Markov-switching VAR models with high-dimensional transition probabilities

George Kapetanios* Stylianos Zlatanos†

February 2026

Abstract

This paper develops a penalised maximum-likelihood estimator for Markov-switching vector autoregressive (VAR) models that allows transition probabilities to depend on a high-dimensional set of predictors. By applying ℓ_1 (Lasso) regularisation to the transition probability coefficients, the approach performs variable selection and aids parameter estimation. We implement a modified Expectation-Maximisation (EM) algorithm that accommodates latent regimes while solving a convex, penalised multinomial logit problem for transition coefficients at each maximisation step. Monte Carlo experiments show that the estimator recovers sparse transition structures as sample size increases, tends to select parsimonious models in moderate samples, and remains effective for regime inference in high-dimensional settings. An empirical application to Growth-at-Risk with 129 macro-financial predictors on US data, demonstrates the estimator's ability to select distinct sets of predictors governing entry into and exit from growth vulnerability regimes.

Keywords: Markov-switching VAR, High-dimensional data, Lasso, Penalised maximum likelihood, EM algorithm

JEL classification: C13, C15, C34, C52, E37

*King's Business School, King's College London.

†King's Business School, King's College London. Email: stylianos.zlatanos@kcl.ac.uk. For helpful comments and suggestions at different stages of this project, we thank Ilias Chronopoulos and Daniele Massacci.

1 Introduction

Markov-switching models have found widespread applications in economics and finance, where many time series exhibit shifts in behaviour or structural relationships over time. In these models, the data-generating process (DGP) is assumed to switch among a finite set of regimes, each corresponding to a distinct configuration of parameters or "mode" of behaviour, such as periods of high volatility or economic contraction.

Regimes are unobserved and inferred from the data, with their evolution governed by a stochastic Markov chain. Thus, regimes can be both temporary and recurring. The seminal work of Hamilton (1989), who used a Markov-switching specification to analyse the U.S. business cycle, has inspired a large body of subsequent research. The switching mechanism can, in principle, affect any subset of model parameters, rendering them state-dependent. Krolzig (1997) extended Hamilton's univariate framework to a multivariate setting based on vector autoregressions (VARs), thereby allowing for regime-switching dynamics and interactions across multiple variables.

In the traditional Markov-switching framework (e.g. Hamilton, 1989), the Markov chain is homogeneous, so transitions occur with probabilities that are constant over time. Diebold, Lee and Weinbach (1994) and Filardo (1994) relaxed this assumption by allowing the transition probabilities to vary over time (TVTP), typically by linking them to observable covariates. Introducing TVTP increases flexibility and lets economic fundamentals (or other predictors) drive the switching mechanism.

We extend the TVTP framework by allowing transition probabilities to depend on a high-dimensional set of predictors and by developing a penalised maximum-likelihood estimator for a Markov-switching VAR with a multinomial logit specification and an ℓ_1 (Lasso) penalty. This approach is particularly useful when the researcher lacks strong prior information about which predictors matter, leading to a large candidate set and curse-of-dimensionality concerns. The Lasso penalty provides a data-driven way to perform variable selection and parameter estimation simultaneously.

This approach offers a distinct alternative to dimensionality reduction. While factor models effectively summarise common variability, they rely on the implicit assumption that idiosyncratic information averages out. However, theoretical work on the "granular hypothesis" has challenged this view in the context of aggregate fluctuations, showing that idiosyncratic shocks to large entities can drive economy-wide volatility (Gabaix, 2011).

If a similar logic applies to regime dynamics, structural shifts may be driven by specific, identifiable leading indicators rather than diffuse common factors. In such settings, factor aggregation inherently obscures the distinct predictive signals. To address this, we develop a penalised estimator designed to identify these drivers from a

high-dimensional set, offering a robust alternative for environments where the switching mechanism is sparse.

Our estimator is characterised by three key features. First, it performs *simultaneous variable selection and estimation*, allowing researchers to remain agnostic about the identity of the relevant predictors. Second, it enhances *interpretability* by isolating the specific leading indicators associated with regime shifts. Third, it achieves *computational scalability* through efficient convex optimisation, making high-dimensional analysis feasible. This approach moves beyond merely predicting regime changes to detecting the sparsity structure of the transition mechanism. By recovering the set of variables active in the transition probability, the estimator highlights empirical candidates for regime determinants, which in applications can assist in evaluating competing explanations for structural change.

To estimate the model, we develop a modified Expectation-Maximisation (EM) algorithm that accommodates both the latent states and the penalised transition probability block. The algorithm retains closed-form updates for the VAR parameters and initial state probabilities, while solving a convex, penalised multinomial logit problem for the transition coefficients at each maximisation step. We calibrate the penalty parameter via an expanding-window cross-validation scheme designed to respect the time-series structure.

Monte Carlo experiments illustrate the finite-sample trade-offs inherent in the estimator. We find that the method successfully recovers sparse transition probability structures as the sample size increases. At intermediate sample sizes, cross-validation prioritises model parsimony—a conservative behaviour that safeguards against overfitting. Crucially, the results show that the estimator maintains robust regime inference even when exact variable selection is challenging due to high dimensionality.

We complement the simulation evidence with an empirical application to Growth-at-Risk on US data, building on the regime-switching framework of Caldara, Cascaldi-Garcia, Cuba-Borda and Loria (2021). Using 129 standardised macro-financial indicators from the FRED-MD database as transition predictors, the estimator selects distinct sets of variables for entry into and exit from a growth vulnerability regime, demonstrating the practical interpretability of the approach.

The remainder of the paper is organised as follows. Section 2 places our contribution within the related literature. Section 3 introduces the Markov-switching VAR framework, which is then estimated in Section 4 using our proposed penalised maximum-likelihood estimator and modified EM algorithm. Sections 5 and 6 present the design and results of the Monte Carlo experiments. Section 7 applies the estimator to US Growth-at-Risk data with 129 macro-financial predictors. Section 8 concludes.

2 Related literature

The Markov-switching (MS) model of Hamilton (1989) marked a paradigm shift in time-series econometrics. By allowing model parameters to follow an unobserved, recurrent, discrete-state Markov process, it captured the pronounced asymmetries and nonlinearities that characterise business cycles and many financial series.[1] Building on the univariate framework, researchers extended MS models to multivariate settings. Kim (1994) generalised the MS specification to a state-space form,[2] while Krolzig (1997) introduced the Markov-switching vector autoregressive (MS-VAR) model. This multivariate extension enabled analysis of state-dependent interactions across variables and addressed questions such as regime shifts in U.S. monetary policy (Sims & Zha, 2006).

The assumption of *homogeneous* (i.e., fixed over time) transition probabilities was relaxed by Filardo (1994) and Diebold et al. (1994), who modelled the switching mechanism as a logistic function of observables, giving rise to time-varying transition probability (TVTP) models. This framework, where the probability of a regime shift depends on economic fundamentals, has found application in areas such as interest-rate dynamics (Gray, 1996) and business-cycle analysis (Ravn & Sola, 1999). While these advances enriched the specification of the switching mechanism, they were typically confined to a small, pre-selected set of covariates.

A distinct but related line of research explores the *endogeneity* of the switching process itself. Most of the literature assumes that the latent state variable is exogenous with respect to the model's disturbances. Kim, Piger and Startz (2008) relax this assumption, developing a probit-based model where the latent variable driving regime changes can be correlated with the structural shocks of the system. Their approach allows for direct tests of endogeneity and provides a framework for cases where unobserved shocks are likely to influence both the outcome variables and the probability of a regime shift. An alternative philosophical approach is found in score-driven models (Bazzi, Blasques, Koopman & Lucas, 2017), where transition probabilities evolve endogenously based on the model's own predictive score, obviating the need for external predictors. We maintain the standard assumption of exogenous predictors but focus on the distinct methodological challenge of selecting them from a large candidate set.

When the set of predictors is large relative to the sample size, the transition-logit parameter space grows faster than available information and standard estimators fail. The literature has pursued two broad strategies to address this problem: dimensionality reduction and variable selection via penalisation. A standard approach to handling

[1] For an accessible survey of regime-switching models, see Hamilton (2018).

[2] See also Kim and Halbert (1999) for a textbook treatment of state-space models with regime switching.

high-dimensional information in this setting is the Markov-switching dynamic factor model (Chauvet, 1998; Diebold & Rudebusch, 1996; Kim, 1994). Recent contributions have extended this framework to large-dimensional settings; for instance, Barigozzi and Massacci (2025) develop an approximate factor model with regime-switching loadings estimated via PCA and EM.

By summarising a large panel of data into a few latent common factors, these models effectively reduce dimensionality. However, this aggregation risks obscuring the specific sources of variation. When a latent factor drives a regime shift, it is often unclear which specific underlying variables are responsible. Our sparsity-based approach contrasts with the view that macroeconomic predictability is inherently "dense"—driven by diffuse signals across many indicators (Giannone, Lenza & Primiceri, 2021). However, while factor aggregation may be optimal for capturing general comovement, we postulate that regime transitions—discrete shifts in the data-generating process—are often triggered by specific, identifiable fundamentals rather than diffuse movements. Our estimator explicitly searches for this sparsity in the non-linear transmission mechanism, distinguishing it from methods that impose density by construction.

An alternative strategy is to perform variable selection directly on the high-dimensional set of potential transition covariates. Penalisation methods, rooted in the foundational work on the Lasso (Tibshirani, 1996), perform variable selection and estimation simultaneously. While the Lasso was originally developed for i.i.d. data (Tibshirani, 1996), its application to dynamic models is grounded in recent theoretical advances. Kock and Callot (2015) established oracle inequalities for high-dimensional vector autoregressions, proving that penalised estimators can recover the correct sparsity pattern even in dependent time-series settings. Leveraging these theoretical foundations, applied work has begun to introduce penalisation into MS-VARs, yet their use has been concentrated on inducing sparsity in the regime-specific VAR parameters (Maung, 2021; Monbet & Ailliot, 2017). The crucial task of applying a penalty to select predictors for the *transition probabilities* themselves has, however, proven to be a significant methodological challenge. Related work in the Bayesian literature has achieved shrinkage using stochastic search variable selection priors, often within factor-augmented models that use a probit specification for the transition mechanism (Huber & Fischer, 2018).

Our work complements this literature by establishing a frequentist framework for high-dimensional regime switching. Unlike factor models, which aggregate information, our penalised estimator preserves the identity of individual predictors, enhancing interpretability. Moreover, by framing the problem as penalised maximum likelihood rather than Bayesian posterior simulation, we avoid the computational burden associated with stochastic search. This allows us to leverage efficient convex optimisation for the trans-

ition block, offering a computationally tractable alternative for high-dimensional applications.

3 Econometric framework

We consider a vector of n variables, $\boldsymbol{y}_t$, that follows a reduced-form VAR model of order p, whose parameters are subject to regime switching. The switching results from a latent first-order Markov chain S_t that can be in one of m regimes. We formulate the model as:

$$\boldsymbol{y}_t = \boldsymbol{c}_{S_t} + \sum_{j=1}^{p} \boldsymbol{\Phi}_j(S_t)\, \boldsymbol{y}_{t-j} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{S_t}) \tag{1}$$

where $\boldsymbol{c}_{S_t}$ is an n-vector of regime-specific intercepts, $\boldsymbol{\Phi}_j(S_t)$ is an $n \times n$ matrix of regime-specific autoregressive coefficients at lag j, and $\boldsymbol{u}_t$ are i.i.d. reduced-form innovations with regime-dependent covariance matrix $\boldsymbol{\Sigma}_{S_t}$. We denote the parameters in regime i as $\boldsymbol{c}_i$, $\boldsymbol{\Phi}_j(i)$, and $\boldsymbol{\Sigma}_i$.

State variable The variable S_t is assumed to be an ergodic first-order Markov chain taking values in $S = \{1, \ldots, m\}$. The state at time $t+1$ depends only on the state at time t (Markov property).

Transition probabilities Let $Z_t \in \mathbb{R}^k$ denote the k-dimensional vector of observed transition predictors, and write the augmented vector with intercept as $\boldsymbol{Z}_t = (1, Z_t')'$. The transition probabilities are time-varying and depend on $\boldsymbol{Z}_t$. We let current predictors drive next-period transitions. Specifically, the probability of moving from state j at time t to state i at time $t+1$ is

$$p_{ji,t} \equiv \Pr(S_{t+1} = i \mid S_t = j,\, \boldsymbol{Z}_t;\, \mathbf{W}) = \frac{\exp(\boldsymbol{Z}_t' \mathbf{w}_{ji})}{\sum_{h=1}^{m} \exp(\boldsymbol{Z}_t' \mathbf{w}_{jh})}\,. \tag{2}$$

We utilise this multinomial logit specification rather than the probit structure common in Bayesian applications (e.g., Huber and Fischer, 2018). While probit models facilitate data augmentation in MCMC samplers, the logit loss function is convex, making it computationally ideal for high-dimensional penalised maximum likelihood estimation.

The corresponding coefficients are $\mathbf{w}_{ji} = (w_{0,ji}, w_{1,ji}, \ldots, w_{k,ji})'$, where $w_{0,ji}$ is the intercept. For each state j, the probabilities must sum to one. For identification, the coefficients for one destination state are set to zero. Throughout, we normalise $\mathbf{w}_{j1} = \mathbf{0}$ for all $j = 1, \ldots, m$. We treat $\boldsymbol{Z}_t$ as exogenous with respect to the VAR, i.e., taken as

given when forming transition probabilities and not contemporaneously affected by the variables entering the VAR.

The probability of the initial state is given by $\pi_i = \Pr(S_{p+1} = i)$, with $\sum_{i=1}^{m} \pi_i = 1$. This vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)'$ is treated as a set of parameters to be estimated.

Throughout, we assume that for each regime i the VAR is stable, that is all eigenvalues of the associated companion matrix lie strictly inside the unit circle. We further assume that the predictor process $\{\boldsymbol{Z}_t\}$ is stationary and ergodic, and that the Markov chain $\{S_t\}$ is irreducible and aperiodic. These conditions ensure existence of moments and ergodicity required for estimation.[3]

3.1 High-dimensional transition probabilities

In the framework above, the full parameter vector lives in a block-structured space that partitions into (i) regime-specific VAR coefficients, (ii) regime-specific covariance parameters, and (iii) transition-probability coefficients governed by a multinomial logit specification. The size of this parameter space grows with (m, n, p) through the VAR and covariance blocks, but only the transition block expands with the number of predictors k. Intuitively, each state contributes destination-specific logit coefficients, so adding predictors inflates the transition layer linearly in k while leaving the VAR and covariance layers unchanged. This imbalance is the essence of our high-dimensional setting: even with moderate m and n, a large candidate set of transition predictors can expand the parameter space and strain unpenalised likelihood-based estimation.

The total parameter count is defined as:

$$K \equiv m(n^2p + n) + m\,\frac{n(n+1)}{2} + m(m-1)(k+1) + (m-1)$$

where the terms correspond to (i) VAR parameters, (ii) covariances, (iii) transition probabilities, and (iv) initial state probabilities. Here, m is the number of regimes, n is the VAR dimension, p is the lag order, and k (TVTP dimension) is the number of transition predictors. When k is large, the transition component $m(m-1)(k+1)$ can dominate the parameter space. This setting corresponds to the sparse design (Experiment A) analysed in Section 5, where $k = 30$ and $m = 2$.

We consider settings where the number of predictors, k, is large relative to the information content of the data. Crucially, the effective sample size for estimating transition probabilities is not the full time series T, but the number of periods spent in a specific origin regime. While this equals T/m in a perfectly balanced sample, macroeconomic regimes are often persistent and asymmetric, meaning the effective observations

[3] Note that for TVTP models, ergodicity generally holds if the transition probabilities are bounded away from 0 and 1, which is satisfied here given bounded predictors.

for a given state can be significantly fewer. Consequently, a moderate predictor set can easily exceed the available degrees of freedom. For instance, in a design with $T = 100$ and $k = 30$, the parameter count (31) may approach or surpass the effective number of transitions from the rarer regime. In such cases, unpenalised maximum likelihood may become ill-posed and highly unstable, leading to perfect separation or diverging coefficient estimates.

We address this scarcity with a penalised maximum-likelihood estimator that adds an ℓ_1 penalty on the transition probability coefficients, inducing sparsity and enabling estimation when k is large. In the multinomial logit specification for transition probabilities, we penalise only slope coefficients. Intercepts remain unpenalised to capture baseline transition rates. For identification, we normalise one destination state per current state by setting $\mathbf{w}_{j1} = \mathbf{0}$ for all j.

We employ an element-wise Lasso penalty. While a Group Lasso approach (Yuan & Lin, 2006) could enforce that a predictor enters the transition probability for all destination states simultaneously, we prefer the flexibility of the element-wise specification, which allows a variable to drive transitions into specific regimes without necessarily affecting others.

