Bayesian Markov-Switching Models with Economic and Financial Applications

Luqi Wang

Faculty of Finance
Bayes Business School
City, University of London

This dissertation is submitted for the degree of
Doctor of Philosophy

August 2024

Preface

This dissertation represents the outcome of my doctoral studies at Faculty of Finance, Bayes Business School, City, University of London and was composed from 2020 to 2024. Financial support for this work was provided by a scholarship from City, University of London.

I wish to express my gratitude to the following individuals for their supports in many ways throughout my doctoral journey: my first supervisor Giovanni Urga, for providing enthusiastic research suggestions, constant support and encouragement; my second supervisor Richard Payne, for providing illuminating discussions during the development of this dissertation; PhD program director Giacinta Cestone, for providing a series of valuable advice about academic job market. I would also like to extend my thanks to all the people who have provided me with their feedback at various seminars and talks.

Finally, I thank my parents, for their unconditional love and my other half, for his never-ending support, love and patience.

Luqi Wang
London, June 2024

Contents

List of figures

List of tables

Introductory remarks

Real-world time series data often change their dynamic behavior, such as means and variances, across different time periods. Sometimes changes may occur in the form of "jumps", which are flash, or in the form of "breaks", which are permanent. A more frequently observed form of these changes is "regime shifts", which are both persistent and recurrent so that data seems to cycle between periods of behavior. Such changes are prevalent in the long-term trends of many macroeconomic and financial series, and are useful to glean a range of economic insights related to the cycling of the economy between recession and expansion periods (see, e.g., Hamilton, 1989; Ang and Bekaert, 2002), "bull" and "bear" markets in equity returns (see, e.g., Guidolin and Timmermann, 2006; Maheu et al., 2012), and various contagion phases experienced by financial institutions over time (see, e.g., Billio et al., 2022). Therefore, modeling and forecasting time series subject to regime shifts constitute a crucial area of econometrics. The dissertation contributes to this area in time series econometrics by advancing the application of Markov-switching models; developing Bayesian inference procedures for efficient model estimation; and providing simulation and empirical applications to understand modern economic and financial systems. The contributions are presented in three pieces of self-contained chapters.

Chapter 1 presents a modeling framework and estimation methods for detecting the regime shifts in the currency-liquidity-timing behavior of globally-diversified funds. Our approach builds upon the recent developments of regression models with endogenous Markov-switching parameters (Hwu et al., 2021; Kim and Kang, 2022), which allows for capturing the potential regime shifts in both direction and strength of funds' timing behavior together with underlying drivers that lead to such shifts. An effective Bayesian inference procedure is implemented for model estimation and selection. By analyzing a sample of 382 international fixed income mutual funds, we find that these globally-diversified funds on average engage in currency liquidity timing by adjusting their currency exposure in response to the underlying liquidity movement; however, this timing behavior exhibits significant regime changes across varying market conditions: funds time currency liquidity negatively (adjust their currency exposure in the opposite direction to the liquidity movement) in normal times but switch to aggressively positive timing (adjust their currency exposure in the same direction to the liquidity movement with increasing aggressivity) in turbulent market conditions. We

present evidence that these regime-switching timing behaviors are possibly driven by currency liquidity deterioration and negative shock on fund returns.

Chapter 2 concerns the problem of measuring connectedness in financial systems, which is central to modern risk management, including market risk (return connectedness and volatility connectedness); credit risk (default connectedness); counter-party and gridlock risk (bilateral and multilateral contractual connectedness); and systemic risk (system-wide connectedness). The literature on connectedness analysis to date has shown some notable empirical features of financial connectedness, such as the mixture of contemporaneous and temporal dependences, high-dimensionality, and regime shifts. However, existing econometric methods at most capture two of the above features. The novel model we introduce in the second chapter, which is refereed to as a Markov-Switching Graphical Structural Vector Autoregressive (MS-GSVAR) model, facilitates a "full sweep" of the list of features. An efficient Bayesian graph inference method is developed to address the computational complexities arising from inference on graph structures in the context of high model dimension, numerous lags, and multiple regimes. Simulation studies validate the effectiveness of the proposed framework in recovering many empirically relevant dependence structures, and in handling large datasets with changing dependence structures. Our model applied to the volatility series of 96 global banks detects different connectedness states, identifies systemically important individuals, and uncovers the frequency-specific source of connectedness, which are relevant to systemic risk management.

Together, the first and second chapter have inspired the focus of the third chapter. Markov-switching panel models face a major challenge in practical implementation, which is determining how many regimes are necessary to adequately characterize the observed data. Existing solutions typically rely on the assumption that the regime dimension is homogeneous in the cross-section. Such an assumption may be restrictive as individuals are likely to be characterized only by one or a subset of regimes identified from the panel. Chapter 3 proposes a general framework to estimate the number of regimes in Markov-switching panel models, allowing possible heterogeneity in cross-sectional regime dimension. We model individual heterogeneity via a binary matrix where its column dimension and configuration indicate respectively the regime dimension of the whole panel and the units. We develop new Bayesian nonparametric inference to jointly estimate the latent binary matrix and the other model parameters. Simulation studies validate the effectiveness of the proposed framework under different panel settings. An application to US state-level macroeconomic indices illustrates the empirical gains of considering likely heterogeneous regime dimension in the cross-section.

Chapter1

Do International Fixed Income Mutual Funds Time Currency Liquidity? Evidence from a Markov-Switching Model*

1.1 Introduction

Globally-diversified funds, though their main business is in international security markets, are subject to nontrivial exposure to currency markets. On the one hand, funds' currency exposure may arise from their underlying local-currency denominated foreign assets as returns on those assets would need to be redenominated in the funds' home currency. On the other hand, funds' currency exposure may arise from their involvement in currency hedging as in this situation funds directly hold currency assets. One may argue that currency exposure can be eliminated if funds hold only home-currency denominated foreign assets or fully currency-hedge their foreign asset positions. Yet doing so is restrictive and expensive (Opie and Riddiough, 2020) and thus zero currency exposure is almost unreal to be achieved by funds trading on a global scale (see observations in, e.g., Sialm and Zhu, 2022). As a result, currency concerns are expected to have a first-order impact on globally-diversified funds' asset allocation.

This paper contributes to the fund timing literature by providing the first formal investigation of the timing skills of the globally-diversified funds in the currency markets. In particular, we take an angle to examine whether funds engaging in international asset allocation strategically adjust their currency exposure in response to the systematic currency

*This chapter has been turned into a working paper (joint with G. Urga) entitled "Optimal N-state endogenous Markov-switching model for currency liquidity timing".

liquidity movement, which we call *Currency liquidity timing*. Among various aspects of timing abilities, we focus on liquidity timing because there is a clear connection between market-wide currency liquidity and the performance of globally-diversified funds. For instance, a decline in currency liquidity might induce higher price impact of trades, thus increasing the volatility of future exchange rate at which return conversion of local-currency denominated foreign assets would take place (Ranaldo and de Magistris, 2022). Besides, liquidity deterioration might cause higher rollover costs, which impair the profitability of currency hedging strategies (Brunnermeier et al., 2008; Mancini et al., 2013). The important implication of these phenomena for asset management is that globally-diversified funds could avoid such liquidity-induced risks if they can reliably anticipate systematic currency liquidity movement and adjust their currency exposure accordingly.

Empirically, we use international fixed income mutual funds as a testing ground to shed initial light on the currency-liquidity-timing behavior of globally-diversified funds. This choice is motivated by Sialm and Zhu (2022) who demonstrate that international fixed income mutual funds often employ dynamic currency hedging strategies and have time-varying currency exposure. Since one possible mechanism behind funds' time-varying risk exposure is timing, international fixed income mutual funds, as a part of globally-diversified funds and with known currency exposure, provide an ideal platform to study currency liquidity timing.

This paper also makes methodological development by introducing a novel Markov-switching model, which integrates a liquidity timing model of Cao et al. (2013) and an N-state endogenous Markov-switching framework of Hwu et al. (2021). Cao et al. (2013) were among the first to develop a method to measure mutual funds' liquidity timing ability. The authors infer timing behavior based on the convexity of fund returns relative to a benchmark factor model, following the literature on market return timing and volatility timing (see, e.g., Treynor and Mazuy, 1966; Ferson and Schadt, 1996; Busse, 1999). Their method has been extended by a growing body of studies in assessing the liquidity timing behavior of funds from different categories (see, e.g., Bodson et al., 2013; Li et al., 2017; Siegmann and Stefanova, 2017; Li et al., 2020a). Markov-switching models have been extensively used in macroeconomics and finance to extract the different regimes, or states, of the data series. In most of these models, two regimes are introduced with a state process determining one of the regimes to take place in each period. Moreover, the state process is typically assumed to be exogenous, meaning that future transitions between regimes are completely determined by the current regime, and does not rely on the realizations of underlying data series or other observable variables. Hwu et al. (2021) relax these restrictive features of the standard Markov-switching models by allowing for N states (i.e., multiple regimes) and endogeneity in the regime transitions (i.e., endogenous state process). Their model has been shown to achieve better forecasting and fitting performances (Kim and Kang, 2022).

Methodologically, we extend the Cao et al. (2013) model of equity market liquidity timing to construct a regression where fund return loadings on the currency risk factors (i.e.,

factor betas) change with currency liquidity. The currency risk factors include a dollar-risk factor and a carry-trade risk factor built in Lustig et al. (2011), consistent with the recent international asset pricing literature (see, e.g., Brusa et al., 2014; Karolyi and Wu, 2021). Our liquidity measure is the average quoted bid-ask spread in the currency spot market, which is extensively used in the literature on currency liquidities (see, e.g., Menkhoff et al., 2012; Karnaukh et al., 2015; Li et al., 2020b). Then, we allow the regression coefficients of beta-liquidity relation subject to regime shifts, which are modeled using Hwu et al. (2021)'s Markov-switching framework. In addition to the underlying fund return series, we allow future transitions between regimes to depend on other observable variables – in our case a real-time measure of currency liquidity level. This choice is motivated by Li et al. (2020a) who find that different market liquidity levels affect the changes in funds' liquidity timing behavior.

Our approach has several advantages over the existing timing models in the literature. First, our model feature expands upon the extensive research that assumes static funds' timing behavior and relies on time-invariant timing models (see, e.g., Treynor and Mazuy, 1966; Cao et al., 2013; Bodnaruk et al., 2019; Bali et al., 2021, among many others). It is reasonable to be suspicious of the results obtained from time-invariant timing models as such model specification, which is usually implemented with Ordinary Least Squares regression (OLS), implicitly assumes that funds either time or not time the market over the time period under investigation, and thus what it really measures is the timing behavior of funds averaged over the entire sample period. Consequently, when funds time the market strategically and intermittently, rather than continuously over the time span of study, the evidence of the potential timing effect in different time periods may be averaged out, impacting upon the significance of the results for funds' timing skills. Our framework overcomes the methodological issue associated with the time-invariant model specification by directly modeling and estimating the dynamics of timing abilities. Second, although greater attention has been directed towards time variation in timing abilities, some important economic insights are still missing in the literature, which include (i) the rich dynamics of timing behavior, and (ii) the drivers (i.e., potential economic mechanisms) underlying such dynamics. Naturally, funds' timing ability may change at any time and in different ways, including changes in the existence (e.g., timing/non-timing), directions (e.g., perverse timing/positive timing), or strengths (e.g., different levels of aggressivity in timing). However, no existing methods are capable of capturing all these rich dynamics of timing abilities in a unified framework. Early methods, such as the sub-period tests used in Cao et al. (2013), are limited to the pre-determined time point at which timing behavior is assumed to change. Recent advances, such as the changepoint approach adopted by Siegmann and Stefanova (2017) and the Hamilton Markov-switching model employed in Li et al. (2020a), are restricted to the prior assumption that timing behavior changes only once or between two regimes. Moreover, funds' motive to change to a specific timing behavior may be driven by various sources of information, arising

from either exogenous market conditions or endogenous fund performance. However, the understanding of such economic mechanisms explaining time variation in timing abilities is still at the early stage. Our framework building on the Hwu et al. (2021)'s method makes it possible to detect any sudden changes in funds' timing behavior, together with further analysis of the potential drivers behind such changes.

Using a sample of 382 international fixed income mutual funds drawn from the CRSP Survivor-Bias-Free US Mutual Fund Database, we analyze the average funds' currency-liquidity-timing behavior based on the equal-weighted portfolio of all fund returns over the period from July 2001 to December 2020. Given that funds may adjust their currency exposure for other reasons except currency liquidity timing, we also perform various robustness checks to distinguish the observed currency liquidity timing from artificial evidence. In summary, our main findings are threefold. First, we find evidence of currency liquidity timing at the aggregate level for the sample funds, which is not explained away by other reasons except currency liquidity timing (e.g., currency market return and volatility timing, liquidity reaction, illiquid holdings, funding constraints, and fund flow). Second, the direction and the strength of the observed currency-liquidity-timing behavior exhibit significant changes across varying market conditions. In particular, funds adjust their currency exposure in the opposite direction to the currency liquidity movement during tranquil market periods, but adjust their currency exposure towards the same direction to the liquidity movement with increasing aggressivity during turbulent market periods. This is evidenced by our baseline results which point to the existence of three distinct states: tranquil market periods are dominated by the *perverse timing* state where the coefficients of beta-liquidity relation are strictly negative; turbulent market periods are dominated by the *weak timing* and *strong timing* states where the coefficients of beta-liquidity relation shift towards largely positive values. In the robustness checks, the evidence of the three distinct states appears weakening but the dynamic pattern that funds' currency-liquidity-timing behavior shifts from a negative beta-liquidity relation towards a positive relation remains robust. We explain the observed dynamic pattern of funds' currency-liquidity-timing behavior based on their currency hedging practices. Third, the systematic currency liquidity level and shocks to fund returns drive to some extent the dynamics of funds' currency-liquidity-timing behavior. Our baseline results show that funds appear to switch from a negative relation between their currency exposure and liquidity towards a positive relation in times of currency liquidity deterioration and negative shock on fund returns. In the robustness checks, the evidence related to shock on fund returns becomes marginal but the results found for the currency liquidity level are materially unchanged.

The remainder of this paper is organized as follows. Section 1.2 discusses the data set used in the paper. Section 1.3 describes the models and methods. Section 1.4 presents the empirical results and Section 1.5 concludes. Technical details and additional empirical results are provided in Appendix 1.

1.2 Data

This section describes the data used in the empirical analysis, including the international fixed income mutual fund data, factor data, and the measure of systematic currency liquidity.

1.2.1 Fund sample

Data on international fixed income mutual funds are obtained from the CRSP Survivor-Bias-Free US Mutual Fund Database for the period July 2001-December 2020. As with Sialm and Zhu (2022), we select funds whose stated objectives indicate that they specialize in international fixed income investments, and exclude passively-managed index funds and ETFs from the sample.[1] For funds with multiple share classes, we aggregate the total net asset[2] (TNA) of individual share classes and calculate the fund-level variables (such as return, expense, and turnover ratio) as the TNA-weighted average across different share classes of the same fund. We retain only the funds with a minimum of 24 monthly returns as well as a minimum of TNA of \$10 million.[3] The data filter applied here follows for example Siegmann and Stefanova (2017) and Bali et al. (2021). Our final sample includes 382 funds.

Panel A of Table 1.1 reports summary statistics of fund characteristics. Age is defined as the number of years between the fund's last performance date and the fund's first offer date. A fund's last performance date is taken to be the latest net asset values (NAV) date across the fund's share classes while a fund's first offer date is taken to be the earliest first offer date across the fund's share classes. TNA is the total net asset in million US dollars, expense is the annual expense ratio in percentages, turnover is the annual turnover ratio in percentages, and return is the monthly return in percentages. Except for age, the fund characteristic reported are first averaged over time for each fund and then averaged across funds. On average, an international fixed income mutual fund has TNA of \$708 million, an annual expense ratio of 0.82%, an age of around 13 years, an annual turnover ratio of 112.34%, and a monthly return of 0.372%.

1.2.2 Factor data and liquidity measure

Recent breakthroughs in international macro-finance by Lustig et al. (2011) have found that the time-series variation in currency returns can be primarily explained by two common

[1] We select funds whose CRSP objective code (as identified by *crsp_obj_cd*) is IF, which cover six Lipper objectives: Emerging Markets Debt Funds (EMD), Emerging Markets Local Currency Funds (EML), Global High Yield Funds (GHY), Global Income Funds (GLI), International Income Funds (INI), and Short World Multi-Market Income Funds (SWM). Among these funds, we exclude funds whose CRSP identifiers "*index_fund_flag*" indicates a B, D or E or "*et_flag*" indicates an ETF or ETN.

[2] In our sample, 0.49% of share class TNA are flagged as missing. We impute missing TNA by using their past values, returns, and a factor adjusted for flow rates as in Ibert et al. (2018).

[3] We include a fund as soon as its inflation-adjusted TNA reached \$10 million. Our inflation index is the Consumer Price Index for All Urban Consumers (CPIAUCSL) series provided by the Federal Reserve Bank of St. Louis' FRED database. The data are available from https://fred.stlouisfed.org/series/CPIAUCSL.

currency risk factors: a dollar-risk factor (RX), and a carry-trade risk factor (HML_FX). These two factors are nearly orthogonal and capture two distinct trading strategies. Particularly, RX is the average portfolio return of a US investor who buys all foreign currencies available in the forward market, while HML_FX is the return in dollars on a zero-cost strategy that goes long in the highest interest rate currencies and short in the lowest interest rate currencies. These two factors have been used to examine global portfolio's exposure to the currency market in the recent international asset pricing literature (see, e.g., Brusa et al., 2014; Karolyi and Wu, 2021). The data of currency risk factors are obtained from the website of Adrien Verdelhan.[4]

To control for other factors associated with international fixed income mutual funds' performances, we follow Sialm and Zhu (2022) and include the following four factors: the hedged global bond market factor (GMF), emerging bond market factor (EMF), term factor (TERM), and credit factor (CREDIT). The data for the hedged global bond market factor, emerging bond market factor, and credit factor are obtained from Bloomberg. The hedged global bond market factor is measured by the return of the Bloomberg Global Aggregate Bond Index USD hedged. The emerging bond market factor is measured by the return of the JPMorgan Emerging Market Bond Index Global. The credit factor is the difference between the returns of the Bloomberg US Aggregate BAA Index and the Bloomberg US Aggregate AAA Index. The data for the term factor is obtained from the Board of Governors of the Federal Reserve System. The term factor is defined as the difference between the monthly change in the ten-year treasury constant maturity yield minus the one-month treasury constant maturity yield.

We measure systematic currency liquidity using the proportional quoted bid-ask spread[5] which is given by

$$L^{(ba)} = (P^A - P^B)/P^M, \tag{1.1}$$

where P^A and P^B are the quoted ask price and bid price, and P^M is the quote midpoint of the bid and ask prices. The measure of systematic currency liquidity is taken as the negative average of equation (1.1) across a basket of n currencies under investigation, which is given by

$$L_{m,t} = -\frac{1}{n}\sum_{i=1}^{n} L_{i,t}^{(ba)}, \tag{1.2}$$

where $L_{i,t}^{(ba)}$, computed in equation (1.1), is the proportional quoted bid-ask spread for currency i at time t, and a low $L_{m,t}$ indicates that the market is illiquid. We use the currency

[4]The data are available from http://web.mit.edu/adrienv/www/Data.html.

[5]We consider the proportional quoted bid-ask spread as our liquidity measure for two reasons. First, this measure is commonly used in the literature on currency liquidities (see, e.g., Menkhoff et al., 2012; Karnaukh et al., 2015; Li et al., 2020b). Second, evidence shows that this measure is highly correlated with other existing liquidity measures for the currency market. For instance, the correlation in Mancini et al. (2013) is 0.853 for the proportional quoted bid-ask spread and price impact, 0.890 for return reversal, 0.954 for effective costs, and 0.949 for price dispersion.

baskets considered in Lustig et al. (2011). The data of bid and ask prices are sourced from Thomson Reuters' Datastream. In addition, because the proportional quoted bid-ask spread is small in magnitude, we follow Li et al. (2020a) and rescale the liquidity measure $L_{m,t}$ by multiplying 1000 to improve numerical stability during the numerical computation in the empirical analysis in Section 1.4.

Panel B of Table 1.1 reports the descriptive statistics of the factor data and the computed systematic currency liquidity measure. Data are constructed in monthly frequency and cover a period from July 2001 to December 2020.

Table 1.1 Descriptive statistics

	Mean	Std. Dev.	P25	P50	P75
Panel A: Fund characteristics					
TNA ($ mil.)	708	1,392	46	202	630
Expense (% p.a.)	0.82	0.33	0.63	0.83	1.03
Age (years)	13.3	8.8	6.5	10.4	18.1
Turnover (% p.a.)	112.34	114.71	52.18	78.91	124.12
Return (% p.m.)	0.372	0.248	0.238	0.381	0.497
Panel B: Risk factors and liquidity measure					
GMF	0.277	0.798	-0.285	0.328	0.834
EMF	0.575	2.581	-0.390	0.788	1.865
TERM	0.153	0.091	0.082	0.163	0.221
CREDIT	0.193	1.646	-0.385	0.267	0.778
HML_FX	0.434	2.247	-0.823	0.501	1.888
RX	0.185	1.775	-0.830	0.239	1.237
L_m	-0.069	0.018	-0.077	-0.066	-0.053

NOTES: This table presents a statistical summary of fund characteristics, risk factors and liquidity measure. Panel A reports the cross-sectional statistics of fund characteristics based on the sample of 382 international fixed income mutual funds obtained from CRSP Survivor-Bias-Free US Mutual Fund Database. Except for age, values of all fund characteristics are first averaged over time for each fund, and then computed the statistics across funds. Panel B reports the descriptive statistics of the factor data and the computed systematic currency liquidity measure. The sample period is from July 2001 to December 2020.

1.3 Methodology

In this section, we first build a static model for assessing the time-invariant currency-liquidity-timing ability, following the existing timing models in the mutual fund literature. Then we incorporate a state process into this model to account for likely shifts in funds' currency-liquidity-timing behavior.

1.3.1 Model for assessing currency liquidity timing

Analysis for mutual funds' liquidity timing ability can be traced back to the work of Cao et al. (2013). To examine how equity mutual funds time aggregate equity market liquidity, Cao et al. (2013) begin with the Carhart (1997) four-factor model, which explains the time

series of equity mutual fund returns as follows

$$R_{p,t} = \alpha_p + \beta_{p,t}^{\text{MKT}} \text{MKT}_t + \sum_{j=1}^{J} \beta_p^j f_t^j + \varepsilon_{p,t}, \tag{1.3}$$

where p represents each fund and t represents each month; $R_{p,t}$ is the return in excess of the risk-free rate (proxied by US one-month Treasury bill rate); MKT_t is the excess return on the market portfolio; f_t^j (J=3 in this case) denote, respectively, the size, value and momentum factors; α_p is the risk-adjusted return; $\beta_{p,t}^{\text{MKT}}$ is the time-varying market beta that measures the dynamic exposure to the market risk; β_p^j are the other factor betas; and $\varepsilon_{p,t}$ is the error term.

To account for liquidity timing, Cao et al. (2013) specify time-varying market beta as a linear function of demeaned market liquidity

$$\beta_{p,t}^{\text{MKT}} = \beta_p^{\text{MKT}} + \lambda_p^{\text{MKT}} (L_{m,t} - \bar{L}_m), \tag{1.4}$$

where β_p^{MKT} captures the fund's average market beta without timing; $L_{m,t}$ is the market liquidity measure; $\bar{L}_m$ is the time series mean of market liquidity measure; and the coefficient λ_p^{MKT} measures fund's liquidity timing ability. The above equation is parallel to the existing specifications of market return timing and volatility timing (see, e.g., Treynor and Mazuy, 1966; Ferson and Schadt, 1996; Busse, 1999), except that the market condition considered here is market liquidity. A positive timing coefficient indicates that the fund has a high (low) market beta during good (poor) market liquidity conditions.

Substituting equation (1.4) to equation (1.3) yields a timing model that Cao et al. (2013) use to estimate the liquidity timing ability of equity mutual funds

$$R_{p,t} = \alpha_p + \beta_p^{\text{MKT}} \text{MKT}_t + \lambda_p^{\text{MKT}} (L_{m,t} - \bar{L}_m) \text{MKT}_t + \sum_{j=1}^{J} \beta_p^j f_t^j + \varepsilon_{p,t}. \tag{1.5}$$

Our timing model follows Cao et al. (2013) and is based on a multi-factor model, which explains the time series of international fixed income mutual fund returns as

$$R_{p,t} = \alpha_p + \beta_{p,t}^{\text{HML\_FX}} \text{HML\_FX}_t + \beta_{p,t}^{\text{RX}} \text{RX}_t + \sum_{j=1}^{J} \beta_p^j f_t^j + \varepsilon_{p,t}, \tag{1.6}$$

where p represents each fund and t represents each month; $R_{p,t}$ is the return in excess of the risk-free rate (proxied by US one-month Treasury bill rate); HML_FX_t and RX_t denote, respectively, the carry and dollar factors; f_t^j (J=4 in this case) denote the set of additional bond market risk factors which are defined in Section 1.2.2; α_p is the risk-adjusted return; $\beta_{p,t}^{\text{HML\_FX}}$ and $\beta_{p,t}^{\text{RX}}$ are the time-varying currency betas that measure the dynamic exposure to the common currency risk factors; β_p^j are the other factor betas; and $\varepsilon_{p,t}$ is the error term.

