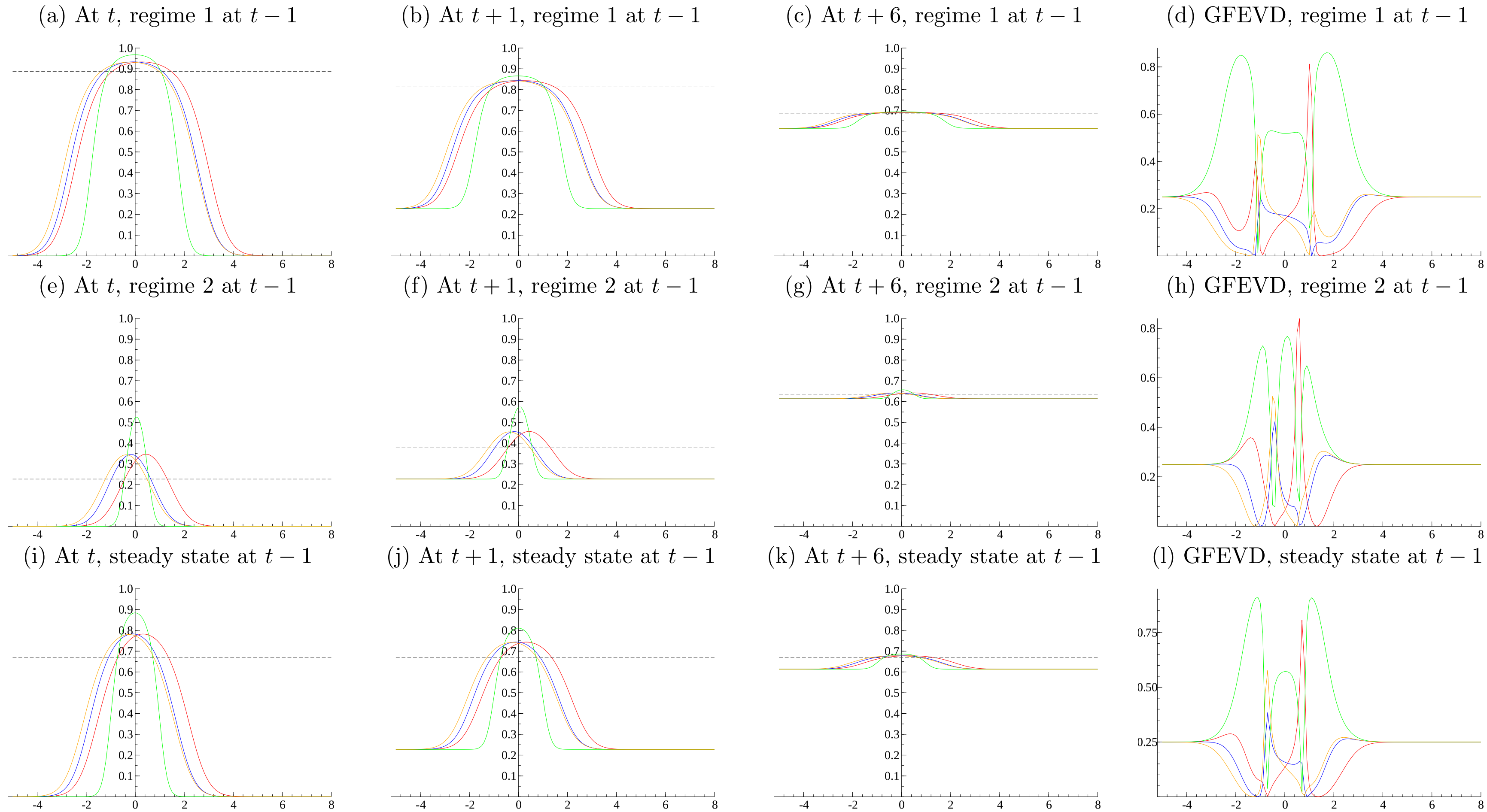
	Regime 1				Regime 2			
	Stocks	Bonds	T-Bill	D/P	Stocks	Bonds	T-Bill	D/P
Stocks	-0.101	-0.097	-0.082	-0.054	0.158	-0.188	-0.051	-0.094
Bonds	0.098	-0.005	-0.042	-0.024	0.154	0.112	-0.134	-0.058
T-Bill	-0.066	-0.028	0.995	0.352	-0.036	-0.024	0.971	0.513
D/P	0.031	-0.010	0.351	0.995	0.103	-0.012	0.501	0.983

(c) Unconditional

	MSIH(2)		VAR(1)				MSIAH(2)-VAR(1)			
	Stocks	Bonds	Stocks	Bonds	T-Bill	D/P	Stocks	Bonds	T-Bill	D/P
Stocks	0.023	-0.013	0.049	-0.151	-0.087	-0.069	0.049	-0.153	-0.080	-0.079
Bonds	-0.013	0.007	0.114	0.062	-0.067	-0.025	0.115	0.061	-0.086	-0.037
T-Bill			-0.072	-0.005	0.983	0.515	-0.068	-0.032	0.987	0.455
D/P			0.043	-0.003	0.507	0.993	0.049	-0.012	0.450	0.991

This table gives regime-conditional and unconditional auto- and cross-correlations with lead 1 for the MSIH(2), VAR(1) and MSIAH(2)-VAR(1) models with parameter values reported in tables 2 to 4. The regime-conditional correlations in panel (a) are calculated from $\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+1}|S_t]$ based on eq. (36), and the conditional volatilities based on proposition 5. The regime-conditional correlations in panel (b) are 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 5. The unconditional correlations are calculated from $\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+1}]$ based on eqs. (25) and (26), and the unconditional volatilities in table 5.

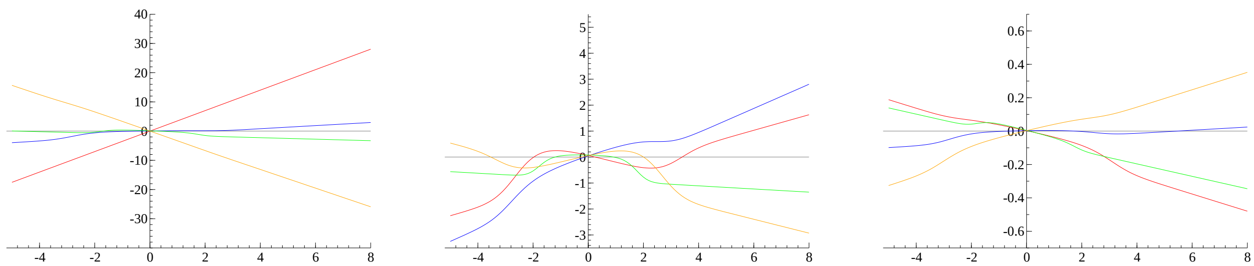
Figure 1: The effect of shocks on the state probabilities in the MSIAH(2)-VAR(1) model



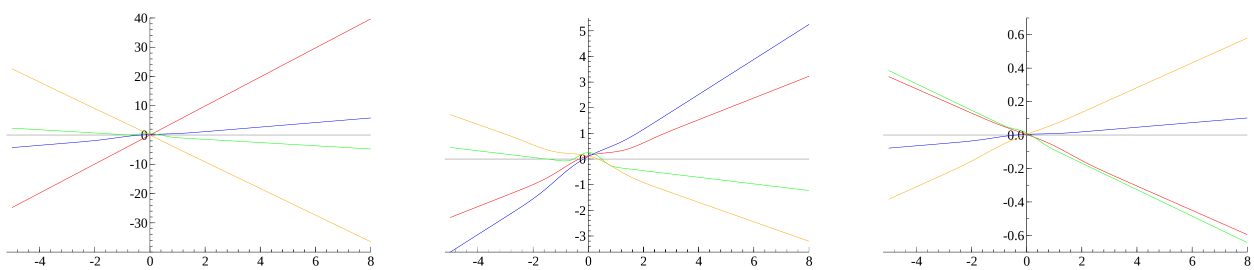
This figure shows the impulse responses for the probability forecasts for regime 1 at time t, $t+1$ and $t+6$, and the Generalized Forecast Error Variance Decomposition (GFEVD) in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The straight dashed gray lines in panels a–c, e–g and i–k give the forecast probability of state 1 in absence of a shock. The solid lines show the probability of state 1 conditional on a shock of δ times the 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, $\Pr\left[S_{t+h} = 1 \middle| S_{t-1}, y_{it} - \mathrm{E}[y_{it}|S_{t-1}] = \delta \operatorname{vol}[y_{it}|S_{t-1}]\right]$. The difference between the dashed and the solid lines corresponds with $GI^{y}_{s_1} = \tilde{\boldsymbol{G}}_{s_1} GI^{y_i}_{\tilde{\boldsymbol{y}}}(h, \delta \operatorname{vol}[y_{it}|S_{t-1}])$ as in eq. (42) in proposition 7. Panels d, h and l give the Generalized Forecast Error Variance Decomposition in eq. (51) for the probability of regime 1 that can attributed to shocks to the different variables as a function of the standardized shock size δ.