4 The penalised maximum likelihood estimator

Notation Before proceeding with the model estimation, it is useful to define the full parameter vector, $\boldsymbol{\theta}$. We partition $\boldsymbol{\theta}$ into components for the VAR dynamics, the error covariances, and the Markov-switching process: $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_{\text{VAR}}^\top, \boldsymbol{\theta}_\Sigma^\top, \boldsymbol{\theta}_\pi^\top)^\top$. The first component, $\boldsymbol{\theta}_{\text{VAR}}$, contains all autoregressive coefficients and intercepts, constructed by stacking the vectorised parameters for each state $s = 1, \ldots, m$. The second, $\boldsymbol{\theta}_\Sigma$, contains the unique elements of the covariance matrices, formed by stacking the vech of each $\boldsymbol{\Sigma}(s)$. The final component, $\boldsymbol{\theta}_\pi$, governs the transition dynamics and is itself partitioned into $\boldsymbol{\theta}_\pi = (\boldsymbol{\theta}_W^\top, \boldsymbol{\pi}^\top)^\top$, where $\boldsymbol{\theta}_W$ stacks all estimable transition coefficients $\mathbf{w}_{ji}$ and $\boldsymbol{\pi}$ is the vector of initial-state probabilities.

The likelihood function In a Markov-switching framework, the goal is to maximise the incomplete-data likelihood (also known as the observed-data likelihood), which is a function of the observed data and model parameters, but not the latent variables. In the context of the MS-VAR measurement equation (1), the incomplete-data likelihood equals the sum, over all possible state sequences from S_{p+1} to S_T, of the joint probability of the observed data and that state sequence.

The exact form of the likelihood function is given by:

$$\mathcal{L}(\mathbf{Y}_T \mid \boldsymbol{\theta}) = \sum_{i_{p+1},\ldots,i_T=1}^{m} \pi_{i_{p+1}} \prod_{t=p+2}^{T} p_{i_{t-1},i_t,t} \prod_{t=p+1}^{T} g(\mathbf{y}_t \mid \mathcal{F}_{t-1};\boldsymbol{\theta}_{i_t}), \tag{3}$$

where $\mathbf{Y}_T = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$ denotes the stacked sample (sequence of observed vectors) and $g(\mathbf{y}_t \mid \mathcal{F}_{t-1};\boldsymbol{\theta}_{S_t})$ is the multivariate normal density of $\mathbf{y}_t$ given the past information set $\mathcal{F}_{t-1}$ and the state-specific parameters $\boldsymbol{\theta}_{S_t}$. (Here $\mathcal{F}_{t-1} = \sigma(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_1)$; for a VAR($p$), $g(\mathbf{y}_t \mid \mathcal{F}_{t-1}) = g(\mathbf{y}_t \mid \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p})$, so conditioning on the last p lags is sufficient.) Under the timing convention $p_{ji,t} = g_{ji}(\boldsymbol{Z}_t)$, current predictors $\boldsymbol{Z}_t$ drive transitions into period $t+1$.

All likelihood statements are made *conditional* on the observed path of predictors $\{\boldsymbol{Z}_t\}_{t=1}^T$. We treat $\boldsymbol{Z}_t$ as predetermined/exogenous with respect to the VAR innovations (as stated above), so the model does not specify a joint law for $(\mathbf{Y}_T, \boldsymbol{Z}_T)$. Estimation therefore targets the conditional incomplete-data likelihood given the realised $\boldsymbol{Z}$ sequence.

The ML estimator Let $\mathcal{L}(\mathbf{Y}_T \mid \boldsymbol{\theta})$ denote the incomplete-data likelihood in (3). Our penalised maximum-likelihood estimator solves

$$\max_{\boldsymbol{\theta}} \ \log \mathcal{L}(\mathbf{Y}_T \mid \boldsymbol{\theta}) \ - \ \lambda \sum_{j=1}^{m} \sum_{i=2}^{m} \|\mathbf{w}_{ji,1k}\|_1, \tag{4}$$

where the regularisation term is a Lasso penalty on the transition probability coefficients from (2). The slope subvector is $\mathbf{w}_{ji,1k} = (w_{1,ji}, \ldots, w_{k,ji})'$, which excludes the intercept, and $\|\cdot\|_1$ denotes the ℓ_1-norm. The penalty applies only to slope coefficients. Intercepts are left unpenalised. Identification is ensured separately by the base-category normalisation $\mathbf{w}_{j1} = \mathbf{0}$ for each current state j.[4]

Directly maximising (4) is difficult because the latent state sequence is unobserved. We therefore use the expectation–maximisation (EM) algorithm, which optimises the same penalised objective indirectly by iterating between (i) taking expectations of the complete-data log-likelihood with respect to the latent states and (ii) maximising this expected criterion (minus the penalty) given current parameters.

4.1 A modified EM algorithm

Direct maximisation of (4) is computationally challenging due to the latent states S_t and the high-dimensional penalty. We therefore implement the EM algorithm of Dempster, Laird and Rubin (1977), which is well suited to Markov-switching models (Hamilton,

[4] For brevity, we write pen($\boldsymbol{\theta}$) for this penalty henceforth.

1990). For a comprehensive exposition (derivations, smoothing recursions, and update rules), see Appendix A.1.

Define the penalised objective as

$$\mathcal{L}_{\text{pen}}(\boldsymbol{\theta}) \equiv \log \mathcal{L}(\mathbf{Y}_T \mid \boldsymbol{\theta}) - \text{pen}(\boldsymbol{\theta}),$$

which is exactly the criterion in (4).

The algorithm consists of two steps that are repeated iteratively. In the expectation (E) step, the expectation of the complete-data log-likelihood is calculated given the current parameter estimates. Next, in the maximisation (M) step, this expectation minus the penalty is maximised to update the parameters. The main innovation of the modified algorithm lies in the M-step for the transition probabilities, where we embed a penalised solver. This design computationally isolates the high-dimensional optimisation task, preserving the closed-form simplicity of the VAR and initial-state updates. While the M-step updates are performed separately, the blocks remain statistically coupled through the smoothed probabilities γ_t computed in the E-step. Consequently, the computational burden of the variable-selection step depends primarily on the dimension of the transition predictors and is independent of the VAR system size. However, errors in variable selection can propagate to the VAR estimates via these smoothed probabilities. For each of 50 random starting parameter vectors $\boldsymbol{\theta}^{(0)}$, the first iteration begins with an E-step, and we monitor convergence using the penalised log-likelihood criterion

$$\frac{|\mathcal{L}_{\text{pen}}^{(r)} - \mathcal{L}_{\text{pen}}^{(r-1)}|}{1 + |\mathcal{L}_{\text{pen}}^{(r-1)}|} < 10^{-8}. \tag{5}$$

For a given sequence of observations $\mathbf{Y}_T$ and states $\mathbf{S}_T = (S_{p+1}, \ldots, S_T)$, the complete-data log-likelihood, conditional on the first p observations and parameters $\boldsymbol{\theta}$, can be decomposed as follows:

$$\begin{aligned} \ell(\mathbf{S}_T, \mathbf{Y}_T; \boldsymbol{\theta}) &= \log \left[\pi_{S_{p+1}} \left(\prod_{t=p+2}^{T} p_{S_{t-1}, S_t, t} \right) \left(\prod_{t=p+1}^{T} g(\mathbf{y}_t \mid \mathcal{F}_{t-1}; \boldsymbol{\theta}_{S_t}) \right) \right] \\ &= \log \pi_{S_{p+1}} + \sum_{t=p+2}^{T} \log p_{S_{t-1}, S_t, t} + \sum_{t=p+1}^{T} \log g(\mathbf{y}_t \mid \mathcal{F}_{t-1}; \boldsymbol{\theta}_{S_t}) \end{aligned} \tag{6}$$

This form separates terms that depend on the VAR parameters, the transition probabilities, and the initial state probabilities.

E-step Since we do not observe the latent process S_t, we work with the conditional expectation of the complete-data log-likelihood. For penalised likelihood, the EM algorithm involves iteratively maximising an auxiliary function, Q, which combines this

expectation with the penalty term (Green, 1990). At iteration (r) of the algorithm, the Q-function is defined as:

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(r-1)}) = E_{\mathbf{S}_T}\left[\ell(\mathbf{S}_T, \mathbf{Y}_T; \boldsymbol{\theta}) \mid \mathbf{Y}_T, \boldsymbol{\theta}^{(r-1)}\right] - \text{pen}(\boldsymbol{\theta}), \tag{7}$$

where $\text{pen}(\boldsymbol{\theta}) = \lambda \sum_{j=1}^{m} \sum_{i=2}^{m} \|\mathbf{w}_{ji,1k}\|_1$ is the Lasso penalty from equation (4). The expectation is taken with respect to the conditional distribution of the latent states $\mathbf{S}_T$, given the observed data $\mathbf{Y}_T$ and the parameter estimates from the previous iteration, $\boldsymbol{\theta}^{(r-1)}$.

Substituting the decomposed log-likelihood from (6) into (7), we get:

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(r-1)}) = & \sum_{i=1}^{m} \Pr(S_{p+1} = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}) \log \pi_i \\ & + \sum_{t=p+2}^{T} \sum_{j=1}^{m} \sum_{i=1}^{m} \Pr(S_t = i, S_{t-1} = j \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}) \log p_{ji,t} \\ & + \sum_{t=p+1}^{T} \sum_{i=1}^{m} \Pr(S_t = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}) \log g(\mathbf{y}_t \mid \mathcal{F}_{t-1}; \boldsymbol{\theta}_i) \\ & - \lambda \sum_{j=1}^{m} \sum_{i=2}^{m} \|\mathbf{w}_{ji,1k}\|_1. \end{aligned} \tag{8}$$

The first term involves the initial state probabilities, weighted by the smoothed probability of being in each state i at time $p+1$. The second term involves the transition probabilities, weighted by the smoothed probability of transitioning from state j to state i at each time t. The third term involves the VAR parameters, weighted by the smoothed probability of being in each state i at each time t.

The smoothed state probabilities, $\Pr(S_t = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)})$, and the joint smoothed probabilities, $\Pr(S_t = i, S_{t-1} = j \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)})$, required in (8) are calculated using the smoothing algorithm of Kim (1994). See Appendix A.1 for implementation details.

M-step In the M-step, we find the parameter values $\boldsymbol{\theta}^{(r)}$ that maximise the Q-function defined in the E-step:

$$\boldsymbol{\theta}^{(r)} = \underset{\boldsymbol{\theta}}{\text{argmax}}\ Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(r-1)}). \tag{9}$$

The decomposition of the Q-function allows us to maximise the terms for each set of parameters separately.

Initial state probabilities

The update for the initial state probabilities $\boldsymbol{\pi}$ is found by maximising the first term

of (8). The solution for each state $i = 1, \ldots, m$ is simply the smoothed probability of being in that state at the beginning of the sample:

$$\hat{\pi}_i^{(r)} = \Pr(S_{p+1} = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}). \tag{10}$$

Transition probability parameters

Let $Z_t \in \mathbb{R}^k$ denote the vector of observed transition predictors that enter the TVTP logits (excluding the intercept, so k is the number of non-intercept predictors). For each current state j, the transition coefficients $\mathbf{W}_j$ are updated by solving the ℓ_1-penalised multinomial logit problem

$$\hat{\mathbf{W}}_j^{(r)} = \underset{\mathbf{W}_j}{\operatorname{argmax}} \left\{ \sum_{t=p+2}^{T} \sum_{i=1}^{m} \Pr(S_t = i, S_{t-1} = j \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}) \, \log p_{ji,t} \; - \; \lambda \sum_{i=2}^{m} \|\mathbf{w}_{ji,1\!\!1_k}\|_1 \right\}, \tag{11}$$

subject to the multinomial logit structure in (2) and the identification restriction $\mathbf{w}_{j1} = \mathbf{0}$. The objective is concave in $\mathbf{W}_j$ (logit likelihood) and the ℓ_1 term is convex, so the problem is a convex program with no closed-form solution. We solve it numerically. See Appendix A.1, eq. (17), for derivation and implementation details.

VAR parameters and covariances

For each state $i = 1, \ldots, m$, the VAR parameters $\boldsymbol{\theta}_i = (\boldsymbol{c}_i, \boldsymbol{\Phi}_1(i), \ldots, \boldsymbol{\Phi}_p(i), \boldsymbol{\Sigma}_i)$ are updated by maximising the third term of (8):

$$\hat{\boldsymbol{\theta}}_i^{(r)} = \underset{\boldsymbol{\theta}_i}{\operatorname{argmax}} \sum_{t=p+1}^{T} \Pr(S_t = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}) \log g(\mathbf{y}_t \mid \mathcal{F}_{t-1}; \boldsymbol{\theta}_i). \tag{12}$$

Let $\mathbf{X}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ be the $(np+1) \times 1$ vector of regressors and let $\mathbf{B}_i = (\boldsymbol{c}_i, \boldsymbol{\Phi}_1(i), \ldots, \boldsymbol{\Phi}_p(i))'$ be the full $(np+1) \times n$ coefficient matrix for state i.

Equivalently, $\mathbf{B}'_i = [\, \boldsymbol{c}_i \;\; \boldsymbol{\Phi}_1(i) \;\; \cdots \;\; \boldsymbol{\Phi}_p(i) \,]$ so that the measurement equation can be written as

$$\mathbf{y}_t = \mathbf{B}'_i \mathbf{X}_t + \mathbf{u}_t,$$

with $\mathbf{X}_t \in \mathbb{R}^{np+1}$, $\mathbf{B}_i \in \mathbb{R}^{(np+1) \times n}$, and $\mathbf{y}_t, \mathbf{u}_t \in \mathbb{R}^n$.

This corresponds to a weighted maximum-likelihood problem, which admits a closed-form weighted least squares (WLS) solution. Denote $\gamma_t(i) = \Pr(S_t = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)})$. The WLS updates are:

$$\hat{\mathbf{B}}_i^{(r)} = \left(\sum_{t=p+1}^{T} \gamma_t(i) \mathbf{X}_t \mathbf{X}'_t \right)^{-1} \left(\sum_{t=p+1}^{T} \gamma_t(i) \mathbf{X}_t \mathbf{y}'_t \right), \tag{13}$$

and the updated covariance matrix is the weighted sample covariance of the residuals $\hat{\mathbf{u}}_{i,t} = \mathbf{y}_t - \hat{\mathbf{B}}_i^{(r)\prime}\mathbf{X}_t$:

$$\hat{\boldsymbol{\Sigma}}_i^{(r)} = \frac{\sum_{t=p+1}^{T} \gamma_t(i)\hat{\mathbf{u}}_{i,t}\hat{\mathbf{u}}_{i,t}'}{\sum_{t=p+1}^{T} \gamma_t(i)}. \tag{14}$$

In summary, the EM algorithm proceeds as follows:

(i) compute filtered and smoothed state probabilities, $\Pr(S_t = i \mid \mathbf{Y}_T)$, and joint probabilities, $\Pr(S_t = i,\, S_{t-1} = j \mid \mathbf{Y}_T)$.

(ii) update the initial-state probabilities $\boldsymbol{\pi}$ using the smoothed distribution at $t = p + 1$.

(iii) update the transition-probability coefficients by solving a penalised multinomial logit for each current state.

(iv) update the VAR parameters via weighted least squares.

These updates constitute a Generalised EM (GEM) algorithm (Green, 1990). While the VAR and initial-state updates (ii, iv) exactly maximise their respective components of the auxiliary function, the transition probability update (iii) involves a numerical solver. Although this subproblem is convex, the use of a numerical optimizer means we may not achieve the exact maximum in finite time, but we are guaranteed to increase the objective. Consequently, the full M-step increases the penalised Q-function, which by standard GEM results implies that the penalised incomplete-data likelihood is non-decreasing across iterations. We monitor the convergence criterion in (5) and, if necessary, repeat an M-step when the observed increase falls below tolerance.

4.2 Selecting the penalty parameter

We tune the regularisation parameter λ via blocked, expanding-window cross-validation (CV), scoring each validation block by the average negative log predictive density with parameters frozen at the training fit.

The penalty applies only to the slope coefficients in the transition-probability (TVTP) block, while intercepts remain unpenalised to capture baseline transition rates. The selection rule is therefore formulated for this mixed-penalty specification.

In brief, for each candidate λ on a pre-scaled grid we estimate the model on an expanding training window and evaluate out-of-sample loss on the subsequent validation block—the average negative log predictive density—using the parameters fixed at the training fit. We then compute, for each λ, a cross-validated score equal to the average loss across validation blocks, with each block weighted by its length. We first identify

the λ that minimises this score. When several values yield scores that are statistically indistinguishable within sampling uncertainty, we choose the least-penalised λ among them. Unlike the standard "one-standard-error" rule which favours sparsity (larger λ), we choose the least penalised value to avoid collapsing to an intercept-only model when the signal is weak. The formal selection criterion appears below.

Indistinguishability often arises because intercepts are unpenalised: for large λ the transition block collapses to an intercept-only specification. When predictors add little signal relative to noise, the CV profile can become nearly flat or largely decreasing in λ, so a pure minimiser may select a boundary solution (excessive sparsity).

This rule explicitly accounts for statistically indistinguishable cases by preferring the least-penalised λ within the set of scores that are not meaningfully worse than the minimum. This safeguard does not alter the cross-validation procedure itself: it leaves the scoring and minimisation unchanged and only governs the tie-break among statistically indistinguishable candidates. In Appendix A.2, we report robustness checks showing that, across Monte Carlo replications, the selected λ distributions are practically indistinguishable with or without this rule (see Figure A1). This confirms that our results are driven by the cross-validation criterion itself rather than the specific tie-breaking mechanism.