In line with the existing timing models in the mutual fund literature, we obtain the currency-liquidity-timing measure from the following equation

$$\begin{aligned}\beta_{p,t}^{\text{HML\_FX}} &= \beta_p^{\text{HML\_FX}} + \mu_p(L_{m,t} - \bar{L}_m),\\ \beta_{p,t}^{\text{RX}} &= \beta_p^{\text{RX}} + \lambda_p(L_{m,t} - \bar{L}_m),\end{aligned} \tag{1.7}$$

where $\beta_p^{\text{HML\_FX}}$ and β_p^{RX} capture the fund's average currency betas without timing. $L_{m,t}$ is the systematic currency liquidity measure; $\bar{L}_m$ is the time series mean of systematic currency liquidity measure; and the coefficients μ_p and λ_p measure fund's currency-liquidity-timing ability. A positive timing coefficient indicates that the fund overweights (underweights) exposure to the corresponding currency risk factor during good (poor) systematic currency liquidity conditions.

Combining equations (1.6) and (1.7) yields a model for assessing currency liquidity timing

$$\begin{aligned}R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_p(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \\ \lambda_p(L_{m,t} - \bar{L}_m)\text{RX}_t + \sum_{j=1}^{J}\beta_p^j f_t^j + \varepsilon_{p,t}.\end{aligned} \tag{1.8}$$

1.3.2 Markov-switching framework

To capture likely shifts in funds' currency-liquidity-timing behavior, we allow the timing coefficients in equation (1.8) to change depending on a state process as follows

$$\begin{aligned}R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \\ \lambda_{p,s_t}(L_{m,t} - \bar{L}_m)\text{RX}_t + \sum_{j=1}^{J}\beta_p^j f_t^j + \sigma\varepsilon_{p,t},\ \varepsilon_{p,t} \sim \text{ i.i.d. } \mathcal{N}(0,1),\end{aligned} \tag{1.9}$$

where $s_t \in \{1,2,...,N\}$ is a discrete regime indicator indicating which of N different regimes is realized at time t; and $\varepsilon_{p,t}$ is the error term assumed to be independent of each other, with a zero mean and variance σ^2, i.e., $\sigma\varepsilon_{p,t} \sim$ i.i.d. $\mathcal{N}(0,\sigma^2)$. The above equation is similar to equation (1.8) but here the timing coefficients μ_{p,s_t} and λ_{p,s_t} are assumed to take different values for each state of the regime.

We formulate the state process s_t based on the Markov-switching model of Hwu et al. (2021). The unique feature of their approach is that they allow for endogeneity in the regime transitions, which means that future transition between regimes depends on the current regime as well as two sources of information: (i) the information coming from some exogenous variables expected to influence the regime transitions, and (ii) the information contained in the regression error as it may include the missing regressors that are correlated with the regime transitions. This is in sharp contrast to the conventional Markov-switching model based on Hamilton (1989), in which the future transition is completely determined by the

current regime. Hwu et al. (2021) show that neglecting endogeneity in regime-switching incurs not only a substantial bias, but also a significant information loss, in estimating model parameters.

In the Markov-switching model of Hwu et al. (2021), s_t takes the following form

$$s_t|s_t^*, s_{t-1} = \begin{cases} 1 & \text{if } 0 = \max\left\{0, s_{1,t}^*, s_{2,t}^*, ..., s_{N-1,t}^*\right\}|s_{t-1} \\ 2 & \text{if } s_{1,t}^* = \max\left\{0, s_{1,t}^*, s_{2,t}^*, ..., s_{N-1,t}^*\right\}|s_{t-1} \\ \vdots & \\ N & \text{if } s_{N-1,t}^* = \max\left\{0, s_{1,t}^*, s_{2,t}^*, ..., s_{N-1,t}^*\right\}|s_{t-1} \end{cases} \tag{1.10}$$

This equation parameterizes N possible outcomes of s_t as the outcome of the values of $N-1$ auxiliary continuous random variables, denoted by $s_t^* = (s_{1,t}^*, s_{2,t}^*, ..., s_{N-1,t}^*)'$.[6]

To incorporate the first source of information into the state process, each of the $N-1$ auxiliary random variables is defined by

$$\begin{aligned} s_{i,t}^* &= \bar{\gamma}_{i,s_{t-1}} + z_t' \gamma_{i,s_{t-1}}^z + \eta_{i,t}, \\ \eta_{i,t} &\sim \text{ i.i.d. } \mathcal{N}(0,1), \\ i &= 1, 2, ..., N-1, \end{aligned} \tag{1.11}$$

where z_t is a vector of exogenous variables affecting the regime transitions. The parameter $\bar{\gamma}_{i,s_{t-1}}$ denotes the average value of the ith auxiliary variable $s_{i,t}^*$. The parameter $\gamma_{i,s_{t-1}}^z$ denotes a vector of coefficients of z_t. For both $\bar{\gamma}_{i,s_{t-1}}$ and $\gamma_{i,s_{t-1}}^z$, the subscript s_{t-1} suggests that the parameters take different values, depending on the past regime. $\eta_{i,t}$ denotes an error term, which is drawn from an independently distributed standard normal distribution.

Based on equation (1.11), the information variables z_t indirectly influences the realization of s_t through $s_{i,t}^*$. Specifically, with a positive $\gamma_{i,s_{t-1}}^z$, a higher value of z_t leads to a higher value of $s_{i,t}^*$. Thus, $s_{i,t}^*$ is more likely to become the maximum among all the $N-1$ auxiliary random variables. As a result, there is an increased probability that the regime at time t shifts to $(i+1)$ according to equation (1.10). The interpretation for a negative $\gamma_{i,s_{t-1}}^z$ is similar. In general, the effects of z_t on the regime transitions are associated with combinations of $\gamma_{i,s_{t-1}}^z$, for $i = 1, 2, ..., N-1$ and $s_{t-1} = 1, 2, ..., N$.

[6] For instance, consider the case of two regimes (i.e., $N = 2$), we have

$$s_t|s_t^*, s_{t-1} = \begin{cases} 1 & \text{if } s_{1,t}^* \leq 0|s_{t-1} \\ 2 & \text{if } s_{1,t}^* > 0|s_{t-1} \end{cases}$$

and consider the case of four regimes (i.e., $N = 4$), we have

$$s_t|s_t^*, s_{t-1} = \begin{cases} 1 & \text{if } (s_{1,t}^* < 0, s_{2,t}^* < 0, s_{3,t}^* < 0)|s_{t-1} \\ 2 & \text{if } (s_{1,t}^* > 0, s_{1,t}^* > s_{2,t}^*, s_{1,t}^* > s_{3,t}^*)|s_{t-1} \\ 3 & \text{if } (s_{2,t}^* > 0, s_{2,t}^* > s_{1,t}^*, s_{2,t}^* > s_{3,t}^*)|s_{t-1} \\ 4 & \text{if } (s_{3,t}^* > 0, s_{3,t}^* > s_{1,t}^*, s_{3,t}^* > s_{2,t}^*)|s_{t-1} \end{cases}$$

To account for endogeneity arising from the regression error term, all the $N-1$ elements in s_t^* and the regression error term $\varepsilon_{p,t}$ are assumed to be generated simultaneously from a joint Gaussian distribution. We denote by θ the collection of model parameters, including all the parameters in equations (1.9) and (1.11) and those in the Gaussian distribution specified below. We denote by $\mathbf{Y}_t$ the observations of the dependent variables through time t, i.e., $\mathbf{Y}_t = \{R_{p,i}\}_{i=1}^{t}$. Then, the joint distribution of s_t^* and $\varepsilon_{p,t}$, given $(s_{t-1}, \mathbf{Y}_{t-1}, z_t, \theta)$, is simply given by

$$\begin{bmatrix} s_t^* \\ \varepsilon_{p,t} \end{bmatrix} | s_{t-1}, \mathbf{Y}_{t-1}, z_t, \theta \sim \mathcal{N}\left(\begin{pmatrix} \bar{\gamma}_{s_{t-1}} + z_t' \gamma_{s_{t-1}}^z \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega & \rho \\ \rho' & 1 \end{pmatrix} \right), \tag{1.12}$$

where $\bar{\gamma}_{s_{t-1}} = (\bar{\gamma}_{1,s_{t-1}}, \bar{\gamma}_{2,s_{t-1}}, ..., \bar{\gamma}_{N-1,s_{t-1}})'$ and $\gamma_{s_{t-1}}^z = (\gamma_{1,s_{t-1}}^z, \gamma_{2,s_{t-1}}^z, ..., \gamma_{N-1,s_{t-1}}^z)'$. ρ is an $(N-1)\times 1$ vector of conditional correlations, i.e., $\rho = (\rho_1, \rho_2, ..., \rho_{N-1})'$, where $\rho_i = \text{corr}(s_{i,t}^*, \varepsilon_{p,t} | s_{t-1}, \mathbf{Y}_{t-1}, z_t)$ is the conditional correlation between the ith auxiliary variable and the regression error term. Ω is an $(N-1)\times(N-1)$ conditional variance-covariance matrix of s_t^*, which can be obtained as follows

$$\Omega = \begin{pmatrix} 1 & \rho_1\rho_2 & \cdots & \rho_1\rho_{N-1} \\ \rho_2\rho_1 & 1 & & \vdots \\ \vdots & & \ddots & \rho_{N-2}\rho_{N-1} \\ \rho_{N-1}\rho_1 & \cdots & \rho_{N-1}\rho_{N-2} & 1 \end{pmatrix}. \tag{1.13}$$

Based on equation (1.12), the presence of some non-zero correlation coefficients in ρ implies endogeneity in the regime transitions. In particular, when ρ_i is positive, the regression error term $\varepsilon_{p,t}$ at time t becomes positively correlated with the ith auxiliary variable. This means that a largely positive $\varepsilon_{p,t}$ is associated with a higher value of $s_{i,t}^*$, and thus results in an increased probability of $s_t = i+1$. The converse is also true. Similar to the case of z_t, endogenous regime shifts are associated with combinations of ρ_i, for $i = 1, 2, ..., N-1$.

In summary, the timing model in equation (1.9) and the Markov-switching model of Hwu et al. (2021) in equations (1.11)-(1.13) constitute the Markov-switching framework used in the empirical analysis. Our framework is estimated with a simulation-based Bayesian procedure. A detailed discussion of each step is provided in Appendix 1.A.

1.4 Results

In this section, we first select the best fitting model among alternative models with different number of regimes and regime-switching restrictions. Then, we analyze the potential drivers that lead to the shifts of funds' currency-liquidity-timing behavior across regimes.

Finally, we study the time path of the identified currency-liquidity-timing regimes. We provide additional robustness checks for the empirical findings in Appendix 1.C.

To our knowledge, literature has explored funds' timing ability at the aggregate level (e.g., Boney et al., 2009), individual level (e.g., Cao et al., 2013), or both (e.g., Cao et al., 2013). In this paper, we focus on an aggregate-level analysis which is implementable under our current modeling framework[7] and sheds initial light on the full picture of the dynamic currency-liquidity-timing behavior among the sample funds. To do so, we estimate the timing coefficients for an equal-weighted portfolio of all funds in the sample, similar as in Boney et al. (2009).

1.4.1 Model selection and identification of regimes

We select the best fitting model specification among variants of the timing model in equation (1.9), which differ across two dimensions: the regime-switching restrictions and the number of regimes. In terms of the regime-switching restrictions, we consider three model specifications: the first and second model specification allow only one timing coefficient to be regime-dependent but the other timing coefficient to be constant, while the third model specification is the most restricted one in which both timing coefficients are regime-dependent. The three model specifications are presented as follows:

Model 1: It is assumed that only the timing coefficient with respect to RX (denoted by λ) changes regimes.

$$\begin{aligned} R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_p(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \\ \lambda_{p,s_t}(L_{m,t} - \bar{L}_m)\text{RX}_t + \sum_{j=1}^{J} \beta_p^j f_t^j + \sigma\varepsilon_{p,t},\ \varepsilon_{p,t} \sim \text{ i.i.d. } \mathcal{N}(0,1), \end{aligned} \tag{1.14}$$

Model 2: It is assumed that only the timing coefficient with respect to HML_FX (denoted by μ) changes regimes.

$$\begin{aligned} R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \\ \lambda_p(L_{m,t} - \bar{L}_m)\text{RX}_t + \sum_{j=1}^{J} \beta_p^j f_t^j + \sigma\varepsilon_{p,t},\ \varepsilon_{p,t} \sim \text{ i.i.d. } \mathcal{N}(0,1), \end{aligned} \tag{1.15}$$

[7] By construction, the proposed model and the associated estimation/selection procedure can be applied to each of the individual funds. However, in practice this would be very time-consuming since the overall model-fitting process has to be repeated for hundreds or thousands of funds. Thus, our current modeling framework is implementable only when the overall model-fitting process can be repeated few times, which is the case favored by an aggregate-level analysis. An individual-level analysis can be a methodological extension and topic of future research, including for example adapting the current modeling framework to a multivariate Markov-switching structure.

Model 3: It is assumed that all the timing coefficients change regimes.

$$R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \lambda_{p,s_t}(L_{m,t} - \bar{L}_m)\text{RX}_t + \sum_{j=1}^{J}\beta_p^j f_t^j + \sigma\varepsilon_{p,t},\ \varepsilon_{p,t} \sim \text{ i.i.d. } \mathcal{N}(0,1), \tag{1.16}$$

For each model specification, we further estimate and compare the specifications with two, three and four regimes. Our motivation draws from the available evidence that funds show stronger liquidity timing ability under certain market conditions (see, e.g., Cao et al., 2013; Luo et al., 2017; Li et al., 2017). Therefore, one may argue that two regimes may not be enough to capture the dynamic features of liquidity timing behavior. To deal with this conjecture, we compare the fitness of the two-regime model against that of the alternative models involving a higher number of regimes. Overall, we have different combinations of regime-switching restrictions and the number of regimes, which finally include nine models.

For a given model, we implement the simulation-based Bayesian method and compute 150,000 iterations starting from the prior distribution of the parameters specified in Appendix 1.A. We discard the first 25,000 iterations and compute results using the remaining 125,000 iterations. In order to reduce the serial correlation of the draws, we keep every 10th draw to compute the posterior moments of the parameters, following Kaufmann (2015). Our posterior estimates are finally derived from an MCMC sample containing 12,500 iterations in total.

We also impose regime identifying restrictions to avoid the label-switching problem. Label-switching is a well-known and fundamental problem in Bayesian estimation of mixture or hidden Markov-switching models. In case that the prior distribution of the model parameters is the same for all states, then both the likelihood and posterior distribution are invariant to permutations of the parameters. This property makes Markov chain Monte Carlo (MCMC) samples simulated from the posterior distribution non-identifiable. See Frühwirth-Schnatter (2001) for further discussion. Specifically, we impose inequality constraints on the timing coefficients associated with different regimes. For Models 1 and 2, we restrict either $\lambda_{p,s_t=1} < \lambda_{p,s_t=2} < ... < \lambda_{p,s_t=N}$ or $\mu_{p,s_t=1} < \mu_{p,s_t=2} < ... < \mu_{p,s_t=N}$, such that the higher number of regimes correspond to the states where funds demonstrate stronger liquidity timing ability with respect to one currency risk factor. For Model 3, we impose similar restrictions on all the timing coefficients, such that the higher number of regimes is characterized by the stronger ability of funds to time all the currency risk factors.[8] These restrictions are

[8]For Model 3, it is possible to impose restrictions only on λ or μ for regime-identification. In this case, funds may employ different timing strategies with respect to the two currency risk factors at the same time. For example, they may show stronger motivation to time one factor but weaker motivation to time the other when switching to the higher number of regimes. We examine whether our identification constraints are too restricted. According to our empirical experiment, in the case of the three-regime, the log marginal likelihoods for Model 3 with restrictions on λ or μ alone are 904.301 and 895.508, respectively. These values are substantially small than that of Model 3 with both restrictions, 920.402. The same results are observed in the case of two-regime and four-regime. Therefore, our restrictions for Model 3 seem to be appropriate.

imposed via rejection sampling, following Kim and Kang (2022). In particular, we accept and update the simulated values of the parameters for which the above restrictions are satisfied. Otherwise, we retain the values of the parameters drawn from the last MCMC iteration.

Because we rely on a Bayesian approach, we can conduct a model selection based on the marginal likelihoods of the models and their ratios, namely the Bayes factors. The Bayes factor, introduced by Kass and Raftery (1995), trades off the improved fit resulting from adding more parameters as the number of regimes grows against the model over-fitting brought by the additional regimes. Following the common practice in the Bayesian literature, we compute the Bayes factor based on the marginal likelihood derived in Appendix 1.A.

$$\mathrm{BF}(i,j) = \frac{p(\mathbf{Y}_T|\mathcal{M}_i)}{p(\mathbf{Y}_T|\mathcal{M}_j)}, \tag{1.17}$$

where $\mathrm{BF}(i,j)$ denotes the Bayes factor in favor of the reference model $\mathcal{M}_i$ over the alternative model $\mathcal{M}_j$; $\mathbf{Y}_t = \{y_i\}_{i=1}^T$ represents the time series of observations; and $p(\mathbf{Y}_T|\mathcal{M})$ denotes the marginal likelihood of the respective model. We set Model 1 with two regimes (hereafter $\mathcal{M}_{1,N=2}$) as the reference model.[9]

Table 1.2 presents the pairwise log-Bayes factor in favor of the reference model $\mathcal{M}_{1,N=2}$ over the nine alternative models, $\mathcal{M}_{1,N=2}$, $\mathcal{M}_{1,N=3}$, $\mathcal{M}_{1,N=4}$, $\mathcal{M}_{2,N=2}$, $\mathcal{M}_{2,N=3}$, $\mathcal{M}_{2,N=4}$, $\mathcal{M}_{3,N=2}$, $\mathcal{M}_{3,N=3}$, and $\mathcal{M}_{3,N=4}$. For the convenience of comparison, we follow Jeffreys (1961, Appendix B) and convert Bayes factor into the logarithm scale (i.e., $\log_{10}(\mathrm{BF}(i,j))$ in equation (1.17)). The log-Bayes factor of the reference model versus itself is thus equal to zero. Kass and Raftery (1995) suggest interpreting the log-Bayes factor between 0 and 0.5 as weak evidence in favor of the reference model, between 1 and 2 as strong evidence, and greater than 2 as decisive evidence. The negative log-Bayes factor of the same magnitude is said to favor the alternative model by the same amount (Jiang et al., 2013).

Given these criteria, the results in Table 1.2 give rise to several interesting observations. First, we look at the log-Bayes factors listed in each row of Table 1.2. Here, we select among the models with different regime-switching restrictions. We find that Bayes factors tend to favor models in which all the timing coefficients are allowed to change regimes. For example, in the case of two regimes (see the first row of Table 1.2), the Bayes factor of 4.0 indicates that model $\mathcal{M}_{2,N=2}$ is less preferred compared to the reference model. In contrast, the Bayes factor of -3.9 provides substantial evidence in favor of model $\mathcal{M}_{3,N=2}$ against the reference model. Similar patterns are observed for the models with three and four regimes. This implies that the regime-switching restriction in Model 3 is most supported by the fund data. Next, we pay attention to each column of Table 1.2. Here, we compare the fitness of models with a different number of regimes. We observe that three regimes are sufficient to describe the shifts in funds' currency-liquidity-timing behavior in all cases. Taking Model 2 (see the

[9] Alternative choices of the reference model (e.g., Model 1 with three regimes, Model 2 with two regimes) leave the conclusions from our model comparison unchanged.

second column of Table 1.2) as an example, the three-regime specification achieve the lowest log-Bayes factors at 3.0. Thus, although all the specifications under Model 2 underperform the reference model in terms of the log-Bayes factors, the improved model fit is much smaller when moving from model $\mathcal{M}_{2,N=3}$ to the reference model. Combining these findings, we observe Model 3 with three regimes ($\mathcal{M}_{3,N=3}$) yields the lowest log-Bayes factor. We therefore set this model specification as the model with the best fit.

Table 1.2 Pairwise log-Bayes factor

N	$\mathcal{M}_{1,N}$	$\mathcal{M}_{2,N}$	$\mathcal{M}_{3,N}$
2	0.0	4.0	-3.9
3	-2.6	3.0	-8.1
4	-1.7	4.6	-7.0

NOTES: This table presents the pairwise log-Bayes factor in favor of the reference model $\mathcal{M}_{1,N=2}$ over the nine alternative models, $\mathcal{M}_{1,N=2}$, $\mathcal{M}_{1,N=3}$, $\mathcal{M}_{1,N=4}$, $\mathcal{M}_{2,N=2}$, $\mathcal{M}_{2,N=3}$, $\mathcal{M}_{2,N=4}$, $\mathcal{M}_{3,N=2}$, $\mathcal{M}_{3,N=3}$, and $\mathcal{M}_{3,N=4}$. Bayes factors are reported in logarithm scale (i.e., $\log_{10}(\text{BF}(i,j))$ in equation (1.17)).

Table 1.3 presents the posterior summary of the previously selected model based on 12,500 MCMC iterations. We report the posterior means, standard errors, 95% credibility intervals (CI) and convergence statistics of the model parameters. The convergence of our MCMC sampler is measured by the inefficiency factors and the p-value of convergence diagnostics test proposed by Geweke (1992). The inefficiency factor is the ratio of the numerical variance of the posterior sample mean to the variance of the hypothetical sample mean from uncorrelated draws. With an inefficiency factor of m, we need to draw MCMC samples m times to generate uncorrelated draws. In Kim and Kang (2022), where the three-regime multivariate endogenous Markov-switching model was estimated for US stock and bond return data, the inefficiency factors were found between 5 and 61 (Kim and Kang, 2022, Tables 14-15). From all panels in Table 1.3, the inefficiency factors are generally in a similar range as reported in Kim and Kang (2022), suggesting that our proposed sampler is highly effective. Geweke's convergence diagnostics test assesses the null hypothesis that the average draws computed with the first 20% and last 40% of the sample of retained draws are statistically equivalent. From all panels in Table 1.3, the p-values are all greater than 0.05, implying that there is no significant evidence against the convergence of the distribution of MCMC samples to the posterior distributions, according to Omori and Watanabe (2008). Overall, these convergence statistics indicate a well-mixing and efficient sampler.

Having determined the number of regimes, we next provide their economic interpretation. We place particular focus on the results contained in Panel B of Table 1.3, which presents the posterior summary of the switching coefficients μ_{s_t} and λ_{s_t} in the selected model. For both timing coefficients, we observe that the 95% CIs nearly do not overlap in the three regimes. This confirms the presence of three distinct currency-liquidity-timing regimes. Based on the posterior estimates, each identified regimes has a clear economic interpretation. Regime 1

can be viewed as a *perverse timing* state[10], given that the posterior means and the 95% CIs of $\mu_{s_t=1}$ and $\lambda_{s_t=1}$ are strictly negative. Regime 2 can be identified as a *weak timing* state as the 95% CIs of $\mu_{s_t=2}$ and $\lambda_{s_t=2}$ are more concentrated in the positive region, though these 95% CIs also include zero. Regime 3 corresponds to a *strong timing* state because the posterior means and the 95% CIs of $\mu_{s_t=3}$ and $\lambda_{s_t=3}$ are strictly positive.