Figure 2: The effects of shocks on expected stock returns in the MSIAH(2)-VAR(1) model

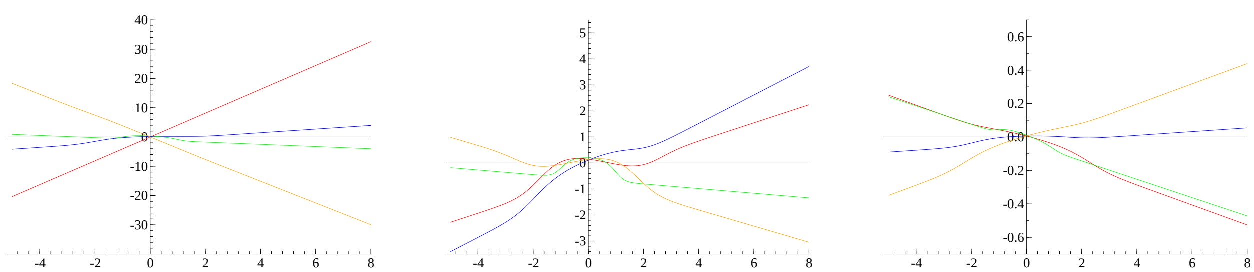
(a) At t, regime 1 at $t-1$ (b) At $t+1$, regime 1 at $t-1$ (c) At $t+6$, regime 1 at $t-1$



(d) At t, regime 2 at $t-1$ (e) At $t+1$, regime 2 at $t-1$ (f) At $t+6$, regime 2 at $t-1$



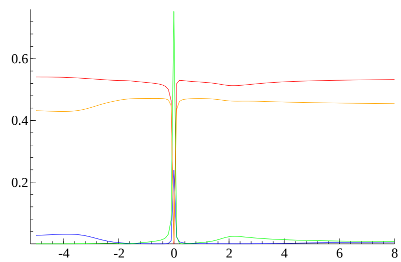
(g) At t, steady state at $t-1$ (h) At $t+1$, steady state at $t-1$ (i) At $t+6$, steady state at $t-1$



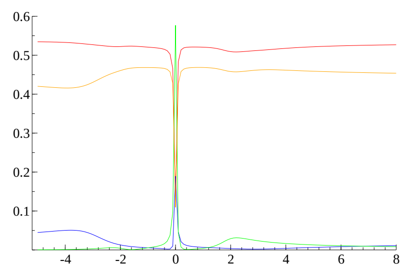
This figure shows the Generalized Impulse Response Functions (GIRF) for stock returns at time t, $t+1$ and $t+6$ in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1 (top row), 2 (middle row) or the steady state (bottom row) at time $t-1$. The lines show the the GIRF as a function of shocks equal to δ times the 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 for horizons $h = 0, 1, 6$ as in eq. (50).

Figure 3: The forecast error variance decomposition of stock returns in the MSIAH(2)-VAR(1) model

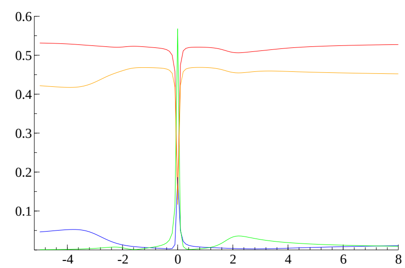
(a) At t, regime 1 at $t-1$



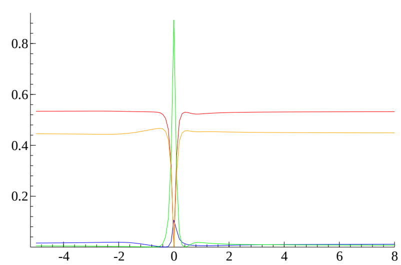
(b) At $t+1$, regime 1 at $t-1$



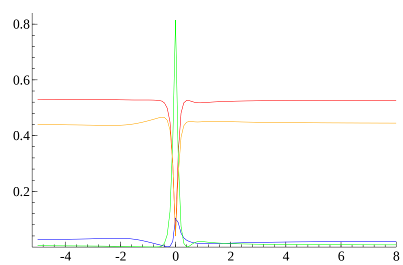
(c) At $t+6$, regime 1 at $t-1$



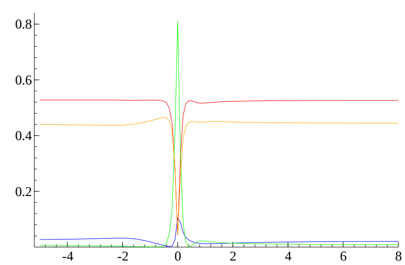
(d) At t, regime 2 at $t-1$



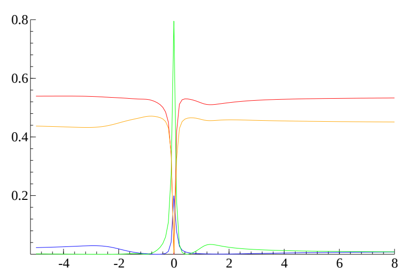
(e) At $t+1$, regime 2 at $t-1$



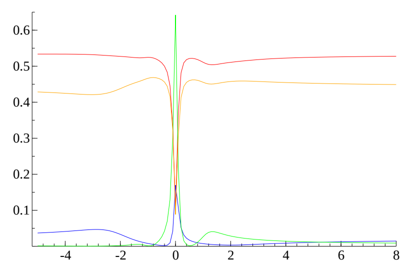
(f) At $t+6$, regime 2 at $t-1$



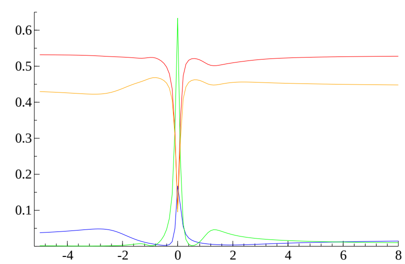
(g) At t, steady state at $t-1$



(h) At $t+1$, steady state at $t-1$



(i) At $t+6$, steady state at $t-1$



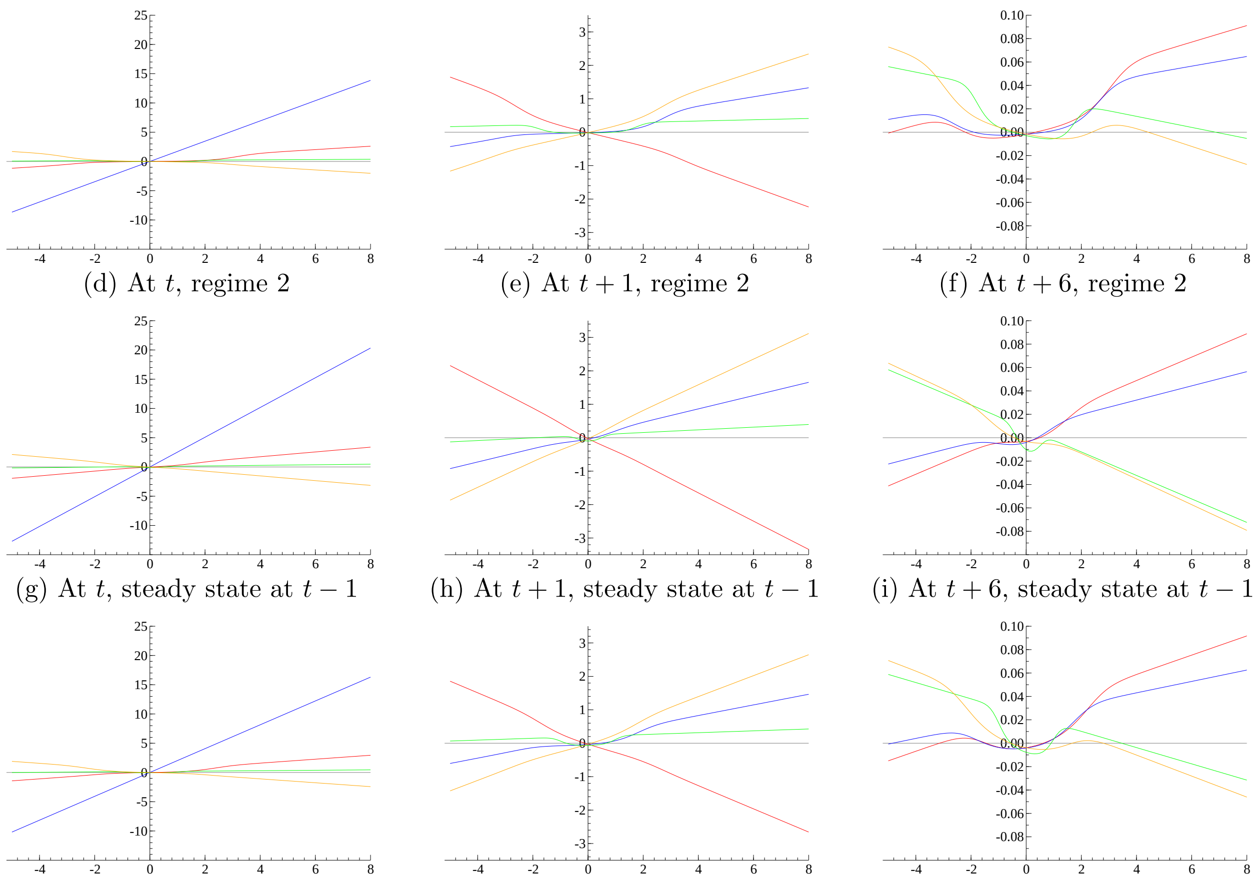
This figure shows the Generalized Forecast Error Variance Decomposition (GFEVD) for stock returns at time t, $t+1$ and $t+6$ in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1 (top row), 2 (middle row) or the steady state (bottom row) at time $t-1$. The lines show the the 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 for horizons $h = 0, 1, 6$ as in eq. (51).