Fold construction Crucially, standard random K-fold cross-validation is invalid in this context as it disrupts the serial dependence structure (Burman, Chow & Nolan, 1994). We instead adopt an expanding-window scheme that preserves the temporal evolution of state probabilities. This ensures that the validation set reflects the continuity of regimes inherent in the data-generating process.

The sample is partitioned into a sequence of training windows T_r and subsequent, contiguous validation blocks V_r. Each fold proceeds as follows: (i) select a training prefix T_r (the forecast origin is the last observation in T_r); (ii) estimate the model on T_r; (iii) score the next block V_r without updating parameters within the block; and (iv) advance the origin to define T_{r+1} and repeat.

Preprocessing is performed using T_r only and applied unchanged within V_r: in particular, we standardise the columns of the predictor matrix $\boldsymbol{Z}$ within T_r (demean and divide by the sample standard deviation) before fitting, and we do not re-standardise inside V_r. For scoring and interpretation on the raw scale, the fitted TVTP coefficients are back-transformed to the original units and applied to raw predictors in validation. This ensures the ℓ_1 penalty treats predictors comparably (scale invariance) while preserving temporal ordering. Validation blocks are contiguous in time and may differ in length. By design, every block strictly follows its training window. In the Monte Carlo simulations, we use five validation blocks by default and set the initial training window

to 50% of the sample (at least $p+1$ observations). The remaining observations are split as evenly as possible across blocks, with the first few blocks receiving one extra observation if necessary.

CV loss We evaluate predictive performance out of sample using expanding training windows followed by contiguous validation blocks. Let these validation blocks be V_r and assign each a weight $w_r = |V_r|/\sum_s |V_s|$ proportional to its length so that longer blocks contribute in proportion to the information they contain. For a given λ, the loss at time $t \in V_r$ is the negative log predictive density, computed with parameters fixed at the fit from the corresponding training window, $\ell_t(\widehat{\boldsymbol{\theta}}_r(\lambda))$. Formally,

$$\ell_t(\widehat{\boldsymbol{\theta}}_r(\lambda)) \;=\; -\log\left(\sum_{i=1}^{m} \alpha_{t|t-1}^{(r,\lambda)}(i)\; g(\mathbf{y}_t \mid \mathcal{F}_{t-1};\, \widehat{\boldsymbol{\theta}}_{r,i}(\lambda))\right),$$

where $\alpha_{t|t-1}^{(r,\lambda)}(i)$ are the one-step-ahead state probabilities obtained by the forward filter under $\widehat{\boldsymbol{\theta}}_r(\lambda)$ (with transitions driven by the observed predictors), and $g(\cdot \mid \mathcal{F}_{t-1}; \widehat{\boldsymbol{\theta}}_{r,i}(\lambda))$ is the state-i Gaussian density implied by the MS-VAR. The transition block affects the mixture weights $\alpha_{t|t-1}^{(r,\lambda)}(i)$ via the fitted multinomial logit in the predictors $\boldsymbol{Z}$, while the VAR block determines the state-conditional densities $g(\cdot)$.[5] At the start of a validation block V_r with first index t_0, the recursion is initialised from the end of the training filter: we take the last filtered probabilities on T_r, $\alpha_{t_0-1|t_0-1}$, and form the one-step prediction $\alpha_{t_0|t_0-1} = \mathbf{P}_{t_0}^{\top}\, \alpha_{t_0-1|t_0-1}$. If filtered probabilities are unavailable, we use a uniform prior over states. We do not smooth using future validation data. The forward filter runs through V_r using only information in $\mathcal{F}_{t-1}$. At validation time t, the one-step weights $\alpha_{t|t-1}$ are computed using predictors dated $t-1$, i.e., transitions into t depend on $\boldsymbol{Z}_{t-1}$, consistent with $p_{ji,t-1} = g(\boldsymbol{Z}_{t-1})$.

We summarise performance on a block by the average loss

$$\overline{\ell}_r(\lambda) \;=\; |V_r|^{-1} \sum_{t \in V_r} \ell_t(\widehat{\boldsymbol{\theta}}_r(\lambda)),$$

and aggregate across blocks to obtain the cross-validated score

$$\widehat{\mathrm{CV}}(\lambda) = \sum_r w_r\, \overline{\ell}_r(\lambda)\,.$$

Lower values of $\widehat{\mathrm{CV}}(\lambda)$ indicate better one-step-ahead predictive fit. The minimising value, $\lambda_{\min}$, is the baseline CV choice before any tie-break is considered. Formally,

$$\lambda_{\min} \;\in\; \arg\min_{\lambda}\, \widehat{\mathrm{CV}}(\lambda),$$

[5] When predictors are standardised within the training window, $\boldsymbol{Z}_{\mathrm{std}} = (\boldsymbol{Z} - \boldsymbol{\mu})/\boldsymbol{\sigma}$ (elementwise), the fitted logit index $\eta = w_0^{\mathrm{std}} + \boldsymbol{w}^{\mathrm{std}\prime}\boldsymbol{Z}_{\mathrm{std}}$ can be expressed on the raw scale as $\eta = w_0^{\mathrm{raw}} + \boldsymbol{w}^{\mathrm{raw}\prime}\boldsymbol{Z}$ with $w_j^{\mathrm{raw}} = w_j^{\mathrm{std}}/\sigma_j$ and $w_0^{\mathrm{raw}} = w_0^{\mathrm{std}} - \sum_j (\mu_j/\sigma_j)\, w_j^{\mathrm{std}}$. We use this back-transformation so that scoring and reporting are on the raw predictor scale while estimation remains scale-invariant.

where Λ denotes the pre-scaled penalty grid.

The δ-band rule To decide whether nearby penalties are statistically indistinguishable from the CV minimum, we employ a tolerance band approach. This strategy adapts the canonical "one-standard-error rule" (Hastie, Tibshirani & Friedman, 2009) by allowing for a tunable tolerance parameter c. We compare each candidate to $\lambda_{\min}$ within each validation block. This paired comparison controls for block-specific difficulty and uses exactly the same validation observations for both candidates. Let $\lambda_{\min}$ denote the minimiser defined above, and define the per-fold differences as

$$d_r(\lambda) = \overline{\ell}_r(\lambda) - \overline{\ell}_r(\lambda_{\min}).$$

Interpret $d_r(\lambda)$ as the excess out-of-sample loss of λ relative to $\lambda_{\min}$ on block r (positive values are worse). We aggregate these per-fold differences with the same block weights to obtain the overall contrast

$$\Delta(\lambda) = \sum_r w_r \, d_r(\lambda), \qquad \widehat{\mathrm{SE}}_\Delta(\lambda) = \sqrt{\frac{\sum_r w_r \left(d_r(\lambda) - \Delta(\lambda)\right)^2}{1 - \sum_r w_r^2}} \sqrt{\sum_r w_r^2}\,.$$

This estimator treats the fold-wise differences as the primary objects, producing a standard error for $\Delta(\lambda)$ (the weighted excess loss) rather than for the level. Using block means respects within-block serial dependence and, together with the weights, provides a stable estimate of uncertainty even when validation blocks have different lengths. We then compare $\Delta(\lambda)$ to a tolerance band based on $\widehat{\mathrm{SE}}_\Delta(\lambda)$: if λ is not meaningfully worse than $\lambda_{\min}$, we prefer the least-penalised such λ. Here, R denotes the number of validation blocks. Let $c \in (0, 1)$ denote a small, pre-specified equivalence margin. Define the tolerance band $\delta(\lambda) = c \cdot \widehat{\mathrm{SE}}_\Delta(\lambda)$, and take as the selected penalty the least-penalised value whose contrast lies within this band:

$$\widehat{\lambda}_\delta \;=\; \min\Big\{\lambda \in \Lambda :\; \Delta(\lambda) \le \delta(\lambda)\Big\}. \tag{15}$$

Equivalently, the rule selects the smallest λ with a standardised contrast

$$T(\lambda) = \frac{\Delta(\lambda)}{\widehat{\mathrm{SE}}_\Delta(\lambda)} \le c.$$

This approach departs from the boundary only when the increase in out-of-sample loss is smaller than a conservative fraction of its sampling uncertainty. Choosing the *smallest* λ within the band guards against over-penalisation when the CV curve is flat. If no candidate satisfies the inequality, the choice reverts to $\lambda_{\min}$.

Penalty grid In all results reported for the simulations, the penalty is evaluated on a short grid scaled at a standard high-dimensional rate, $\lambda = \kappa\sqrt{\frac{\log k}{T}}$ (with k the number of non-intercept transition predictors), with $\kappa \in \{0.01, 0.06, 0.32, 0.56, 1.00\}$. This coarse grid is chosen for computational efficiency in Monte Carlo experiments. In applications, a finer grid (or local refinement around the minimum) can be used without changing the procedure.

5 Monte Carlo experiments

In a series of Monte Carlo experiments, we study the finite-sample behaviour of our penalised maximum-likelihood estimator for MS-VARs with time-varying transition probabilities. To keep the focus on the transition-probability component, the data-generating process holds the VAR dynamics fixed and targets average regime durations, while we systematically vary the sample size (T), the number of transition predictors (k), and the degree of sparsity in the transition block. The regularisation strength is selected by expanding-window cross-validation, and we summarise performance using variable-selection measures, block-specific RMSEs, and probability scoring rules for one-step-ahead regime probabilities (QPS and LPS).

5.1 Data-generating process

Model We work with a two-state MS-VAR of order $p = 1$ for $n = 4$ variables. Let S_t denote the regime at time t. Conditional on being in regime $i \in \{1, 2\}$, the process evolves according to

$$\mathbf{y}_t = \boldsymbol{c}_i + \Phi\,\mathbf{y}_{t-1} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i).$$

The regime intercepts are fixed across designs at $\boldsymbol{c}_1 = -\mathbf{1}_4$ and $\boldsymbol{c}_2 = 0.5\,\mathbf{1}_4$, implying a higher mean in state 2. Both regimes share the same autoregressive matrix, $\Phi = 0.5\,\mathbf{I}_4$, which we hold constant across replications. The innovation covariances are $\boldsymbol{\Sigma}_1 = \mathrm{diag}(3.0, \ldots, 3.0)$ and $\boldsymbol{\Sigma}_2 = \mathrm{diag}(0.75, \ldots, 0.75)$, so state 2 is the lower-volatility regime. This asymmetry mirrors findings in the Growth-at-Risk literature that emphasise downside risk in low-volatility expansions (e.g., Adrian, Boyarchenko and Giannone, 2019; Caldara et al., 2021).

Transition probabilities The regime evolves according to time-varying transition probabilities specified in Eq. (2), using predictors Z_t. In the two-state case we adopt

the standard normalisation $\mathbf{w}_{j1} = \mathbf{0}$, so that

$$p_{j2,t} = \frac{\exp(\boldsymbol{Z}_t' \mathbf{w}_{j2})}{1 + \exp(\boldsymbol{Z}_t' \mathbf{w}_{j2})}, \qquad j \in \{1, 2\},$$

where $\boldsymbol{Z}_t = (1, Z_t')$ includes a leading one (intercept). The sparsity pattern is imposed on the slope elements of $\mathbf{w}_{j2}$ (the intercept is unpenalised) and is set by the experiment (Section 5.2). To control average persistence, we calibrate the intercept components of $\mathbf{w}_{12}$ and $\mathbf{w}_{21}$ so that the expectation over the stationary distribution of Z_t yields $E_{Z_t}[p_{12,t}] = E_{Z_t}[p_{21,t}] = 1/12$. The calibrated transition-probability parameters are held fixed within each design across replications and sample sizes T, while realised durations vary with the draw and with T.

Predictors Let $Z_t \in \mathbb{R}^k$ denote the k-dimensional vector of transition predictors (excluding the intercept). When forming the transition probabilities in Eq. (2), we augment with a leading one so that $\boldsymbol{Z}_t = (1, Z_t')$. Each element of the predictor vector Z_t is generated as a stationary AR(1) process with persistence $\rho = 0.5$. We partition the k predictors into approximately five equal-sized blocks (block size $= \lceil k/5 \rceil$). Within each block, contemporaneous correlations are set to $\rho_b = 0.5$ and across blocks to 0, yielding a block-diagonal correlation matrix. This structure mimics the clustering often observed in macroeconomic datasets and tests the estimator's ability to identify active predictors in the presence of correlation, a known challenge for Lasso-type methods. Predictors are left on their raw scale within the DGP.

For signal sizing, we draw the slope subvector of the transition-probability coefficients once per experiment from $[-1.5, -0.5] \cup [0.5, 1.5]$ and then rescale it so that, for each origin state j, $\mathrm{Var}(Z_t' \mathbf{w}_{j2,1:k}) \in [3.0, 6.0]$. This prevents degenerate logits while matching the intended dispersion.

5.2 Experimental design

We consider sample sizes $T \in \{100, 200, 500\}$ and predictor dimensions in the transition probabilities $k \in \{10, 20, 30\}$, where k denotes the number of non-intercept transition predictors. For each (T, k) configuration we perform 500 replications.

Experiment A: Sparse TVTP This experiment studies selection, estimation, and probability scoring when transition probabilities are sparse. For each origin state, 10% of the k predictors are active in the corresponding transition logit. Active indices and signs are drawn independently by state and are fixed across replications within each (T, k). All remaining design elements follow Section 5.1: VAR parameters and covariances are held fixed, the predictor correlation structure uses a block-diagonal

correlation with $\rho_b = 0.5$, active coefficients are sized as described under Predictors, and intercepts are calibrated by duration targeting.

5.3 Evaluation metrics

Variable selection and sparsity We report the following measures:

- *True model included:* share of replications in which all truly active predictors are included in the selected model. It equals 1 only when every true predictor is selected. It does not penalise extra (false-positive) selections.
- *Selected variables:* $\frac{1}{R}\sum_{r=1}^{R}\|\widehat{W}^{(r)}\|_0$, where R is the number of replications. This captures average model size (sparsity). Lower values indicate more parsimonious selections. Higher values suggest over-selection.
- *Recall on W:* Recall $= \dfrac{\#\{\text{true positives in } \widehat{W}\}}{\#\{\text{nonzeros in } W^\star\}}$. This captures sensitivity to true signals (fraction of active coefficients recovered). It does not penalise inclusion of irrelevant predictors.

All counts are computed on slope coefficients only (intercepts are unpenalised) and exclude the normalised base destination state ($i = 1$).

Estimation accuracy Let θ stack all model parameters and let component blocks be VAR intercept/lag matrices, covariance parameters Σ, and TVTP parameters W. Define

$$\text{RMSE}_\theta \;=\; \sqrt{\frac{1}{d_\theta}\sum_{j=1}^{d_\theta}(\widehat{\theta}_j - \theta_j)^2}, \qquad \text{RMSE}_{\text{VAR}},\ \text{RMSE}_\Sigma,\ \text{RMSE}_W$$

defined analogously on the corresponding subvectors, with RMSE_W computed on W (slopes/intercepts), not on implied probabilities.

Probability scoring For the one-step-ahead probability $p_t = \Pr(S_t = 2 \mid \mathcal{F}_{t-1})$ and indicator $y_t = \mathbf{1}\{S_t = 2\}$, we report:

$$\text{QPS} \;=\; \frac{100}{T}\sum_{t=1}^{T}(p_t - y_t)^2, \qquad \text{LPS} \;=\; -\frac{100}{T}\sum_{t=1}^{T}\Big[\, y_t \log p_t + (1 - y_t)\log(1 - p_t)\,\Big].$$

Both are strictly proper. Lower is better. (QPS is the Brier score scaled by 100, and LPS is the negative average log score.)

5.4 Simulation design and implementation

For each configuration (T, k) and experiment, we run 500 replications. Predictors Z_t are generated from their stationary distribution. Active sets are drawn once per configuration and held fixed across replications. The intercept components of $\mathbf{w}_{12}$ and $\mathbf{w}_{21}$ (and the corresponding slope vectors) are calibrated once per experiment by the duration-targeting and signal-sizing procedures in Section 5.1 and then held fixed across sample sizes T and replications. We initialise EM from 50 random starts, begin with an E-step, and declare convergence when the relative change in the penalised log-likelihood falls below 10^{-8}. Predictors are standardised before the logit updates. To prevent label switching, we align regimes ex post by the mean of the first variable and volatility before computing all metrics.

6 Results

6.1 Experiment A

Experiment A evaluates finite-sample performance under sparse time-varying transition probabilities. We examine whether the penalised EM algorithm accurately recovers state-conditional VAR parameters, correctly identifies active transition predictors, and produces well-calibrated regime probability forecasts.

Estimation accuracy Table 2a reveals that root mean squared errors (RMSE) decline with sample size across all parameter blocks. The largest absolute reductions occur in the transition-probability coefficients, while the VAR dynamics and residual covariances also improve markedly as T increases.