Table 1.3 Estimation results of the selected model $\mathcal{M}_{3,N=3}$

	s_t/s_{t-1}	Mean	s.e.	95% CI		Ineff	p-val
Panel A: Non-switching coefficients in the timing model							
α (%)		0.083	0.011	0.082	0.124	68.900	0.816
β^{GMF}		0.054	0.009	0.052	0.091	68.467	0.718
β^{EMF}		0.065	0.003	0.065	0.074	61.332	0.922
β^{TERM}		1.479	0.071	1.195	1.496	70.630	0.691
β^{CREDIT}		0.108	0.009	0.105	0.144	69.826	0.679
$\beta^{\text{HML\_FX}}$		-0.001	0.005	-0.019	0.000	68.773	0.783
β^{RX}		0.179	0.007	0.150	0.180	71.379	0.717
σ		0.004	0.000	0.004	0.005	1.932	0.296
Panel B: Switching coefficients in the timing model							
μ_{s_t}	1	-0.076	0.018	-0.136	-0.072	69.515	0.672
	2	0.086	0.038	-0.032	0.096	71.740	0.669
	3	0.633	0.117	0.283	0.663	71.959	0.674
λ_{s_t}	1	-0.813	0.091	-0.836	-0.544	69.586	0.677
	2	0.874	0.246	-0.072	0.939	73.790	0.658
	3	1.767	0.269	0.834	1.838	72.721	0.662
Panel C: Parameters in the Markov model							
$\bar{\gamma}_{1,s_{t-1}}$	1	-1.342	0.198	-1.671	-1.022	6.355	0.730
	2	1.150	0.199	0.825	1.481	5.190	0.693
	3	-0.136	0.208	-0.486	0.201	1.203	0.942
$\gamma^z_{1,s_{t-1}}$	1	0.026	0.255	-0.399	0.442	2.453	0.846
	2	-0.445	0.271	-0.886	0.003	5.983	0.693
	3	-0.179	0.295	-0.670	0.301	1.439	0.896
$\bar{\gamma}_{2,s_{t-1}}$	1	-2.001	0.220	-2.362	-1.643	3.508	0.533
	2	-0.645	0.200	-0.974	-0.318	2.209	0.677
	3	1.295	0.211	0.948	1.640	2.218	0.650
$\gamma^z_{2,s_{t-1}}$	1	-0.167	0.278	-0.627	0.287	2.409	0.733
	2	-0.199	0.278	-0.663	0.256	1.653	0.842
	3	-0.277	0.281	-0.741	0.183	1.450	0.758
ρ_1		-0.431	0.251	-0.682	0.203	52.937	0.676
ρ_2		0.009	0.208	-0.353	0.339	14.468	0.624

NOTES: This table presents the posterior summary of the selected model $\mathcal{M}_{3,N=3}$, i.e., the three-regime timing model in equation (1.16). We report results for an equal-weighted portfolio of all sample funds for the period of July 2001-December 2020. Results are based on 12,500 MCMC iterations. s_t/s_{t-1} are, respectively, the regime indicator at time t and $t-1$, Mean is the posterior mean, s.e. is the posterior standard error, 95% CI is the 95% credibility interval, Ineff is the inefficiency factor, and p-val is Geweke (1992) p-value.

[10]Timing the market in the opposite way is referred to as *perverse timing* by the literature (see, e.g., Ferson and Schadt, 1996; Boney et al., 2009; Alda et al., 2015).

1.4.2 Endogenous regime transitions

After identifying the best fitted timing model, we now investigate the potential drivers that lead to the shifts of funds' currency-liquidity-timing behavior across regimes.

We begin with the first source of information coming from some exogenous variables expected to influence the regime transitions. Following Li et al. (2020a), we investigate how different levels of systematic currency liquidity affect the changes in the state process. We consider the exogenous variable to be a binary variable representing the current level of systematic currency liquidity. Specifically, we assume that z_t in equation (1.11) takes either 1 or 0 based on whether the current level of systematic currency liquidity at time t above/below its historical mean

$$z_t = \begin{cases} 1 & \text{if } L_{m,t} > \bar{L}_m \\ 0 & \text{otherwise} \end{cases} \tag{1.18}$$

Hence, the parameter $\gamma^z_{i,s_{t-1}}$ in equation (1.11) captures the regime-dependent effect of the current level of systematic currency liquidity on the ith auxiliary variable $s^*_{i,t}$.

Panel C of Table 1.3 presents the posterior summary of the parameters $\gamma^z_{i,s_{t-1}}$, for $i = 1,2,...,N-1$ and $s_{t-1} = 1,2,...,N$. In our previously selected three-regime timing model (i.e., $N = 3$), the parameters $\gamma^z_{i,s_{t-1}}$ consist of two elements $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$. Each of the element takes different values, depending on which of the three regimes was realized previously, i.e., $s_{t-1} = 1,2,3$. We observe that the posterior means of $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$ for $s_{t-1} = 1,2,3$ are generally negative and the corresponding 95% CIs lie more in the negative region. Besides, as shown in Figure 1.1, while we assume the distributions of $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$ are centered on zero a priori, the posterior distributions of the parameters appear to shift towards negative region. Overall, the results indicate a minor negative effect of the current level of systematic currency liquidity on the state process.

To further illustrate the effects of systematic currency liquidity on the transition distribution, we compute the unconditional transition probability matrix formulated in Appendix 1.B. In doing so, we characterize the state process using the information coming from the currency market liquidity conditions alone, regardless of the information contained in the regression error term. For the scenario where the systematic currency liquidity is low (i.e., $z_t = 0$), we can obtain the unconditional transition probability matrix

$$\mathbf{P}_t(z_t = 0) = \begin{pmatrix} p_{11,t} & p_{12,t} & p_{13,t} \\ p_{21,t} & p_{22,t} & p_{23,t} \\ p_{31,t} & p_{32,t} & p_{33,t} \end{pmatrix} = \begin{pmatrix} 0.890 & 0.089 & 0.022 \\ 0.092 & 0.828 & 0.080 \\ 0.054 & 0.139 & 0.807 \end{pmatrix}, \tag{1.19}$$

where $p_{ij,t} = \Pr(s_t = j | s_{t-1} = i, z_t = 0, \theta)$ is the unconditional transition probability from regime i at time $t-1$ to j at time t, given that $z_t = 0$.

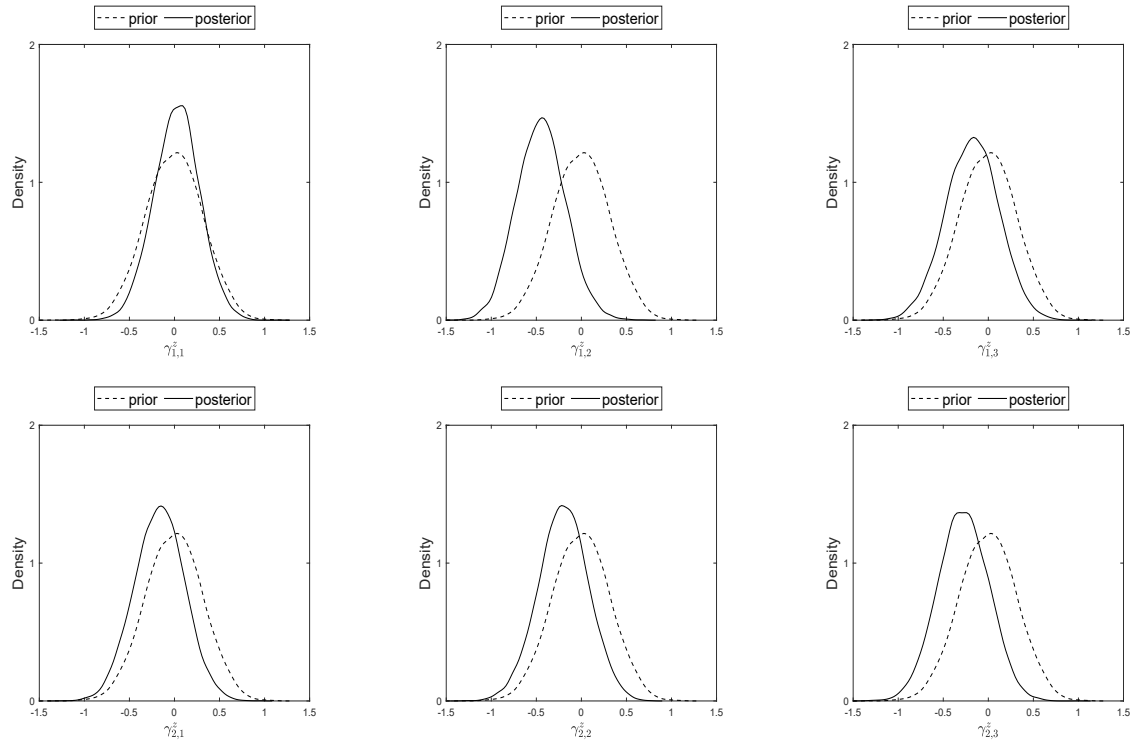


Figure 1.1 Prior-posterior distributions of the parameters $\gamma^z_{i,s_{t-1}}$. This figure plots the marginal posterior distributions (solid line) of the parameters $\gamma^z_{i,s_{t-1}}$, where $i = 1,2$ and $s_{t-1} = 1,2,3$, against their prior distributions (dashed line) for the selected model $\mathcal{M}_{3,N=3}$, i.e., the three-regime timing model in equation (1.16). Results are based on 12,500 MCMC iterations.

Similarly, when the current currency market liquidity level is high (i.e., $z_t = 1$), we can obtain the unconditional transition probability matrix

$$\mathbf{P}_t(z_t = 1) = \begin{pmatrix} p_{11,t} & p_{12,t} & p_{13,t} \\ p_{21,t} & p_{22,t} & p_{23,t} \\ p_{31,t} & p_{32,t} & p_{33,t} \end{pmatrix} = \begin{pmatrix} 0.892 & 0.094 & 0.014 \\ 0.192 & 0.719 & 0.088 \\ 0.096 & 0.143 & 0.761 \end{pmatrix}, \tag{1.20}$$

where $p_{ij,t} = \Pr(s_t = j | s_{t-1} = i, z_t = 1, \theta)$ is the unconditional transition probability from regime i at time $t-1$ to j at time t, given that $z_t = 1$.

Based on equation (1.19), when the overall currency market is relatively illiquid, the international fixed income mutual funds stay in the *weak timing* and *strong timing* states, with probabilities of 0.828 and 0.807, respectively. In the meantime, they switch from the *weak timing* and *strong timing* states to the *perverse timing* state with probabilities of 0.092 and 0.054, respectively. Based on equation (1.20), when the overall currency market is relatively liquid, the international fixed income mutual funds stay in the *weak timing* and *strong timing* states, with probabilities of 0.719 and 0.761, respectively. In this case, they switch from the *weak timing* and *strong timing* states to the *perverse timing* state with probabilities of

0.192 and 0.096, respectively. Obviously, the lower level of systematic currency liquidity is associated with a higher probability that the international fixed income mutual funds will continue to stay in the *weak timing* and *strong timing* states, and hence a lower probability that they will switch to the *perverse timing* state. This explains how the current level of systematic currency liquidity adversely affects the state process. Nevertheless, marginal differences are observed between the two unconditional transition probability matrices, which also points to a minor effect of the systematic currency liquidity on the regime transitions.

Next, we turn to the second source of information. At the bottom of Panel C in Table 1.3, we report the posterior summary of the parameters ρ_i for $i = 1, 2, ..., N-1$. According to equation (1.12), these parameters represent the conditional correlations between the regression error term $\varepsilon_{p,t}$ and the ith auxiliary variables $s^*_{i,t}$. In our previously selected three-regime timing model (i.e., $N = 3$), the parameters ρ_i are composed by two elements ρ_1 and ρ_2. We observe that the posterior mean of ρ_1 is sizable and displays a negative sign, while that of ρ_2 is almost zero. This is confirmed in Figure 1.2 where the posterior distribution of ρ_1 has larger mass in the negative region while that of ρ_2 is centered around zero. Overall, the results indicate between $s^*_{1,t}$ and $\varepsilon_{p,t}$, but a zero correlation between $s^*_{2,t}$ and $\varepsilon_{p,t}$.

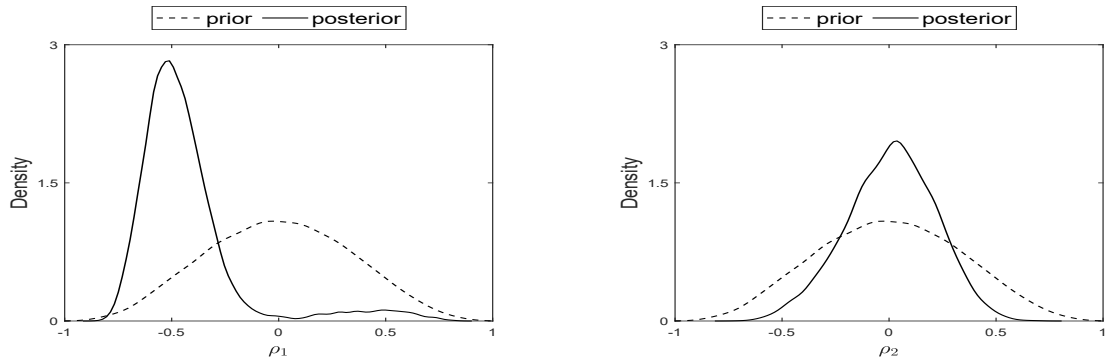


Figure 1.2 Prior-posterior distributions of the parameters ρ_i. This figure plots the marginal posterior distributions (solid line) of the parameters ρ_i, where $i = 1, 2$, against their prior distributions (dashed line) for the selected model $\mathcal{M}_{3,N=3}$, i.e., the three-regime timing model in equation (1.16). Results are based on 12,500 MCMC iterations.

To explain the conditional correlations in more detail, we compute the conditional transition probability matrix formulated in Appendix 1.B. We further eliminate the transition dependence on z_t (i.e., current level of systematic currency liquidity) by imposing zero values on the parameters $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$, for $s_{t-1} = 1, 2, 3$. In doing so, we characterize the state process using the information coming from the regression error term only. We randomly generate 12,500 artificial $\varepsilon_{p,t}$ covering a wide range from -10 to 10. Figure 1.3 plots conditional transition probabilities against alternative realizations of $\varepsilon_{p,t}$. Focusing on the diagonal entries, we find that the conditional probability of continuing in the *weak timing* state ($p_{22,t}$) increases rapidly as $\varepsilon_{p,t}$ moves from a positive value toward a negative value.

The persistence probabilities of the *perverse timing* ($p_{11,t}$) and the *strong timing* states ($p_{33,t}$) however display opposite shifts as opposed to $p_{22,t}$ when $\varepsilon_{p,t}$ is varied. Similarly, there is a sharp increase in the probabilities of moving from other regimes to the *weak timing* state ($p_{12,t}$ and $p_{32,t}$), but a gradual decrease in the probabilities of moving out of the *weak timing* state ($p_{21,t}$ and $p_{23,t}$) as $\varepsilon_{p,t}$ falls to a largely negative value. These patterns of probabilities suggest that larger negative values of $\varepsilon_{p,t}$ are associated with an increased likelihood of the *weak timing* state occurring relative to both *perverse timing* and *strong timing* states. Such a negative relationship between the regression error term and the *weak timing* state is reflected in the sign of parameter ρ_1. We also observe that the variation of the conditional probabilities associated solely with the *perverse timing* and *strong timing* states ($p_{13,t}$ and $p_{31,t}$) is not as dramatic as that of the probabilities involving the *weak timing* state. This implies that the realization of $\varepsilon_{p,t}$ seems to be unrelated to the likelihood of the *strong timing* state occurring relative to the *perverse timing* state. Such a weak relationship between the regression error term and the *strong timing* state can be possibly attributed to the insignificance of parameter ρ_2.

Overall, the results show that both sources of information affect to some extent the probabilities that funds switch from one currency-liquidity-timing regime to others. We identify the current level of systematic currency liquidity as one potential financial force that drives to some extent funds' motives for currency liquidity timing. This finding confirms what reported in Li et al. (2020a) based on the global hedge funds data. One possible explanation is that when sudden shocks of currency liquidity crisis occur, systematic currency liquidity starts to become an important consideration for the international fixed income mutual funds, who then observe and time the liquidity level as closely as possible to reduce the currency exposure if possible. We also observe the negative relationship between the regression error term and the probability of the *weak timing* state occurring. In our setting, regression error term can be viewed as shocks to the international fixed income mutual fund returns. Therefore, the regime shifts from other states to the *weak timing* state seem to interact with a negative shock to the fund returns. Sialm and Zhu (2022) argue that international fixed income mutual funds may face higher costs of outflows when holding illiquid assets and are thus more incentivized to hedge their currency risks in response to their poor performance. This coincides with our finding and explains why funds' currency-liquidity-timing behavior is sensitive to downside returns.

1.4.3 Time path of the currency liquidity timing regimes

We now provide a graphical analysis to visualize the identified currency-liquidity-timing regimes. To do so, we classify any time period t to be the *perverse timing*, *weak timing* or *strong timing* state if the posterior probability of being in a particular state is greater than

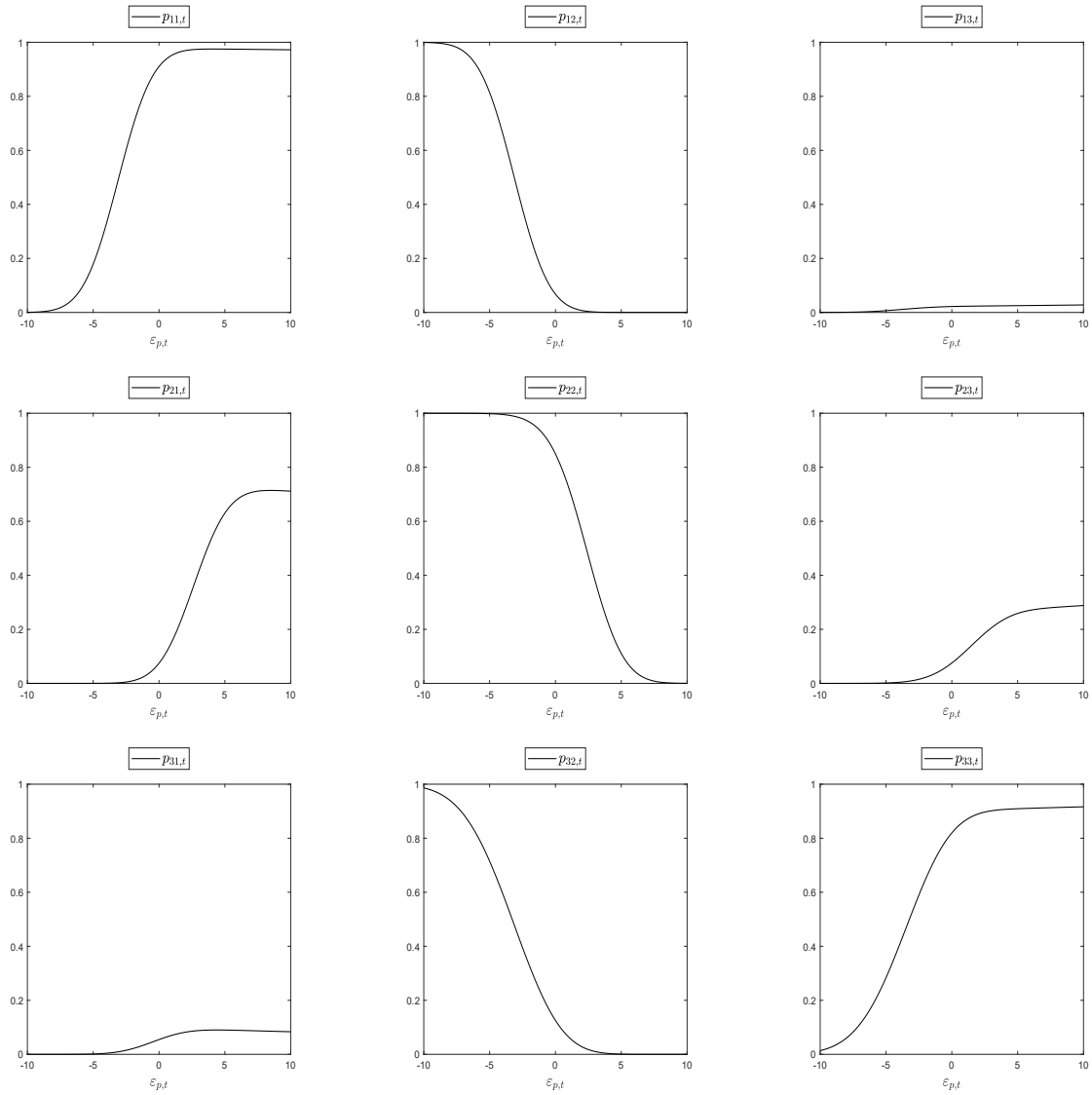


Figure 1.3 Conditional transition probabilities against alternative realizations of $\varepsilon_{p,t}$. This figure plots the transition probabilities (solid line) conditional on the regression error term $\varepsilon_{p,t}$ inside the selected model $\mathcal{M}_{3,N=3}$, i.e., the three-regime timing model in equation (1.16). The x-axis measures alternative values of $\varepsilon_{p,t} \in [-10, 10]$. The y-axis measures alternative values of conditional transition probabilities $p_{ij,t} \in [0, 1]$, where $p_{ij,t} = \Pr(s_t = j | s_{t-1} = i, \varepsilon_{p,t}, \theta)$ is the conditional transition probability from regime i at time $t-1$ to j at time t, given $\varepsilon_{p,t}$. Results are based on 12,500 MCMC iterations.

50% at time t.[11] Figure 1.4 plots shaded areas associated with the periods of funds being in a particular currency-liquidity-timing regime. For the 19-year sample period and using a

[11]We choose to adopt a threshold of 50% for the presentation of results for two reasons. First, a threshold of 50% is extensively used in the Markov-switching literature (see, e.g., Chan et al., 2011; Jutasompakorn et al., 2014) in determining the regime. A particular regime is identified by the smoothed posterior probability exceeding 50%. Second, using a threshold of 50% minimizes the occurrence of type I and type II errors which involve selecting too low and too high threshold, respectively. An inappropriately low threshold may result

50% threshold probability, we observe that international fixed income mutual funds spend 61.11%, 18.80% and 3.42% of days in the *perverse timing*, *weak timing* and *strong timing* states respectively. There are 16.67% of days where we cannot determine the regime as each regime has smoothed probability of less than 50%.

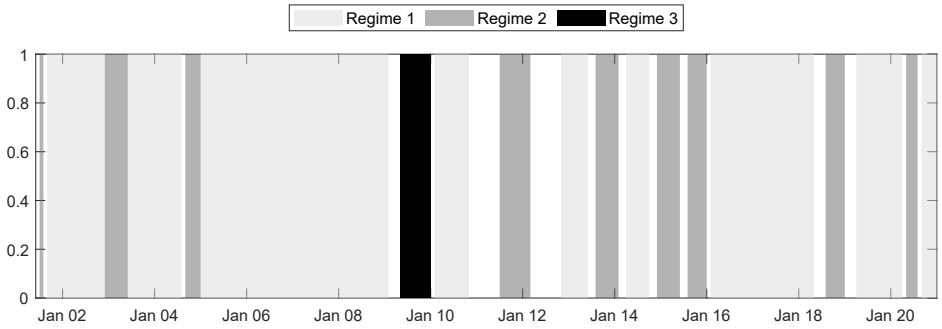


Figure 1.4 Time path of the currency liquidity timing regimes. This figure plots shaded areas associated with the periods of funds being in a particular regime. The light gray bars, dark gray bars, and black bar correspond to the periods of funds being in regimes 1-3, respectively. The empty white bars correspond to the periods during which the model cannot determine the regime as each regime has smoothed probability of less than 50%. Based on the selected model $\mathcal{M}_{3,N=3}$, regimes 1-3 can be viewed as *perverse timing*, *weak timing* and *strong timing* states, respectively. The sample period is from July 2001 to December 2020.

From Figure 1.4, we find that international fixed income mutual funds normally stay in the *perverse timing* state during which the overall financial market is calm. Interestingly, they unusually switch to the *weak timing* and *strong timing* states during which the overall financial market becomes more turbulent. We observe nine periods of *weak timing* state: July 2001-August 2001, December 2002-May 2003, September 2004-December 2004, July 2011-February 2012, August 2013-January 2014, December 2014-May 2015, August 2015-December 2015, August 2018-December 2018, May 2020-July 2020, and one period of *strong timing* state: May 2009-December 2009. The *weak timing* state covers the early 2000s recession, several rounds of US Quantitative Easing (QE) programs from 2009 to 2015, and the Covid-19 crisis in early 2020. The *strong timing* state corresponds to the aftermath of the sub-prime crisis; for example, the credit crisis with Greece's Bailout taking place after 2009.