Figure 4: The effects of shocks on expected bond returns in the MSIAH(2)-VAR(1) model

(a) At t, regime 1 (b) At $t+1$, regime 1 (c) At $t+6$, regime 1

(d) At t, regime 2 (e) At $t+1$, regime 2 (f) At $t+6$, regime 2

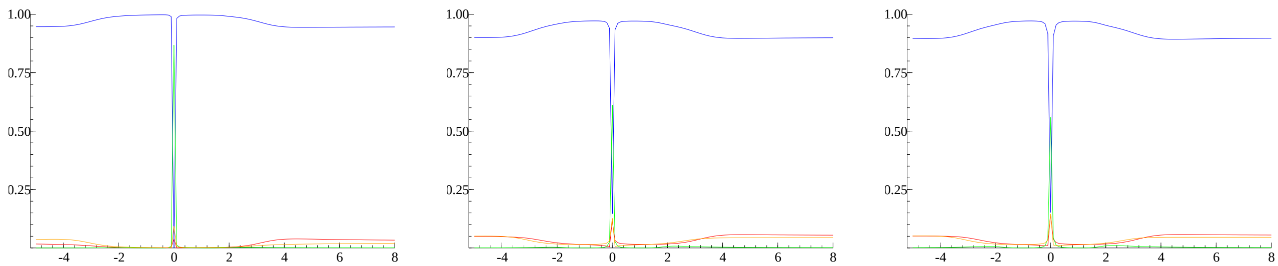
(g) At t, steady state at $t-1$ (h) At $t+1$, steady state at $t-1$ (i) At $t+6$, steady state at $t-1$



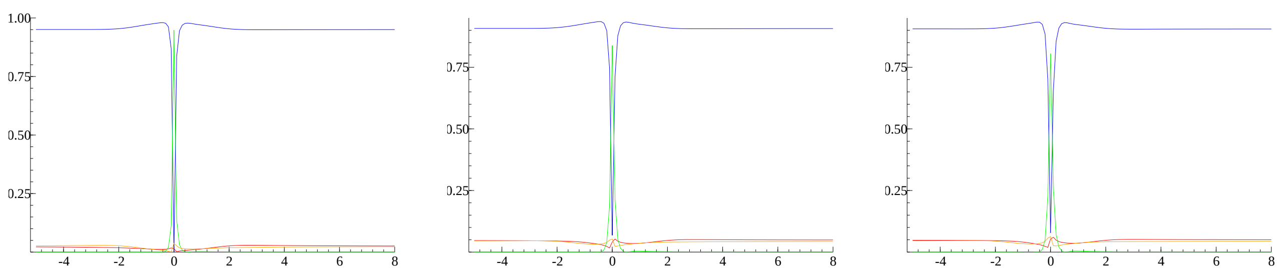
This figure shows the Generalized Impulse Response Functions (GIRF) for bond returns at time t, $t+1$ and $t+6$ in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1 (top row), 2 (middle row) or the steady state (bottom row) at time $t-1$. The lines show the the GIRF as a function of shocks equal to δ times the forecast standard deviation of bond returns in red, bond returns in blue, the T-bill rate in green and the D/P ratio in orange for horizons $h = 0, 1, 6$ as in eq. (50).

Figure 5: The forecast error variance decomposition of bond returns in the MSIAH(2)-VAR(1) model

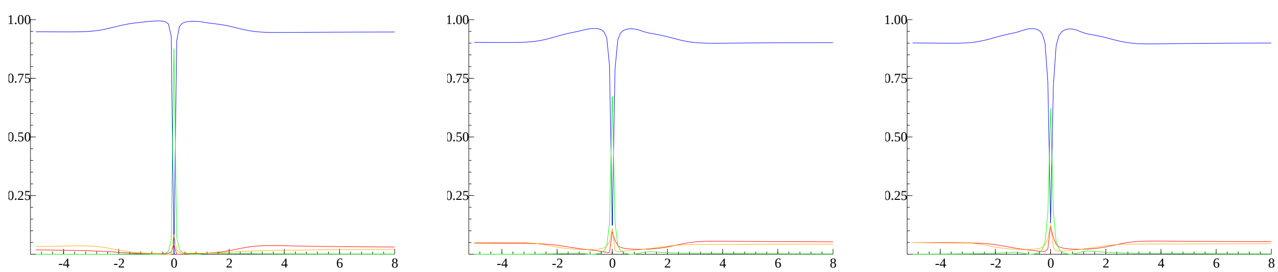
(a) At t, regime 1 at $t-1$ (b) At $t+1$, regime 1 at $t-1$ (c) At $t+6$, regime 1 at $t-1$



(d) At t, regime 2 at $t-1$ (e) At $t+1$, regime 2 at $t-1$ (f) At $t+6$, regime 2 at $t-1$



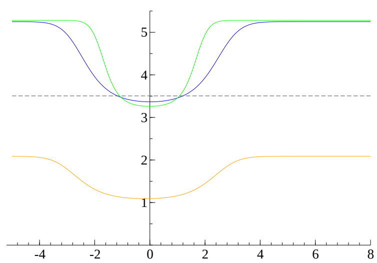
(g) At t, steady state at $t-1$ (h) At $t+1$, steady state at $t-1$ (i) At $t+6$, steady state at $t-1$



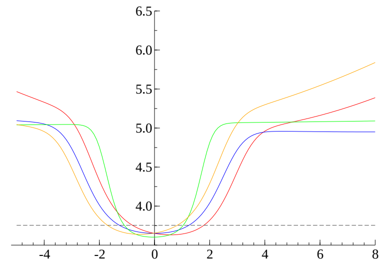
This figure shows the Generalized Forecast Error Variance Decomposition (GFEVD) for bond returns at time t, $t+1$ and $t+6$ in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1 (top row), 2 (middle row) or the steady state (bottom row) at time $t-1$. The lines show the the GFEVD as a function of shocks equal to δ times the time t forecast standard deviation of bond returns in red, bond returns in blue, the T-bill rate in green and the D/P ratio in orange for horizons $h = 0, 1, 6$ as in eq. (51).

Figure 6: The effects of shocks on the forecast volatility of stock returns in the MSIAH(2)-VAR(1) model

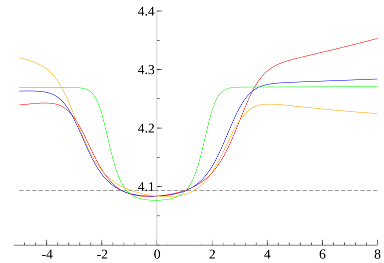
(a) Response at t, regime 1



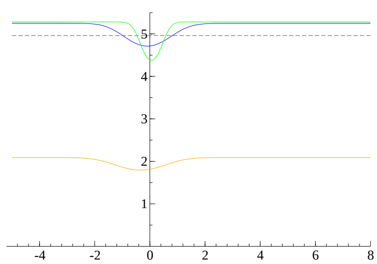
(b) Response at $t+1$, regime 1



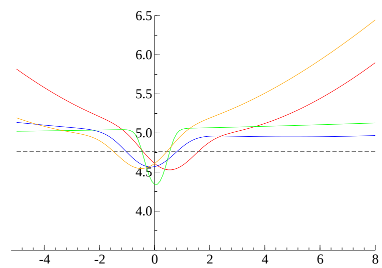
(c) Response at $t+6$, regime 1



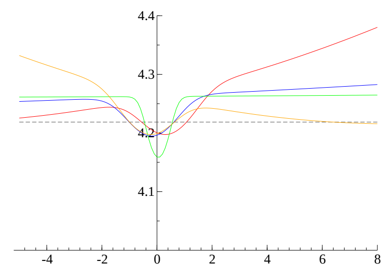
(d) Response at t, regime 2



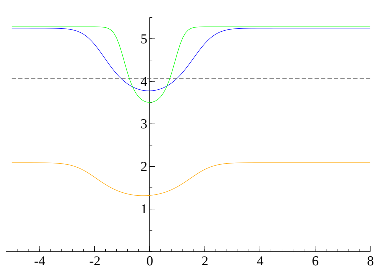
(e) Response at $t+1$, regime 2



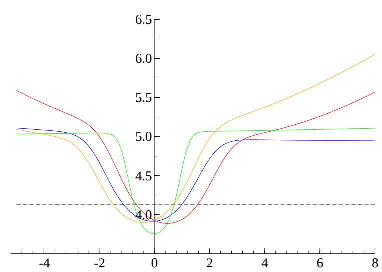
(f) Response at $t+6$, regime 2



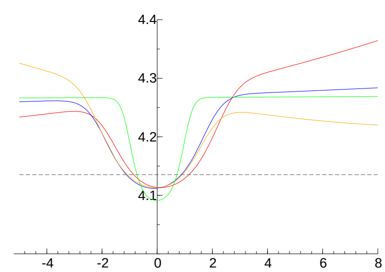
(g) Response at t, steady state at $t-1$