Two mechanisms may drive this pattern. First, as T increases, sharper regime inference from the forward–backward smoother improves the precision of the weighted moment conditions for state-conditional parameters. Second, our penalty specification $\lambda \propto \sqrt{(\log k)/T}$ imposes diminishing shrinkage on active TVTP coefficients, accelerating the reduction of bias in W relative to the unpenalised VAR and covariance estimates. This differential convergence is consistent with our design choice to separate penalty structures by parameter block: the ℓ_1 penalty can exploit sparsity in the transition layer without altering standard EM updates for the VAR and covariance parameters.

Increasing the predictor dimension k produces opposing effects on different parameter blocks. Under the maintained sparsity assumption, enlarging k raises the fraction of coefficients set exactly to zero by the DGP, which the ℓ_1 penalty efficiently exploits: RMSE_W decreases monotonically in k at each T. While the reduction in RMSE_W is

partly mechanical due to the sparsity of the true parameter vector, the key question is whether a higher-dimensional transition predictor space destabilises estimation of the VAR and covariance parameters. The simulation results indicate that it does not: RMSE_{VAR} is virtually identical across all k at $T = 200$ and $T = 500$, differing by at most 0.001 between $k = 10$ and $k = 30$, and RMSE_{Σ} shows comparably small cross-k variation. This stability suggests that the penalised estimator effectively insulates VAR and covariance estimation from the dimensionality of the transition predictor space, even when variable selection in the transition layer remains imperfect.

Penalty selection behaviour Figure 1 displays the cross-validated out-of-sample loss (negative log predictive density) as a function of the penalty parameter λ for each sample size and predictor dimension. For a given λ, we plot the mean CV loss across Monte Carlo replications; the shaded band shows ± 1 Monte Carlo standard error of that mean. Each curve's minimum within the grid is marked by a dot.

The figure reveals that penalty-selection behaviour evolves systematically with sample size. At $T = 100$, the mean loss curves decline over the entire λ grid, so heavier penalisation continues to reduce out-of-sample loss on the examined range. This suggests that, at this sample size, estimation variance dominates and the risk-minimising penalty lies at or beyond the upper end of our grid.[6] At $T = 200$, the curves remain downward sloping but flatten in the interior, generating near-ties among candidate penalties and concentrating choices on relatively large λ values. By $T = 500$, each curve exhibits a clear U-shape with an interior minimum, located at a smaller penalty than the grid-boundary choices favoured at $T = 100$ and $T = 200$.

This evolution reflects the standard bias–variance trade-off in penalised estimation. At small T, aggressive shrinkage improves predictive performance because it strongly reduces variance and the grid does not yet extend into the region where shrinkage-induced bias dominates. As T grows, variance declines and bias becomes relatively more costly, so cross-validation shifts toward weaker penalties that preserve signal in active coefficients. The transition from boundary-optimal penalties at small T to interior-optimal penalties at large T is reflected most clearly in model size: Table 2b shows that the average number of selected variables reaches its minimum at $T = 200$ for every k, then rises again at $T = 500$. Recall does not follow the same pattern uniformly across dimensions—it increases from $T = 100$ to $T = 200$ for $k = 10$ and $k = 20$, but dips for $k = 30$ before rising sharply at $T = 500$. Taken together, these results indicate an intermediate conservative phase driven primarily by stronger sparsity, with delayed

[6] We do not extend the grid to higher penalties because such values produce nearly intercept-only transition models with limited practical interest, and the computational cost of fitting heavily penalised models across many Monte Carlo replications is substantial without yielding additional insights.

recall gains in the highest-dimensional design. This behaviour accords with theory on penalised M-estimators, where optimal penalty intensity must decline with sample size to preserve consistency while maintaining variance control.

Variable selection and sparsity recovery The metric reported in the table as "True model included"—that is, the share of replications in which all true predictors are included in the selected model—peaks at $T = 500$. This criterion does not penalise false positives, so it should be read jointly with model size and recall. Even under this weaker notion of recovery, it becomes increasingly rare for the selected model to include the full true active set as the predictor dimension k grows, illustrating how demanding recovery of all active predictors remains in finite samples (Table 2b).

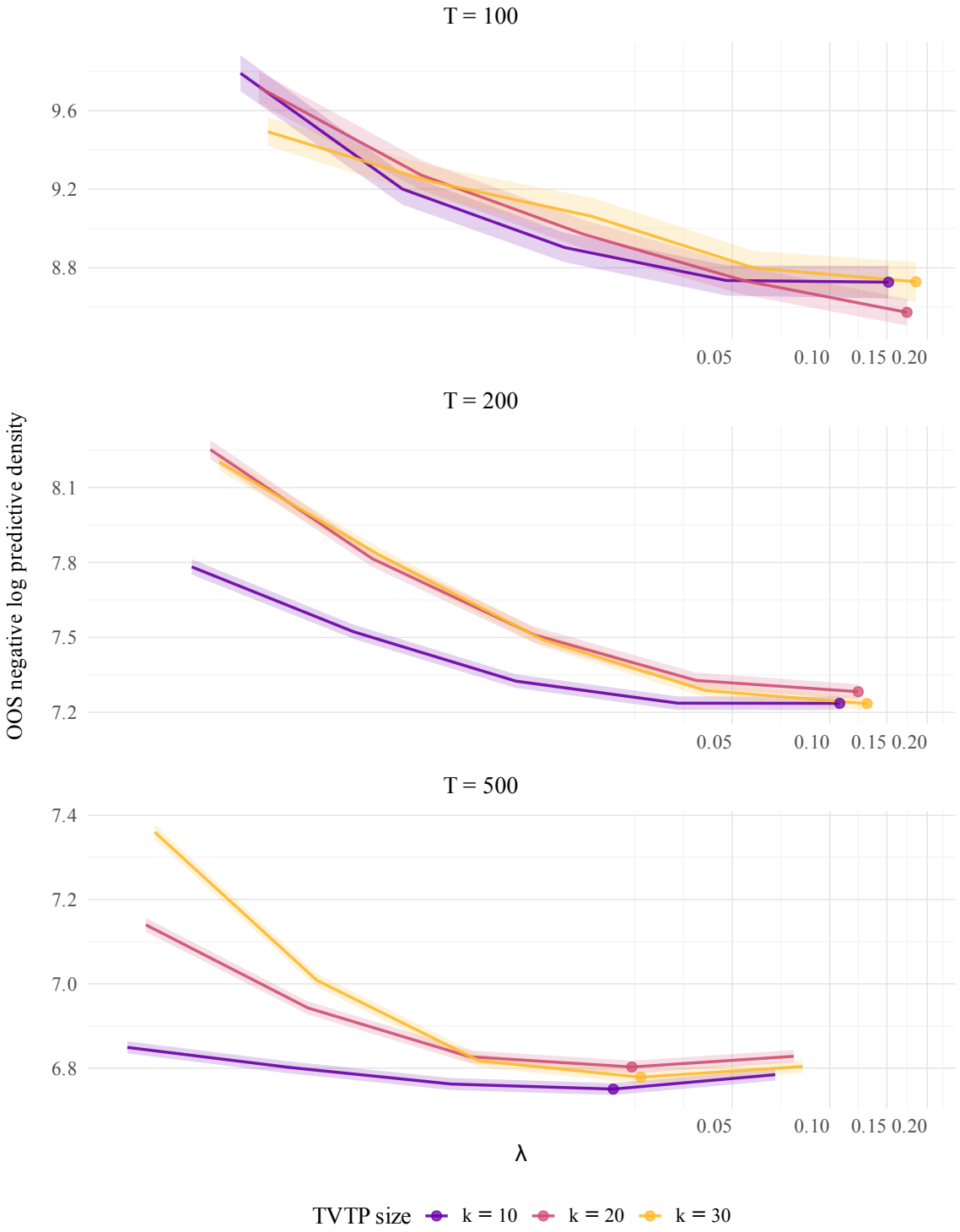
The shift from boundary solutions at small samples to interior solutions at $T = 500$ directly determines model size. Table 2b shows that average model size is smallest at $T = 200$ and larger at both endpoints, mirroring the evolution of the penalty selection described above. At $T = 200$, cross-validation concentrates at higher penalties with less dispersion, producing sparser models. At $T = 500$, weaker optimal penalties allow more coefficients to survive shrinkage, expanding the active set even as selection accuracy improves.

Recovery of the true active set—measured by recall—shows a more heterogeneous pattern than model size. For $k = 10$ and $k = 20$, recall rises from $T = 100$ to $T = 200$, albeit only modestly for $k = 20$; for $k = 30$, it dips at $T = 200$ before increasing strongly at $T = 500$. The clearest signature of the "conservative intermediate phase" is therefore the trough in model size at $T = 200$, with recall gains becoming uneven as the dimension of the predictor space increases.

At $T = 100$, estimation variance is high; the penalty selection alternates between very lean fits and noisier ones, leading to weak inclusion of all true predictors and unstable variable selection. At $T = 200$, the estimator suppresses many spurious inclusions and yields the sparsest average models in every configuration, but this conservatism can also prune true signals when the design is most demanding. That effect is mild for $k = 10$ and $k = 20$, where recall still improves, but it is visible for $k = 30$, where recall falls before recovering. Finally, at $T = 500$, the signal dominates the noise, allowing the estimator to relax the penalty, expand the active set, and achieve much stronger inclusion of all true predictors and higher recall across all k.

The aggregate metrics above summarise selection outcomes but do not reveal how consistently the estimator performs variable selection across replications. Figure 2 provides this complementary perspective by displaying the sparsity patterns in estimated transition probability coefficients for each replication, visualising the estimator's capacity to discriminate active from inactive predictors under varying sample sizes and

Figure 1: Cross-validated out-of-sample negative log predictive density by sample size and TVTP dimension.



Note: Panels correspond to $T \in \{100, 200, 500\}$. *Solid lines denote the mean across Monte Carlo replications and shaded bands show* $\pm$ *one Monte Carlo standard error of the mean at each* λ. *Dots mark the minimum of the mean curve for each TVTP size. All panels share a common logarithmic axis for the effective penalty* $\lambda = \kappa\sqrt{(\log k)/T}$.

predictor dimensions. Each panel corresponds to one (T, k) configuration and displays results for both regimes as stacked heatmaps. Within each heatmap, rows represent predictors (excluding the unpenalised intercept), columns represent Monte Carlo replications, and tile colour indicates whether the coefficient is estimated as non-zero (green) or exactly zero (blue). Red horizontal bands mark the positions of truly active predictors.

Three patterns emerge. First, active-predictor rows (marked by red bands) show increasing shares of non-zero tiles as T rises, becoming nearly solid green by $T = 500$, indicating consistent recovery of true signals. Second, inactive-predictor rows remain predominantly blue and grow cleaner with T, demonstrating effective suppression of spurious selections. Third, higher predictor dimension k introduces more scattered green tiles on inactive rows at $T = 100$ and $T = 200$—most pronounced at $k = 30$—but these largely disappear by $T = 500$.

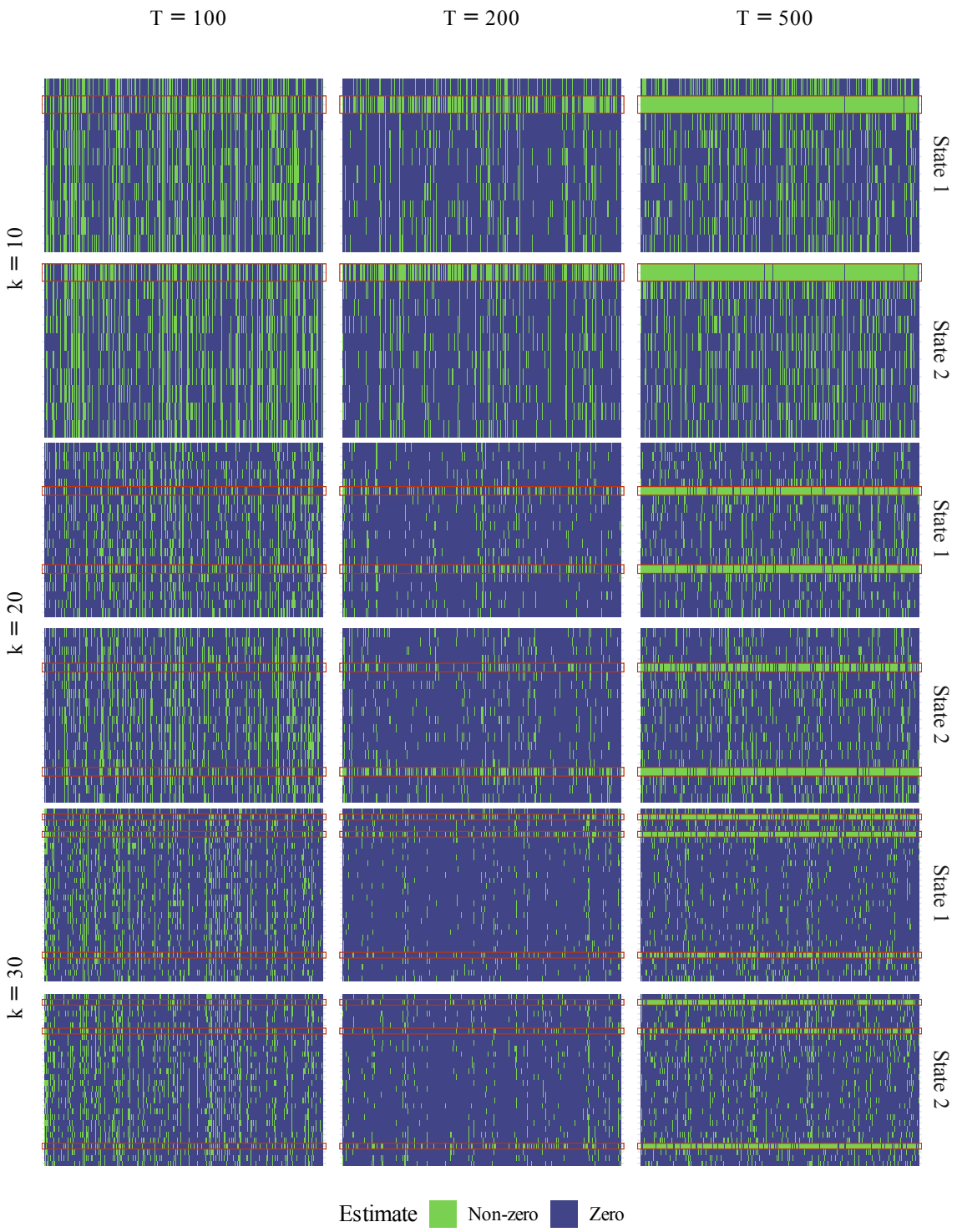
Notably, the $T = 200$ panels exhibit a visible reduction in green tiles along active rows relative to $T = 500$, and in the higher-dimensional configurations also relative to $T = 100$, mirroring the mid-sample trough in model size and the weaker active-signal recovery in those settings. This pattern reflects the tighter penalty concentration at $T = 200$ combined with improved signal detection at $T = 500$. Overall, variable-selection performance improves with sample size, overcoming the intermediate conservative phase to achieve much more reliable signal recovery and effective noise suppression by $T = 500$.

Finally, we assess whether the imposed correlation ($\rho_b = 0.5$) leads to variable swapping. While low-dimensional settings show stable selection, higher dimensions ($k = 20, 30$) initially challenge the estimator to distinguish active predictors from correlated substitutes. At $T = 100$, we observe some leakage where inactive proxies are selected. This substitutability, however, is transient: by $T = 500$, selection concentrates decisively on the true signals. Furthermore, $T = 200$ is associated with fewer non-zero estimates in the higher-dimensional designs, which is consistent with less frequent selection of correlated proxy predictors.

One-step probability scoring Having established the estimator's ability to recover the underlying transition structure, we now examine how these selection properties translate into predictive performance for the latent regimes. To assess forecast accuracy, we compute one-step-ahead quadratic probability scores (QPS) and log probability scores (LPS) for regime predictions. Table 2c reports mean scores across replications.

Forecast performance improves monotonically with sample size for all predictor dimensions. Both scores are substantially higher at $T = 100$ than at larger samples, consistent with the variable-selection results: the Lasso has insufficient information at short samples to reliably distinguish true from spurious predictors, resulting in noisy

Figure 2: TVTP sparsity patterns across configurations.



Note: Each small panel corresponds to one (T, k) cell. Within each panel, rows are transition predictors (excluding the unpenalised intercept w_1) and columns are Monte Carlo replications. Red horizontal bands mark truly active predictors. Tiles show whether the estimated coefficient is non-zero (green) or zero (blue) in each replication.

filtered probabilities. By $T = 200$, both scores drop substantially—by roughly 60–65 % relative to $T = 100$—and a further, more modest improvement of around 10–15 % is achieved at $T = 500$, as the penalty increasingly isolates the true transition signal.

Across predictor dimensions, scores at $T = 200$ and $T = 500$ are similar but not identical. In both cases, $k = 10$ yields the best average scores, while $k = 20$ and $k = 30$ are slightly worse and close to one another. The same ranking is already visible at $T = 100$, consistent with the harder selection problem when noise variables are more numerous.