The following mechanism emerges from our findings. When the overall financial market is calm, international fixed income mutual funds possibly do not care about their currency exposure as exchange rate fluctuations are relatively stable. Therefore, they tend to only hedge their currency exposure when the hedging costs are low but leave their currency exposure unhedged when the hedging costs are high. Since hedging costs are adversely correlated with liquidity, the above behavior implies that funds' currency exposure decreases due to hedging in times of increased currency liquidity; by contrast, funds' currency exposure

in an overestimated number of a particular regime, whereas an overly stringent threshold could cause an underestimated number of a particular regime.

increases due to less hedging in times of decreased currency liquidity. As a result, the *perverse timing* state is associated with the tranquil periods and under this circumstance funds adjust their currency exposure in the opposite direction to the liquidity movement. Conversely, when the overall financial market becomes more turbulent, international fixed income mutual funds are possibly concerned about their currency exposure as exchange rate fluctuations are relatively volatile. Therefore, they tend to hedge their currency exposure even though the hedging costs are high. Moreover, in episodes of global financial stress, currency liquidity is likely to worsen jointly with the liquidity of underlying assets. Consequently, funds' risk appetite falls and they seek to rebalance their global portfolio more frequently than they do during the tranquil periods, inducing higher currency hedging activities. The above behavior implies that the further reduction in currency liquidity, the more increase in currency hedging and thus the more decrease in funds' currency exposure. As a result, the *weak timing* and *strong timing* states are associated with the turmoil periods and under these circumstances funds adjust their currency exposure towards the same direction to the liquidity movement.

1.5 Conclusions

This paper provided the first formal investigation of the currency-liquidity-timing behavior of globally-diversified funds. Using a sample of international fixed income mutual funds as a testing ground, we analyzed aggregate timing behavior based on the equal-weighted portfolio of all funds over the period from July 2001 to December 2020. To capture dynamics in funds' timing behavior and the potential drivers behind such dynamics, we proposed a novel Markov-switching model that allows for multiple regimes and endogeneity in the regime transitions. We found that the sample funds on average engage in currency liquidity timing by adjusting their currency exposure in response to the systematic currency liquidity movement. Interestingly, this timing behavior appeared to change with market conditions, in that funds time currency liquidity negatively (adjust their currency exposure in the opposite direction to the liquidity movement) in normal times but switch to aggressively positive timing (adjust their currency exposure in the same direction to the liquidity movement with increasing aggressivity) in turbulent market conditions. We observed that the dynamics of currency liquidity timing are possibly driven by currency liquidity deterioration and negative shocks to fund returns. We proposed a potential explanation of the dynamic currency-liquidity-timing behavior based on funds' currency hedging practices.

There are many possible avenues for future work. First, an important next step is to dig deeper into whether funds with distinct fund characteristics (e.g., age, size, expense ratio, turnover ratio) differ in their dynamic currency-liquidity-timing behavior. In this respect, one way is to estimate timing coefficients directly for funds grouped by characteristics (see, e.g., Boney et al., 2009; Siegmann and Stefanova, 2017), while an alternative is to run a cross-sectional regression of timing coefficients estimated for the individual funds on certain fund

characteristics (see, e.g., Cao et al., 2013; Chen and Liang, 2007). Under our current modeling framework, the former is implementable whereas the latter calls for a multivariate extension of the proposed Markov-switching structure that can be applied directly to individual funds. Second, a study concerning the implications of dynamic currency-liquidity-timing behavior on fund performance is likely to be promising. For example, by applying the proposed method to fund subgroups sorted by fund performance, the new insight we may bring to this issue is whether funds with superior performance and those with inferior performance behave differently to time currency liquidity.

Appendix 1.A

This appendix gives the technical details of the Bayesian Markov chain Monte Carlo (MCMC) algorithm utilized to estimate the model. We first discuss the specification of prior distributions in Appendix 1.A.1. We then outline the posterior sampling procedure in Appendix 1.A.2. Finally, we provide a detailed derivation of the marginal likelihoods used for model comparison in Appendix 1.A.3. Appendices 1.A.2 and 1.A.3 follow closely Kim and Kang (2022, Appendices A and B) who develop an efficient MCMC algorithm for the estimation of the multivariate endogenous Markov-switching models. We adjust their formulations to accommodate for our model structure.

We use the following notations in this appendix. Consider the timing model in equation (1.9) of the paper, we denote by $\mathbf{Y}_t = \{y_i\}_{i=1}^t$ and $\mathbf{X}_t = \{x_i\}_{i=1}^t$ the observations of the dependent variables and covariates through time t. Besides, we denote by β a vector containing all the non-switching and switching coefficients in the timing model other than σ. Consider the Markov model in equation (1.11) of the paper, we denote by $\mathbf{S}_t = \{s_i\}_{i=0}^t$, and $\mathbf{S}_t^* = \{s_i^*\}_{i=1}^t$ the time series of the regime indicators and the auxiliary variables through time t. Then, we denote by $\mathbf{Z}_t = \{z_i\}_{i=1}^t$ the time series of exogenous information variables through time t. Moreover, we denote by γ a vector containing all the elements in $\gamma_{i,s_{t-1}}$, where $\gamma_{i,s_{t-1}} = (\bar{\gamma}_{i,s_{t-1}}, \gamma^z_{i,s_{t-1}})$ for $i = 1,2,...,N-1$ and $s_{t-1} = 1,2,...,N$. Consider the Gaussian distribution in equation (1.12) of the paper, we denote by ρ a vector of conditional correlations, i.e., $\rho = \{\rho_i\}_{i=1}^{N-1}$. Finally, we let $\theta = \{\beta, \gamma, \rho, \sigma^2\}$ be the collection of all parameters to be estimated in the model.

1.A.1 Prior distributions

We specify a standard set of prior distributions for the model parameters, as shown in Table 1.A.1. We assume that the model parameters are mutually independent, *a priori*.

The prior on β is assumed to be Gaussian with the mean of $\bar{\beta}$ and variance $\bar{V}_\beta$. We set the hyper-parameters of $\bar{\beta}$ and $\bar{V}_\beta$ to be rather uninformative to avoid the case in which the regimes are identified by the prior information rather than by the data, i.e., $\bar{\beta} = 0_{K\times 1}$, $\bar{V}_\beta = I_K$, where $0_{K\times 1}$ is a K-dimensional vector of zeros, I_K is a $K \times K$ identity matrix and K denotes the number of coefficients contained in β.

The prior on γ is also assumed to be Gaussian. Specifically, for each $i = 1,2,...,N-1$, we assume that the vector $(\gamma_{i,s_{t-1}=1}, \gamma_{i,s_{t-1}=2}, ..., \gamma_{i,s_{t-1}=N})'$ are normally distributed with the mean of $\bar{\Gamma}_i$ and variance $\bar{V}_{\Gamma_i}$. As for $\bar{\gamma}_{i,s_{t-1}}$, we impose prior means to be regime-specific and set the hyper-parameters similar as in Kim and Kang (2022), who show that this prior choice is successful in capturing a certain persistence of each regime. Besides, we set the prior variances of $\bar{\gamma}_{i,s_{t-1}}$ to have tight hyper-parameters, which are equal to 0.05 across regimes. As for $\gamma^z_{i,s_{t-1}}$, we impose priors centered on zero and with sharp shrinking (i.e., prior variances

equal to 0.1). These priors show that we are largely uninformative about the effect of the exogenous information variable on the regime transitions.

The prior on σ^2 is assumed to be inverse Gamma with the shape parameter ν_0 and the scale parameter R_0. The formulations of ν_0 and R_0 imply that the prior mean for σ^2 is concentrated around the standard OLS estimate of the error variance for the regression, i.e., $\nu_0 = R_0/\hat{s}+1$, $R_0 = \hat{s}(\hat{s}^2/\sigma^2_{\hat{s}})+1$, where $\hat{s}$ denotes the residual standard error obtained via the OLS estimation and $\sigma^2_{\hat{s}}$ denotes the variance of the residual standard error. We set $\sigma^2_{\hat{s}}$ equal to 0.25, which seems to be appropriate to reflect an ex ante plausible range for the values of $\hat{s}$.

Finally, we assume a beta prior for $0.5\times(\rho_i+1)$, for $i = 1, 2, ..., N-1$. We specify the same hyper-parameters as in Kim and Kang (2022) such that the prior means for the correlation coefficients ρ_i are equal to zero.

Table 1.A.1 Prior specifications

Parameter	Prior distribution	Hyper-parameters
β	$\mathcal{N}(\bar{\beta}, \bar{V}_\beta)$	$\bar{\beta} = 0_{K\times 1}$, $\bar{V}_\beta = I_K$
γ	$\mathcal{N}(\bar{\Gamma}_i, \bar{V}_{\Gamma_i})$	N = 2: $\bar{\Gamma}_1 = (\bar{\gamma}_{1,1}, \gamma^z_{1,1}, \bar{\gamma}_{1,2}, \gamma^z_{1,2})$' = (-1.5, 0, 1.5, 0)',
		$\bar{V}_{\Gamma_1} = I_N \otimes$ diag(0.05,0.1)
		N = 3: $\bar{\Gamma}_1 = (\bar{\gamma}_{1,1}, \gamma^z_{1,1}, \bar{\gamma}_{1,2}, \gamma^z_{1,2}, \bar{\gamma}_{1,3}, \gamma^z_{1,3})$'
		= (-1.5, 0 1.5, 0, 0, 0)', $\bar{V}_{\Gamma_1} = I_N \otimes$ diag(0.05,0.1)
		$\bar{\Gamma}_2 = (\bar{\gamma}_{2,1}, \gamma^z_{2,1}, \bar{\gamma}_{2,2}, \gamma^z_{2,2}, \bar{\gamma}_{2,3}, \gamma^z_{2,3})$'
		= (-2, 0, -0.5, 0, 1.5, 0)', $\bar{V}_{\Gamma_2} = I_N \otimes$ diag(0.05,0.1)
		N = 4: $\bar{\Gamma}_1 = (\bar{\gamma}_{1,1}, \gamma^z_{1,1}, \bar{\gamma}_{1,2}, \gamma^z_{1,2}, \bar{\gamma}_{1,3}, \gamma^z_{1,3}, \bar{\gamma}_{1,4}, \gamma^z_{1,4})$'
		= (-2.5, 0, 1.5, 0, -1.5, 0, 0, 0)', $\bar{V}_{\Gamma_1} = I_N \otimes$ diag(0.05,0.1)
		$\bar{\Gamma}_2 = (\bar{\gamma}_{2,1}, \gamma^z_{2,1}, \bar{\gamma}_{2,2}, \gamma^z_{2,2}, \bar{\gamma}_{2,3}, \gamma^z_{2,3}, \bar{\gamma}_{2,4}, \gamma^z_{2,4})$'
		= (-2, 0, 0.5, 0, 1.5, 0, -0.5, 0)', $\bar{V}_{\Gamma_2} = I_N \otimes$ diag(0.05,0.1)
		$\bar{\Gamma}_3 = (\bar{\gamma}_{3,1}, \gamma^z_{3,1}, \bar{\gamma}_{3,2}, \gamma^z_{3,2}, \bar{\gamma}_{3,3}, \gamma^z_{3,3}, \bar{\gamma}_{3,4}, \gamma^z_{3,4})$'
		= (-1.5, 0, 0, 0, -0.5, 0, 1.5, 0)', $\bar{V}_{\Gamma_3} = I_N \otimes$ diag(0.05,0.1)
σ^2	$\mathcal{IG}(\nu_0, R_0)$	$\nu_0 = R_0/\hat{s}+1$, $R_0 = \hat{s}(\hat{s}^2/\sigma^2_{\hat{s}})+1$
$0.5\times(\rho_i+1)$	$Beta(a_0, b_0)$	$a_0 = 4$, $b_0 = 4$

NOTES: This table presents the prior specifications of the model parameters for the empirical application. All of the model parameters, apart from γ, are given common hyper-parameters across regimes.

1.A.2 Posterior sampling

Let $\theta_{-\beta}$ denote the collection of model parameters, excluding β. Similarly, let $\theta_{-\gamma}$, $\theta_{-\rho}$, and $\theta_{-\sigma^2}$ denote the collections of model parameters except γ, ρ, and σ^2, respectively. Our posterior sampling procedure is summarized as follows:

Algorithm 1: Posterior Sampling Procedure

Step 1: Sample $\mathbf{S}_T|\mathbf{Y}_T,\mathbf{X}_T,\mathbf{Z}_T,\theta$.
Step 2: Sample $\mathbf{S}_T^*|\mathbf{S}_T,\mathbf{Y}_T,\mathbf{X}_T,\mathbf{Z}_T,\theta$.
Step 3: Sample $\beta|\mathbf{S}_T,\mathbf{S}_T^*,\mathbf{Y}_T,\mathbf{X}_T,\mathbf{Z}_T,\theta_{-\beta}$.
Step 4: Sample $\gamma|\mathbf{S}_T,\mathbf{S}_T^*,\mathbf{Y}_T,\mathbf{Z}_T,\theta_{-\gamma}$.
Step 5: Sample $\rho|\mathbf{S}_T,\mathbf{S}_T^*,\mathbf{Y}_T,\mathbf{X}_T,\mathbf{Z}_T,\theta_{-\rho}$.
Step 6: Sample $\sigma^2|\mathbf{S}_T,\mathbf{S}_T^*,\mathbf{Y}_T,\mathbf{X}_T,\mathbf{Z}_T,\theta_{-\sigma^2}$.

The details of each step are provided in the following subsections.

1.A.2.1 Sampling regime indicators

We sample the time series of the regime indicators from its full conditional distribution based on the multi-move method proposed by Chib (1998). This method is implemented in two stages. In the first stage, we carry out a forward recursion to obtain the filtered probability, denoted by $\Pr(s_t|\mathbf{Y}_t,\mathbf{X}_t,\mathbf{Z}_t,\theta)$ for $t=1,2,...,T$. In the second stage, we operate a backward recursion, which is initialized by sampling the regime at time T, s_T, from its filtered probability $\Pr(s_T|\mathbf{Y}_T,\mathbf{X}_T,\mathbf{Z}_T,\theta)$. We then sample the remaining regimes s_t for $t=T-1,T-2,...,1,0$, using its full conditional distribution, denoted by $\Pr(s_t|s_{t+1},\mathbf{Y}_{t+1},\mathbf{X}_t,\mathbf{Z}_t,\theta)$. We describe the details as follows:

Algorithm 2: Multi-Move Sampling of the Regimes
First stage: Forward Recursion
Step 1: At $t=0$, compute the initial filtered probability. Following the usual practice (see, e.g., Hamilton, 1989; Hwu et al., 2021; Kim and Kang, 2022), we use the unconditional probabilities of the initial regime s_0 as the initial filtered probability. Specifically, we denote by P the unconditional transition probability matrix, in which the (i,j)th element is $\Pr(s_t=j|s_{t-1}=i,z_t,\theta)$ obtained in equations (1.B.9) and (1.B.11). Following Amisano and Fagan (2013), we assume $z_t=0$ for $t=1,2,..,T$. Given P, the initial filtered probability is obtained as

$$p(s_0=j|\theta)=(\bar{P}'\bar{P})^{-1}\bar{P}'\begin{bmatrix}0_{N\times 1}\\1\end{bmatrix}\text{ where }\bar{P}=\begin{bmatrix}I_N-P'\\1_N'\end{bmatrix}, \tag{1.A.1}$$

where N represents the number of regimes, I_N is an $N\times N$ identity matrix, and $1_N'$ is an $N\times 1$ vector of ones. After initializing the filtered probability at time 0, we implement Steps 2 to 6 recursively for $t=1,2,...,T$.
Step 2: For $j,q=1,2,...,N$, compute the conditional joint density of (y_t,s_t),

$$p(y_t,s_t=j|s_{t-1}=q,\mathbf{Y}_{t-1},\mathbf{X}_t,\mathbf{Z}_t,\theta)=\mathcal{N}(y_t|x_t'\beta_j,\sigma^2)\Pr(s_t=j|s_{t-1}=q,\varepsilon_{p,t},z_t,\theta), \tag{1.A.2}$$

where $\Pr(s_t=j|s_{t-1}=q,\varepsilon_{p,t},\theta)$ is the conditional transition probability derived from equations (1.B.4) and (1.B.6).

Step 3: For $j = 1, 2, ..., N$, compute the joint predictive density of (y_t, s_t),

$$p(y_t, s_t = j|\mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta) = \sum_{q=1}^{N} \left[p(y_t, s_t = j|s_{t-1} = q, \mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta) \Pr(s_{t-1} = q|\mathbf{Y}_{t-1}, \mathbf{X}_{t-1}, \mathbf{Z}_{t-1}, \theta) \right], \tag{1.A.3}$$

where $\Pr(s_{t-1} = q|\mathbf{Y}_{t-1}, \mathbf{X}_{t-1}, \mathbf{Z}_{t-1}, \theta)$ is the filtered probability at time $t-1$.
Step 4: Compute the predictive density,

$$p(y_t|\mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta) = \sum_{j=1}^{N} p(y_t, s_t = j|\mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta). \tag{1.A.4}$$

Step 5: For $j, q = 1, 2, ..., N$, compute the joint conditional density of (s_t, s_{t-1}),

$$\Pr(s_t = j, s_{t-1} = q|\mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta) = \frac{p(y_t, s_t = j|s_{t-1} = q, \mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta) \Pr(s_{t-1} = q|\mathbf{Y}_{t-1}, \mathbf{X}_{t-1}, \mathbf{Z}_{t-1}, \theta)}{p(y_t|\mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta)}. \tag{1.A.5}$$

Step 6: For $j = 1, 2, ..., N$, compute the filtered probability,

$$\Pr(s_t = j|\mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta) = \sum_{q=1}^{N} \Pr(s_t = j, s_{t-1} = q|\mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta). \tag{1.A.6}$$

Second stage: Backward Recursion
Step 1: Sample the regime at time T, s_T, from its filtered probability $\Pr(s_T|\mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta)$ and repeat Step 2 for $t = T-1, T-2, ..., 1$.
Step 2: For $j, q = 1, 2, ..., N$, sample $s_t = j$ using the full conditional distribution,

$$\Pr(s_t = j|s_{t+1} = q, \mathbf{Y}_{t+1}, \mathbf{X}_t, \mathbf{Z}_t, \theta) = \frac{\Pr(s_{t+1} = q, s_t = j|\mathbf{Y}_{t+1}, \mathbf{X}_{t+1}, \mathbf{Z}_{t+1}, \theta)}{\Pr(s_{t+1} = q|\mathbf{Y}_{t+1}, \mathbf{X}_{t+1}, \mathbf{Z}_{t+1}, \theta)}, \tag{1.A.7}$$

where $\Pr(s_{t+1} = q, s_t = j|\mathbf{Y}_{t+1}, \mathbf{X}_{t+1}, \mathbf{Z}_{t+1}, \theta)$ is the joint conditional density of (s_{t+1}, s_t) and $\Pr(s_{t+1} = q|\mathbf{Y}_{t+1}, \mathbf{X}_{t+1}, \mathbf{Z}_{t+1}, \theta)$ is the filtered probability of s_{t+1}. Both are obtained in Steps 5 and 6 of the forward recursion, respectively.

1.A.2.2 Sampling auxiliary variables

We sample the time series of the auxiliary variables from its full conditional distribution, denoted by $p(s_t^*|s_t, s_{t-1}, \mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta)$. As shown in equation (1.10) of the paper and equation (1.B.2), s_t^* is composed by $\left\{ s_{i,t}^* \right\}_{i=1}^{N-1}$, which are mutually independent given the error term $\varepsilon_{p,t}$. Therefore, the full conditional distribution $p(s_t^*|s_t, s_{t-1}, \mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta)$ is simply given by the product of $\left\{ p(s_{i,t}^*|s_t, s_{t-1}, \mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta) \right\}_{i=1}^{N-1}$. For $h = 0, 1, ..., i-1, i+1, ..., N-1$, the full

conditional density of $p(s^*_{i,t}|s_t, s_{t-1}, \mathbf{Y}_t, \mathbf{X}_t, \mathbf{Z}_t, \theta)$ is proportional to

$$\begin{aligned}&\mathcal{N}(s^*_{i,t}|\bar{\gamma}_{i,s_{t-1}} + z'_t\gamma^z_{i,s_{t-1}} + e'_i\rho\varepsilon_{p,t}, 1-\rho_i^2)\times \\ &[\mathbf{I}(s^*_{i,t} > s^*_{h,t}|s_t)\Pr(\mathbf{I}(s^*_{i,t} > s^*_{h,t}|s_t)) + \mathbf{I}(s^*_{i,t} \le s^*_{h,t}|s_t)\Pr(\mathbf{I}(s^*_{i,t} \le s^*_{h,t}|s_t))],\end{aligned} \tag{1.A.8}$$

where e_i is a vector of zeros with a one in the ith position and $\mathbf{I}(\cdot)$ is an indicator function.

The above equation has different functional forms depending on the realization of s_t. Consider the case that $s_t = j$, where $j = 1$, all auxiliary variables should be negative. This reflects in equation (1.A.8) that $\mathbf{I}(s^*_{q,t} > s^*_{0,t}|s_t) = 0$, $\Pr(\mathbf{I}(s^*_{q,t} > s^*_{0,t}|s_t)) = 0$, $\mathbf{I}(s^*_{q,t} \le s^*_{0,t}|s_t) = 1$, and $\Pr(\mathbf{I}(s^*_{q,t} \le s^*_{0,t}|s_t)) = 1$ for $q = 1, 2, ..., N-1$, $s^*_{0,t} = 0$. Thus, the full conditional density of $s^*_{q,t}$ is given by

$$\mathcal{N}(s^*_{q,t}|\bar{\gamma}_{1,s_{t-1}} + z'_t\gamma^z_{1,s_{t-1}} + e'_1\rho\varepsilon_{p,t}, 1-\rho_1^2) \times \mathbf{I}(s^*_{q,t} \le s^*_{0,t}|s_t). \tag{1.A.9}$$

This indicates that $s^*_{q,t}$ are sampled from the truncated conditional normal distributions over $(-\infty,0]$

$$s^*_{q,t}|s_t = 1, s_{t-1}, \varepsilon_{p,t}, z_t, \theta \sim \mathcal{TN}_{(-\infty,0]}(\bar{\gamma}_{q,s_{t-1}} + z'_t\gamma^z_{q,s_{t-1}} + e'_q\rho\varepsilon_{p,t}, 1-\rho_q^2). \tag{1.A.10}$$

For $2 \le s_t = j+1 \le N$, the jth auxiliary variables should be positive. Thus, the full conditional density is given by

$$\mathcal{N}(s^*_{j,t}|\bar{\gamma}_{j,s_{t-1}} + z'_t\gamma^z_{j,s_{t-1}} + e'_j\rho\varepsilon_{p,t}, 1-\rho_j^2) \times \mathbf{I}(s^*_{j,t} > s^*_{0,t}|s_t). \tag{1.A.11}$$

This indicates that the jth auxiliary variables are sampled from the truncated conditional normal distributions over $(0, \infty]$

$$s^*_{j,t}|s_t = j+1, s_{t-1}, \varepsilon_{p,t}, z_t, \theta \sim \mathcal{TN}_{(0,\infty]}(\bar{\gamma}_{j,s_{t-1}} + z'_t\gamma^z_{j,s_{t-1}} + e'_j\rho\varepsilon_{p,t}, 1-\rho_j^2). \tag{1.A.12}$$

Besides, the jth auxiliary variables should be greater than the other auxiliary variables. Thus, given the sample of $s^*_{j,t}$, the full conditional density of the remaining variables is given by

$$\mathcal{N}(s^*_{h,t}|\bar{\gamma}_{h,s_{t-1}} + z'_t\gamma^z_{h,s_{t-1}} + e'_h\rho\varepsilon_{p,t}, 1-\rho_h^2) \times \mathbf{I}(s^*_{h,t} < s^*_{j,t}|s_t), \tag{1.A.13}$$

for $h = 1, 2, ..., j-1, j+1, ..., N-1$. This indicates that $s^*_{h,t}$ are sampled from the truncated conditional normal distributions over $(-\infty, s^*_{j,t}]$

$$s^*_{h,t}|s_t = j+1, s_{t-1}, s^*_{j,t}, \varepsilon_{p,t}, z_t, \theta \sim \mathcal{TN}_{(-\infty,s^*_{j,t}]}(\bar{\gamma}_{h,s_{t-1}} + z'_t\gamma^z_{h,s_{t-1}} + \rho'_h\varepsilon_{p,t}, 1-\rho_h^2). \tag{1.A.14}$$

We repeat the above procedures for $t = 1, 2, ..., T$.