(h) Response at $t+1$, steady state at $t-1$

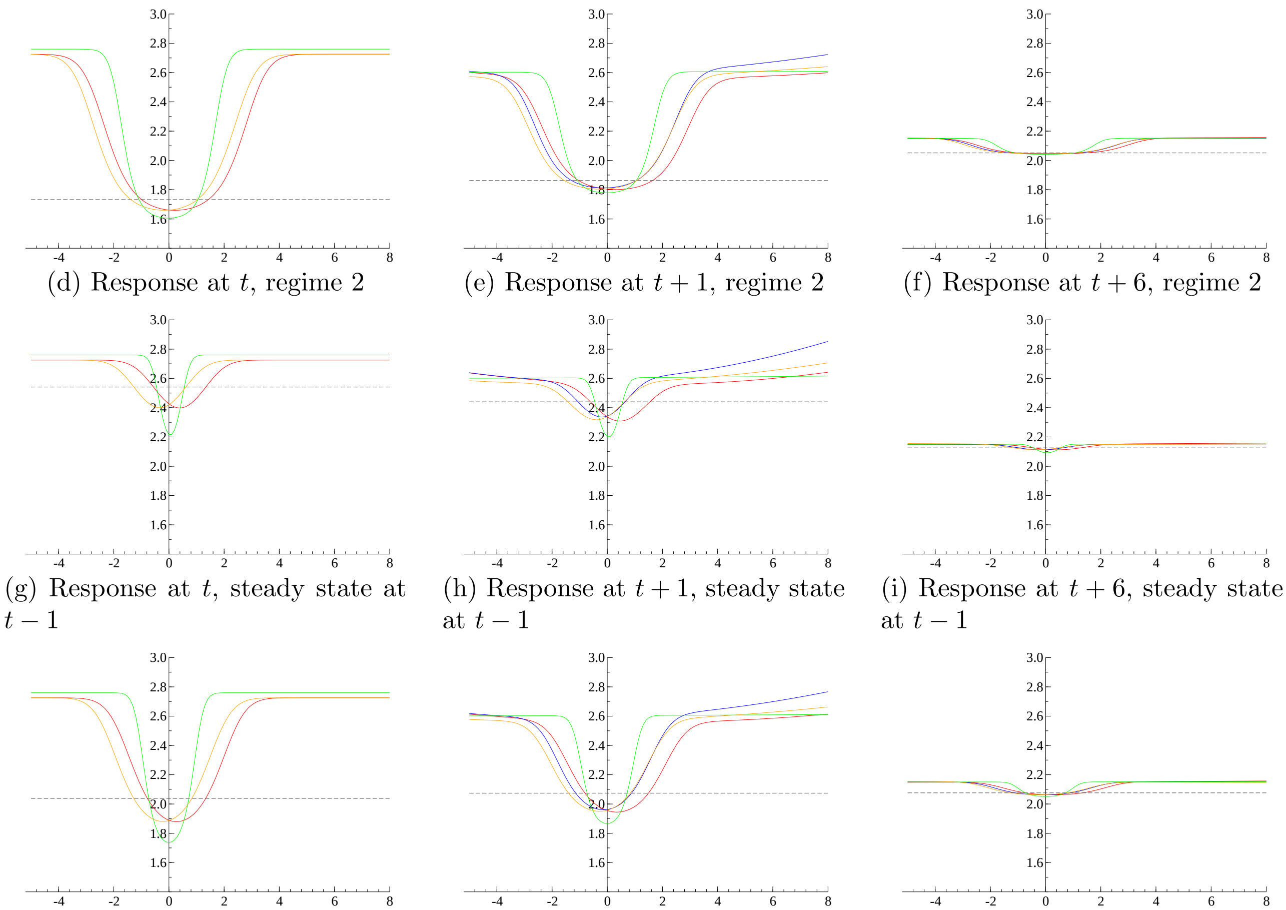


(i) Response at $t+6$, steady state at $t-1$



This figure shows the impulse responses for the forecast volatility of stock returns at time t, $t+1$ and $t+6$ in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The straight dashed gray lines give the forecast volatility without a shock. The solid lines show the forecast volatility conditional on a shock of δ times the forecast standard deviation of stock returns $i=1$ in red, bond returns $i=2$ in blue, the T-bill rate $i=3$ in green and the D/P ratio in orange, $\mathrm{vol}\big[y_{1,t+h}\big|S_{t-1}, y_{it}-\mathrm{E}[y_{it}|S_{t-1}] = \delta\,\mathrm{vol}[y_{it}|S_{t-1}]\big]$. The difference between the dashed and the solid lines gives the volatility impulse function.

Figure 7: The effects of shocks on the forecast volatility of bond returns in the MSIAH(2)-VAR(1) model



(a) Response at t, regime 1

(b) Response at $t+1$, regime 1

(c) Response at $t+6$, regime 1

(d) Response at t, regime 2

(e) Response at $t+1$, regime 2

(f) Response at $t+6$, regime 2

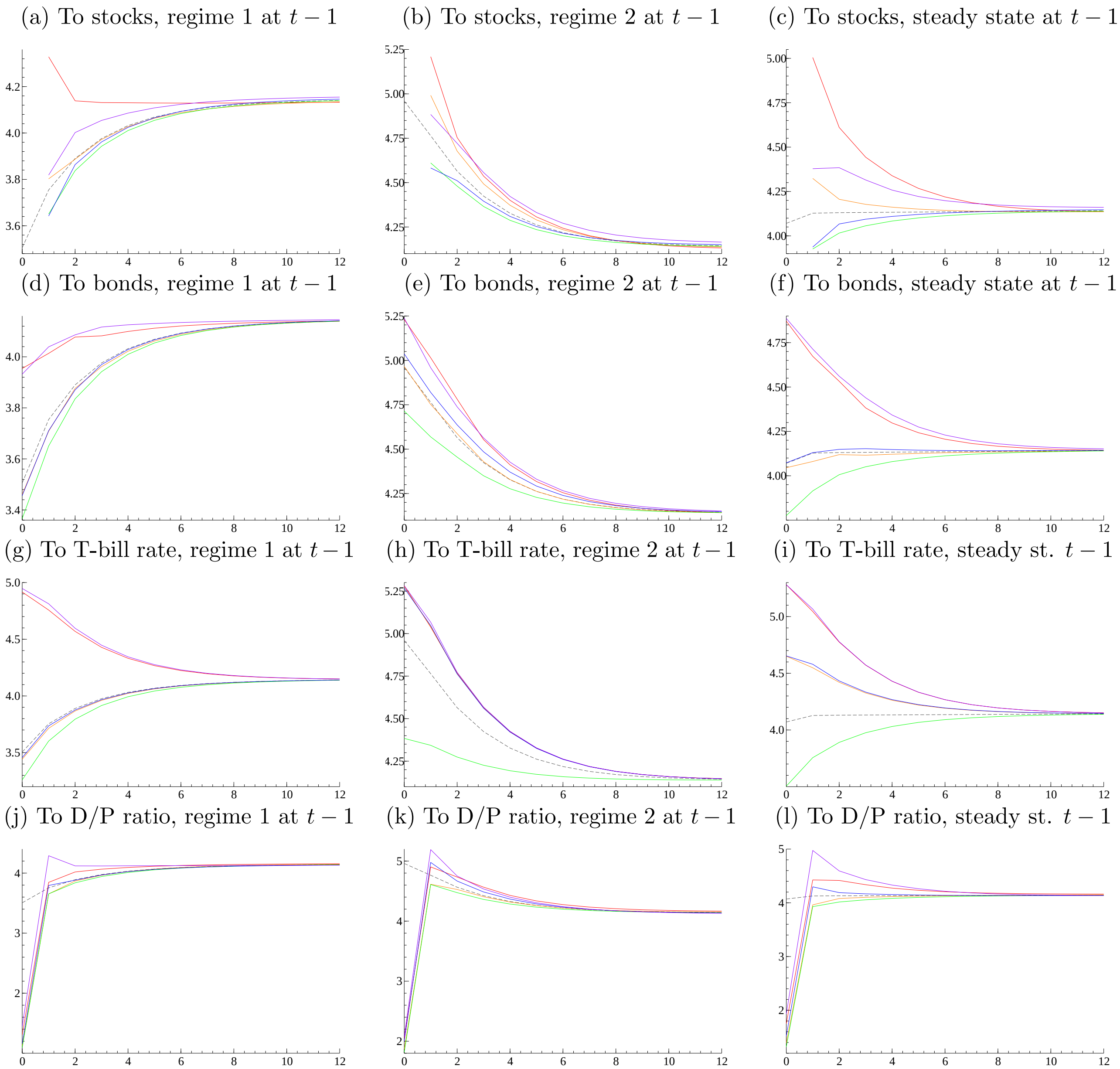
(g) Response at t, steady state at $t-1$

(h) Response at $t+1$, steady state at $t-1$

(i) Response at $t+6$, steady state at $t-1$

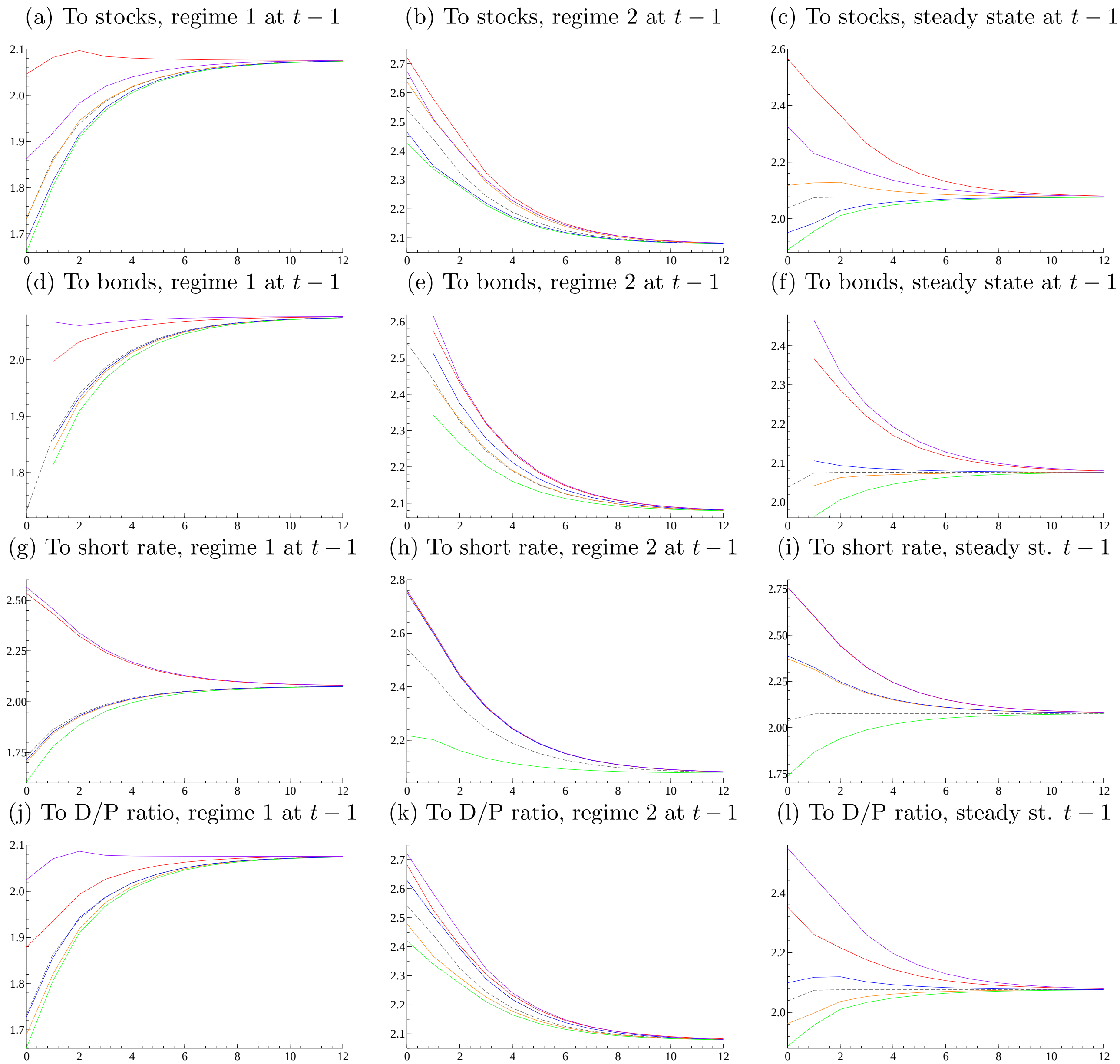
This figure shows the impulse responses for the forecast volatility of bond returns at time t, $t+1$ and $t+6$ in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The straight dashed gray lines give the forecast volatility without a shock. The solid lines show the forecast volatility conditional on a shock of δ times the forecast standard deviation of stock returns $i=1$ in red, bond returns $i=2$ in blue, the T-bill rate $i=3$ in green and the D/P ratio in orange, $\mathrm{vol}\big[y_{2,t+h}\big|S_{t-1}, y_{it}-\mathrm{E}[y_{it}|S_{t-1}] = \delta\,\mathrm{vol}[y_{it}|S_{t-1}]\big]$. The difference between the dashed and the solid lines gives the volatility impulse function.