Experiment A provides evidence that, in sparse TVTP designs of this type, the penalised estimator delivers good regime forecasts and reasonably small parameter RMSEs at moderate to large sample sizes, while facing the expected challenges of recovering the full set of active predictors when k is large. Estimation error declines with T across all parameter blocks, with the largest absolute improvements in the penalised transition coefficients where the ℓ_1 penalty exploits sparsity effectively. These results favour the use of the proposed procedure for density forecasting applications where interpretable, data-driven transition determinants are desired. However, applications that require near-complete recovery of the active predictors may still need supplementary post-selection diagnostics, particularly when the predictor dimension is large relative to the sample size.

7 Empirical application: Growth-at-Risk with high-dimensional predictors

To illustrate the estimator on real data, we apply it to a Growth-at-Risk (GaR) setting. GaR quantifies downside risk to future GDP growth by characterising the conditional distribution of growth outcomes. In the regime-switching approach, tail risk arises from the probability of entering or remaining in a low-growth, high-volatility state, with the predictive distribution determined by time-varying mixture weights across regimes.[7]

This application builds on the regime-switching framework of Caldara et al. (2021), who model tail risk through time-varying mixture weights driven by two estimated factors. We apply their framework to a two-regime MS-VAR with time-varying transition probabilities, replacing their low-dimensional factor-based predictors with the penalised high-dimensional transition probabilities developed in Section 4. This allows direct identification of which macro-financial indicators drive regime entry and exit.

[7] For the quantile regression approach to GaR, see Adrian et al. (2019). The regime-switching framework provides an alternative structural interpretation in which downside risk is driven by discrete shifts in the data-generating process.

Table 1: Monte Carlo results for Experiment A

	Sample size T		
k	100	200	500
RMSE_{VAR}			
10	0.325	0.143	0.080
20	0.345	0.145	0.081
30	0.334	0.143	0.081
RMSE_{Σ}			
10	0.533	0.261	0.155
20	0.565	0.263	0.157
30	0.538	0.267	0.159
RMSE_{W}			
10	1.174	0.534	0.303
20	0.878	0.432	0.297
30	0.779	0.377	0.291
RMSE_{θ}			
10	0.703	0.319	0.182
20	0.664	0.312	0.206
30	0.637	0.301	0.218

(a) RMSE metrics

	Sample size T		
k	100	200	500
True model included			
10	0.280	0.455	0.986
20	0.044	0.068	0.648
30	0.008	0.004	0.228
Selected variables			
10	5.902	3.569	5.676
20	7.901	3.386	8.146
30	10.163	3.607	9.484
Recall on W			
10	0.400	0.527	0.988
20	0.234	0.259	0.854
30	0.197	0.169	0.711

(b) Variable-selection metrics

	Sample size T		
k	100	200	500
QPS			
10	20.26	7.36	6.32
20	22.45	8.28	7.08
30	21.22	8.27	7.13
LPS			
10	68.85	26.22	22.08
20	76.75	29.39	24.59
30	74.04	29.31	24.74

(c) Probability scoring metrics

Note: Panel A reports RMSE metrics (2a), Panel B reports variable-selection metrics (2b), and Panel C reports one-step probability scoring metrics (2c) across sample sizes and transition-predictor dimensions.

7.1 Data and model specification

The VAR comprises three variables: forward GDP growth, a macro factor, and a financial factor. The growth target is the 12-month forward average of year-over-year growth in the Brave-Butters-Kelley (BBK) Monthly GDP index.[8] The macro and financial factors are the first principal components of four real activity and four financial conditions indicators, respectively.[9] Taken together, the three variables closely mirror the factor-based series estimated in Caldara et al. (2021), exhibiting similar dynamics.

The transition predictor vector $\boldsymbol{Z}_t$ comprises 129 standardised indicators from the FRED-MD database (McCracken & Ng, 2016).[10] The estimation sample runs from January 1995 to December 2022, over which we fit an MS-VAR(1) with two regimes capturing low growth-high volatility and high growth-low volatility states.

While Section 4.2 develops a general cross-validation framework for penalty selection, for this empirical application we adopt an alternative calibration approach. We set $\lambda = 6.16 \times 10^{-3}$ based on companion specifications that prioritise out-of-sample predictive density performance and probability integral transform adequacy (Rossi & Sekhposyan, 2019).[11] This choice selects 43 unique variables, with 26 active in the entry equation and 24 in the exit equation. Before examining which specific predictors drive regime transitions, we first establish that the model successfully identifies economically meaningful regimes.

7.2 Regime identification

The estimated regime parameters reveal pronounced asymmetries between the normal and vulnerable states. In the GDP growth equation, the vulnerable regime features a substantially lower conditional growth mean and innovation volatility approximately five times higher than in the normal regime. Volatility is also elevated in the financial and macro factor equations (Table B1 in Appendix B). Contemporaneous linkages between the factors and growth also differ across regimes, with stronger sensitivity to both macro and financial conditions in the vulnerable state.

[8] The growth series is detrended using a backward-looking 10-year moving average.

[9] The macro factor is constructed from ISM Manufacturing New Orders, Industrial Production, Retail Sales excluding Motor Vehicles, and Initial Jobless Claims rate. The financial factor uses the Excess Bond Premium, TED Spread, VXO, and Commercial Paper-Treasury Spread. PCA loadings are estimated on a pre-pandemic window (through December 2019) and factor scores are projected on these fixed loadings.

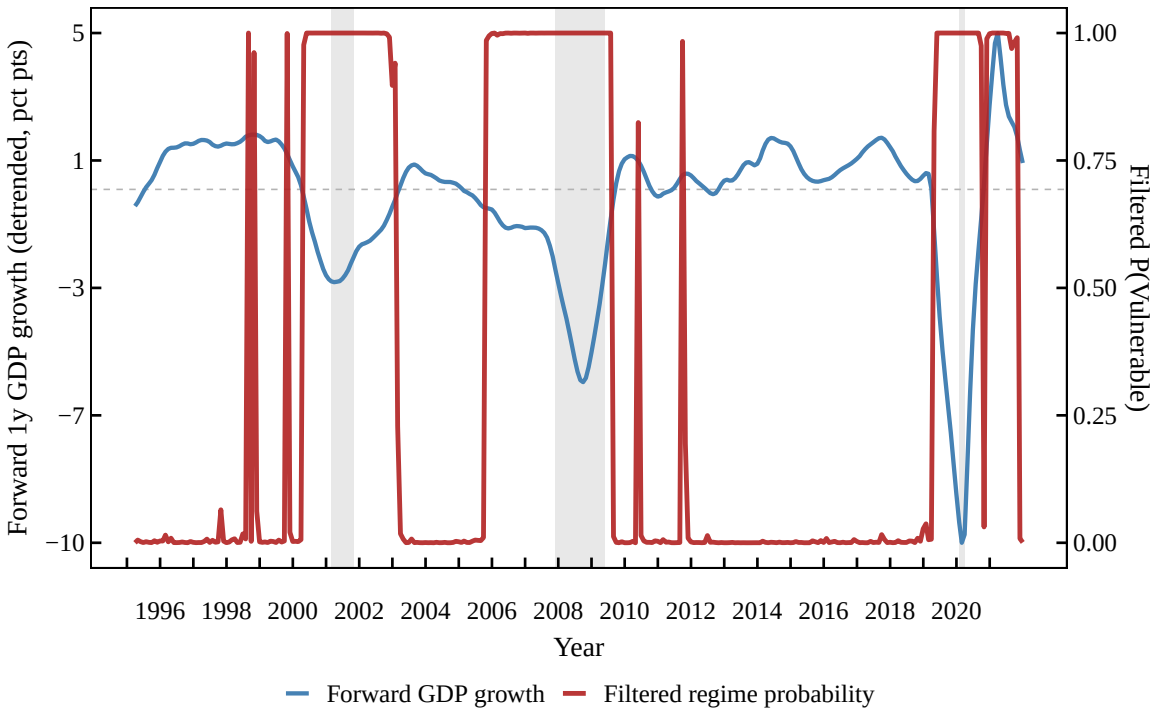
[10] The 129 predictors span eight economic categories: equity markets, yield curve and rates, real activity, labour market, housing, money and credit, prices and wages, and external indicators. See Appendix B for the full list.

[11] Full calibration details and diagnostics are documented in Appendix B.3.

Figure 3 plots the filtered vulnerable-regime probability against realised forward GDP growth. The model identifies three major vulnerable episodes, each corresponding to NBER recessions in the sample: the dot-com bust (2001–2003), the Global Financial Crisis (2008–2009), and the COVID-19 pandemic (2020–2021). Across all three episodes, the filtered probability tracks forward growth movements closely, rising sharply as growth deteriorates and declining as recovery takes hold.

The dot-com episode illustrates the distinction between NBER recession dating and the model's regime classification: the model remained in the vulnerable state until February 2003, well after the NBER recession ended in November 2001, as forward growth remained negative throughout this period. Brief spikes in the vulnerable-regime probability also occur during stress episodes that do not materialise into persistent regime changes. Spikes in 2010 and 2011—possibly reflecting European sovereign debt stress—quickly reverted as forward growth remained positive.

Figure 3: Forward GDP growth and filtered vulnerable-regime probability.



Note: The dashed horizontal line marks zero growth. Grey shaded areas denote NBER recession periods. Estimation sample: 1995:02–2021:12.

7.3 Variable selection and transition dynamics

The preferred specification selects 26 predictors for entry into the vulnerable regime and 24 for exit. Table 3 reports the corresponding non-zero coefficients in the two logit transition equations, organised by transition direction and economic category. Only seven predictors appear in both equations, indicating that entry and exit are governed by largely distinct subsets of the information set. Both transition intercepts are strongly negative (-4.29 for entry and -4.85 for exit), implying low baseline monthly switching probabilities when predictors are near their historical means.

Entry into vulnerability Several forward-looking financial indicators are among the most prominent entry predictors. The yield curve slope—measured as the 10-year Treasury minus federal funds rate spread—enters with a sizable negative coefficient, so a flatter or inverted curve raises the probability of transitioning into the vulnerable regime. This pattern is consistent with the established recession-forecasting role of term spreads (Estrella & Hardouvelis, 1991; Estrella & Mishkin, 1998; Rudebusch & Williams, 2009; Wright, 2006). Implied equity volatility (VXO) also enters positively, and the S&P P/E ratio enters negatively, indicating higher entry risk when market-implied volatility rises and equity valuations fall. These variables are commonly interpreted as forward-looking measures of uncertainty and risk appetite (Bekaert, Hoerova & Lo Duca, 2013), and have been linked to adverse macroeconomic outcomes (Bloom, 2009) and downside risk to GDP growth (Adrian et al., 2019).

Real activity and labour market indicators complete the entry equation. Industrial production measures enter with negative coefficients, indicating that weakening manufacturing activity raises entry probability. Long-duration unemployment (27+ weeks) enters positively, consistent with the view that persistent labour market slack elevates transition risk. Among prices, the PCE Price Index enters positively, potentially capturing late-cycle price pressures that precede regime switches.

Notably absent from the selected entry predictors are several credit spread measures prominent in the financial conditions literature. Neither the Excess Bond Premium (EBP) of Gilchrist and Zakrajšek (2012) nor the TED spread survives Lasso selection. This reflects the conditional nature of penalised selection: once the yield curve slope and other predictors enter, credit spreads contribute little additional information for regime transitions.

Exit from vulnerability Real activity indicators dominate the exit equation. Industrial production of durable materials enters with the largest positive coefficient, indicating that stronger manufacturing output growth is associated with a higher prob-

ability of transitioning back to the normal regime. The remaining real activity measures (personal income, personal consumption, and other industrial production components) are selected with coefficients close to zero, so the exit signal is concentrated in this manufacturing component.

Labour market predictors are also selected, with a clear role for unemployment persistence. The help-wanted index—a vacancy indicator—enters positively, indicating that stronger labour demand is associated with higher exit probability. Unemployment of 15+ weeks enters with a large negative coefficient, indicating that increases in long-duration unemployment are associated with a lower probability of exit.

Financial indicators play a smaller role in the exit equation than in entry. The yield curve slope that dominates entry is not selected for exit. Both volatility measures (VIX and VXO) enter with positive coefficients. Housing indicators—specifically starts and permits in the Northeast and South—are selected for exit but not entry. Overall, exit rests primarily on real activity fundamentals, with financial indicators playing a secondary role.

Figure 4 displays the estimated time-varying transition probabilities. The entry probability (upper panel) reaches near-unity during the Global Financial Crisis and the pandemic onset, while the exit probability (lower panel) spikes during recovery phases. The asymmetry between the two panels is consistent with the different predictor compositions: entry reflects primarily financial stress indicators that can deteriorate rapidly, while exit depends on real activity measures that typically recover more gradually.

Regularisation-path stability The broad-based selection pattern is robust to the regularisation choice. Among the predictors selected in the preferred specification, 23 of 26 entry predictors and 16 of 24 exit predictors remain active in at least five of the ten penalty levels, suggesting systematic predictive content rather than dependence on a single regularisation choice. Sign reversals are rare: only approximately 10% of ever-selected predictors flip sign across the penalty grid. The key entry/exit asymmetry documented above is similar for nearby penalty levels. Appendix B reports the full diagnostic grid.

Table 3: Estimated transition probability parameters

Entry: Normal → Vulnerable

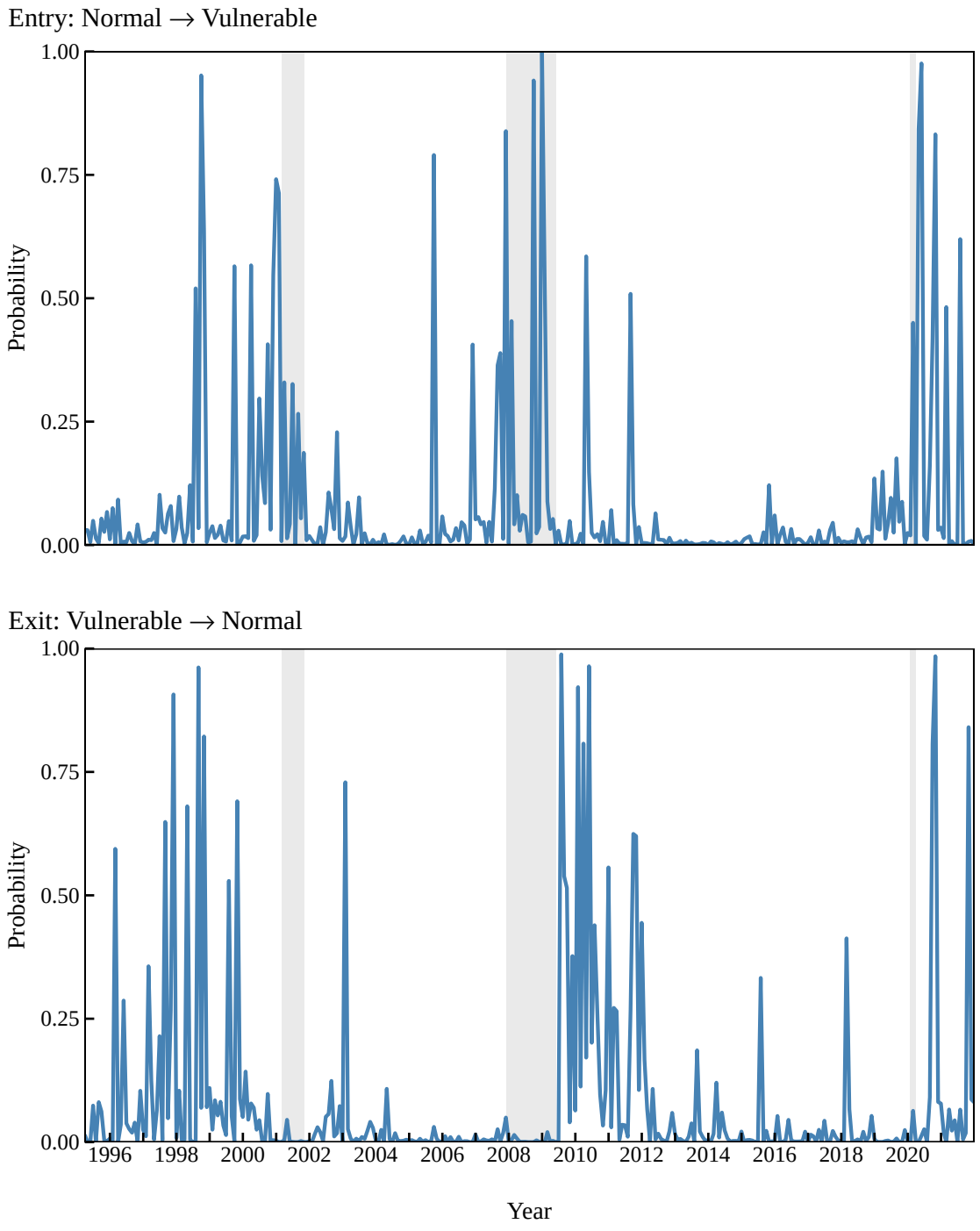
Variable	Coef.
Intercept	−4.29
Equity markets	
VXO Volatility Index	0.65
S&P P/E Ratio	−0.36
Yield curve & rates	
10Y-FFR Spread	−1.04
10-Year Treasury Rate	−0.58
Real activity	
IP: Durable Consumer Goods	−0.10
IP: Nondurable Materials	−0.40
IP: Materials	−0.01
Inventories to Sales Ratio	−0.16
Labor market	
Unemployed 15-26 Weeks	−0.07
Unemployed 27+ Weeks	0.59
Initial Claims	−0.13
Durable Goods Employment	−0.25
Construction Employment	0.004
Housing	
Housing Starts: Northeast	−0.03
Money & credit	
Consumer Motor Vehicle Loans	−1.09
Total Consumer Loans	0.13
Prices & wages	
PCE Price Index	0.95
PPI: Metals	−0.39
CPI: Apparel	−0.10
CPI: Medical Care	0.02
CPI: Durables	0.06
Avg Hourly Earnings: Construction	0.35
Avg Hourly Earnings: Manufacturing	−0.18
External	
Japan/U.S. Exchange Rate	−0.14
U.S./U.K. Exchange Rate	0.08
Canada/U.S. Exchange Rate	0.13

Exit: Vulnerable → Normal

Variable	Coef.
Intercept	−4.85
Equity markets	
VXO Volatility Index	0.32
CBOE VIX	1.04
S&P P/E Ratio	−0.22
Yield curve & rates	
5-Year Treasury Rate	0.48
Real activity	
Real Personal Income	−0.06
Real PCE	−0.01
IP: Durable Materials	2.31
IP: Residential Utilities	−0.06
IP: Fuels	−0.02
Labor market	
Help-Wanted Index	0.18
Unemployed <5 Weeks	0.29
Unemployed 5-14 Weeks	0.01
Unemployed 15+ Weeks	−1.35
Unemployed 27+ Weeks	0.20
Retail Trade Employment	−0.02
Housing	
Housing Starts: Northeast	−0.41
Building Permits: Northeast	−0.05
Building Permits: South	−0.34
Money & credit	
M1 Money Stock	0.07
Consumer Motor Vehicle Loans	0.22
Prices & wages	
CPI: Durables	0.18
Avg Hourly Earnings: Goods	−0.33
Oil Price	−0.26
External	
Japan/U.S. Exchange Rate	−0.19

Note: Baseline model estimation with Lasso-penalized coefficients ($\lambda = 6.16 \times 10^{-3}$). Entry coefficients are $\mathbf{w}_{12}$; exit coefficients are $\mathbf{w}_{21}$. Positive coefficients increase the respective transition probability. See Table B3 for variable mnemonics and descriptions. Sample: 1995:02–2021:12.