1.A.2.3 Sampling β

We sample β from its full conditional distribution. To begin with, we use the joint distribution of $(s_t^*, \varepsilon_{p,t})$ in equation (1.12) of the paper to derive the conditional distribution of $\varepsilon_{p,t}$, which is given by

$$\varepsilon_{p,t}|s_t^*, s_{t-1}, \mathbf{Y}_{t-1}, z_t, \theta \sim \mathcal{N}(\rho'\Omega^{-1}(s_t^* - \bar{\gamma}_{s_{t-1}} - z_t'\gamma_{s_{t-1}}^z), 1 - \rho'\Omega^{-1}\rho). \tag{1.A.15}$$

It follows that the conditional distribution of y_t is given by

$$y_t|\mathbf{S}_t, \mathbf{S}_t^*, \mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta \sim \mathcal{N}(x_t'\beta_{s_t} + \sigma\rho'\Omega^{-1}(s_t^* - \bar{\gamma}_{s_{t-1}} - z_t'\gamma_{s_{t-1}}^z), \sigma(1 - \rho'\Omega^{-1}\rho)\sigma'). \tag{1.A.16}$$

This can be rewritten as

$$y_t^*|\mathbf{S}_t, \mathbf{Y}_{t-1}, \mathbf{X}_t, \theta \sim \mathcal{N}(x_t'\beta_{s_t}, \sigma(1 - \rho'\Omega^{-1}\rho)\sigma'), \tag{1.A.17}$$

where $y_t^* = y_t - \sigma\rho'\Omega^{-1}(s_t^* - \bar{\gamma}_{s_{t-1}} - z_t'\gamma_{s_{t-1}}^z)$. equation (1.A.17) suggests that the conditional distribution of y_t^* is linear to β_{s_t}. Therefore, given the normal prior distribution of β_{s_t}, the full conditional distribution of $\beta = (\beta_{s_t=1}', \beta_{s_t=2}', ..., \beta_{s_t=N}')'$ can be written as

$$\beta|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\beta} \sim \mathcal{N}(B_1 A, B_1), \tag{1.A.18}$$

with

$$\begin{aligned}
X_t &= \left[(\mathbf{I}(s_t = 1)\ \mathbf{I}(s_t = 2)\ \cdots\ \mathbf{I}(s_t = N)) \otimes x_t'\right]', \\
B_1 &= (\bar{V}_\beta^{-1} + \sum_{t=1}^{T} X_t(\sigma(1 - \rho'\Omega^{-1}\rho)\sigma')^{-1}X_t')^{-1}, \\
A &= (\bar{V}_\beta^{-1}\bar{\beta} + \sum_{t=1}^{T} X_t(\sigma(1 - \rho'\Omega^{-1}\rho)\sigma')^{-1}y_t^*).
\end{aligned}$$

1.A.2.4 Sampling γ

We sample γ from its full conditional distribution. To begin with, we rewrite equation (1.B.2) as

$$s_{i,t}^*|s_{t-1}, \varepsilon_{p,t}, z_t, s_t, \theta \sim \mathcal{N}(\bar{\gamma}_{i,s_{t-1}} + z_t'\gamma_{i,s_{t-1}}^z + e_i'\rho\varepsilon_{p,t}, 1 - \rho_i^2). \tag{1.A.19}$$

It follows that

$$s_{i,t}^* - e_i'\rho\varepsilon_{p,t}|s_{t-1}, \varepsilon_{p,t}, z_t, s_t, \theta \sim \mathcal{N}(\bar{\gamma}_{i,s_{t-1}} + z_t'\gamma_{i,s_{t-1}}^z, 1 - \rho_i^2). \tag{1.A.20}$$

The above equation suggests that the conditional distribution of $s_{i,t}^* - e_i'\rho\varepsilon_{p,t}$ is linear to $\bar{\gamma}_{i,s_{t-1}}$, $\gamma_{i,s_{t-1}}^z$, and the vector $\gamma_{i,s_{t-1}}$ defined at the beginning of this appendix. Therefore,

for each $i = 1, 2, ..., N-1$, the full conditional distribution of $\gamma = (\gamma_{i,1}, \gamma_{i,2}, ..., \gamma_{i,N})'$ can be written as

$$\gamma | \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{Y}_T, \mathbf{Z}_T, \theta_{-\gamma} \sim \mathcal{N}(B_{i,1} A_i, B_{i,1}), \tag{1.A.21}$$

with

$$\begin{aligned}
\mathbf{I}_t &= \left[(\mathbf{I}(s_{t-1} = 1)\ \mathbf{I}(s_{t-1} = 2)\ \cdots\ \mathbf{I}(s_{t-1} = N)) \otimes x_t' \right]', \\
x_t &= [1, z_t]', \\
B_{i,1} &= (\sum_{t=1}^{T} (1 - \rho_i^2)^{-1} \mathbf{I}_t \mathbf{I}_t' + \bar{V}_{\Gamma_i}^{-1})^{-1}, \\
A_i &= (\sum_{t=1}^{T} (1 - \rho_i^2)^{-1} \mathbf{I}_t (s_{i,t}^* - e_i' \rho \varepsilon_{p,t}) + \bar{V}_{\Gamma_i}^{-1} \bar{\Gamma}_i).
\end{aligned}$$

1.A.2.5 Sampling ρ

We sample ρ from its full conditional distribution $p(\rho | \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\rho})$, which is proportional to $p(\mathbf{S}_T^*, \mathbf{Y}_T | \mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta) \times \pi(\rho)$. The prior density $\pi(\rho)$ is computed as

$$\pi(\rho) = \prod_{i=1}^{N-1} \text{Beta}((\rho_i + 1)/2 | a_0, b_0), \tag{1.A.22}$$

where $\text{Beta}(\cdot | a_0, b_0)$ is the beta distribution with the hyper-parameters (a_0, b_0). From equation (1.A.16), the joint conditional density $p(\mathbf{S}_T^*, \mathbf{Y}_T | \mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta)$ is given by

$$\begin{aligned}
p(\mathbf{S}_T^*, \mathbf{Y}_T | \mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta) &= \prod_{t=1}^{T} p(y_t, s_t^* | s_t, s_{t-1}, \mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta) \\
&= \prod_{t=1}^{T} \mathcal{N}\left(\begin{bmatrix} s_t^* \\ \sigma^{-1}(y_t - x_t' \beta_{s_t}) \end{bmatrix} | \begin{pmatrix} \bar{\gamma}_{s_{t-1}} + z_t' \gamma_{s_{t-1}}^z \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega & \rho \\ \rho' & 1 \end{pmatrix} \right).
\end{aligned} \tag{1.A.23}$$

However, the computation of $p(\rho | \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\rho})$ is not feasible because Ω is non-linear to ρ, as shown in equation (1.13) of the paper. Therefore, we sample ρ by using a Metropolis-Hastings algorithm, described as follows:

Algorithm 3: Sampling ρ Through a Metropolis–Hastings Algorithm
Step 1: Let $\rho^{(g-1)}$ denote the $(g-1)$th MCMC draw of ρ. At each MCMC iteration, repeat the following steps.
Step 2: Maximize the log full conditional density with respect to ρ with the constraint

$$1 - \rho' \Omega^{-1} \rho > 0. \tag{1.A.24}$$

Step 3: Find its mode, $\bar{\rho}$, and compute the inverse of the negative hessian matrix of the log full conditional density at $\bar{\rho}$

$$V_p = (-\frac{\partial^2 \ln\{p(\mathbf{S}_T^*, \mathbf{Y}_T|\mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \bar{\rho}, \theta_{-\rho})\pi(\bar{\rho})\}}{\partial\rho\partial\rho'})^{-1}. \tag{1.A.25}$$

Step 4: Draw a candidate, denoted by ρ^*, from the multivariate normal distribution with mean $\bar{\rho}$, variance V_p, and domain$(-1,1)$

$$\rho^* \sim \mathcal{N}_{(-1,1)}(\bar{\rho}, V_p). \tag{1.A.26}$$

Step 5: If ρ^* violates the constraint in equation (1.A.24) it is rejected immediately. Otherwise, calculate the acceptance probability

$$\begin{aligned} &\alpha(\rho^{(g-1)}, \rho^*|\mathbf{Y}_T, \mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\rho}) \qquad (1.A.27)\\ &= \min\left\{1, \frac{p(\mathbf{S}_T^*, \mathbf{Y}_T|\mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \rho^*, \theta_{-\rho}) \times \text{Beta}(\rho^*|a_0, b_0) \times \mathcal{N}_{(-1,1)}(\rho^{(g-1)}|\bar{\rho}, V_p)}{p(\mathbf{S}_T^*, \mathbf{Y}_T|\mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \rho^{(g-1)}, \theta_{-\rho}) \times \text{Beta}(\rho^{(g-1)}|a_0, b_0) \times \mathcal{N}_{(-1,1)}(\rho^*|\bar{\rho}, V_p)}\right\}. \end{aligned}$$

Step 6: Generate a uniform random number $u \in [0,1]$. if $u \leq \alpha(\rho^{(g-1)}, \rho^*|\mathbf{Y}_T, \mathbf{S}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\rho})$, then accept the candidate draw and set $\rho^{(g)} = \rho^*$. Otherwise, reject the candidate draw and set $\rho^{(g)} = \rho^{(g-1)}$.

1.A.2.6 Sampling σ^2

We sample σ^2 from its full conditional distribution $p(\sigma^2|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\sigma^2})$, which is proportional to $p(\mathbf{Y}_T|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \theta) \times \mathcal{IG}(\sigma^2|\nu_0, R_0)$, where

$$\begin{aligned} p(\mathbf{Y}_T|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \theta) &= \prod_{t=1}^{T} p(y_t|s_t, s_t^*, \mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \theta)\\ &= \prod_{t=1}^{T} \mathcal{N}(y_t^*|x_t'\beta_{s_t}, \sigma(1-\rho'\Omega^{-1}\rho)\sigma'). \end{aligned} \tag{1.A.28}$$

However, the computation of $p(\sigma^2|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\sigma^2})$ is not feasible because σ appears in both the conditional mean and the variance of the observations, as shown in equation (1.A.16). Therefore, similar to ρ, we sample σ^2 by using a Metropolis–Hastings Algorithm, described as follows:

Algorithm 4: Sampling σ^2 Through a Metropolis–Hastings Algorithm
Step 1: Let $\sigma^{2,(g-1)}$ denote the $(g-1)$th MCMC draw of σ^2. At each MCMC iteration, repeat the following steps.
Step 2: Maximize the log full conditional density with respect to σ^2 with the positive constraint.

Step 3: Find its mode, $\bar{\sigma}^2$, and compute the inverse of the negative hessian matrix of the log full conditional density at $\bar{\sigma}^2$

$$V_{\sigma^2} = (-\frac{\partial^2 \ln\{p(\mathbf{Y}_T|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \bar{\sigma}^2, \theta_{-\sigma^2})\pi(\bar{\sigma}^2)\}}{\partial\sigma^2\partial(\sigma^2)'})^{-1}. \tag{1.A.29}$$

Step 4: Draw a candidate, denoted by $\sigma^{2,*}$, from the multivariate normal distribution with mean $\bar{\sigma}^2$, variance V_{σ^2}, and the positive definite constraint

$$\sigma^{2,*} \sim \mathcal{N}(\bar{\sigma}^2, V_{\sigma^2}). \tag{1.A.30}$$

Step 5: If $\sigma^{2,*}$ violates the positive constraint it is rejected immediately. Otherwise, calculate the acceptance probability

$$\begin{aligned} &\alpha(\sigma^{2,(g-1)}, \sigma^{2,*}|\mathbf{Y}_T, \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\sigma^2}) \qquad (1.A.31)\\ &= \min\left\{1, \frac{p(\mathbf{Y}_T|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \sigma^{2,*}, \theta_{-\sigma^2}) \times \mathcal{IG}(\sigma^{2,*}|\nu_0, R_0) \times \mathcal{N}(\sigma^{2,(g-1)}|\bar{\sigma}^2, V_{\sigma^2})}{p(\mathbf{Y}_T|\mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \sigma^{2,(g-1)}, \theta_{-\sigma^2}) \times \mathcal{IG}(\sigma^{2,(g-1)}|\nu_0, R_0) \times \mathcal{N}(\sigma^{2,*}|\bar{\sigma}^2, V_{\sigma^2})}\right\}. \end{aligned}$$

Step 6: Generate a uniform random number $u \in [0,1]$. if $u \leq \alpha(\sigma^{2,(g-1)}, \sigma^{2,*}|\mathbf{Y}_T, \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \theta_{-\sigma^2})$, then accept the candidate draw and set $\sigma^{2,(g)} = \sigma^{2,*}$. Otherwise, reject the candidate draw and set $\sigma^{2,(g)} = \sigma^{2,(g-1)}$.

1.A.3 Marginal likelihood computation

The log marginal likelihood of model $\mathcal{M}$, denoted by $\log p(\mathbf{Y}_T|\mathcal{M})$, is computed as

$$\log p(\mathbf{Y}_T|\mathcal{M}) = \log p(\mathbf{Y}_T|\hat{\theta}, \mathcal{M}) + \log \pi(\hat{\theta}|\mathcal{M}) - \log \pi(\hat{\theta}|\mathbf{Y}_T, \mathcal{M}), \tag{1.A.32}$$

where $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\rho}, \hat{\sigma}^2)$ is the posterior mode. The first term of equation (1.A.32) is equal to the sum of the log conditional densities of the observations given in equation (1.A.4). That is,

$$\log p(\mathbf{Y}_T|\hat{\theta}, \mathcal{M}) = \sum_{t=1}^{T} \log p(y_t|\mathbf{Y}_{t-1}, \mathbf{X}_t, \mathbf{Z}_t, \hat{\theta}, \mathcal{M}). \tag{1.A.33}$$

The second term is the prior density at $\hat{\theta}$, which is equal to the sum of all log prior distributions of the parameters, as defined in Appendix 1.A.1. The third term is the log posterior density of $\hat{\theta}$, which can be obtained from a marginal conditional decomposition proposed by Chib and Jeliazkov (2001),

$$\begin{aligned} \log \pi(\hat{\theta}|\mathbf{Y}_T, \mathcal{M}) = &\log \pi(\hat{\beta}|\mathbf{Y}_T, \hat{\theta}_{-\beta}, \mathcal{M}) + \log \pi(\hat{\gamma}|\mathbf{Y}_T, \hat{\sigma}^2, \hat{\rho}, \mathcal{M}) \\ &+ \log \pi(\hat{\rho}|\mathbf{Y}_T, \hat{\sigma}^2, \mathcal{M}) + \log \pi(\hat{\sigma}^2|\mathbf{Y}_T, \mathcal{M}). \end{aligned} \tag{1.A.34}$$

The details of each term of equation (1.A.34) are discussed in the following subsections.

1.A.3.1 $\hat{\beta}$ conditional density

The first term $\pi(\hat{\beta}|\mathbf{Y}_T, \hat{\theta}_{-\beta}, \mathcal{M})$ of equation (1.A.34) is computed by integrating the joint density $\pi(\hat{\beta}, \mathbf{S}_T, \mathbf{S}_T^*|\mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \hat{\theta}_{-\beta}, \mathcal{M})$ over $(\mathbf{S}_T, \mathbf{S}_T^*)$. That is,

$$\begin{aligned}\pi(\hat{\beta}|\mathbf{Y}_T, \hat{\theta}_{-\beta}, \mathcal{M}) &= \int \pi(\hat{\beta}|\mathbf{Y}_T, \hat{\theta}_{-\beta}, \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \mathcal{M}) p(\mathbf{S}_T, \mathbf{S}_T^*|\mathbf{Y}_T, \mathbf{X}_T, \mathbf{Z}_T, \hat{\theta}_{-\beta}, \mathcal{M}) d(\mathbf{S}_T, \mathbf{S}_T^*) \\ &\approx n_1^{-1} \sum_{g=1}^{n_1} \mathcal{N}(\hat{\beta}|B_1^{(g)} A^{(g)}, B_1^{(g)}),\end{aligned} \tag{1.A.35}$$

with

$$\begin{aligned} y_t^{*,(g)} &= y_t - \hat{\sigma}\hat{\rho}'\hat{\Omega}^{-1}(s_t^{*,(g)} - \bar{\gamma}_{s_{t-1}^{(g)}} - z_t'\gamma^z_{s_{t-1}^{(g)}}), \\ B_1^{(g)} &= (\bar{V}_\beta^{-1} + \sum_{t=1}^{T} X_t(\hat{\sigma}(1 - \hat{\rho}'\Omega^{-1}\hat{\rho})\hat{\sigma}')^{-1} X_t')^{-1}, \\ A^{(g)} &= (\bar{V}_\beta^{-1}\bar{\beta} + \sum_{t=1}^{T} X_t(\hat{\sigma}(1 - \hat{\rho}'\Omega^{-1}\hat{\rho})\hat{\sigma}')^{-1} y_t^{*,(g)}). \end{aligned}$$

where the superscript (g) denotes the gth draw from reduced MCMC runs and n_1 denotes the MCMC size beyond the burn-in. Note that $(\beta, \mathbf{S}_T, \mathbf{S}_T^*)$ are simulated from the reduced MCMC run while (γ, ρ, σ^2) are fixed at their mode, $(\hat{\gamma}, \hat{\rho}, \hat{\sigma}^2)$.

1.A.3.2 $\hat{\gamma}$ conditional density

The second term $\pi(\hat{\gamma}|\mathbf{Y}_T, \hat{\sigma}^2, \hat{\rho}, \mathcal{M})$ of equation (1.A.34) is computed in a similar way of $\hat{\beta}$ conditional density, which is given by

$$\begin{aligned}\pi(\hat{\gamma}|\mathbf{Y}_T, \hat{\sigma}^2, \hat{\rho}, \mathcal{M}) &= \int \pi(\hat{\gamma}|\mathbf{Y}_T, \hat{\rho}, \hat{\sigma}^2, \beta, \mathbf{S}_T, \mathbf{S}_T^*, \mathbf{X}_T, \mathbf{Z}_T, \mathcal{M}) p(\beta, \mathbf{S}_T, \mathbf{S}_T^*|\mathbf{Y}_T, \hat{\rho}, \hat{\sigma}^2, \mathbf{X}_T, \mathbf{Z}_T, \mathcal{M}) d(\beta, \mathbf{S}_T, \mathbf{S}_T^*) \\ &\approx n_1^{-1} \sum_{g=1}^{n_1} \mathcal{N}(\hat{\gamma}|B_{i,1}^{(g)} A_i^{(g)}, B_{i,1}^{(g)}),\end{aligned} \tag{1.A.36}$$

with

$$B_{i,1}^{(g)} = (\sum_{t=1}^{T}(1-\hat{\rho}_i^2)^{-1}\mathbf{I}_t\mathbf{I}_t' + \bar{V}_{\Gamma_i}^{-1})^{-1},$$
$$\varepsilon_t^{(g)} = \hat{\sigma}^{-1}(y_t - x_t'\beta_{s_t^{(g)}}^{(g)}),$$
$$A_i^{(g)} = (\sum_{t=1}^{T}(1-\hat{\rho}_i^2)^{-1}\mathbf{I}_t(s_{i,t}^{*,(g)} - e_i'\hat{\rho}\varepsilon_{p,t}^{(g)}) + \bar{V}_{\Gamma_i}^{-1}\bar{\Gamma}_i).$$

Here, $(\gamma, \beta, \mathbf{S}_T, \mathbf{S}_T^*)$ are simulated from the reduced MCMC run while (ρ, σ^2) are fixed at their mode, $(\hat{\rho}, \hat{\sigma}^2)$.

1.A.3.3 $\hat{\rho}$ conditional density

The third term $\pi(\hat{\rho}|\mathbf{Y}_T, \hat{\sigma}^2, \mathcal{M})$ of equation (1.A.34) is computed based on the algorithm of Chib and Jeliazkov (2001). We first write $\pi(\hat{\rho}|\mathbf{Y}_T, \hat{\sigma}^2, \mathcal{M})$ as

$$\pi(\hat{\rho}|\mathbf{Y}_T, \hat{\sigma}^2, \mathcal{M}) = \frac{\int \alpha(\rho^{(g)}, \hat{\rho}|\mathbf{Y}_T, \hat{\sigma}^2) \times \mathcal{N}(\hat{\rho}|\bar{\rho}, V_\rho) \times \pi(\rho|\mathbf{Y}_T, \hat{\sigma}^2) d\rho}{\int \alpha(\hat{\rho}, \rho|\mathbf{Y}_T, \hat{\sigma}^2) \times \mathcal{N}(\rho|\bar{\rho}, V_\rho) d\rho}. \tag{1.A.37}$$

It follows that a simulation-consistent estimate of $\pi(\hat{\rho}|\mathbf{Y}_T, \hat{\sigma}^2, \mathcal{M})$ is given by

$$\frac{n_1^{-1}\sum_{g=1}^{n_1}\left[\alpha(\rho^{(g)}, \hat{\rho}|\mathbf{Y}_T, \mathbf{S}_T^{(g)}, \beta^{(g)}, \gamma^{(g)}, \hat{\sigma}^2) \times \mathcal{TN}_{(-1,1)}(\hat{\rho}|\bar{\rho}, V_\rho)\right]}{n_1^{-1}\sum_{g=1}^{n_1}\left[\alpha(\hat{\rho}, \rho^{(g)}|\mathbf{Y}_T, \mathbf{S}_T^{(g)}, \beta^{(g)}, \gamma^{(g)}, \hat{\sigma}^2)\right]}. \tag{1.A.38}$$

For the numerator, $(\mathbf{S}_T^{(g)}, \beta^{(g)}, \gamma^{(g)}, \rho^{(g)})$ are drawn via the reduced MCMC runs, where σ^2 is fixed at $\hat{\sigma}^2$. For the denominator, $(\mathbf{S}_T^{(g)}, \beta^{(g)}, \gamma^{(g)})$ are simulated from the reduced MCMC runs with the fixed value $(\hat{\rho}, \hat{\sigma}^2)$, and $\rho^{(g)}$ is sampled from its proposal distribution, $\mathcal{N}(\bar{\rho}, V_\rho)$.

1.A.3.4 $\hat{\sigma}^2$ conditional density

Similar to the conditional density of $\hat{\rho}$, the last term $\pi(\hat{\sigma}^2|\mathbf{Y}_T, \mathcal{M})$ of equation (1.A.34) is approximated as

$$\frac{n_1^{-1}\sum_{g=1}^{n_1}\left[\alpha(\sigma^{2,(g)}, \hat{\sigma}^2|\mathbf{Y}_T, \mathbf{S}_T^{(g)}, \theta_{-\sigma^2}^{(g)}) \times \mathcal{IG}(\hat{\sigma}^2|\bar{\sigma}^2, V_{\sigma^2})\right]}{n_1^{-1}\sum_{g=1}^{n_1}\left[\alpha(\hat{\sigma}^2, \sigma^{2,(g)}|\mathbf{Y}_T, \mathbf{S}_T^{(g)}, \theta_{-\sigma^2}^{(g)})\right]}. \tag{1.A.39}$$

$(\mathbf{S}_T^{(g)}, \sigma^{2,(g)}, \theta_{-\sigma^2}^{(g)})$ in the numerator are the outputs generated from the reduced MCMC runs. $(\mathbf{S}_T^{(g)}, \theta_{-\sigma^2}^{(g)})$ in the denominator are simulated from the reduced MCMC runs with the fixed value $\hat{\sigma}^2$.

Appendix 1.B

This appendix derives the transition probabilities of the regimes, which are a generalization of Hwu et al. (2021, Appendix B).

As shown in equation (1.12) of the paper, the unconditional distribution of s_t^* is given by

$$s_t^*|s_{t-1}, \mathbf{Y}_{t-1}, z_t, \theta \sim \mathcal{N}(\bar{\gamma}_{s_{t-1}} + z_t'\gamma_{s_{t-1}}^z, \Omega). \tag{1.B.1}$$

Similarly, the conditional distribution of s_t^* is given by

$$s_t^*|s_{t-1}, \mathbf{Y}_{t-1}, \varepsilon_{p,t}, z_t, \theta \sim \mathcal{N}(\bar{\gamma}_{s_{t-1}} + z_t'\gamma_{s_{t-1}}^z + \rho\varepsilon_{p,t}, R), \tag{1.B.2}$$

where

$$R = \begin{pmatrix} 1-\rho_1^2 & 0 & \cdots & 0 \\ 0 & 1-\rho_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1-\rho_{N-1}^2 \end{pmatrix}. \tag{1.B.3}$$

This suggests that s_t^* are mutually uncorrelated, conditioned on $\varepsilon_{p,t}$.

Given s_t^*, we now compute the transition probabilities of the regimes. We first derive the conditional transition probabilities of the regimes, denoted by $\Pr(s_t|s_{t-1}, \varepsilon_{p,t}, z_t, \theta)$. Next, we derive the unconditional transition probabilities of the regimes, denoted by $\Pr(s_t|s_{t-1}, z_t, \theta)$.