Figure 8: The effects of shocks on the forecast volatility of stock returns over time



This figure shows the impulse responses for the forecast volatility of stock returns over time in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The dashed gray lines give the forecast volatility without a shock. The solid lines show the forecast volatility conditional on a shock of δ times the forecast standard deviation to stock returns (first row), bond returns (second row), the T-bill rate (third row), and the D/P ratio (fourth row), with δ equal to -2 (red line), -1 (orange line), 0 (green line), 1 (blue line), and 2 (violet line). The difference between the dashed and the solid lines gives the volatility impulse function. The forecast horizon runs from 0 through 12 months.

Figure 9: The effects of shocks on the forecast volatility of bond returns over time



(a) To stocks, regime 1 at $t-1$ (b) To stocks, regime 2 at $t-1$ (c) To stocks, steady state at $t-1$

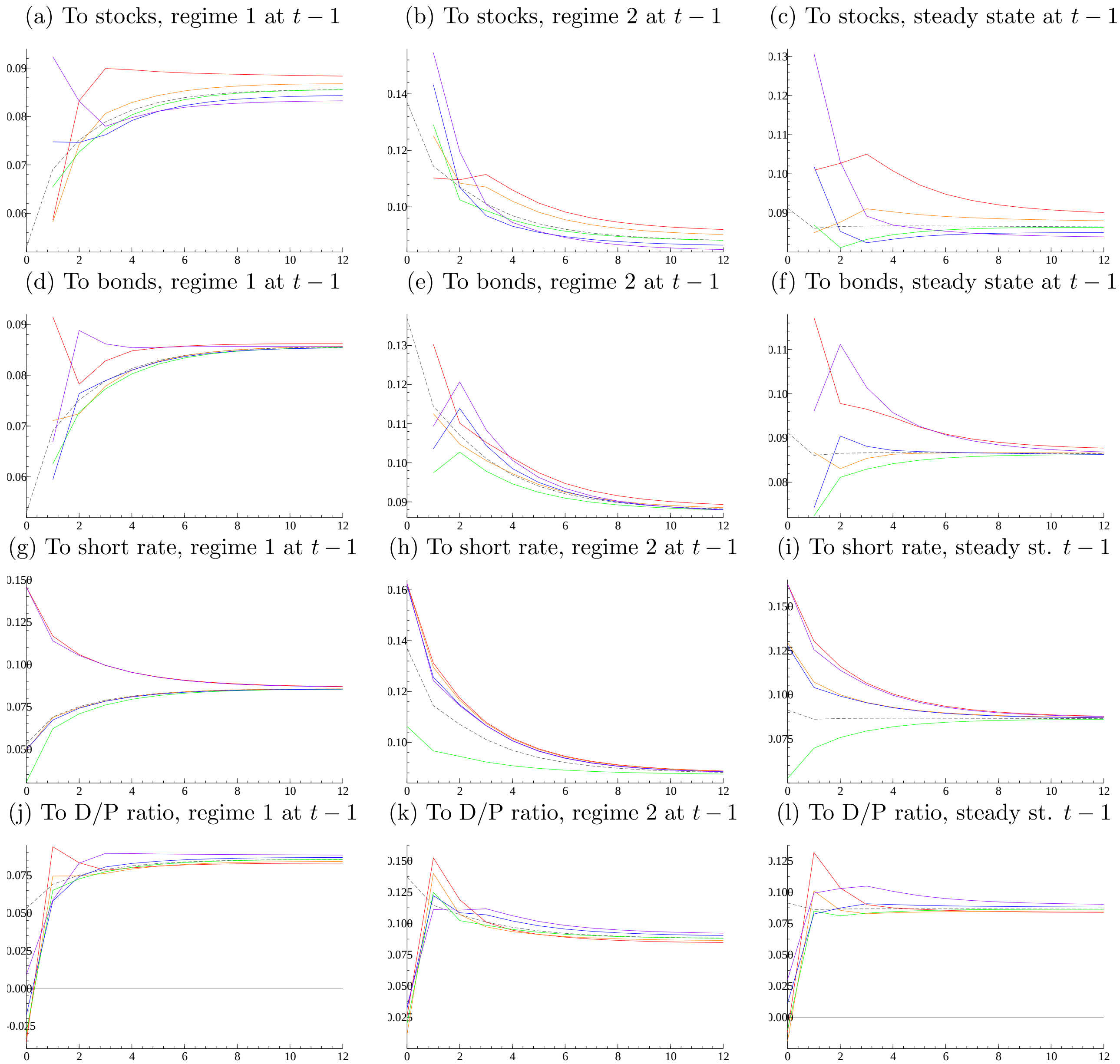
(d) To bonds, regime 1 at $t-1$ (e) To bonds, regime 2 at $t-1$ (f) To bonds, steady state at $t-1$

(g) To short rate, regime 1 at $t-1$ (h) To short rate, regime 2 at $t-1$ (i) To short rate, steady st. $t-1$

(j) To D/P ratio, regime 1 at $t-1$ (k) To D/P ratio, regime 2 at $t-1$ (l) To D/P ratio, steady st. $t-1$

This figure shows the impulse responses for the forecast volatility of bond returns over time in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The dashed gray lines give the forecast volatility without a shock. The solid lines show the forecast volatility conditional on a shock of δ times the forecast standard deviation to stock returns (first row), bond returns (second row), the T-bill rate (third row), and the D/P ratio (fourth row), with δ equal to -2 (red line), -1 (orange line), 0 (green line), 1 (blue line), and 2 (violet line). The difference between the dashed and the solid lines gives the volatility impulse function. The forecast horizon runs from 0 through 12 months.

Figure 10: The effects of shocks on the forecast correlation of stock and bond returns over time



(a) To stocks, regime 1 at $t-1$ (b) To stocks, regime 2 at $t-1$ (c) To stocks, steady state at $t-1$

(d) To bonds, regime 1 at $t-1$ (e) To bonds, regime 2 at $t-1$ (f) To bonds, steady state at $t-1$

(g) To short rate, regime 1 at $t-1$ (h) To short rate, regime 2 at $t-1$ (i) To short rate, steady st. $t-1$

(j) To D/P ratio, regime 1 at $t-1$ (k) To D/P ratio, regime 2 at $t-1$ (l) To D/P ratio, steady st. $t-1$

This figure shows the impulse responses for the forecast correlation of stock and bond returns over time in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The dashed gray lines give the forecast correlation without a shock. The solid lines show the forecast correlation conditional on a shock of δ times the forecast standard deviation to stock returns (first row), bond returns (second row), the T-bill rate (third row), and the D/P ratio (fourth row), with δ equal to -2 (red line), -1 (orange line), 0 (green line), 1 (blue line), and 2 (violet line). The difference between the dashed and the solid lines gives the correlation impulse function. The forecast horizon runs from 0 through 12 months.