Figure 4: Estimated time-varying transition probabilities.



Note: Upper panel: entry probability (Normal → Vulnerable). Lower panel: exit probability (Vulnerable → Normal). Grey shaded areas denote NBER recession periods. Estimation sample: 1995:02–2021:12.

8 Conclusion

The central challenge addressed in this paper is the dimensionality bottleneck in regime-switching models. While time-varying transition probability (TVTP) specifications allow economic fundamentals to drive regime changes, the curse of dimensionality has historically limited these models to a small set of pre-selected predictors. The standard remedy—dynamic factor models—overcomes this by summarising high-dimensional data into latent indices, but at the cost of obscuring the specific predictive drivers of structural shifts.

This loss of granular detail is non-trivial. Theoretical insights into the granular nature of aggregate fluctuations suggest that they often arise from idiosyncratic origins rather than solely from diffuse common factors. While we do not claim to empirically establish that regime transitions are granular, this perspective serves as a conceptual motivation for our approach. To explore whether regime transitions follow a similar logic, researchers require an estimator that can handle high-dimensional datasets without imposing aggregation by construction. Our paper addresses this challenge by developing a penalised maximum-likelihood estimator that maintains the identity of individual predictors, providing a method well-suited to sparse transition structures.

Monte Carlo experiments illustrate the estimator's finite-sample properties, revealing a critical trade-off between model parsimony and regime calibration. Notably, we identify a "conservative intermediate phase" at medium sample sizes, where the data-driven penalty prioritises sparsity to safeguard against overfitting. This is accompanied by better probability-scoring performance, including lower log probability scores, even when recall remains imperfect, suggesting that regime inference can remain reasonably robust to imperfect variable selection in the transition block.

For practitioners, this framework offers a distinct complement to factor analysis. Unlike factor models, which obscure individual signals behind latent indices, our penalised estimator facilitates interpretability by isolating the specific leading indicators that trigger a regime switch. The empirical application to US Growth-at-Risk reinforces this point. Applied to 129 macro-financial predictors from the FRED-MD database, the estimator selects 26 predictors for entry into the vulnerable regime and 24 for exit, with only seven appearing in both equations. Entry into vulnerability is driven primarily by forward-looking financial indicators—notably the yield curve slope and equity volatility—while exit depends on real activity and labour market recovery. This asymmetry between entry and exit determinants, which is robust across nearby penalty levels, provides granular insight that other approaches would obscure.

A Technical Appendix

A.1 A Modified EM Algorithm for Gaussian Errors

In this section we present the EM algorithm and the steps followed for the estimator proposed in Section 4.1.

E-step details: filtering, smoothing, and expected log-likelihood

Let m be the number of regimes and define the time–t transition matrix

$$\mathbf{P}_t = [p_{ji,t}]_{j,i=1}^{m}, \qquad p_{ji,t} = \Pr(S_{t+1} = i \mid S_t = j, \boldsymbol{Z}_t; \mathbf{W}),$$

so that each row sums to one. For the observation density, write

$$f_i(\mathbf{y}_t) \equiv g(\mathbf{y}_t \mid \mathcal{F}_{t-1}; \boldsymbol{\theta}_i).$$

We use three state-belief vectors: predicted probabilities $\boldsymbol{\alpha}_{t|t-1} = \Pr(S_t = \cdot \mid \mathbf{Y}_{t-1}; \boldsymbol{\theta})$, filtered probabilities $\boldsymbol{\alpha}_{t|t} = \Pr(S_t = \cdot \mid \mathbf{Y}_t; \boldsymbol{\theta})$, and smoothed probabilities $\boldsymbol{\gamma}_t = \Pr(S_t = \cdot \mid \mathbf{Y}_T; \boldsymbol{\theta})$. Element-wise (Hadamard) multiplication and division are denoted by $\odot$ and $\oslash$, and $\mathbf{1}_m$ is an $m \times 1$ vector of ones.

We initialise the filter at $t = p + 1$ with

$$\boldsymbol{\alpha}_{p+1|p} = \boldsymbol{\pi},$$

the vector of initial state probabilities from the main text.

Forward filter (Hamilton, 1989) Prediction (for $t = p + 2, \ldots, T$):

$$\boldsymbol{\alpha}_{t|t-1} = \mathbf{P}_t^{\top} \, \boldsymbol{\alpha}_{t-1|t-1}.$$

Update:

$$\boldsymbol{\alpha}_{t|t} = \frac{\boldsymbol{f}_t \odot \boldsymbol{\alpha}_{t|t-1}}{\mathbf{1}_m^{\top}(\boldsymbol{f}_t \odot \boldsymbol{\alpha}_{t|t-1})}, \qquad \boldsymbol{f}_t = (f_1(\mathbf{y}_t), \ldots, f_m(\mathbf{y}_t))^{\top}.$$

Backward smoother (Kim, 1994) Initialise $\boldsymbol{\gamma}_T = \boldsymbol{\alpha}_{T|T}$ and, for $t = T-1, \ldots, p+1$,

$$\boldsymbol{\gamma}_t = \boldsymbol{\alpha}_{t|t} \odot \left(\frac{\mathbf{P}_{t+1} \, \boldsymbol{\gamma}_{t+1}}{\boldsymbol{\alpha}_{t+1|t}} \right).$$

Joint smoothed probabilities Define

$$\Xi_t(i, j) = \Pr(S_t = i, S_{t-1} = j \mid \mathbf{Y}_T; \boldsymbol{\theta}), \qquad t = p + 2, \ldots, T.$$

A convenient matrix expression is

$$\Xi_t = \operatorname{diag}(\boldsymbol{\alpha}_{t-1|t-1}) \, \mathbf{P}_t \, \operatorname{diag}(\boldsymbol{\gamma}_t \oslash \boldsymbol{\alpha}_{t|t-1}),$$

so that $\Xi_t(i, j)$ is the element in row j, column i.

Expected complete-data log-likelihood Using the decomposition in (6), the E-step quantity is

$$E_{\mathbf{S}_T}\Big[\ell(\mathbf{S}_T, \mathbf{Y}_T; \boldsymbol{\theta}) \mid \mathbf{Y}_T, \boldsymbol{\theta}^{(\ell-1)}\Big] = \sum_{i=1}^{m} \gamma_{p+1}(i)\, \log \pi_i + \sum_{t=p+2}^{T} \sum_{j=1}^{m} \sum_{i=1}^{m} \Xi_t(i,j)\, \log p_{ji,t} + \sum_{t=p+1}^{T} \sum_{i=1}^{m} \gamma_t(i)\, \log g(\mathbf{y}_t \mid \mathcal{F}_{t-1}; \boldsymbol{\theta}_i).$$

and subtracting $\text{pen}(\boldsymbol{\theta})$ gives the Q-function in (8).

Log-likelihood evaluation After convergence, the incomplete-data log-likelihood is

$$\ell(\mathbf{Y}_T \mid \boldsymbol{\theta}) = \sum_{t=p+1}^{T} \log\big(\mathbf{1}_m^\top(\boldsymbol{f}_t \odot \boldsymbol{\alpha}_{t|t-1})\big)\,.$$

M-step details: separate maximisations

Because the expected complete-data log-likelihood in (8) decomposes over parameter blocks, each set can be updated separately in the M-step.

Initial state probabilities Maximising the first term of (8) gives

$$\hat{\pi}_i^{(r)} = \gamma_{p+1}(i), \qquad i = 1, \ldots, m, \tag{16}$$

where $\gamma_{p+1}(i) = \Pr(S_{p+1} = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)})$ is the smoothed probability at the start of the effective sample.

Transition probability parameters For each current state j, the second term of (8) plus the penalty yields

$$\hat{\mathbf{W}}_j^{(r)} = \underset{\mathbf{W}_j}{\text{argmax}} \left\{ \sum_{t=p+2}^{T} \sum_{i=1}^{m} \Xi_t(i,j)\, \log p_{ji,t} \;-\; \lambda \sum_{i=2}^{m} \|\mathbf{w}_{ji,1:k}\|_1 \right\}, \tag{17}$$

subject to the multinomial logit structure in (2) and the identification restriction $\mathbf{w}_{j1} = \mathbf{0}$. The objective is concave in $\mathbf{W}_j$ (equivalently, the negative is convex) and the ℓ_1 term is convex, so the problem is a convex program with no closed-form solution. We solve it numerically using a standard convex optimizer (e.g., interior-point or proximal-gradient with soft-thresholding).

VAR parameters and covariances The third term of (8) separates by regime. Let

$$\gamma_t(i) = \Pr(S_t = i \mid \mathbf{Y}_T; \boldsymbol{\theta}^{(r-1)}), \qquad \mathbf{X}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})',$$

and let $\mathbf{B}_i$ collect $(\boldsymbol{c}_i, \boldsymbol{\Phi}_1(i), \ldots, \boldsymbol{\Phi}_p(i))'$. Equivalently, $\mathbf{B}'_i = [\, \boldsymbol{c}_i \;\; \boldsymbol{\Phi}_1(i) \;\; \cdots \;\; \boldsymbol{\Phi}_p(i) \,]$ and $\mathbf{y}_t = \mathbf{B}'_i \mathbf{X}_t + \mathbf{u}_t$, with $\mathbf{X}_t \in \mathbb{R}^{np+1}$ and $\mathbf{B}_i \in \mathbb{R}^{(np+1)\times n}$. Then the weighted least squares updates are

$$\hat{\mathbf{B}}_i^{(r)} = \left(\sum_{t=p+1}^{T} \gamma_t(i)\, \mathbf{X}_t \mathbf{X}'_t \right)^{-1} \left(\sum_{t=p+1}^{T} \gamma_t(i)\, \mathbf{X}_t \mathbf{y}'_t \right), \tag{18}$$

and with residuals $\hat{\mathbf{u}}_{i,t} = \mathbf{y}_t - \hat{\mathbf{B}}_i^{(r)\prime} \mathbf{X}_t$,

$$\hat{\boldsymbol{\Sigma}}_i^{(r)} = \frac{\sum_{t=p+1}^{T} \gamma_t(i)\, \hat{\mathbf{u}}_{i,t} \hat{\mathbf{u}}'_{i,t}}{\sum_{t=p+1}^{T} \gamma_t(i)}, \qquad i = 1, \ldots, m. \tag{19}$$

These three blocks complete the M-step. Together with the E-step recursions, they define the modified EM algorithm described in Section 4.1.

A.2 Robustness of the δ-band rule

In Section 4.2, we introduced the δ-band rule to favour parsimony when multiple penalty parameters yield statistically indistinguishable predictive performance. Specifically, we select the smallest λ within one standard error of the minimum cross-validation loss. This choice guards against excessive sparsity (i.e., collapsing to an intercept-only model) when the signal is weak but statistically indistinguishable from a sparser alternative.

Figure A1 compares the frequency of selected λ values under the standard minimum-CV rule versus the δ-band rule across 500 Monte Carlo replications. The panels display the results for varying sample sizes ($T \in \{100, 200, 500\}$) and predictor dimensions ($k \in \{10, 20, 30\}$).

The histograms reveal that the distributions of selected penalties are nearly identical for both rules in this controlled simulation environment. However, we retain the δ-band rule as a theoretical safeguard for empirical applications. While the clean Monte Carlo DGP produces a clear convex loss curve, real-world data is often noisier with flatter minima, making the δ-band rule a necessary robustness check to ensure parsimony when the signal is ambiguous. The fact that it aligns with the minimum-CV rule here confirms that it does not distort the results when the signal is clear, but its utility lies in its protection against overfitting in less idealised settings.

Figure A1: Penalty selection frequency under minimum-CV and δ-band rules.

Note: The panels display the distribution of selected effective penalty values $\lambda = \kappa\sqrt{(\log k)/T}$ *across 500 replications.*

B Application data and model selection

B.1 Data construction

The forward GDP growth target is constructed from the Brave-Butters-Kelley (BBK) Monthly GDP series. We convert the annualised month-over-month growth into a level index, compute year-over-year growth rates, and average months $t+1$ through $t+12$ to form the forward growth target $\tilde{y}_t$. Slow-moving trends are removed using a one-sided 10-year backward-looking moving average, $\bar{y}_t^{10y} = \frac{1}{120}\sum_{j=0}^{119} y_{t-j}$, yielding the detrended target $\tilde{y}_t = y_t - \bar{y}_t^{10y}$. Pre-1995 observations initialise this moving average. The forward construction reserves January 2022 through December 2022 for computing targets through December 2021.

The macro and financial factors are extracted as described in Section 7.1. The macro factor is the first principal component of four real activity indicators (ISM Manufacturing New Orders, Industrial Production, Retail Sales excluding Motor Vehicles, and Initial Jobless Claims rate). The financial factor is the first principal component of four financial conditions indicators (Excess Bond Premium, TED Spread, VXO, and Commercial Paper-Treasury Spread). PCA loadings are estimated on a pre-pandemic window (through December 2019) and factor scores are projected on these fixed loadings. Signs are normalised so that an increase in m_t signals stronger real activity, while an increase in f_t signals tighter financial conditions.

The 129 transition predictors are drawn from the FRED-MD database (McCracken & Ng, 2016), transformed to stationarity according to their release codes, and standardised to zero mean and unit variance over the estimation window. Factor principal components are excluded from the predictor set, ensuring that the transition equations depend directly on observable indicators. Table B3 provides the complete list of transition predictors grouped by economic category.

B.2 Regime parameter estimates

Table B1 reports the regime-specific parameter estimates for the reported L08 specification used in the main application.

Table B1: Regime-specific parameter estimates (semi-structural)

	Normal Regime	Vulnerable Regime
GDP Growth, Δy_t (annualized %)		
Intercept (α_y)	0.73	−1.45
Volatility (σ_y)	0.51	2.45
Financial factor (β_{0y})	0.21	−0.34
Macro factor (γ_{0y})	0.15	0.58
Financial Factor, f_t		
Intercept (α_f)	−0.11	0.16
Volatility (σ_f)	0.33	1.06
Macro factor (η_f)	−0.03	−0.10
Macro Factor, m_t		
Intercept (α_m)	0.09	−0.10
Volatility (σ_m)	0.71	1.26

Note: Semi-structural parameters from MS-VAR with recursive identification (ordering: $m_t \to f_t \to \Delta y_t$). All equation parameters are regime-dependent. α, σ denote intercepts and volatilities. β_{0y}, γ_{0y} capture contemporaneous responses of GDP to financial and macro factors. η_f captures the contemporaneous response of the financial factor to the macro factor. Baseline model with Lasso penalty $\lambda = 6.16 \times 10^{-3}$. Sample: 1995:02–2021:12.

B.3 Penalty selection

The penalty parameter λ is selected from a grid of ten logarithmically spaced values from 10^{-4} to 2×10^{-2} based on pseudo-out-of-sample evaluation of the predictive density. As a necessary condition, models must pass the Rossi and Sekhposyan (2019) HAC-robust PIT uniformity test at the 5% level. Among calibrated specifications, models are ranked using a composite score that weights left-tail quantile score (40%), weighted quantile score (20%), coverage accuracy (15%), PIT calibration (15%), and regime separation (10%).