We start by using the parameterization in equation (1.10) of the paper and the conditional distribution of s_t^* in equation (1.B.2) to derive the conditional transition probabilities of the regimes. For $s_t = 1$, all auxiliary variables in equation (10) should be negative. Thus, given the conditional independence in equation (1.B.2), the conditional transition probability is equal to the product of the probabilities that $s_{i,t}^*$ is negative for all $i = 1, 2, ..., N-1$. Formally, we have

$$\begin{aligned} \Pr(s_t = 1|s_{t-1}, \varepsilon_{p,t}, z_t, \theta) &= \Pr(s_{1,t}^* < 0, s_{2,t}^* < 0, ..., s_{N-1,t}^* < 0|s_{t-1}, \varepsilon_{p,t}, z_t, \theta) \\ &= \prod_{i=1}^{N-1} \Phi\left(\frac{-\bar{\gamma}_{i,s_{t-1}} - z_t'\gamma_{i,s_{t-1}}^z - e_i'\rho\varepsilon_{p,t}}{\sqrt{1-\rho_i^2}}\right), \end{aligned} \tag{1.B.4}$$

where $\Phi(\cdot)$ is the cumulative density function (CDF) of the standard normal distribution, and e_i is a vector of zeros with a one in the ith position.

Then, for the case of $s_t = j \in \{2, ..., N\}$, the $j-1$th auxiliary variable should be positive and greater than the other auxiliary variables. For simplicity of exposition, we denote by $S_{-(j-1),t}^* = (s_{0,t}^*, s_{1,t}^*, ..., s_{j-2,t}^*, s_{j,t}^*, ..., s_{N-1,t}^*)'$ an $(N-1)$-dimensional vector of auxiliary variables, excluding the $j-1$th element $s_{j-1,t}^*$. Formally, the conditional transition probability

is given by

$$\begin{aligned}
&\Pr(s_t = j | s_{t-1}, \varepsilon_{p,t}, z_t, \theta) \\
&\quad = \Pr\left(\begin{array}{l} s^*_{j-1,t} - s^*_{0,t} > 0, s^*_{j-1,t} - s^*_{1,t} > 0, ..., s^*_{j-1,t} - s^*_{j-2,t} > 0, \\ s^*_{j-1,t} - s^*_{j,t} > 0, s^*_{j-1,t} - s^*_{j+1,t} > 0, ..., s^*_{j-1,t} - s^*_{N-1,t} > 0 \end{array} \middle| \, s_{t-1}, \varepsilon_{p,t}, z_t, \theta \right) \\
&\quad = \Pr\left(S^*_{-(j-1),t} - s^*_{j-1,t} < 0_{(N-1)\times 1} | s_{t-1}, \varepsilon_{p,t}, z_t, \theta \right).
\end{aligned} \tag{1.B.5}$$

Because equation (1.B.2) indicates that $S^*_{-(j-1),t} - s^*_{j-1,t}$ follows an $(N-1)$-dimensional multivariate normal distribution, we can rewrite equation (1.B.5) as

$$\Pr(s_t = j | s_{t-1}, \varepsilon_{p,t}, z_t, \theta) = F(0_{(N-1)\times 1} | c_{j-1,t}, V_{j-1}), \tag{1.B.6}$$

where $F(\cdot | c_{j-1,t}, V_{j-1})$ denotes the CDF of the multivariate normal distribution with mean $c_{j-1,t}$ and variance-covariance V_{j-1}. We derive $c_{j-1,t}$ as follows

$$c_{j-1,t} = \begin{pmatrix} -\bar{\gamma}_{j-1,s_{t-1}} - z_t' \gamma^z_{j-1,s_{t-1}} - e_{j-1}' \rho \varepsilon_{p,t} \\ \bar{\gamma}_{1,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{1,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) + (e_1 - e_{j-1})' \rho \varepsilon_{p,t} \\ \vdots \\ \bar{\gamma}_{j-2,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{j-2,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) + (e_{j-2} - e_{j-1})' \rho \varepsilon_{p,t} \\ \bar{\gamma}_{j,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{j,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) + (e_j - e_{j-1})' \rho \varepsilon_{p,t} \\ \bar{\gamma}_{j+1,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{j+1,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) + (e_{j+1} - e_{j-1})' \rho \varepsilon_{p,t} \\ \vdots \\ \bar{\gamma}_{N-1,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{N-1,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) + (e_{N-1} - e_{j-1})' \rho \varepsilon_{p,t} \end{pmatrix}, \tag{1.B.7}$$

and V_{j-1} is given by

$$V_{j-1} = \begin{pmatrix} 1-\rho^2_{j-1} & 1-\rho^2_{j-1} & \cdots & 1-\rho^2_{j-1} \\ 1-\rho^2_{j-1} & 2-\rho^2_{j-1}-\rho^2_1 & & \vdots \\ \vdots & & \ddots & 1-\rho^2_{j-1} \\ 1-\rho^2_{j-1} & \cdots & 1-\rho^2_{j-1} & 2-\rho^2_{j-1}-\rho^2_{N-1} \end{pmatrix}. \tag{1.B.8}$$

Similarly, we employ the parameterization in equation (1.10) of the paper and the unconditional distribution of s^*_t in equation (1.B.1) to compute the unconditional transition probabilities of the regimes. For $s_t = 1$, all auxiliary variables in equation (10) should be

negative. That is,

$$\begin{aligned}\Pr(s_t = 1|s_{t-1}, z_t, \theta) &= \Pr(s^*_{1,t} < 0, s^*_{2,t} < 0, ..., s^*_{N-1,t} < 0|s_{t-1}, z_t, \theta) \\ &= \Pr(s^*_t < 0_{(N-1)\times 1}|s_{t-1}, z_t, \theta) \\ &= F(0_{(N-1)\times 1}|\bar{\gamma}_{s_{t-1}} + z_t'\gamma^z_{s_{t-1}}, \Omega).\end{aligned} \tag{1.B.9}$$

Then, for the case of $s_t = j \in \{2, ..., N\}$, the auxiliary variable $s^*_{j-1,t}$ should be positive and greater than all elements of $S^*_{-(j-1),t}$. Formally, we have

$$\begin{aligned}\Pr(s_t = j|s_{t-1}, z_t, \theta) &= \Pr\left(s^*_{j-1,t} - S^*_{-(j-1),t} > 0_{(N-1)\times 1}|s_{t-1}, z_t, \theta\right) \\ &= \Pr\left(S^*_{-(j-1),t} - s^*_{j-1,t} < 0_{(N-1)\times 1}|s_{t-1}, z_t, \theta\right).\end{aligned} \tag{1.B.10}$$

Based on equation (1.B.1), we can rewrite equation (1.B.10) as

$$\Pr(s_t = j|s_{t-1}, z_t, \theta) = F(0_{(N-1)\times 1}|\bar{c}_{j-1,t}, \bar{V}_{j-1}), \tag{1.B.11}$$

where the mean $\bar{c}_{j-1,t}$ is given by

$$c_{j-1,t} = \begin{pmatrix} -\bar{\gamma}_{j-1,s_{t-1}} - z_t'\gamma^z_{j-1,s_{t-1}} \\ \bar{\gamma}_{1,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{1,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) \\ \vdots \\ \bar{\gamma}_{j-2,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{j-2,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) \\ \bar{\gamma}_{j,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{j,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) \\ \bar{\gamma}_{j+1,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{j+1,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) \\ \vdots \\ \bar{\gamma}_{N-1,s_{t-1}} - \bar{\gamma}_{j-1,s_{t-1}} + z_t'(\gamma^z_{N-1,s_{t-1}} - \gamma^z_{j-1,s_{t-1}}) \end{pmatrix}, \tag{1.B.12}$$

and the variance-covariance $\bar{V}_{j-1}$ is formulated as

$$\bar{V}_{j-1} = \begin{pmatrix} 1 & 1-\rho_1\rho_{j-1} & \cdots & 1-\rho_{N-1}\rho_{j-1} \\ 1-\rho_1\rho_{j-1} & 2(1-\rho_1\rho_{j-1}) & & \vdots \\ \vdots & & \ddots & 1+\rho_{N-1}\rho_{N-2} \\ & & & -\rho_{N-1}\rho_{j-1}-\rho_{N-2}\rho_{j-1} \\ 1-\rho_{N-1}\rho_{j-1} & \cdots & 1+\rho_{N-1}\rho_{N-2} & 2(1-\rho_{N-1}\rho_{j-1}) \\ & & -\rho_{N-1}\rho_{j-1}-\rho_{N-2}\rho_{j-1} & \end{pmatrix}. \tag{1.B.13}$$

Appendix 1.C

This appendix presents additional robustness checks for our empirical findings obtained from the selected three-regime timing model in equation (1.16) of the paper.

1.C.1 Controlling for market return and volatility timing

Our timing model focuses on the adjustment of fund currency betas in response to systematic currency liquidity. However, fund managers may use a sophisticated set of timing strategies to hedge risks in financial markets, such as return timing and volatility timing (see, e.g., Chen and Liang, 2007; Bodson et al., 2013). In the literature, there is some evidence that the currency market liquidity has a positive relation with currency market return and a negative relation with currency market volatility (see, e.g., Melvin and Taylor, 2009; Menkhoff et al., 2012; Mancini et al., 2013). Thus, it is possible that the evidence on currency liquidity timing can be partially attributed to fund managers' activities of timing currency market return or currency market volatility. To address this concern, we control for market-return and volatility timing in our selected three-regime timing model.

Table 1.C.1 presents the posterior summary of the selected three-regime timing model with controls for market-return and volatility timing. From Panel B, we observe the evidence of *perverse timing* becomes weaker in regime 1 as the 95% CIs of $\mu_{s_t=1}$ and $\lambda_{s_t=1}$ are not strictly negative. Similarly, we find the evidence of *weak timing* becomes indecisive in regime 2 as the 95% CI of one of the timing coefficients appears to be strictly positive. These findings agree with Cao et al. (2013) who document that controlling for market-return and volatility timing reduces the significance of negative liquidity timing. Moreover, there is virtually no difference between the results from Panel C of Table 1.C.1 and those reported in Table 1.3 of the paper. Specifically, the current level of systematic currency liquidity and shock to fund returns highlight their pivotal roles in driving regime switches of funds' currency liquidity timing behavior. Figure 1.C.1a plots the time path of the currency liquidity timing regimes after controlling for return and volatility timing. Comparing with its counterpart in Figure 1.4 of the paper, we can see that the captured switches of currency liquidity timing behavior are similar. Overall, though the evidence of *perverse timing* is not as strong as previously observed, both market-return and volatility timing do not affect the regime-switching behavior of currency liquidity timing among the international fixed income mutual funds.

1.C.2 Liquidity timing versus liquidity reaction

Cao et al. (2013) argue that fund managers may also change their portfolio exposures based on lagged values of liquidity. If fund managers use observed liquidity in month $t-1$ to derive a predictable component of liquidity and adjust their fund beta accordingly, they have no timing skill but simply react to public information (Ferson and Schadt, 1996). Given

this conjecture, our evidence on currency liquidity timing might rather reflect fund managers' response to lagged currency market liquidity conditions. To distinguish liquidity timing skill from liquidity reaction, we follow Cao et al. (2013) and adapt our selected three-regime timing model to the model including both liquidity timing and liquidity reaction terms.

Table 1.C.2 presents the posterior summary of the selected three-regime timing model with controls for liquidity reaction. From Panel A, we observe significant evidence of liquidity reaction among international fixed income mutual funds as the 95% CIs of the liquidity reaction coefficients $\psi^{\text{HML\_FX}}$ and ψ^{RX} do not include zero. However, Panel B shows that the inference about regime-dependent currency liquidity timing ability remains largely unchanged even when we take liquidity reaction into account. Specifically, regime switches from regime 1 to regime 3 are characterized by a stronger tendency of funds to positively time the carry-trade and dollar risk factors. From Panel C, we find that the current level of systematic currency liquidity continues to show adverse effects on the regime transitions, which are reflected in the overall negative estimates of $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$. Unlike the original results, the effect of regression error term on the regime transitions becomes less evident as the 95% CIs of ρ_1 and ρ_2 are nearly centered around zero. Figure 1.C.1b plots the time path of the currency liquidity timing regimes after controlling for liquidity reaction. We observe the occurrence of the *weak timing* and *strong timing* states becomes less frequent than those in Figure 1.4 of the paper, implying that some periods of positive currency liquidity timing are partly results of funds' reaction to past currency liquidity conditions. Overall, these findings suggest a certain level of liquidity reaction among the international fixed income mutual funds, but funds' currency liquidity timing ability cannot be fully attributed to liquidity reaction.

1.C.3 Controlling for illiquid holdings

Choi et al. (2022) show strong evidence of persistent staleness in the pricing of fixed income mutual funds, possibly due to the illiquidity of funds' holdings. This phenomenon leads to thin or nonsynchronous trading that may bias estimates of fund beta (see, e.g., Scholes and Williams, 1977; Dimson, 1979). In their study on bond mutual funds market timing, Chen et al. (2010) find that if the extent of stale pricing is related to the market factor (a case they call systematic stale pricing), then the inference about timing ability can also be biased. Cao et al. (2013) and Osinga et al. (2021) address systematic stale pricing when measuring the timing ability of hedge funds, and they find that controlling for this bias is important. In the same spirit of Cao et al. (2013) and Osinga et al. (2021), we reexamine the currency liquidity timing skill by augmenting our selected three-regime timing model with lagged values of currency risk factors and interaction terms between lagged values of currency risk factors and that of systematic currency liquidity.

Table 1.C.3 presents the posterior summary of the selected three-regime timing model with controls for illiquid holdings. From Panel A, we observe that HML_FX_{t-1}, HML_FX_{t-2},

RX_{t-1} and RX_{t-2} enter the regression significantly, indicating illiquid holdings in international fixed income mutual funds. From Panel B, we find the 95% CIs of μ_{s_t} and λ_{s_t} in the three regimes overlap more substantially relative to those reported in Table 1.3 of the paper, implying that the evidence of distinct regimes is somewhat weaker when controlling for funds' illiquid holdings. The results of $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$ in Panel C do not change materially. The evidence of endogeneity in the regime transitions, as shown in ρ_1, is marginal in this case. Figure 1.C.1c plots the time path of the currency liquidity timing regimes after controlling for illiquid holdings. We note that the captured switches of liquidity timing behavior are still highly comparable to those in Figure 1.4 of the paper.

1.C.4 Controlling for funding constraints

Financial leverage is commonly believed to play a trivial role in the performance of mutual funds, since these funds generally make very limited use of debt financing (see, e.g., Almazan et al., 2004; Boguth and Simutin, 2018; Fricke and Wilke, 2020). However, recent developments in the bond mutual fund sector challenge this traditional view. Given low investment returns, debt levels in bond funds have increased after the financial crisis (Avalos et al., 2015). Vivar et al. (2023) show that bond mutual funds' use of leverage, normally provided by prime brokers through short-term funding, exposes funds to the risk of sudden margin calls that can force them to liquidate positions. It is possible that the our sample funds' reduction of currency risk exposure during liquidity crisis periods is merely caused by a deterioration of funding liquidity because prime brokers have cut funding or increased borrowing costs. In this case, tightened funding constraints might be misidentified as liquidity timing. To control for the impact of funding constraints, we use the TED measure, which is measured by the difference between the three-month London Interbank Offered Rate (LIBOR) and the three-month Treasury bill rate. The TED spread reflects the market perceived counterparty default risk and a wider TED spread indicates an increase of counterparty default risk. This may impact the cost that the prime broker would charge when providing higher borrowing leverages to the mutual funds. We follow Cao et al. (2013) and Luo et al. (2017) and include an additional interaction term between the TED spread and currency risk factors in our selected three-regime timing model.

Table 1.C.4 presents the posterior summary of the selected three-regime timing model with controls for funding constraints. Unlike the original results, here we observe more pronounced overlap among the 95% CIs of μ_{s_t} and λ_{s_t} in the three regimes, implying that the pattern of distinct regimes is less evident when controlling for funds' funding constraints. The results in Panel C of Table 1.C.4 are again highly similar to those in Table 1.3 of the paper, though the evidence of endogeneity in the regime transitions captured by ρ_1 is marginal in this case. Figure 1.C.1d plots the time path of the currency liquidity timing regimes after

controlling for funding constraints. We note that the captured switches of liquidity timing behavior are highly comparable to those in Figure 1.4 of the paper.

1.C.5 Controlling for fund flow

In addition to changes in financial leverage, external funding constraints can be caused by investor redemptions, which present another mechanism by which a fund's market exposure could change rapidly. It is possible that when bonds do well, money flows into international fixed income mutual funds. The managers may take some time to allocate new money and thus it appears that the funds have more cash holdings regardless of the manager's actual investment skill. In the opposite case, when the market declines, investors withdraw their capital, the managers may unwind their positions, leading to a decrease in market exposure. Therefore, fund flow can bias the true timing ability of fund managers, which is called the "cash-flow" hypothesis in the literature (see, e.g., Warther, 1995; Ferson and Schadt, 1996; Edelen, 1999; Alda et al., 2015). To address this issue, we follow Cao et al. (2013) and examine a subsample of funds having low fund-flow volatility. Specifically, we examine funds whose monthly flow volatility is below the median level of peer funds. Funds with low flow volatility should be less affected by investor flows. Following prior research (see, e.g., Sirri and Tufano, 1998), we measure fund flows as the percentage change in total net assets (TNA) after adjusting for fund returns. The subsample contains 185 funds (or 48% of the overall sample), which is not exactly 50% of the entire sample because some funds do not report TNA.

Table 1.C.5 presents the posterior summary of the selected three-regime timing model applied to funds whose monthly flow volatility is lower than the median level of peer funds. Figure 1.C.1e plots the time path of the currency liquidity timing regimes based on the sub-sample funds. Overall, the analysis with the selected sample of funds leads to results similar to those obtained when using the full sample, suggesting that fund flows cannot explain our findings of currency liquidity timing behavior.

1.C.6 Discussion

We draw several conclusions from the above robustness checks. First, various controls appear to show some foreseeable impacts on the identification of the regimes and regime endogeneity. Specifically, we observe the weakening evidence of the three distinct regimes in the various robustness checks. Moreover, we observe the diminishing evidence that shocks to fund returns drive the shifts in regimes when controls are in place. These observations are expected in our robustness checks because a number of extra controls which are added to the original regression tend to capture part of the explanatory power of the variables already included and leave few information in the regression error term. Thus, it is natural to

anticipate a reduction in the significance of timing coefficients[12] as well as that of conditional correlations between the regression error term and regime transitions. Second, the main findings of the paper are materially unchanged under all the robustness tests: (i) the evidence on currency liquidity timing behavior among the sample funds is not explained away by other reasons except currency liquidity timing. This is supported by the observation that for both timing coefficients, their 95% CIs in the three regimes, although include zero, are not exactly centered around zero; (ii) the dynamic pattern that currency liquidity timing behavior shifts from a negative beta-liquidity relationship towards a positive relationship remains robust under different controls. This is evidenced by the observation that for both timing coefficients, their 95% CIs which cover a larger negative region in regime 1 appear to move towards the positive region in regimes 2-3; (iii) systematic currency liquidity continues to show minor adverse effects on the regime transitions in funds' timing behavior. This is reflected by the overall negative estimates of $\gamma^z_{1,s_{t-1}}$ and $\gamma^z_{2,s_{t-1}}$.

[12]Similar observations have been reported, for instance, by Chen et al. (2010) where the timing coefficients which are significantly negative in the original regression are found to appear neutral to weakly positive with several controls for non-timing-related nonlinearity.

Table 1.C.1 Estimation results: Robustness 1

	s_t/s_{t-1}	Mean	s.e.	95% CI		Ineff	p-val
Panel A: Non-switching coefficients in the timing model							
α (%)		0.093	0.056	-0.013	0.209	67.796	0.616
β^{GMF}		0.094	0.025	0.027	0.122	62.741	0.601
β^{EMF}		0.059	0.013	0.025	0.067	57.632	0.607
β^{TERM}		1.581	0.247	1.129	2.079	66.398	0.614
β^{CREDIT}		0.106	0.034	0.061	0.184	67.802	0.541
$\beta^{\text{HML\_FX}}$		0.033	0.021	-0.006	0.047	70.074	0.072
β^{RX}		0.140	0.018	0.130	0.177	65.711	0.217
$\psi^{\text{HML\_FX}}$		-1.422	0.349	-1.653	-0.516	64.826	0.574
ψ^{RX}		-0.216	0.516	-1.254	0.081	65.589	0.351
$\delta^{\text{HML\_FX}}$		-0.113	0.145	-0.448	0.049	64.781	0.465
δ^{RX}		0.033	0.199	-0.095	0.469	69.850	0.093
σ		0.004	0.000	0.004	0.004	12.100	0.950
Panel B: Switching coefficients in the timing model							
μ_{s_t}	1	0.075	0.081	-0.105	0.120	65.196	0.230
	2	0.142	0.087	-0.025	0.311	63.225	0.695
	3	0.544	0.205	0.087	0.798	69.615	0.306
λ_{s_t}	1	-0.870	0.427	-1.130	0.118	71.188	0.149
	2	0.610	0.254	0.035	1.009	63.467	0.892
	3	1.264	0.407	0.843	2.497	62.678	0.628
Panel C: Parameters in the Markov model							
$\bar{\gamma}_{1,s_{t-1}}$	1	-1.366	0.212	-1.726	-1.025	12.591	0.120
	2	1.178	0.199	0.852	1.504	4.546	0.526
	3	-0.134	0.208	-0.474	0.206	2.053	0.618
$\gamma^z_{1,s_{t-1}}$	1	0.102	0.288	-0.391	0.556	17.813	0.202
	2	-0.301	0.274	-0.751	0.151	9.110	0.772
	3	-0.165	0.292	-0.650	0.314	2.130	0.647
$\bar{\gamma}_{2,s_{t-1}}$	1	-1.974	0.223	-2.337	-1.606	6.009	0.462
	2	-0.637	0.203	-0.972	-0.305	3.691	0.053
	3	1.270	0.213	0.923	1.624	3.210	0.470
$\gamma^z_{2,s_{t-1}}$	1	-0.093	0.292	-0.581	0.378	4.841	0.139
	2	-0.237	0.277	-0.696	0.220	4.293	0.320
	3	-0.305	0.288	-0.781	0.164	2.622	0.097
ρ_1		-0.315	0.252	-0.635	0.186	44.187	0.193
ρ_2		0.011	0.238	-0.374	0.402	16.746	0.363

NOTES: This table presents the posterior summary of the selected three-regime timing model with controls for market-return and volatility timing. The model is given by

$$R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \lambda_{p,s_t}(L_{m,t} - \bar{L}_m)\text{RX}_t + \psi_p^{\text{HML\_FX}}\text{HML\_FX}_t^2$$
$$+ \psi_p^{\text{RX}}\text{RX}_t^2 + \delta_p^{\text{HML\_FX}}(\sigma_{m,t} - \bar{\sigma}_m)\text{HML\_FX}_t + \delta_p^{\text{RX}}(\sigma_{m,t} - \bar{\sigma}_m)\text{RX}_t + \sum_{j=1}^{J}\beta_p^j f_t^j + \sigma\varepsilon_{p,t},$$

where p represents each fund and t represents each month; $R_{p,t}$ is the return in excess of the risk-free rate; HML_FX_t and RX_t denote, respectively, the carry and dollar factors; f_t^j (J=4 in this case) denote the set of additional bond market risk factors; $L_{m,t}$ is the systematic currency liquidity measure and $\bar{L}_m$ is the time series mean of systematic currency liquidity measure; $\sigma_{m,t}$ is the Menkhoff et al. (2012) currency volatility measure and $\bar{\sigma}_m$ is the time series mean of currency market volatility; α_p is the risk-adjusted return; the coefficients μ_{p,s_t} and λ_{p,s_t} measure the regime-dependent fund's currency liquidity timing ability; the coefficients $\psi_p^{\text{HML\_FX}}$ and ψ_p^{RX} measure fund's currency market-return timing ability; the coefficients $\delta_p^{\text{HML\_FX}}$ and δ_p^{RX} measure fund's currency volatility timing ability; and $\varepsilon_{p,t}$ is the error term. We report results for an equal-weighted portfolio of all sample funds for the period of July 2001-December 2020. Results are based on 12,500 MCMC iterations. s_t/s_{t-1} are, respectively, the regime indicator at time t and $t-1$, Mean is the posterior mean, s.e. is the posterior standard error, 95% CI is the 95% credibility interval, Ineff is the inefficiency factor, and p-val is Geweke (1992) p-value.