A Proofs for Section 2 (MSVAR models and their moments)

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

Proposition 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (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*

$$\boldsymbol{y}_t^* = \boldsymbol{C}\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Phi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\varepsilon}_t^*, \tag{A.5}$$

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 eq. (3)*. $\tilde{\boldsymbol{y}}_t$ follows the process*

$$\tilde{\boldsymbol{y}}_t = \begin{pmatrix} \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Phi}}\tilde{\boldsymbol{y}}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t, \tag{A.6}$$

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

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

The multiplication of the coefficient matrices $\boldsymbol{C}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Lambda}$ with $\boldsymbol{s}_t$ ensures that the correct regime-dependent coefficients are selected. Substitution of eq. (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 (5) 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 eqs.* (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*

$$\boldsymbol{z}_t = \boldsymbol{\gamma}_{S_t} + \boldsymbol{\omega}_{S_t} + \boldsymbol{\Psi}_{S_t}\boldsymbol{y}_{t-1} + \boldsymbol{\Upsilon}_{S_t}\boldsymbol{z}_{t-1} + \boldsymbol{\zeta}_t, \tag{A.11}$$

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

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

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

$$\boldsymbol{z}_t^* = (\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{P}\boldsymbol{s}_{t-1} + \boldsymbol{\Psi}(\boldsymbol{P} \otimes \boldsymbol{I}_n)\boldsymbol{y}_{t-1}^* + \boldsymbol{\Upsilon}(\boldsymbol{P} \otimes \boldsymbol{I}_{n^2})\boldsymbol{z}_{t-1}^* + \boldsymbol{\zeta}_t^*, \tag{A.12}$$

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

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

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

$$\tilde{\boldsymbol{z}}_t = \begin{pmatrix} \boldsymbol{z}_t^* \\ \boldsymbol{y}_t^* \\ \boldsymbol{s}_t \end{pmatrix} = \tilde{\boldsymbol{\Upsilon}}\tilde{\boldsymbol{z}}_{t-1} + \tilde{\boldsymbol{\zeta}}_t, \tag{A.13}$$

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. Morover, $\boldsymbol{\zeta}_t$, $\boldsymbol{\zeta}_t^*$ and $\tilde{\boldsymbol{\zeta}}_t$ are MDS.

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

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

and working out the multiplication gives eq. (11). We use that $\boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t = (\boldsymbol{\Lambda}_{S_t} \otimes \boldsymbol{\Lambda}_{S_t})(\boldsymbol{\varepsilon}_t \otimes \boldsymbol{\varepsilon}_t) = \operatorname{vec}(\boldsymbol{\Lambda}_{S_t}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'\boldsymbol{\Lambda}'_{S_t})$ 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 $\boldsymbol{z}^*_t$ 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 eq. (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 eq. (12). 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 eqs.* (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 $\tilde{\boldsymbol{z}}_{t+k,t} = \big(\boldsymbol{z}^{*\prime}_{t+k,t}, \boldsymbol{y}^{*\prime}_{t+k,t}, \boldsymbol{s}'_{t+k}\big)'$ follows*

$$\tilde{\boldsymbol{z}}_{t+k,t} = \begin{pmatrix} \tilde{\boldsymbol{\Phi}} \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P} \end{pmatrix} \tilde{\boldsymbol{z}}_{t+k-1,t} + \begin{pmatrix} \tilde{\boldsymbol{\varepsilon}}_{t+k} \otimes \boldsymbol{y}_t \\ \boldsymbol{u}_{t+k} \end{pmatrix}, \quad k \geq 1 \tag{A.18}$$

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

Proof. We use that

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

where we have substitued eq. (6) 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 eq. (18). 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 $k \geq 1, l \geq 1$. □

Lemma 1. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (1) *and* (2)*. Let $\boldsymbol{Q}$ be the transition matrix of the time-reversed Markov chain of the proces S_t, and let $\boldsymbol{\Upsilon} = \mathrm{bdiag}_{j=1}^{m}(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j)$. 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 $\boldsymbol{Q} = \mathrm{diag}(\bar{\boldsymbol{\xi}})\boldsymbol{P}' \,\mathrm{diag}(\bar{\boldsymbol{\xi}})^{-1}$ to find

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

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 $\big(\mathrm{diag}(\bar{\boldsymbol{\xi}}) \otimes \boldsymbol{I}_{n^2}\big)^{-1}$ is diagonal, which allows to switch the order of multiplication. □

Proposition 4. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (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 $\boldsymbol{\mu} = \big(\boldsymbol{\mu}_1', \dots \boldsymbol{\mu}_m'\big)'$. Then*

$$
\boldsymbol{\mu} = \big(\boldsymbol{I}_{nm} - \boldsymbol{\Phi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\big)^{-1}\boldsymbol{c}, \tag{A.28}
$$

where $\boldsymbol{c} = \big(\boldsymbol{c}_1', \dots, \boldsymbol{c}_m'\big)'$, $\boldsymbol{\Phi} = \mathrm{bdiag}_{i=1}^{m}(\boldsymbol{\Phi}_i)$, 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

$$
\mathrm{E}[\boldsymbol{y}_t | \boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}, \tag{A.29}
$$

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

$$
\mathrm{E}[\tilde{\boldsymbol{y}}_{t+h} | \boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{A.30}
$$

with $\tilde{\boldsymbol{y}}_{t+h}$ and $\tilde{\boldsymbol{\Phi}}$ defined in proposition 1 and $\tilde{\boldsymbol{G}}_{\boldsymbol{y}}$ as in (7)*.*

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 1 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 eqs.* (1) *and* (2)*, and assume that 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 $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \dots \boldsymbol{\mu}_m')'$ and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \dots \boldsymbol{\nu}_m')'$. Then*

$$\boldsymbol{\nu} = \big(\boldsymbol{I}_{n^2m} - \boldsymbol{\Upsilon}(\boldsymbol{Q}' \otimes \boldsymbol{I}_{n^2})\big)^{-1}\big(\boldsymbol{\gamma} + \boldsymbol{\omega} + \boldsymbol{\Psi}(\boldsymbol{Q}' \otimes \boldsymbol{I}_n)\boldsymbol{\mu}\big), \tag{A.31}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_m')'$ *with* $\boldsymbol{\gamma}_j = \boldsymbol{c}_j \otimes \boldsymbol{c}_j$, $\boldsymbol{\omega} = (\boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_m')'$ *with* $\boldsymbol{\omega}_j = \mathrm{vec}(\boldsymbol{\Sigma}_j)$, $\boldsymbol{\Upsilon} = \mathrm{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{\Phi}_j)$, $\boldsymbol{\Psi} = \mathrm{bdiag}_{j=1}^m(\boldsymbol{\Phi}_j \otimes \boldsymbol{c}_j + \boldsymbol{c}_j \otimes \boldsymbol{\Phi}_j)$, $\boldsymbol{Q}$ *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*

$$\mathrm{E}[\boldsymbol{z}_t|\boldsymbol{\xi}_t] = (\boldsymbol{\xi}_t' \otimes \boldsymbol{I}_{n^2})\boldsymbol{\nu}. \tag{A.32}$$

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*

$$\mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{\Upsilon}}^h \begin{pmatrix} \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_{n^2}\big)\boldsymbol{\nu} \\ \big(\mathrm{diag}(\boldsymbol{\xi}_t) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu} \\ \boldsymbol{\xi}_t \end{pmatrix}, \tag{A.33}$$

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

Proof. Use eq. (11) 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 1 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 eqs.* (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*

$$\mathrm{E}[\boldsymbol{z}_{t+h+k,t+h}|\boldsymbol{\xi}_t] = \tilde{\boldsymbol{H}}_{\boldsymbol{z}} \begin{pmatrix} \tilde{\boldsymbol{\Phi}}^k \otimes \boldsymbol{I}_n & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P}^k \end{pmatrix} \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|\boldsymbol{\xi}_t], \tag{A.34}$$

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 eq. (19) in the last equality. □

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

$$\mathrm{vec}\big(\mathrm{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t]\big) = (\boldsymbol{\iota}_m' \otimes \boldsymbol{I}_{n^2})\big(\boldsymbol{C}\boldsymbol{P}^k\big(\mathrm{diag}(\boldsymbol{\xi}_t) - \boldsymbol{\xi}_t\boldsymbol{\xi}_t'\big) \otimes \boldsymbol{I}_n\big)\boldsymbol{\mu}. \tag{A.37}$$

Proof. The derivation of eq. (37) follows as

$$
\begin{aligned}
&\operatorname{vec}\big(\operatorname{Cov}[\boldsymbol{y}_t, \boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t]\big) \\
&= \mathrm{E}[\boldsymbol{y}_{t+k} \otimes \boldsymbol{y}_t|\boldsymbol{\xi}_t] - \mathrm{E}[\boldsymbol{y}_{t+k}|\boldsymbol{\xi}_t] \otimes \mathrm{E}[\boldsymbol{y}_t|\boldsymbol{\xi}_t] \\
&= \boldsymbol{H}_{\boldsymbol{z}}\,\mathrm{E}[\boldsymbol{z}_{t+k,t}|\boldsymbol{\xi}_t] - \boldsymbol{G}_y\,\mathrm{E}[\boldsymbol{y}^*_{t+k}|\boldsymbol{\xi}_t] \otimes \boldsymbol{G}_y\,\mathrm{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 responses 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 eqs.* (1) *and* (2)*, and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{y}}_t$ be defined by eq.* (38) *and the results in proposition 1. Let the vector $\boldsymbol{y}_{t-1}$ be part of I_{t-1} or calculated as* $\mathrm{E}[\boldsymbol{y}_{t-1}|I_{t-1}]$*. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of I_{t-1} or calculated as* $\mathrm{E}[\boldsymbol{s}_{t-1}|I_{t-1}]$*. 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*

$$
GI^{\boldsymbol{u}}_{\tilde{\boldsymbol{y}}}(h, \boldsymbol{u}_t, I_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \emptyset, \boldsymbol{u}_t, I_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t \end{pmatrix}. \tag{B.40}
$$

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

$$
GI^{\varepsilon_i}_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, I_{t-1}) = GI_{\tilde{\boldsymbol{y}}}(h, \varepsilon_{it}, \emptyset, I_{t-1}) = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m \end{pmatrix}. \tag{B.41}
$$

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

$$
GI^{y_i}_{\tilde{\boldsymbol{y}}}(h, \eta_{it}, I_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|y_{it}, I_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{y}}_{t+h}|I_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{\varepsilon}^*_t|y_{it}, I_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|y_{it}, I_{t-1}] \end{pmatrix}. \tag{B.42}
$$