Table B2 reports diagnostics across the penalty grid. Three models fail the calibration filter: L01 and L02 at the dense end, and L10 at the sparse end. Among the calibrated specifications (L03–L09), the composite score identifies L08 ($\lambda = 6.16 \times 10^{-3}$, 43 active variables, PIT $p = 0.154$) as the preferred model.

Table B2: Penalty parameter grid and model selection

Label	λ ($\times 10^{-3}$)	Active vars	PIT p-value	LTQS	Cov90 (%)	VolRatio	Composite score
L01	0.10	92	0.035	1.092	42.2	3.69	—
L02	0.18	86	0.041	1.086	43.7	3.63	—
L03	0.32	75	0.181	1.000	40.7	4.48	0.641
L04	0.58	66	0.201	0.990	40.3	4.51	0.625
L05	1.05	67	0.277	0.975	40.3	4.84	0.551
L06	1.90	65	0.126	0.928	43.3	4.54	0.537
L07	3.42	50	0.146	0.798	58.9	4.92	0.233
L08	**6.16**	**43**	**0.154**	**0.735**	**60.8**	**4.94**	**0.122**
L09	11.10	18	0.100	0.744	67.7	4.94	0.161
L10	20.00	11	0.036	0.718	74.9	4.93	—

Note: LTQS is the average quantile score at $\tau \in \{0.05, 0.10\}$; lower is better. Cov90 is empirical coverage of 90% prediction intervals (nominal 90%). VolRatio is $\sigma_{\text{vuln}}/\sigma_{\text{normal}}$. Models with PIT p-value < 0.05 fail calibration and are excluded from the composite score ranking (marked "—"). Bold indicates the selected specification. The active-variable count reports unique predictors selected across both transition equations. Estimation sample: 1995:02–2021:12.

B.4 Predictor selection across the penalty grid

Figures B1–B4 display heatmaps of the estimated TVTP coefficients across the penalty grid. Each row corresponds to a candidate predictor; each column corresponds to a penalty level (L01 through L10, with increasing λ from left to right). Cell shading indicates coefficient magnitude and sign: blue for positive coefficients (increasing transition probability) and red for negative coefficients (decreasing transition probability), with intensity proportional to magnitude. White cells indicate coefficients shrunk to exactly zero by the Lasso penalty.

The heatmaps illustrate the sparsity pattern as λ increases: at low penalty levels (left columns), many predictors receive non-zero coefficients; as the penalty strengthens (right columns), coefficients are progressively shrunk to zero, leaving only the most robust predictors active. The selected model L08 (highlighted column) retains a moderate number of predictors that survive across a range of penalty values, suggesting stable selection.

Regularisation-path interpretation The regularisation path provides a robustness check for the high-dimensional Lasso transition mechanism: moving from low to high

λ, the TVTP block becomes substantially sparser, with many coefficients set exactly to zero and only a small subset remaining active at the strongest penalties. This pattern is visible in both transition equations, but is more pronounced for the entry logit (Good $\rightarrow$ Bad), which contracts from roughly fifty active slope coefficients at low penalties (L01–L04) to only a handful at L10. The key implication is that the transition mechanism is not driven by an idiosyncratic choice of λ: a core set of economically interpretable predictors continues to be selected over a wide region of the λ grid. In particular, the yield curve slope (T10YFFM) remains selected for 9 out of 10 λ levels in the Good $\rightarrow$ Bad equation and consistently enters with the same sign, while volatility indicators (e.g., VXO) and selected real activity measures also survive across multiple penalties. Taken together, the heatmaps show that the Lasso estimator delivers a stable, interpretable ranking of transition predictors as the model is regularised from a relatively dense specification toward a highly sparse one.

Sign stability and sign reversals Across the full grid, sign changes are rare relative to the number of predictors that are ever selected: only about one in ten predictors exhibits any sign reversal across λ (8 out of 83 for Good $\rightarrow$ Bad; 9 out of 87 for Bad $\rightarrow$ Good, excluding intercepts). Moreover, when a sign reversal occurs, it is typically a one-off event: most of the flipping predictors change sign only once across the grid (5 out of 8 in Good $\rightarrow$ Bad; 6 out of 9 in Bad $\rightarrow$ Good), and none changes sign more than twice. Consistent with this, the dominant predictors that persist across a wide range of λ levels maintain a stable sign whenever selected—for example, T10YFFM is always negative in the Good $\rightarrow$ Bad equation across the λ levels where it is active, while VXO is always positive in both transition equations. This combination of sparse shrinkage, persistence of core predictors, and largely stable coefficient signs is consistent with a robust economic interpretation of the Lasso-based transition mechanism and indicates that the qualitative drivers of entry into (and exit from) the vulnerable regime are not sensitive to the particular regularisation level.

Table B3: List of variables used as transition probability predictors

Variable	Description	Tcode
Yield curve slope		
T10YFFM	10-Year Treasury C Minus FEDFUNDS	1
T1YFFM	1-Year Treasury C Minus FEDFUNDS	1
T5YFFM	5-Year Treasury C Minus FEDFUNDS	1
TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	1

Continued on next page

Table B3 – *Continued from previous page*

Variable	Description	Tcode
TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	1
Credit spreads		
AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	1
BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	1
EBP	Excess Bond Premium (Gilchrist-Zakrajsek)	1
TED	TED Spread (3m LIBOR - 3m Treasury)	1
Equity volatility & valuations		
S.P.500	S&P's Common Stock Price Index: Composite	5
S.P.PE.ratio	S&P's Composite Common Stock: Price-Earnings Ratio	5
S.P.div.yield	S&P's Composite Common Stock: Dividend Yield	2
VIXCLSx	CBOE Volatility Index (VIX)	1
VXO	VXO Volatility Index (old VIX)	1
Money & credit aggregates		
BOGMBASE	Monetary Base (St. Louis Adjusted)	6
BUSLOANS	Commercial and Industrial Loans	6
CONSPI	Nonrevolving consumer credit to Personal Income	2
DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	6
DTCTHFNM	Total Consumer Loans and Leases Outstanding	6
INVEST	Securities in Bank Credit at All Commercial Banks	6
M1SL	M1 Money Stock	6
M2REAL	Real M2 Money Stock	5
M2SL	M2 Money Stock	6
NONBORRES	Reserves Of Depository Institutions	7
NONREVSL	Total Nonrevolving Credit	6
REALLN	Real Estate Loans at All Commercial Banks	6
TOTRESNS	Total Reserves of Depository Institutions	6
Policy & market rate		
AAA	Moody's Seasoned Aaa Corporate Bond Yield	2
BAA	Moody's Seasoned Baa Corporate Bond Yield	2
FEDFUNDS	Effective Federal Funds Rate	2
GS1	1-Year Treasury Rate	2
GS10	10-Year Treasury Rate	2

Continued on next page

Table B3 – *Continued from previous page*

Variable	Description	Tcode
GS5	5-Year Treasury Rate	2
TB3MS	3-Month Treasury Bill	2
TB6MS	6-Month Treasury Bill	2
Real activity		
ACOGNO	New Orders for Consumer Goods	5
AMDMNOx	New Orders for Durable Goods	5
AMDMUOx	Unfilled Orders for Durable Goods	5
ANDENOx	New Orders for Nondefense Capital Goods	5
BUSINVx	Total Business Inventories	5
CMRMTSPLx	Real Manu. and Trade Industries Sales	5
CUMFNS	Capacity Utilization: Manufacturing	2
DPCERA3M086SBEA	Real personal consumption expenditures	5
HOUST	Housing Starts: Total New Privately Owned	4
HOUSTMW	Housing Starts, Midwest	4
HOUSTNE	Housing Starts, Northeast	4
HOUSTS	Housing Starts, South	4
HOUSTW	Housing Starts, West	4
INDPRO	IP Index	5
IP	Industrial Production Index	5
IPB51222S	IP: Residential Utilities	5
IPBUSEQ	IP: Business Equipment	5
IPCONGD	IP: Consumer Goods	5
IPDCONGD	IP: Durable Consumer Goods	5
IPDMAT	IP: Durable Materials	5
IPFINAL	IP: Final Products (Market Group)	5
IPFPNSS	IP: Final Products and Nonindustrial Supplies	5
IPFUELS	IP: Fuels	5
IPMANSICS	IP: Manufacturing (SIC)	5
IPMAT	IP: Materials	5
IPNCONGD	IP: Nondurable Consumer Goods	5
IPNMAT	IP: Nondurable Materials	5
ISRATIOx	Total Business: Inventories to Sales Ratio	2
NAPMNEWO	ISM Manufacturing New Orders Index	1^*

Continued on next page

Table B3 – *Continued from previous page*

Variable	Description	Tcode
PERMIT	New Private Housing Permits (SAAR)	4
PERMITMW	New Private Housing Permits, Midwest (SAAR)	4
PERMITNE	New Private Housing Permits, Northeast (SAAR)	4
PERMITS	New Private Housing Permits, South (SAAR)	4
PERMITW	New Private Housing Permits, West (SAAR)	4
RETAILx	Retail and Food Services Sales	5
RPI	Real Personal Income	5
UMCSENTx	Consumer Sentiment Index	2
W875RX1	Real personal income ex transfer receipts	5
Labor market		
AWHMAN	Avg Weekly Hours: Manufacturing	1
AWOTMAN	Avg Weekly Overtime Hours: Manufacturing	2
CE16OV	Civilian Employment	5
CES0600000007	Avg Weekly Hours: Goods-Producing	1
CES1021000001	All Employees: Mining and Logging: Mining	5
CLAIMSx	Initial Claims	5
CLF16OV	Civilian Labor Force	5
DMANEMP	All Employees: Durable goods	5
HWI	Help-Wanted Index for United States	2
HWIURATIO	Ratio of Help Wanted/No. Unemployed	2
MANEMP	All Employees: Manufacturing	5
NDMANEMP	All Employees: Nondurable goods	5
PAYEMS	Total Nonfarm Payroll Employment	5
SRVPRD	All Employees: Service-Providing Industries	5
UEMP15OV	Civilians Unemployed - 15 Weeks & Over	5
UEMP15T26	Civilians Unemployed for 15-26 Weeks	5
UEMP27OV	Civilians Unemployed for 27 Weeks and Over	5
UEMP5TO14	Civilians Unemployed for 5-14 Weeks	5
UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	5
UEMPMEAN	Average Duration of Unemployment (Weeks)	2
UNRATE	Civilian Unemployment Rate	2
USCONS	All Employees: Construction	5
USFIRE	All Employees: Financial Activities	5

Continued on next page

Table B3 – *Continued from previous page*

Variable	Description	Tcode
USGOOD	All Employees: Goods-Producing Industries	5
USGOVT	All Employees: Government	5
USTPU	All Employees: Trade, Transportation & Utilities	5
USTRADE	All Employees: Retail Trade	5
USWTRADE	All Employees: Wholesale Trade	5
Wages & prices		
CES0600000008	Avg Hourly Earnings: Goods-Producing	6
CES2000000008	Avg Hourly Earnings: Construction	6
CES3000000008	Avg Hourly Earnings: Manufacturing	6
CPIAPPSL	CPI: Apparel	6
CPIAUCSL	CPI: All Items	6
CPIMEDSL	CPI: Medical Care	6
CPITRNSL	CPI: Transportation	6
CPIULFSL	CPI: All Items Less Food	6
CUSR0000SA0L2	CPI: All items less shelter	6
CUSR0000SA0L5	CPI: All items less medical care	6
CUSR0000SAC	CPI: Commodities	6
CUSR0000SAD	CPI: Durables	6
CUSR0000SAS	CPI: Services	6
DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	6
DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	6
DSERRG3M086SBEA	Personal Cons. Exp: Services	6
OILPRICEx	Crude Oil, spliced WTI and Cushing	6
PCEPI	Personal Cons. Expend.: Chain Index	6
PPICMM	PPI: Metals and metal products	6
WPSFD49207	PPI: Finished Goods	6
WPSFD49502	PPI: Finished Consumer Goods	6
WPSID61	PPI: Intermediate Materials	6
WPSID62	PPI: Crude Materials	6
External		
EXCAUSx	Canada / U.S. Foreign Exchange Rate	5
EXJPUSx	Japan / U.S. Foreign Exchange Rate	5
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	5

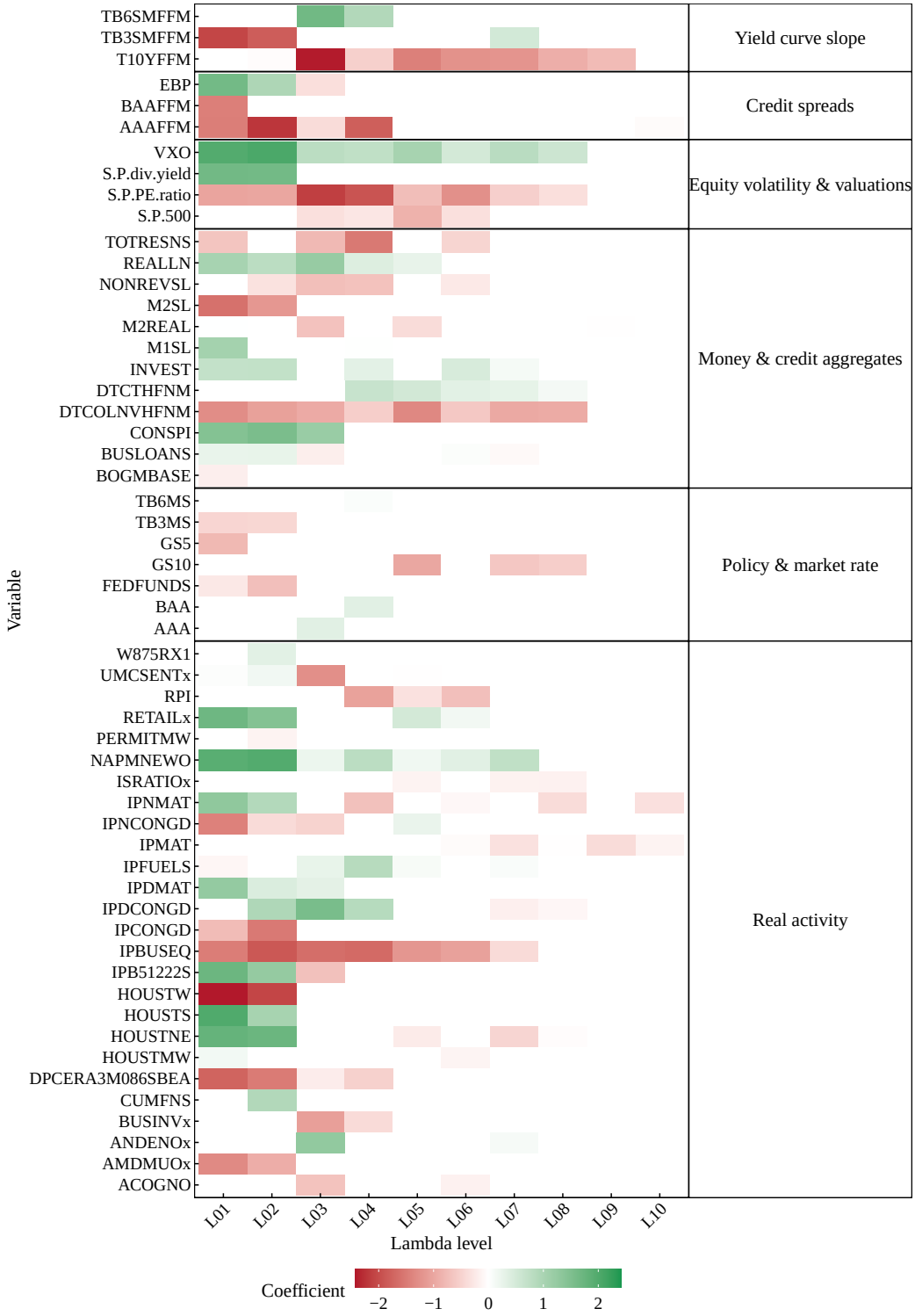
Continued on next page

Table B3 – *Continued from previous page*

Variable	Description	Tcode
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	5
TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index: Advanced Foreign Economies	5