Table 1.C.2 Estimation results: Robustness 2

	s_t/s_{t-1}	Mean	s.e.	95% CI		Ineff	p-val
Panel A: Non-switching coefficients in the timing model							
α (%)		0.049	0.007	0.044	0.049	53.281	0.937
β^{GMF}		0.110	0.014	0.071	0.113	71.062	0.750
β^{EMF}		0.065	0.008	0.067	0.071	72.798	0.776
β^{TERM}		1.495	0.045	1.497	1.504	66.598	0.751
β^{CREDIT}		0.098	0.008	0.070	0.100	69.139	0.703
$\beta^{\text{HML\_FX}}$		0.020	0.014	0.016	0.067	73.609	0.586
β^{RX}		0.146	0.006	0.145	0.145	69.095	0.827
$\psi^{\text{HML\_FX}}$		-0.071	0.038	-0.199	-0.065	71.733	0.822
ψ^{RX}		0.107	0.019	0.050	0.112	66.746	0.625
σ		0.004	0.000	0.004	0.005	20.831	0.540
Panel B: Switching coefficients in the timing model							
μ_{s_t}	1	0.033	0.349	-0.930	0.154	74.053	0.542
	2	0.238	0.270	-0.394	0.328	72.121	0.566
	3	0.382	0.098	0.352	0.677	70.956	0.598
λ_{s_t}	1	-1.136	0.104	-1.145	-0.807	64.076	0.882
	2	-0.060	0.297	-0.154	0.753	72.767	0.586
	3	2.086	0.371	0.831	2.210	73.605	0.559
Panel C: Parameters in the Markov model							
$\bar{\gamma}_{1,s_{t-1}}$	1	-1.531	0.211	-1.882	-1.184	3.903	0.589
	2	1.299	0.205	0.960	1.638	2.071	0.631
	3	-0.134	0.207	-0.470	0.206	0.964	0.998
$\gamma^z_{1,s_{t-1}}$	1	-0.036	0.285	-0.513	0.424	3.023	0.826
	2	-0.272	0.281	-0.731	0.191	2.023	0.867
	3	-0.129	0.289	-0.611	0.347	1.031	0.951
$\bar{\gamma}_{2,s_{t-1}}$	1	-1.921	0.214	-2.272	-1.570	2.843	0.627
	2	-0.585	0.204	-0.919	-0.251	1.428	0.797
	3	1.218	0.210	0.880	1.567	2.979	0.706
$\gamma^z_{2,s_{t-1}}$	1	-0.052	0.291	-0.531	0.419	2.126	0.788
	2	-0.165	0.279	-0.627	0.294	1.446	0.919
	3	-0.315	0.288	-0.785	0.170	1.357	0.099
ρ_1		0.183	0.253	-0.245	0.588	8.403	0.399
ρ_2		-0.196	0.190	-0.473	0.147	18.487	0.603

NOTES: This table presents the posterior summary of the selected three-regime timing model with controls for liquidity reaction. The model is given by

$$R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}\tilde{L}_{m,t}\text{HML\_FX}_t + \lambda_{p,s_t}\tilde{L}_{m,t}\text{RX}_t + \psi_p^{\text{HML\_FX}}(L_{m,t-1}-\bar{L}_m)\text{HML\_FX}_t + \psi_p^{\text{RX}}(L_{m,t-1}-\bar{L}_m)\text{RX}_t + \sum_{j=1}^{J}\beta_p^j f_t^j + \sigma\varepsilon_{p,t},$$

where p represents each fund and t represents each month; $R_{p,t}$ is the return in excess of the risk-free rate; HML_FX_t and RX_t denote, respectively, the carry and dollar factors; f_t^j (J=4 in this case) denote the set of additional bond market risk factors; $L_{m,t-1}$ is the one-month lagged systematic currency liquidity measure and represents past information; $\bar{L}_m$ is the time series mean of systematic currency liquidity measure prior to month $t-1$; $\tilde{L}_{m,t}$ is the innovation in systematic currency liquidity from an AR(2) process and represents the unpredictable component of systematic currency liquidity; α_p is the risk-adjusted return; the coefficients μ_{p,s_t} and λ_{p,s_t} measure the regime-dependent fund's currency liquidity timing ability; the coefficients $\psi_p^{\text{HML\_FX}}$ and ψ_p^{RX} measure funds' currency liquidity reaction; and $\varepsilon_{p,t}$ is the error term. We report results for an equal-weighted portfolio of all sample funds for the period of July 2001-December 2020. Results are based on 12,500 MCMC iterations. s_t/s_{t-1} are, respectively, the regime indicator at time t and $t-1$, Mean is the posterior mean, s.e. is the posterior standard error, 95% CI is the 95% credibility interval, Ineff is the inefficiency factor, and p-val is Geweke (1992) p-value.

Table 1.C.3 Estimation results: Robustness 3

	s_t/s_{t-1}	Mean	s.e.	95% CI		Ineff	p-val
Panel A: Non-switching coefficients in the timing model							
α (%)		0.043	0.050	-0.034	0.130	38.988	0.545
β^{GMF}		0.104	0.039	0.051	0.167	32.501	0.898
β^{EMF}		0.087	0.026	0.048	0.140	51.273	0.627
β^{TERM}		0.940	0.263	0.604	1.376	37.155	0.971
β^{CREDIT}		0.090	0.028	0.053	0.134	37.551	0.847
$\beta^{\text{HML\_FX}}$		0.009	0.017	-0.016	0.038	40.691	0.949
$\beta^{\text{HML\_FX,lag1}}$		0.035	0.012	0.015	0.052	30.044	0.999
$\beta^{\text{HML\_FX,lag2}}$		0.024	0.016	0.001	0.047	41.755	0.771
β^{RX}		0.144	0.022	0.117	0.184	42.984	0.808
$\beta^{\text{RX,lag1}}$		0.089	0.015	0.065	0.115	29.498	0.682
$\beta^{\text{RX,lag2}}$		0.113	0.015	0.088	0.135	33.118	0.841
μ^{lag1}		-0.049	0.053	-0.144	0.030	24.496	0.628
μ^{lag2}		0.032	0.052	-0.064	0.099	29.107	0.971
λ^{lag1}		-0.071	0.068	-0.191	0.041	27.322	0.502
λ^{lag2}		-0.091	0.083	-0.208	0.048	47.423	0.121
σ		0.004	0.000	0.003	0.004	29.215	0.262
Panel B: Switching coefficients in the timing model							
μ_{s_t}	1	-0.065	0.054	-0.161	0.019	20.287	0.749
	2	0.008	0.075	-0.100	0.119	30.371	0.783
	3	0.444	0.368	-0.016	1.052	50.267	0.266
λ_{s_t}	1	-0.442	0.490	-1.200	0.400	49.222	0.614
	2	0.477	0.299	0.077	0.999	38.849	0.535
	3	0.881	0.500	0.228	1.656	39.513	0.320
Panel C: Parameters in the Markov model							
$\bar{\gamma}_{1,s_{t-1}}$	1	-1.475	0.223	-1.841	-1.109	10.034	0.317
	2	1.247	0.210	0.902	1.594	6.666	0.002
	3	-0.121	0.207	-0.461	0.218	1.796	0.959
$\gamma^z_{1,s_{t-1}}$	1	0.009	0.305	-0.507	0.501	13.706	0.132
	2	-0.239	0.291	-0.712	0.241	8.208	0.824
	3	-0.179	0.285	-0.657	0.285	2.570	0.817
$\bar{\gamma}_{2,s_{t-1}}$	1	-2.012	0.217	-2.370	-1.656	3.783	0.638
	2	-0.627	0.207	-0.961	-0.286	4.338	0.922
	3	1.314	0.213	0.963	1.663	2.369	0.854
$\gamma^z_{2,s_{t-1}}$	1	-0.075	0.294	-0.556	0.405	4.346	0.689
	2	-0.181	0.280	-0.646	0.282	4.570	0.757
	3	-0.266	0.290	-0.730	0.221	3.785	0.782
ρ_1		-0.140	0.293	-0.593	0.376	29.756	0.294
ρ_2		-0.021	0.262	-0.464	0.400	16.031	0.184

NOTES: This table presents the posterior summary of the selected three-regime timing model with controls for illiquid holdings. The model is given by

$$R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \lambda_{p,s_t}(L_{m,t} - \bar{L}_m)\text{RX}_t + \sum_{k=1}^{2}[\beta_p^{\text{HML\_FX,lagk}}\text{HML\_FX}_{t-k}$$

$$+ \beta_p^{\text{RX,lagk}}\text{RX}_{t-k} + \mu_p^{\text{lagk}}(L_{m,t-k} - \bar{L}_m)\text{HML\_FX}_{t-k} + \lambda_p^{\text{lagk}}(L_{m,t-k} - \bar{L}_m)\text{RX}_{t-k}] + \sum_{j=1}^{J}\beta_p^j f_t^j + \sigma\varepsilon_{p,t},$$

where p represents each fund; t, $t-1$ and $t-2$ represent the current time period, one-month lagged period and two-month lagged period. For each time period t, $t-1$ and $t-2$, $\bar{L}_m$ is the time series mean of systematic currency liquidity measure prior to this period. The coefficients μ_{p,s_t} and λ_{p,s_t} measure the regime-dependent fund's currency liquidity timing ability. We report results for an equal-weighted portfolio of all sample funds for the period of September 2001-December 2020. The data on July 2001 and August 2001 are used to compute the initial two lags of variables. Results are based on 12,500 MCMC iterations. s_t/s_{t-1} are, respectively, the regime indicator at time t and $t-1$, Mean is the posterior mean, s.e. is the posterior standard error, 95% CI is the 95% credibility interval, Ineff is the inefficiency factor, and p-val is Geweke (1992) p-value.

Table 1.C.4 Estimation results: Robustness 4

	s_t/s_{t-1}	Mean	s.e.	95% CI		Ineff	p-val
Panel A: Non-switching coefficients in the timing model							
α (%)		0.091	0.053	0.001	0.152	54.089	0.174
β^{GMF}		0.086	0.036	0.027	0.132	49.320	0.794
β^{EMF}		0.046	0.033	0.008	0.098	66.551	0.555
β^{TERM}		1.268	0.275	0.896	1.692	51.587	0.183
β^{CREDIT}		0.147	0.039	0.090	0.185	62.802	0.507
$\beta^{\text{HML\_FX}}$		0.026	0.024	-0.008	0.071	57.331	0.784
β^{RX}		0.102	0.030	0.059	0.156	50.905	0.857
ψ		-0.056	0.041	-0.122	0.030	54.449	0.518
δ		0.124	0.057	0.015	0.212	54.891	0.850
σ		0.004	0.000	0.004	0.005	20.795	0.106
Panel B: Switching coefficients in the timing model							
μ_{s_t}	1	-0.047	0.080	-0.209	0.036	51.559	0.973
	2	0.075	0.119	-0.135	0.282	55.659	0.763
	3	0.353	0.216	-0.020	0.551	58.401	0.911
λ_{s_t}	1	-0.514	0.432	-0.935	0.248	58.886	0.701
	2	0.607	0.369	0.073	1.086	52.492	0.039
	3	1.808	0.687	0.507	2.719	57.618	0.860
Panel C: Parameters in the Markov model							
$\bar{\gamma}_{1,s_{t-1}}$	1	-1.429	0.215	-1.784	-1.079	9.843	0.955
	2	1.207	0.210	0.865	1.553	6.525	0.672
	3	-0.146	0.212	-0.492	0.198	2.199	0.501
$\gamma^z_{1,s_{t-1}}$	1	0.034	0.303	-0.471	0.523	16.348	0.256
	2	-0.310	0.291	-0.777	0.184	10.075	0.624
	3	-0.192	0.291	-0.674	0.281	3.224	0.507
$\bar{\gamma}_{2,s_{t-1}}$	1	-1.944	0.225	-2.315	-1.571	7.945	0.276
	2	-0.622	0.203	-0.954	-0.290	3.637	0.963
	3	1.263	0.209	0.920	1.606	3.633	0.558
$\gamma^z_{2,s_{t-1}}$	1	-0.101	0.293	-0.591	0.374	5.367	0.725
	2	-0.189	0.277	-0.645	0.264	4.257	0.907
	3	-0.286	0.284	-0.752	0.180	2.364	0.852
ρ_1		-0.221	0.327	-0.690	0.342	49.806	0.684
ρ_2		-0.040	0.249	-0.475	0.355	27.149	0.032

NOTES: This table presents the posterior summary of the selected three-regime timing model with controls for funding constraints. The model is given by

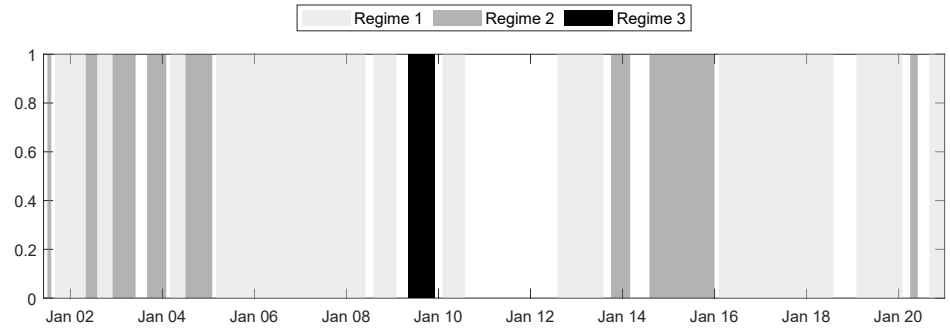
$$R_{p,t} = \alpha_p + \beta_p^{\text{HML\_FX}}\text{HML\_FX}_t + \beta_p^{\text{RX}}\text{RX}_t + \mu_{p,s_t}(L_{m,t} - \bar{L}_m)\text{HML\_FX}_t + \lambda_{p,s_t}(L_{m,t} - \bar{L}_m)\text{RX}_t + \psi_p\text{HML\_FX}_t\text{TED}_t + \delta_p\text{RX}_t\text{TED}_t + \sum_{j=1}^{J}\beta_p^j f_t^j + \sigma\varepsilon_{p,t},$$

where p represents each fund and t represents each month; $R_{p,t}$ is the return in excess of the risk-free rate; HML_FX_t and RX_t denote, respectively, the carry and dollar factors; f_t^j (J=4 in this case) denote the set of additional bond market risk factors; $L_{m,t}$ is the systematic currency liquidity measure and $\bar{L}_m$ is the time series mean of systematic currency liquidity measure; α_p is the risk-adjusted return; the coefficients μ_{p,s_t} and λ_{p,s_t} measure the regime-dependent fund's currency liquidity timing ability; the coefficients ψ_p and δ_p measure the impact of TED; and $\varepsilon_{p,t}$ is the error term. We report results for an equal-weighted portfolio of all sample funds for the period of July 2001-December 2020. Results are based on 12,500 MCMC iterations. s_t/s_{t-1} are, respectively, the regime indicator at time t and $t-1$, Mean is the posterior mean, s.e. is the posterior standard error, 95% CI is the 95% credibility interval, Ineff is the inefficiency factor, and p-val is Geweke (1992) p-value.

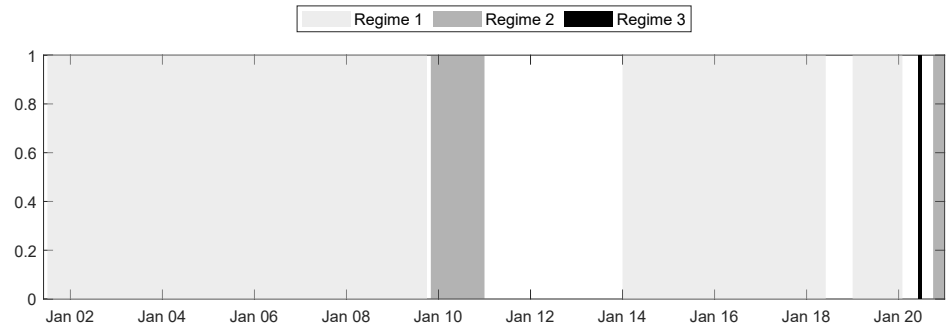
Table 1.C.5 Estimation results: Robustness 5

	s_t/s_{t-1}	Mean	s.e.	95% CI		Ineff	p-val
Panel A: Non-switching coefficients in the timing model							
α (%)		0.101	0.035	0.041	0.133	66.694	0.923
β^{GMF}		0.106	0.040	0.043	0.134	70.030	0.477
β^{EMF}		0.038	0.026	0.016	0.073	71.125	0.420
β^{TERM}		1.140	0.187	0.931	1.518	65.999	0.000
β^{CREDIT}		0.154	0.026	0.099	0.196	68.623	0.241
$\beta^{\text{HML\_FX}}$		0.040	0.030	-0.006	0.066	72.893	0.386
β^{RX}		0.136	0.025	0.075	0.157	70.137	0.706
σ		0.004	0.000	0.004	0.004	13.418	0.714
Panel B: Switching coefficients in the timing model							
μ_{s_t}	1	-0.146	0.046	-0.208	-0.083	52.807	0.399
	2	-0.058	0.043	-0.119	0.026	53.661	0.157
	3	0.067	0.064	0.031	0.216	58.327	0.410
λ_{s_t}	1	-0.673	0.273	-0.896	-0.223	69.366	0.282
	2	0.158	0.149	-0.046	0.405	56.299	0.839
	3	1.267	0.443	0.498	1.592	70.494	0.479
Panel C: Parameters in the Markov model							
$\bar{\gamma}_{1,s_{t-1}}$	1	-1.477	0.207	-1.822	-1.142	4.095	0.872
	2	1.282	0.204	0.947	1.618	3.123	0.880
	3	-0.184	0.203	-0.518	0.147	1.962	0.928
$\gamma^z_{1,s_{t-1}}$	1	-0.015	0.284	-0.482	0.454	3.866	0.929
	2	-0.118	0.289	-0.598	0.346	12.195	0.458
	3	-0.180	0.288	-0.651	0.293	2.375	0.977
$\bar{\gamma}_{2,s_{t-1}}$	1	-1.891	0.216	-2.251	-1.543	4.049	0.993
	2	-0.610	0.206	-0.956	-0.276	4.478	0.431
	3	1.287	0.204	0.947	1.624	4.556	0.380
$\gamma^z_{2,s_{t-1}}$	1	-0.046	0.294	-0.536	0.440	4.310	0.959
	2	-0.193	0.276	-0.654	0.255	5.052	0.436
	3	-0.297	0.285	-0.759	0.173	2.642	0.393
ρ_1		-0.018	0.267	-0.402	0.488	36.944	0.532
ρ_2		0.012	0.227	-0.384	0.369	26.580	0.370

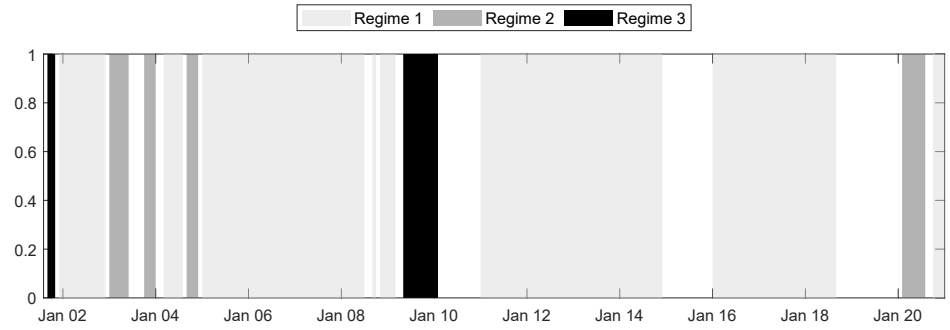
NOTES: This table presents the posterior summary of the selected three-regime timing model applied to funds whose monthly flow volatility is lower than the median level of peer funds. We report results for an equal-weighted portfolio of the subsample funds with low flow volatility for the period of July 2001-December 2020. Results are based on 12,500 MCMC iterations. s_t/s_{t-1} are, respectively, the regime indicator at time t and $t-1$, Mean is the posterior mean, s.e. is the posterior standard error, 95% CI is the 95% credibility interval, Ineff is the inefficiency factor, and p-val is Geweke (1992) p-value.



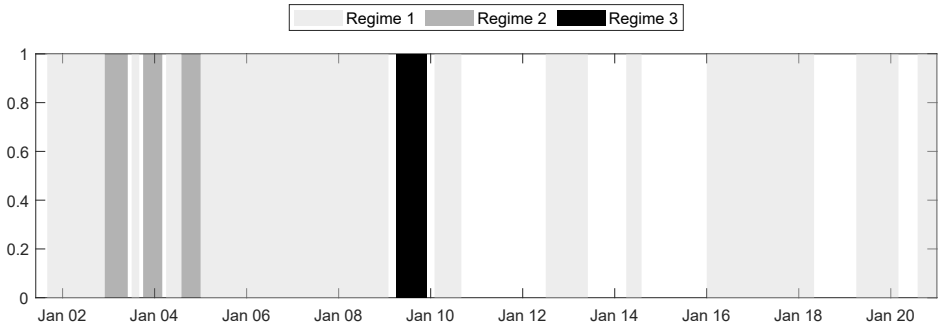
(a) Robustness 1



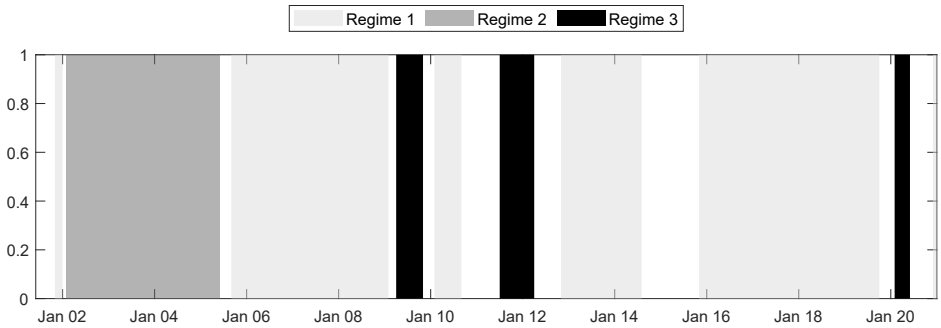
(b) Robustness 2



(c) Robustness 3



(d) Robustness 4



(e) Robustness 5

Figure 1.C.1 Time path of the currency liquidity timing regimes. This figure plots shaded areas associated with the periods of funds being in a particular regime. The light gray bars, dark gray bars, and black bar correspond to the periods of funds being in regimes 1-3, respectively. The empty white bars correspond to the periods during which the model cannot determine the regime as each regime has smoothed probability of less than 50%. The sample period for Robustness 3 is from September 2001 to December 2020 because the data on July 2001 and August 2001 are used to compute the initial two lags of variables. The sample period for the remaining robustness tests is from July 2001 to December 2020.

Chapter2

Markov-Switching Graphical Structural VAR†

2.1 Introduction

This paper concerns the problem of measuring connectedness in financial systems, which is central to many areas of research, including risk management, portfolio allocation, and business cycle analysis. Our main contribution is to propose a novel Markov-Switching Graphical Structural Vector Autoregressive (MS-GSVAR) model, which simultaneously captures some salient empirical features of financial connectedness, such as the mixture of contemporaneous and temporal dependences, high-dimensionality, and abrupt structural changes.

The literature on connectedness analysis to date has shown some stylized facts of financial connectedness. First, financial connectedness can arise from either the contemporaneous (instantaneous) or temporal (lagged) dependences among financial variables. This is evidenced by a handful of papers detecting connectedness based on contemporaneous dependence (see, e.g., Acharya et al., 2012; Adrian and Brunnermeier, 2016; Acharya et al., 2017), and also indicated by much of the studies using temporal dependence for measuring connectedness (see, e.g., Billio et al., 2012; Diebold and Yılmaz, 2014, 2015; Baruník and Křehlík, 2018). Second, the scale of financial connectedness can be very large. This is reflected by a growing body of literature using large datasets (between 10 and 100 series) to analyze financial connectedness, systemic risk and contagion effects (see Billio et al., 2019, for a review). Third, the sparsity and strength of financial connectedness may undergo abrupt structural changes. This is particularly pointed out by a recent research stream exploring the topological structure of financial networks (see, e.g., Billio et al., 2021, 2022).

†This chapter has been turned into a working paper (joint with G. Urga) entitled "Markov-Switching Graphical Structural VAR".

Many methodological advances have been made in capturing one of the above features, with only a few efforts attempting to capture more than one relevant features. For joint modeling of the contemporaneous and temporal dependences, some authors combine a Bayesian graphical approach with vector autoregressive (VAR) models (see, e.g., Ahelegbey et al., 2016a; Casarin et al., 2020; Ahelegbey and Giudici, 2022). For measurement of connectedness with large datasets, see for example Billio et al. (2019), who propose a Bayesian nonparametric Lasso VAR (BNP-Lasso VAR) model to extract the Granger causality networks of a high-dimensional time series through clustering and shrinking effects. For extensions concerning time variation in connectedness, see for example, Markov-switching models (Billio et al., 2016; Casarin et al., 2018; Billio et al., 2022) and Bayesian time-varying parameter vector autoregressions (TVP-VAR) (Ciccarelli and Rebucci, 2007; Guidolin et al., 2019; Poon and Zhu, 2022). With regards to the modeling of contemporaneous and temporal dependences in high-dimensional settings, Barigozzi and Brownlees (2019) apply the dimension reduction technique LASSO into a large VAR system to extract both the Granger causality and partial correlation structures. In respect of handling the structural changes in both the contemporaneous and temporal dependences, Ahelegbey et al. (2021) propose a change-point technique for turning point detection in a Bayesian graphical piece-wise VAR model that approximates the contemporaneous and temporal dependences between financial variables. As for measuring the time-varying connectedness in large-scale financial systems, Massacci (2017) and Massacci (2023) propose regime changes in high-dimensional factor models through a threshold mechanism and measure connectedness based on covariance matrices. Moreover, Bianchi et al. (2019) assume a latent Markov chain process driving covariance restrictions in high-dimensional SUR models and study connectedness based on network.