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

$$\mathrm{E}[\boldsymbol{s}_t|I_{t-1}] = \boldsymbol{P}\boldsymbol{\xi}_{t-1},$$

$$\mathrm{E}[\boldsymbol{s}_t|y_{it}, I_{t-1}] = \frac{1}{\boldsymbol{f}'\boldsymbol{P}\boldsymbol{\xi}_{t-1}}\boldsymbol{f} \odot \boldsymbol{P}\boldsymbol{\xi}_{t-1}, ,$$

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*

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

with last term

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

and

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

Proof. Equation (40) follows from

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

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

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

Equation (41) follows from

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

In this case, the first expectation reduces to

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

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

In the derivations for eq. (42), $\mathrm{E}[\boldsymbol{s}_t|I_{t-1}]$ follows from the definition of the regime process, and $\mathrm{E}[\boldsymbol{s}_t|y_{it}, I_{t-1}]$ 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

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

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

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

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

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

□

Proposition 8. *Let $\boldsymbol{y}_t$ follow the MS-VAR process as specified in eqs.* (1) *and* (2)*, and let the Generalized Impulse Response Function for $\tilde{\boldsymbol{z}}_t$ be defined by eq.* (43) *and the results in proposition 2. Let the vector $\boldsymbol{y}_{t-1}$ be part of I_{t-1} or calculated as* $\mathrm{E}[\boldsymbol{y}_{t-1}|I_{t-1}]$*. Let the vector with regime probabilities $\boldsymbol{\xi}_{t-1}$ be part of I_{t-1} or calculated as* $\mathrm{E}[\boldsymbol{s}_{t-1}|I_{t-1}]$*. 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*

$$GI_{\tilde{\boldsymbol{z}}}^{\boldsymbol{u}}(h, \boldsymbol{u}_t, I_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \emptyset, \boldsymbol{u}_t, I_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h\begin{pmatrix}(\boldsymbol{\Gamma} + \boldsymbol{\Omega})\boldsymbol{u}_t + \boldsymbol{\Psi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) + \boldsymbol{\Upsilon}(\boldsymbol{u}_t \otimes \boldsymbol{z}_{t-1}) \\ \boldsymbol{C}\boldsymbol{u}_t + \boldsymbol{\Phi}(\boldsymbol{u}_t \otimes \boldsymbol{y}_{t-1}) \\ \boldsymbol{u}_t\end{pmatrix}. \tag{B.45}$$

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

$$GI_{\tilde{\boldsymbol{z}}}^{\varepsilon_i}(h, \varepsilon_{it}, I_{t-1}) = GI_{\tilde{\boldsymbol{z}}}(h, \varepsilon_{it}, \emptyset, I_{t-1}) = \tilde{\boldsymbol{\Upsilon}}^h\begin{pmatrix}\mathrm{E}[\boldsymbol{\zeta}_t^*|\varepsilon_{it}, I_{t-1}] \\ \varepsilon_{it}\boldsymbol{\Lambda}(\boldsymbol{P}\boldsymbol{\xi}_{t-1} \otimes \boldsymbol{e}_i) \\ \boldsymbol{0}_m\end{pmatrix}, \tag{B.46}$$

with

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

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

$$GI_{\tilde{\boldsymbol{z}}}^{y_i}(h, \eta_{it}, I_{t-1}) = \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|y_{it}, I_{t-1}] - \mathrm{E}[\tilde{\boldsymbol{z}}_{t+h}|I_{t-1}] = \tilde{\boldsymbol{\Phi}}^h \begin{pmatrix} \mathrm{E}[\boldsymbol{z}_t^*|y_{it}, I_{t-1}] - \mathrm{E}[\boldsymbol{z}_t^*|I_{t-1}] \\ \mathrm{E}[\boldsymbol{\varepsilon}_t^*|y_{it}, I_{t-1}] \\ \mathrm{E}[\boldsymbol{u}_t|y_{it}, I_{t-1}] \end{pmatrix}, \quad \text{(B.47)}$$

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

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

and

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

with

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

Proof. Equation (45) follows from

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

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

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

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

Equation (46) follows from

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

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

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

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

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

C Additional empirical results

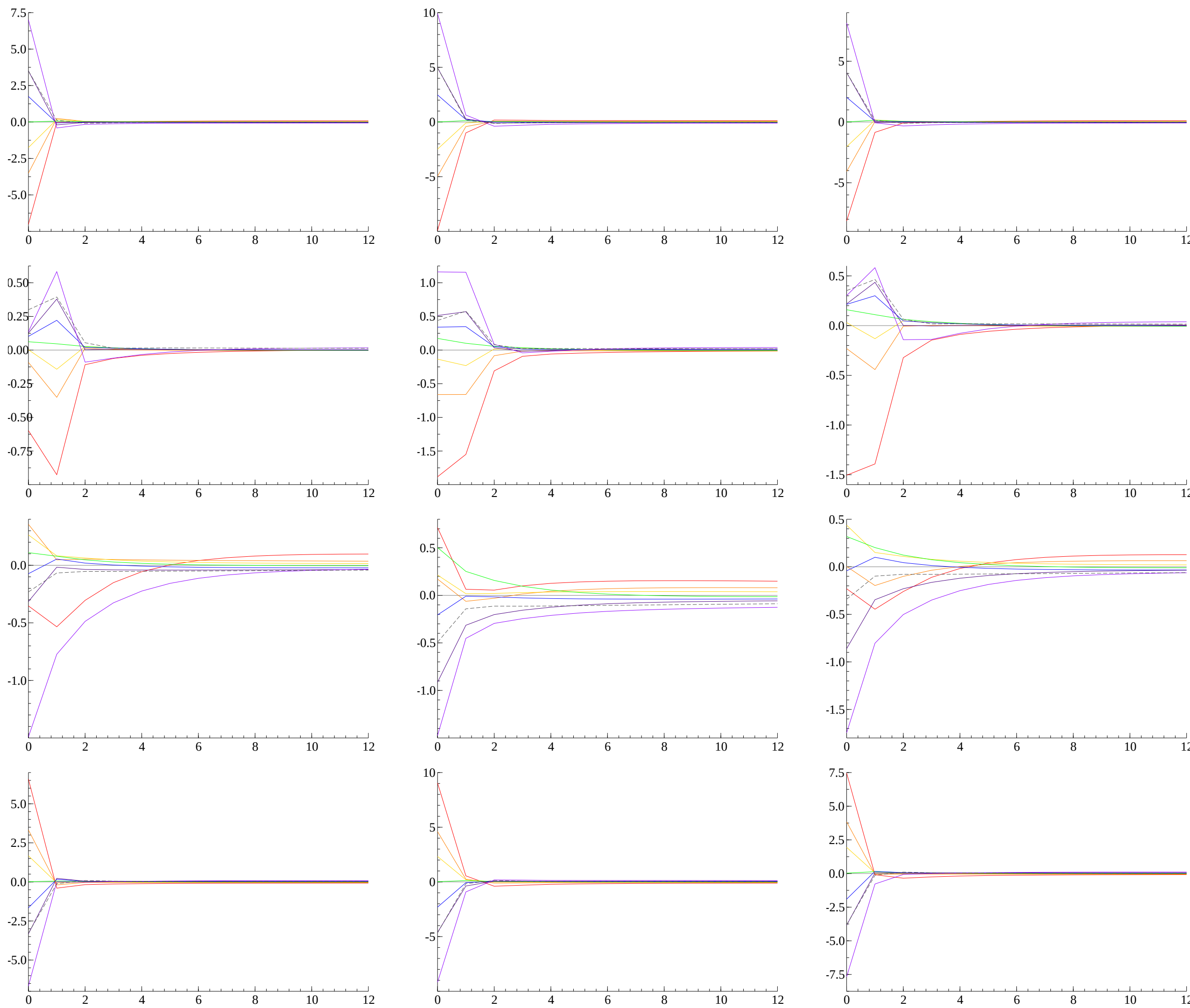
[Figure C.1 about here.]

[Figure C.2 about here.]

[Figure C.3 about here.]