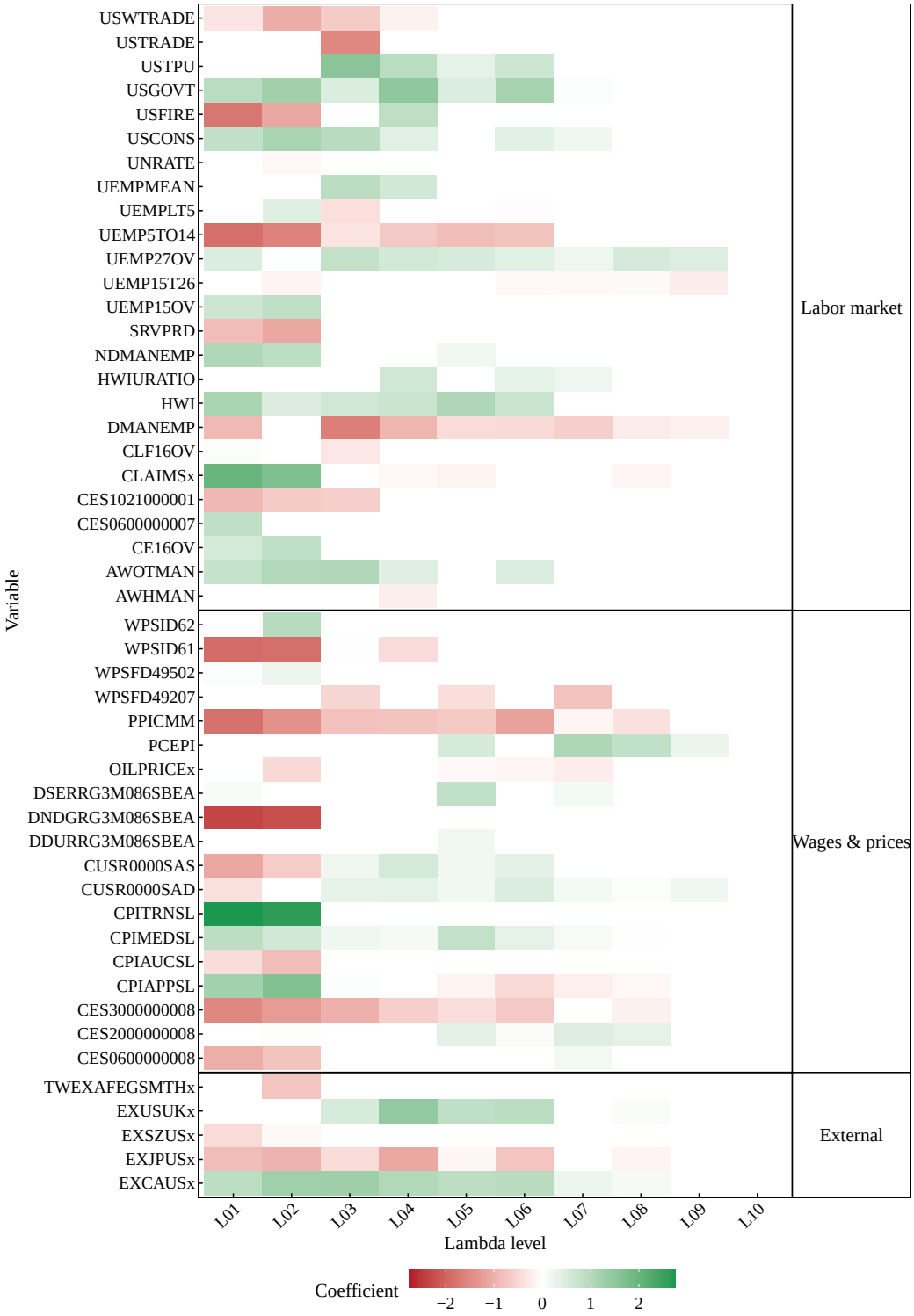
Note: Transformation codes: 1 = level, 2 = first difference, 4 = log, 5 = log difference, 6 = second log difference, 7 = first difference of percent change. *NAPMNEWO is a diffusion index centered at 50 (above 50 indicates improvement, below 50 indicates deterioration). Sample: 1995:02–2021:12.

Figure B1: TVTP coefficient heatmap: Entry transition (Normal → Vulnerable), Part 1.



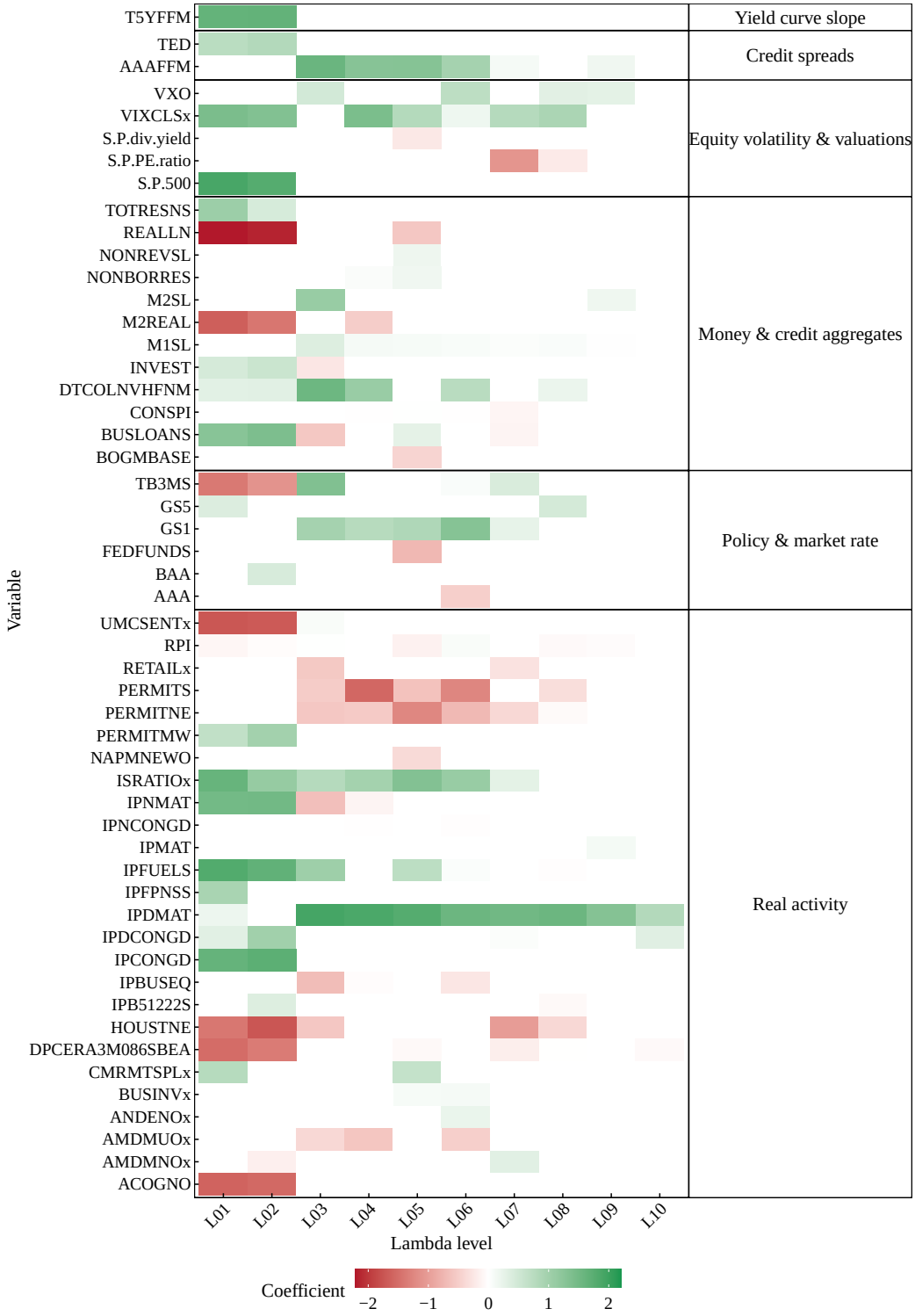
Note: Each row corresponds to a candidate predictor and each column corresponds to a penalty level (L01 through L10, with increasing λ from left to right). Cell shading indicates coefficient sign and magnitude (blue for positive coefficients and red for negative coefficients). White cells indicate coefficients shrunk to exactly zero by the Lasso penalty.

Figure B2: TVTP coefficient heatmap: Entry transition (Normal → Vulnerable), Part 2.



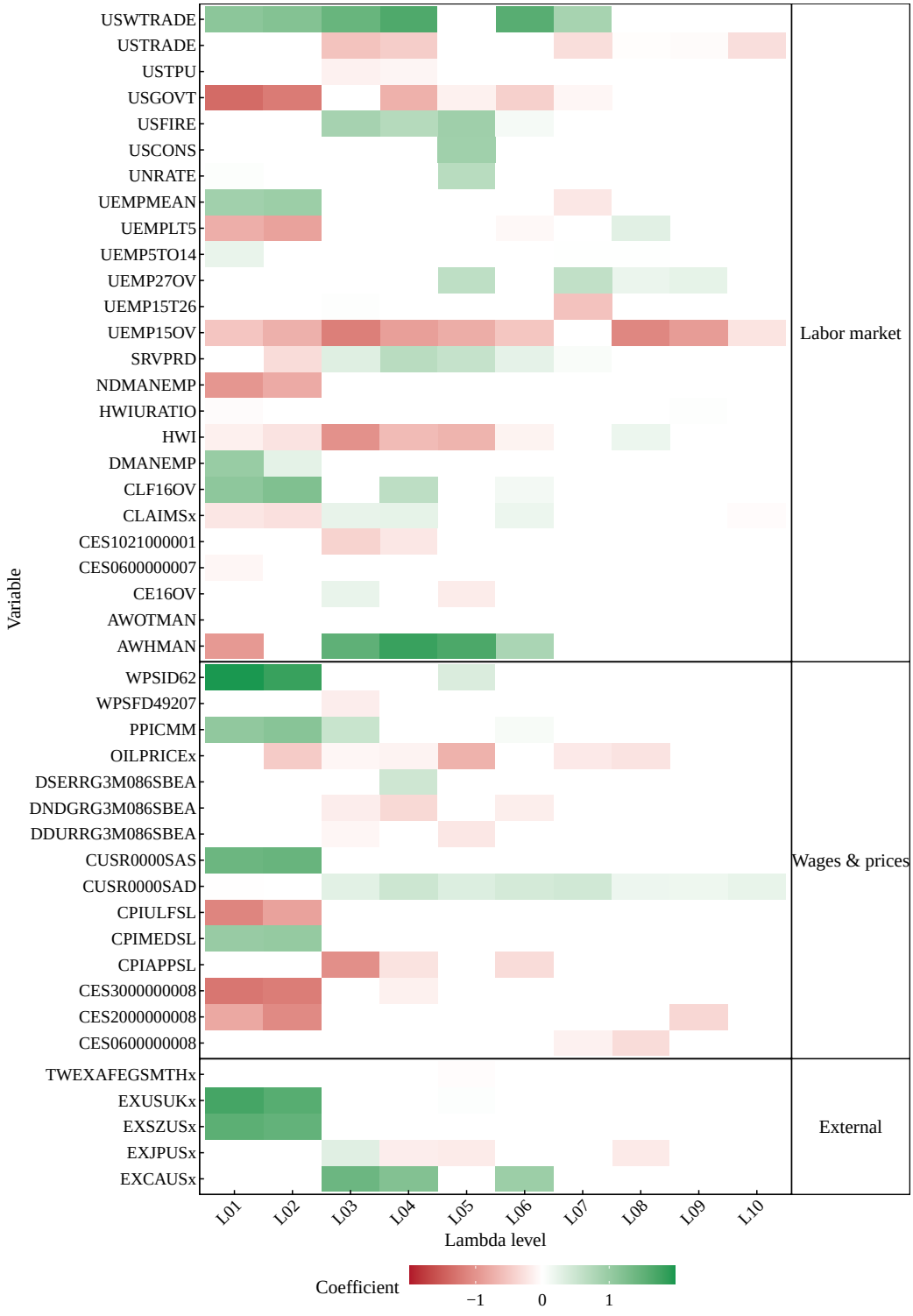
Note: Each row corresponds to a candidate predictor and each column corresponds to a penalty level (L01 through L10, with increasing λ from left to right). Cell shading indicates coefficient sign and magnitude (blue for positive coefficients and red for negative coefficients). White cells indicate coefficients shrunk to exactly zero by the Lasso penalty.

Figure B3: TVTP coefficient heatmap: Exit transition (Vulnerable → Normal), Part 1.



Note: Each row corresponds to a candidate predictor and each column corresponds to a penalty level (L01 through L10, with increasing λ from left to right). Cell shading indicates coefficient sign and magnitude (blue for positive coefficients and red for negative coefficients). White cells indicate coefficients shrunk to exactly zero by the Lasso penalty.

Figure B4: TVTP coefficient heatmap: Exit transition (Vulnerable → Normal), Part 2.



Note: Each row corresponds to a candidate predictor and each column corresponds to a penalty level (L01 through L10, with increasing λ from left to right). Cell shading indicates coefficient sign and magnitude (blue for positive coefficients and red for negative coefficients). White cells indicate coefficients shrunk to exactly zero by the Lasso penalty.

References

Adrian, T., Boyarchenko, N. & Giannone, D. (2019). Vulnerable Growth. *American Economic Review*, *109*(4), 1263–1289.

Barigozzi, M. & Massacci, D. (2025). Modelling Large Dimensional Datasets with Markov Switching Factor Models. *Journal of Econometrics*, *247*, 105919.

Bazzi, M., Blasques, F., Koopman, S. J. & Lucas, A. (2017). Time-Varying Transition Probabilities for Markov Regime Switching Models. *Journal of Time Series Analysis*, *38*(3), 458–478.

Bekaert, G., Hoerova, M. & Lo Duca, M. (2013). Risk, Uncertainty and Monetary Policy. *Journal of Monetary Economics*, *60*(7), 771–788.

Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica*, *77*(3), 623–685.

Burman, P., Chow, E. & Nolan, D. (1994). A cross-validatory method for dependent data. *Biometrika*, *81*(2), 351–358.

Caldara, D., Cascaldi-Garcia, D., Cuba-Borda, P. & Loria, F. (2021). *Understanding Growth-at-Risk: A Markov-Switching Approach*. SSRN.

Chauvet, M. (1998). An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Diebold, F. X., Lee, J.-H. & Weinbach, G. C. (1994). Regime Switching with Time-Varying Transition Probabilities. In *Business Cycles: Durations, Dynamics, and Forecasting* (Vol. 1, pp. 144–165). Princeton University Press.

Diebold, F. X. & Rudebusch, G. D. (1996). Measuring Business Cycles: A Modern Perspective. *The Review of Economics and Statistics*, *78*(1), 67–77.

Estrella, A. & Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance*, *46*(2), 555–576.

Estrella, A. & Mishkin, F. S. (1998). Predicting U.S. Recessions: Financial Variables as Leading Indicators. *The Review of Economics and Statistics*, 17.

Filardo, A. J. (1994). Business-Cycle Phases and Their Transitional Dynamics. *Journal of Business & Economic Statistics*, *12*(3), 299–308.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, *79*(3), 733–772.

Giannone, D., Lenza, M. & Primiceri, G. E. (2021). Economic Predictions with Big Data: The Illusion of Sparsity. *Econometrica*, *89*(5), 2409–2437.

Gilchrist, S. & Zakrajšek, E. (2012). Credit Spreads and Business Cycle Fluctuations. *American Economic Review*, *102*(4), 1692–1720.

Gray, S. F. (1996). Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics*, *42*(1), 27–62.

Green, P. J. (1990). On Use of the EM for Penalized Likelihood Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, *52*(3), 443–452.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 357–384.

Hamilton, J. D. (1990). Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, *45*(1–2), 39–70.

Hamilton, J. D. (2018). Why You Should Never Use the Hodrick-Prescott Filter. *The Review of Economics and Statistics*, *100*(5), 831–843.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd). Springer.

Huber, F. & Fischer, M. M. (2018). A Markov Switching Factor-Augmented VAR Model for Analyzing US Business Cycles and Monetary Policy. *Oxford Bulletin of Economics and Statistics*, *80*(3), 575–604.

Kim, C.-J. (1994). Dynamic Linear Models with Markov-switching. *Journal of Econometrics*, *60*(1), 1–22.

Kim, C.-J. & Halbert, D. C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA, USA: MIT Press.

Kim, C.-J., Piger, J. & Startz, R. (2008). Estimation of Markov Regime-Switching Regression Models with Endogenous Switching. *Journal of Econometrics*, *143*(2), 263–273.

Kock, A. B. & Callot, L. (2015). Oracle Inequalities for High Dimensional Vector Autoregressions. *Journal of Econometrics*, *186*(2), 325–344.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions*. Springer.

Maung, K. (2021). *Estimating High-Dimensional Markov-switching VARs*.

McCracken, M. W. & Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Monbet, V. & Ailliot, P. (2017). Sparse Vector Markov Switching Autoregressive Models. Application to Multivariate Time Series of Temperature. *Computational Statistics & Data Analysis*, *108*, 40–51.

Ravn, M. O. & Sola, M. (1999). Business Cycle Dynamics: Predicting Transitions with Macrovariables. In P. Rothman (Ed.), *Nonlinear Time Series Analysis of Economic and Financial Data* (pp. 231–265).

Rossi, B. & Sekhposyan, T. (2019). Alternative Tests for Correct Specification of Conditional Predictive Densities. *Journal of Econometrics*, *208*(2), 638–657.

Rudebusch, G. D. & Williams, J. C. (2009). Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve. *Journal of Business & Economic Statistics*, *27*(4), 492–503.

Sims, C. A. & Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *The American Economic Review*, *96*(1), 54–81.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Wright, J. H. (2006). *The Yield Curve and Predicting Recessions* (Finance and Economics Discussion Series No. 2006-07). Board of Governors of the Federal Reserve System.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67.