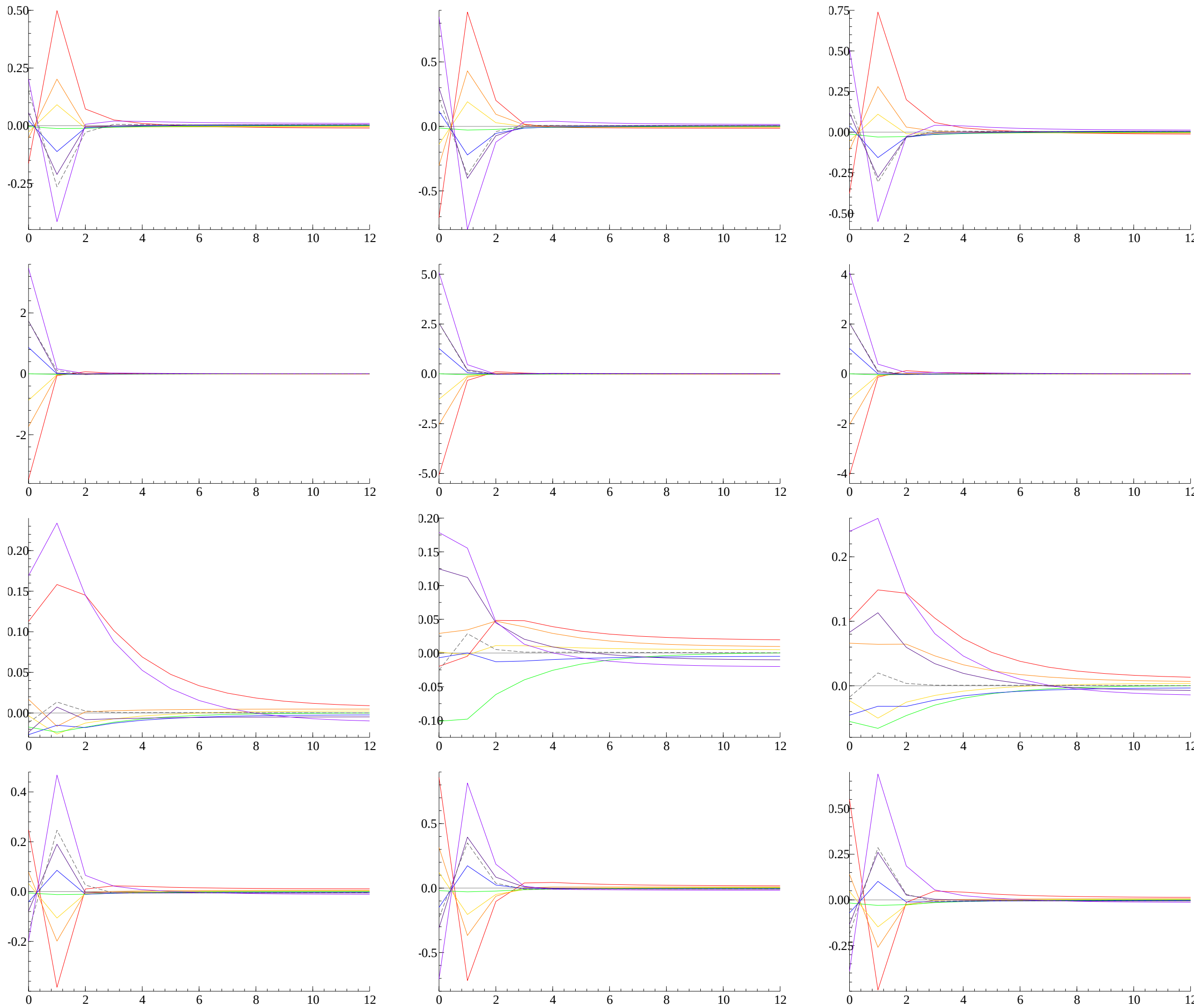
[Figure C.4 about here.]

Figure C.1: The effects of shocks on expected stock returns over time



This figure shows the impulse responses for forecast stock returns over time in the VAR(1) model and in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The colored lines show the effect of a shock of δ times the forecast standard deviation to stock returns (first row), bond returns (second row), the T-bill rate (third row), and the D/P ratio (fourth row) as in eq. (50), with δ equal to -2 (red), -1 (orange), -0.5 (yellow), 0 (green), 0.5 (blue), 1 (indigo) and 2 (violet), and the horizon from 0 through 12 months. The dashed gray lines gives the effect of a one standard deviation shock in the VAR(1) model.

Figure C.2: The effects of shocks on expected bond returns over time

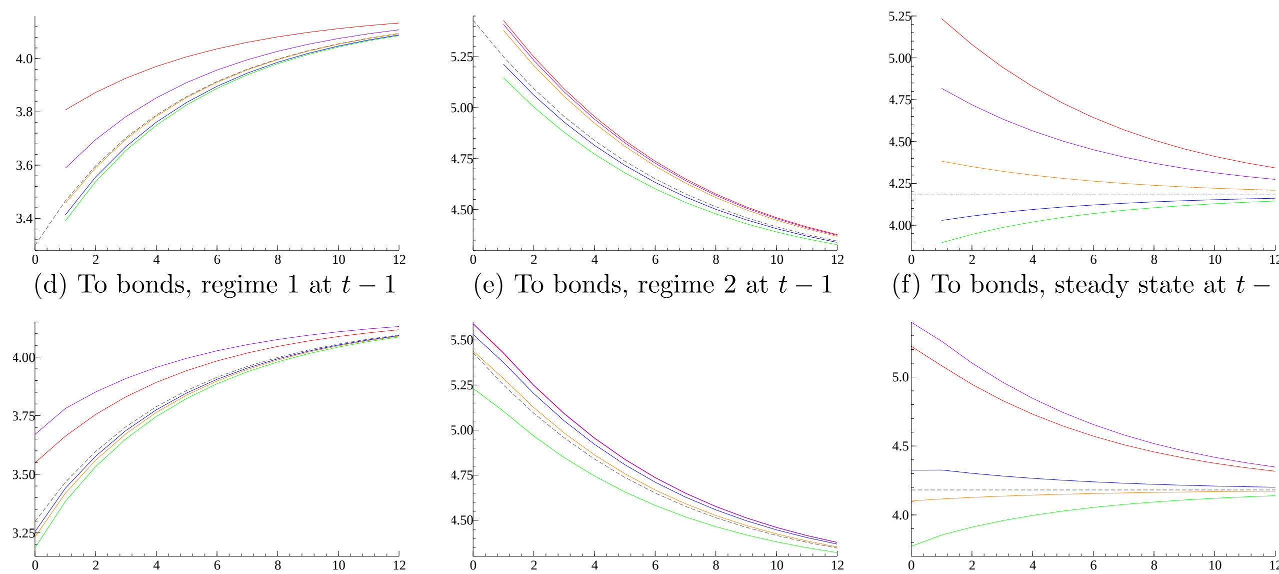


This figure shows the impulse responses for forecast bond returns over time in the VAR(1) model and in the MSIAH(2)-VAR(1) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The colored lines show the effect of a shock of δ times the forecast standard deviation to stock returns (first row), bond returns (second row), the T-bill rate (third row), and the D/P ratio (fourth row) as in eq. (50), with δ equal to -2 (red), -1 (orange), -0.5 (yellow), 0 (green), 0.5 (blue), 1 (indigo) and 2 (violet), and the horizon from 0 through 12 months. The dashed gray lines gives the effect of a one standard deviation shock in the VAR(1) model.

Figure C.3: The effects of shocks on the forecast volatility of stock returns over time in the MSIH(2) model

(a) To stocks, regime 1 at $t-1$ (b) To stocks, regime 2 at $t-1$ (c) To stocks, steady state at $t-1$

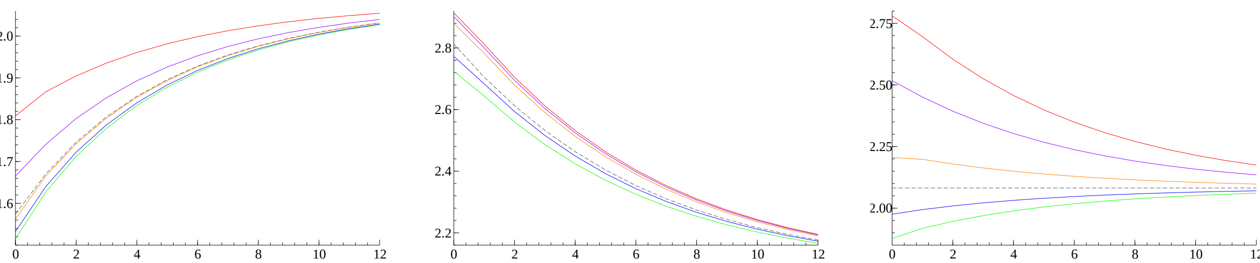
(d) To bonds, regime 1 at $t-1$ (e) To bonds, regime 2 at $t-1$ (f) To bonds, steady state at $t-1$



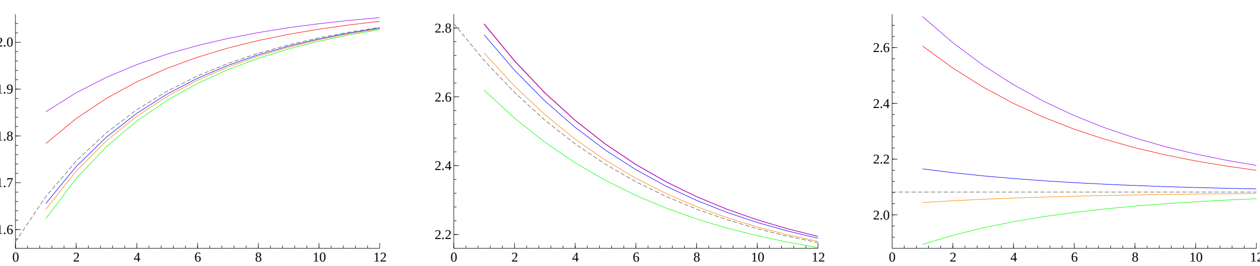
This figure shows the impulse responses for the forecast volatility of stock returns over time in the MSIH(2) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The dashed gray lines give the forecast volatility without a shock. The solid lines show the forecast volatility conditional on a shock of δ times the forecast standard deviation to stock returns (first row) and bond returns (second row), with δ equal to -2 (red line), -1 (orange line), 0 (green line), 1 (blue line), and 2 (violet line). The difference between the dashed and the solid lines gives the volatility impulse function. The forecast horizon runs from 0 through 12 months.

Figure C.4: The effects of shocks on the forecast volatility of bond returns over time in the MSIH(2) model

(a) To stocks, regime 1 at $t-1$ (b) To stocks, regime 2 at $t-1$ (c) To stocks, steady state at $t-1$



(d) To bonds, regime 1 at $t-1$ (e) To bonds, regime 2 at $t-1$ (f) To bonds, steady state at $t-1$



This figure shows the impulse responses for the forecast volatility of bond returns over time in the MSIH(2) model, conditional on the regime process being in regime 1, 2 or the steady state at time $t-1$. The dashed gray lines give the forecast volatility without a shock. The solid lines show the forecast volatility conditional on a shock of δ times the forecast standard deviation to stock returns (first row) and bond returns (second row), with δ equal to -2 (red line), -1 (orange line), 0 (green line), 1 (blue line), and 2 (violet line). The difference between the dashed and the solid lines gives the volatility impulse function. The forecast horizon runs from 0 through 12 months.