The model we put forward facilitates an integrated analysis of all the aforementioned features of financial connectedness. The usefulness of our model is reflected in many empirical aspects. First, one can implement the model to detect different states of system connectivity and date turning points of the connectedness level over the study period. Such information can be viewed as early warning signals of how contagion-prone the system will be. Second, the model allows us to quantify the risk transmission and reception capacity of each entity in the system, which assists to identify systemically important individuals. Third, our framework allows for a decomposition of the overall connectedness into the contemporaneous and temporal components. This means one can look deeper into the frequency-specific source of connectedness (e.g., the instantaneous, short-term or long-term connectedness), which helps to monitor the accumulation of systemic risk. Fourth, the model is fast and easy to implement for multivariate time series with dimension up to around 100 series. Thus, the model may find application in a wide spectrum of research where connectedness analysis on large financial datasets is of interest.

Methodologically, we build on the expanding literature on Bayesian Graphical model for multivariate systems (see, e.g., Ahelegbey et al., 2016a,b; Giudici and Spelta, 2016; Bianchi

et al., 2019; Billio et al., 2019; Casarin et al., 2020; Agudze et al., 2022). In particular, Ahelegbey et al. (2016a) combine a Bayesian graphical approach with structural vector autoregressive (SVAR) models, which allows to capture the contemporaneous and temporal dependences among multivariate time series under directed acyclic graphs. Their method also presents an automatic way to achieve dimensionality reduction since relevant variables that need to be included in an SVAR can be identified from the graphical representation of the model. We extend the framework of Ahelegbey et al. (2016a) by incorporating Markov-switching dynamics into the graph structures and coefficients of an SVAR system. The resulting model is referred to as a Markov-Switching Graphical Structural Vector Autoregressive (MS-GSVAR) model, which integrates an SVAR to capture both contemporaneous and temporal dependences, a graphical representation to tackle high-dimensionality, and a hidden Markov chain process to account for structural changes in graph structures and coefficients. In addition, we propose an identification scheme for different states of connectedness by exploiting the centrality properties of a weighted directed graph that unionizes the contemporaneous and temporal dependence structures in each state. Specifically, we provide a graph centrality measure in the spirit of the *out* version of the Pagerank score, proposed by Zhang et al. (2022) for nodes of weighted directed graphs. Such score assigns relative centrality to each node according to its risk-transmitting ability, which accounts not only for the number and strength of its out-going edges, but also for the centrality of other nodes it directs towards. We formally show that the average node score for our weighted directed graphs is equal under different states. Given this property, a comparable centrality measure of the graph for each state can be developed using the average score of the leading risk-transmitting nodes (i.e., those nodes with scores exceeding the average node score). Different states are then separated according to the state-specific graph centrality, which by construction quantifies the risk-propagation level of the graph. Since systemic risk by its minimal definition is more associated with risk propagation than with absorption (Benoit et al., 2017), the states of connectedness identified by our model are naturally linked to the states of systemic risk.

Another contribution of this paper is to address the computational complexity arising from inference on graphs in the context of high model dimension, numerous lags, and multiple states. In our MS-GSVAR, the number of graphs to be estimated depends on the number of lag orders and states specified for the model, and the time needed for learning each graph increases super-exponentially with the number of variables in the model. To provide a high quality estimation with a reasonable computational time, we develop a graph inference method building on a hybrid Markov chain Monte Carlo (Hybrid-MCMC) algorithm introduced by Kuipers et al. (2022) for high-dimensional Bayesian graphical models. The attractive property of Hybrid-MCMC is a combination of two well-established graph inference methods, namely the PC algorithm of Spirtes et al. (2000) and the Order-MCMC algorithm of Friedman and Koller (2003). Such property exhibits several advantages over the Structure-MCMC algorithm

(Madigan et al., 1995; Giudici and Castelo, 2003) applied most often in the literature on Bayesian Graphical model for multivariate systems.

Through simulations, we demonstrate that our graph search method relying on Hybrid-MCMC provides more reliable graph inference in much less time than Structure-MCMC in all simulation settings we consider. The improved accuracy and efficiency gain achieved by our method are substantial particularly in the high-dimensional settings. In the context of a large dataset with changing dependence structures, we show that our model produces more accurate parameter estimates and better replicates the simulated data than a conventional Markov-Switching specification without the graphical representation. Moreover, standard convergence statistics support the effectiveness of our model inference procedure. In a real-data illustration, we apply the model to measure the volatility connectedness in a global banking system consisting of 96 banks. Our model provides a data-driven detection of different connectedness states that best describe the historical evolution of the global banking volatility network. In addition, the systemic risk profiles of individual banks, featured by their risk transmission and reception capacity, is readily inferred from the model output. The model also offers deeper insights about the frequency-specific source of volatility connectedness, explaining the build-up of the overall volatility connectedness and identifying the primary role played by the contemporaneous dependence.

The remainder of this paper is organized as follows. Section 2.2 introduces the model specification. Section 2.3 provides an identification scheme for different states of connectedness. Section 2.4 discusses the model inference procedure. Section 2.5 provides an empirical illustration and Section 2.6 concludes. Simulation results, technical details and additional empirical discussions are provided in Appendix 2.

2.2 Model specification

Consider a connectedness analysis for $Y_t = (Y_{1,t}, Y_{2,t}..., Y_{n,t})'$, an n-dimensional multivariate time series. To jointly capture the contemporaneous and temporal dependences among series, our baseline formulation relies on a pth-order SVAR

$$Y_t = B_0 Y_t + \sum_{i=1}^{p} B_i Y_{t-i} + \varepsilon_t, \quad t = 1, ..., T, \tag{2.1}$$

where B_0 is an $n \times n$ structural matrix with zero diagonals that determines the contemporaneous dependence; B_i with $1 \leq i \leq p$ is an $n \times n$ autoregressive matrix that specifies the lag-i temporal dependence conditional on the contemporaneous dependence; ε_t is an n-dimensional vector of structural errors with diagonal covariance matrix Σ_ε, that is $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, \Sigma_\varepsilon)$.

The number of coefficients in (2.1) is $n(np+n)$, which can be large for high-dimensional multivariate time series. To work with a parsimonious model specification, we employ a

graphical representation of the SVAR proposed by Ahelegbey et al. (2016a). Their graphical representation is concerned with representing the contemporaneous and temporal dependence structures underlying an SVAR system through directed acyclic graphs (DAGs).[1] A DAG is expressed by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of directed edges that contains no directed cycles. Given an SVAR(p) model as in (2.1), the set of nodes $\mathcal{V}$ is $\{1, ..., n, n+1, ..., 2n, ..., np+1, ..., np+n\}$, where each element corresponds to a component of Y_t and its lagged observations (i.e., $\{Y_{1,t}, ..., Y_{n,t}, Y_{1,t-1}, ..., Y_{n,t-1}, ..., Y_{1,t-p}, ..., Y_{n,t-p}\}$), while the set of directed edges $\mathcal{E}$ is a subset of $\mathcal{V} \times \mathcal{V}$ such that the pair $(r, l) \in \mathcal{E}$ if and only if the components r and l are linked by an directed edge. Ahelegbey et al. (2016a) show that a natural graphical representation of the SVAR model is based on the union of two different types of DAG: the first DAG contains directed edges representing contemporaneous (instantaneous) dependences, whereas the second contains directed edges denoting temporal (lagged) dependences. We label the two graphs respectively as the contemporaneous and temporal graphs with the following definition.

Definition 2.2.1 (Contemporaneous graph)**.** Under a structural matrix in (2.1), the associated graph is $\mathcal{G}_0 = (\mathcal{V}_0, \mathcal{E}_0)$ with $\mathcal{V}_0 = \{1, ..., n\}$ and

$$\mathcal{E}_0 = \{(r, l) \in \mathcal{V}_0 \times \mathcal{V}_0 : A_{rl,i} \neq 0 \text{ for } i = 0\}, \tag{2.2}$$

where $r, l \in \mathcal{V}_0$ and $r \neq l$; $A_{rl,0}$ is the entry of an $n \times n$ binary matrix (or so called the adjacency matrix) A_0, and encodes the presence of an edge $(r, l) \in \mathcal{E}_0$.

Remark 2.2.1. In this graph, each element of $\mathcal{V}_0$ corresponds to a component of Y_t (i.e., $\{Y_{1,t}, ..., Y_{n,t}\}$), while an edge $(r, l) \in \mathcal{E}_0$ exists if and only if $Y_{l,t}$ instantaneously causes $Y_{r,t}$; that is the SVAR coefficients $B_{rl,0}$ is not null.

Remark 2.2.2. The Markov equivalence principle of contemporaneous directed graphs (Andersson et al., 1997) states that two or more graphs with similar correlation structures but different edge directions may have the same marginal likelihood. This implies that Ahelegbey et al. (2016a)'s contemporaneous graph does not provide a unique solution to the SVAR identification problem. Therefore, following the suggestion of Ahelegbey et al. (2016a), the researcher should choose among the graphs in the equivalent classes by using other arguments (e.g., economic theory or other external sources of information).

Definition 2.2.2 (Temporal graph)**.** Under an autoregressive matrix in (2.1), the associated graph is $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$ with $\mathcal{V}_i = \{1, ..., n, ni+1, ..., ni+n\}$ and

$$\mathcal{E}_i = \{(r, l) \in \mathcal{V}_i \times \mathcal{V}_i : A_{rl,i} \neq 0 \text{ for } 1 \leq i \leq p\}, \tag{2.3}$$

[1] Formal definitions of graph theory concepts are provided in Appendix 2.A.

where $r,l \in \mathcal{V}_i$ and $1 \leq r \leq n$, $ni+1 \leq l \leq ni+n$; $A_{rl,i}$ is the entry of an $n \times n$ binary matrix (or so called the adjacency matrix) A_i for $1 \leq i \leq p$, and encodes the presence of an edge $(r,l) \in \mathcal{E}_i$.

Remark 2.2.3. In this graph, each element of $\mathcal{V}_i$ corresponds to a component of Y_t and its observations at lag i (i.e., $\{Y_{1,t},...,Y_{n,t},Y_{1,t-i},...,Y_{n,t-i}\}$), while an edge $(r,l) \in \mathcal{E}_i$ exists if and only if $Y_{l,t-i}$ Granger causes $Y_{r,t}$; that is the SVAR coefficients $B_{rl,i}$ is not null.

Based on definitions 2.2.1-2.2.2, the coefficient matrices of the SVAR in (2.1) can be reparameterized as

$$B_i = A_i \circ \Phi_i, \quad 0 \leq i \leq p, \tag{2.4}$$

where A_i is an $n \times n$ binary adjacency matrix; operator $\circ$ is the element-by-element Hadamard's product; Φ_i is an $n \times n$ matrix of coefficients such that its entry represents the value of the effect of $Y_{l,t-i}$ on $Y_{r,t}$. Thus, (2.1) can be reformulated as

$$Y_t = (A_0 \circ \Phi_0)Y_t + \sum_{i=1}^{p}(A_i \circ \Phi_i)Y_{t-i} + \varepsilon_t, \quad t = 1,...,T \tag{2.5}$$

where $(A_0 \circ \Phi_0)$ and $(A_i \circ \Phi_i)$ are the graphical model coefficient matrices whose nonzero elements describe the value associated with the contemporaneous and temporal dependences, respectively. Clearly, non-zero (zero) entries in A_i can be interpreted as indicators of relevant (not-relevant) variables to be included (excluded) in (from) the model. These zero entries restrictions allow us to avoid estimating all the coefficients in the SVAR model. It thus presents an automatic way to achieve dimensionality reduction and variable selection.

To account for potential structural changes, we extend the framework of Ahelegbey et al. (2016a) by incorporating Markov-switching dynamics into the graph structures and coefficients of an SVAR system. The resulting model is referred to as a Markov-Switching Graphical Structural Vector Autoregressive (MS-GSVAR) model. The MS(K)-GSVAR(p) model with K states and p lags is specified as

$$Y_t = (A_0(s_t) \circ \Phi_0(s_t))Y_t + \sum_{i=1}^{p}(A_i(s_t) \circ \Phi_i(s_t))Y_{t-i} + \varepsilon_t, \quad t = 1,...,T \tag{2.6}$$

where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,\Sigma_\varepsilon(s_t))$; s_t is a discrete Markov chain process with states $1,...,K$, i.e., $s_t \in \{1,...,K\}$. The properties of s_t are fully governed by an $K \times K$ matrix of transition probabilities P. The (i,j)th element of P is the probability of switching to state j at time t, given that at time $t-1$ the process is in state i, $p_{ij} = \Pr(s_t = j | s_{t-1} = i)$, for $i,j \in \{1,...,K\}$, and $\sum_{j=1}^{K} p_{ij} = 1$. Our model specification in (2.6) portrays a switching dependence structure. The state-specific adjacency matrices and coefficient matrices record respectively changes in the edge existence and the edge strength of the contemporaneous and temporal dependence structures. Moreover, the state-dependent feature of the covariance matrix addresses potential

heteroskedasticity biases that are typical of correlation-based measures, as discussed in Forbes and Rigobon (2002).

The estimatable reduced-form representation of (2.6) can be expressed as

$$Y_t = \sum_{i=1}^{p} B_i^*(s_t) Y_{t-i} + \mu_t, \quad t = 1, ..., T \tag{2.7}$$

where $B_i^*(s_t) = B(s_t)(A_i(s_t) \circ \Phi_i(s_t))$ is an $n \times n$ matrix of reduced-form coefficients, with $B(s_t) = (I_n - A_0(s_t) \circ \Phi_0(s_t))^{-1}$, and I_n denoting an $n \times n$ identity matrix; $\mu_t = B(s_t)\varepsilon_t$ is an *iid* $n \times 1$ vector of reduced-form errors, with covariance matrix $\Sigma_\mu(s_t) = B(s_t)\Sigma_\varepsilon(s_t)B(s_t)'$, i.e., $\mu_t \sim$ i.i.d. $\mathcal{N}(0, \Sigma_\mu(s_t))$.

Throughout the paper, we denote by $Y_{\tau:t} = (Y_\tau, ..., Y_t)$ and $S_{\tau:t} = (s_\tau, ..., s_t)$ the sequence of observations and states between time τ and t. We let $A = (A_{0:p}(1), ..., A_{0:p}(K))$ be the collections of state-specific adjacency matrices, with $A_{0:p}(k)$ denoting the collection of lag-specific adjacency matrices for the state k. We let $\Theta = (\Theta(1), ..., \Theta(K))$ be the collections of state-specific model parameters, with $\Theta(k) = (B_0(k), B_{1:p}^*(k), \Omega_\varepsilon(k), \Omega_\mu(k), p(k))$ denoting the collections of parameters for the state k, where $B_0(k)$ is a vector that collects structural coefficients, $B_{1:p}^*(k)$ is a vector that collects lag-specific reduced-form coefficients, $\Omega_\varepsilon(k) = \Sigma_\varepsilon(k)^{-1}$ is the precision matrix of the structural errors, $\Omega_\mu(k) = \Sigma_\mu(k)^{-1}$ is the precision matrix of the reduced-form errors, and $p(k)$ is the kth row of the transition matrix P.

2.3 States identification scheme

We propose an identification scheme for different states of connectedness by exploiting the centrality properties of a graph that unionizes the contemporaneous and temporal dependences in each state. In our MS-GSVAR model, the information about the existence of a linkage between pairs of nodes in state k is encoded by the adjacency matrices $A_{0:p}(k)$, while the strength of this linkage is described by the absolute value of the corresponding structural coefficient in matrices $B_{0:p}(k)$. This allows us to define a null-diagonal matrix $\tilde{A}(k) \in \mathbb{R}^{n \times n}$, with (r, l)th entry given by

$$\tilde{A}_{rl}(k) = \begin{cases} \sum_{i=0}^{p} |B_{rl,i}(k)| & \text{if } \sum_{i=0}^{p} A_{rl,i}(k) \neq 0 \\ 0 & \text{if } \sum_{i=0}^{p} A_{i,rl}(k) = 0 \end{cases}, \quad r, l = 1, ..., n, \ \forall r \neq l, \tag{2.8}$$

that is the weight $\tilde{A}_{rl}(k)$ assigned to each pair of nodes $(Y_{r,t}, Y_{l,t})$ of Y_t is equal to a sum of the structural and autoregressive coefficients in absolute values, if there is a contemporaneous or temporal directed edge between them in state $k \in K$. Thus, $\tilde{A}(k)$ can be considered as an adjacency matrix of a weighted directed graph which unionizes the contemporaneous and temporal dependences for the state k.

The connectivity features of $\tilde{A}(k)$ can be characterized by node centrality, a concept developed in graph theory to assess who occupies critical positions in a graph, and to identify systemically important individuals. Classical node centrality measures include degree, closeness, betweenness, eigenvector centrality, Katz centrality, and PageRank (see Newman, 2010, for a review). However, no measures are ideal for weighted directed graphs. Degree centrality simply counts the number of connections each node maintains, without effectively discriminating the relative importance of these connections with respect to the whole graph structure. Closeness and betweenness centrality measures are based on the shortest paths between one node and all other nodes in the graph. These measures implicitly assume simplistic and pre-determined paths, though in real world the transition mechanism of economic shocks are unlikely to be restricted to follow specific paths. Eigenvector centrality, which accounts for all possible paths connecting two nodes, turns out to be trivial for directed graphs as it gives non-zero centralities only to nodes that fall in strongly connected graphs (i.e., a graph structure where each node is reachable from every other node). Katz centrality generalizes the concept of eigenvector centrality to directed graphs that are not strongly connected. Despite that, this measure has one undesirable feature: if a node with high centrality score has edges pointing to many others then those others also get high centrality score, regardless of whether themselves have high centrality score. A popular variant of Katz centrality is PageRank, with which all nodes adjacent to a high-centrality node will only receive a portion of that node's centrality score. Although this measure complements Katz centrality to provide more reliable centrality evaluation for directed graphs, it does not account for edge weight in definition and, thus, unable to deal with weighted directed graphs.

Our node centrality measure is based on weighted PageRank (WPR) generalized by Zhang et al. (2022) for weighted directed graphs. WPR gives each node in a weighted directed graph two different centrality scores, the weighted OutPageRank (WOutPR) and the weighted InPageRank (WInPR) scores. From a financial viewpoint, nodes with high WOutPR scores are the influencers (risk transmitters), and high WInPR scoring nodes are strongly influenced by others (risk receivers). Since systemic risk by its minimal definition is more associated with risk propagation than with absorption (Benoit et al., 2017), we use the *out* version of Zhang et al. (2022)'s Pagerank score, namely the WOutPR score, to identify different states of connectedness which are thus naturally linked to states of systemic risk. The relative WOutPR score for the lth node in the kth state graph $\tilde{A}(k)$, denoted by $\omega_l^{\text{out}}(k)$, can be defined as[2]

$$\omega_l^{\text{out}}(k) = \gamma \sum_{r=1}^{n} \left(\theta \frac{\tilde{A}_{rl}(k)}{s_r^{\text{in}}} + (1-\theta) \frac{\mathbb{I}(\tilde{A}_{rl}(k) \neq 0)}{d_r^{\text{in}}} \right) \omega_r^{\text{out}}(k) + \frac{1-\gamma}{n}, \tag{2.9}$$

[2]The WInPR score, $\omega_l^{\text{in}}(k)$, can be derived by replacing $\tilde{A}_{rl}(k)$, s_r^{in}, and d_r^{in} in (2.9) by the transpose $(\tilde{A}_{rl}(k))^{\mathrm{T}}$, $s_r^{\text{out}} = \sum_{l=1}^{n} (\tilde{A}_{rl}(k))^{\mathrm{T}}$, and $d_r^{\text{out}} = \sum_{l=1}^{n} \mathbb{I}\big((\tilde{A}_{rl}(k))^{\mathrm{T}} \neq 0\big)$.

where $\mathbb{I}(\cdot)$ is an indicator function that takes a value of one when $\tilde{A}_{rl}(k)$ is nonnull and zero otherwise; $s_r^{\text{in}} = \sum_{l=1}^{n} \tilde{A}_{rl}(k)$ is the total weight of edges towards the rth node from any other nodes in the graph; $d_r^{\text{in}} = \sum_{l=1}^{n} \mathbb{I}(\tilde{A}_{rl}(k) \neq 0)$ is the total number of edges towards the rth node from any other nodes in the graph; $\gamma \in [0,1)$ is a dampening factor, that is the portion of score that is transferred by a node; $\theta \in [0,1]$ is a tuning parameter balancing between the relative importance of edge weights with respect to the number of edges; we choose $\gamma = 0.85$ and $\theta = 0.5$ throughout the paper, as recommended in Zhang et al. (2022).

Proposition 2.3.1. For any specified state k, $\omega_l^{\text{out}}(k) \in (0,1)$ and $\sum_{l=1}^{n} \omega_l^{\text{out}}(k) = 1$.

Proof of Proposition 2.3.1. Consider nodes in a graph as states in a Markov chain. Let E be the transition matrix of the associated Markov chain for WOutPR, with (r,l)th entry given by

$$e_{rl} = \begin{cases} \theta \tilde{A}_{rl}(k)/s_r^{\text{in}} + (1-\theta)\mathbb{I}(\tilde{A}_{rl}(k) \neq 0)/d_r^{\text{in}} & \text{if } d_r^{\text{in}} \neq 0 \\ 1/n & \text{if } d_r^{\text{in}} = 0 \end{cases} \tag{2.10}$$

Notice that

$$\sum_{l=1}^{n} \left(\theta \frac{\tilde{A}_{rl}(k)}{s_r^{\text{in}}} + (1-\theta) \frac{\mathbb{I}(\tilde{A}_{rl}(k) \neq 0)}{d_r^{\text{in}}} \right) = \theta + (1-\theta) = 1. \tag{2.11}$$

That is matrix E is non-negative and column stochastic. Let $W = (\omega_1^{\text{out}}, \omega_2^{\text{out}}, ..., \omega_n^{\text{out}})'$ be a column vector collecting the WOutPR for each node. In matrix terms, (2.9) is equivalent to

$$\begin{aligned} W &= \gamma E W + (1-\gamma)\mathbf{1}^* \\ &= (\gamma E + (1-\gamma)J)W \\ &= E^* W, \end{aligned} \tag{2.12}$$

where $\mathbf{1}^* = \mathbf{1}/\|\mathbf{1}\|_1$ is the normalization of an all-ones vector with $\|\cdot\|_1$ denoting the L^1-norm; J is an n-dimensional matrix whose columns given by $\mathbf{1}^*$. It is obvious that J is also column stochastic, rendering that E^* is strictly positive and column stochastic. By the Perron–Frobenius theorem[3], the largest eigenvalue of E^* is equal to one, and the unique solution to (2.12) is the corresponding eigenvector. Such a eigenvector is positive and the sum of its entries equal to one. This completes the proof.

From proposition 2.3.1, the node's WOutPR score for any state-specific graph $\tilde{A}(k)$ is bounded between zero and one, with the average equal to $1/n$. Given this property, a comparable centrality measure of the graph for each state can be developed using the average score of the leading risk-transmitting nodes (i.e., those nodes with their own scores exceeding $1/n$). In this aspect, the relative WOutPR score of the kth state graph $\tilde{A}(k)$, denoted by

[3]See, e.g., Horn and Johnson (2013), Theorem 8.2.5, 8.2.6, for details and proofs.

$q^{\text{out}}(\tilde{A}(k))$, can be defined as

$$q^{\text{out}}(\tilde{A}(k)) = \frac{1}{\sum_{l=1}^{n} \mathbb{I}(\omega_l^{\text{out}}(k) > \frac{1}{n})} \sum_{l=1}^{n} \omega_l^{\text{out}}(k) \mathbb{I}(\omega_l^{\text{out}}(k) > \frac{1}{n}). \tag{2.13}$$

where $\mathbb{I}(\cdot)$ is an indicator function that takes a value of one when $\omega_l^{\text{out}}(k) > 1/n$ and zero otherwise. (2.13) suggests that a graph $\tilde{A}(k)$ of state k would get a high WOutPR score with respect to graphs of other states if it has a small number of nodes with WOutPR scores that greatly exceed the average. Intuitively, the measure we propose considers the whole system (graph) to be more contagious when it is more vulnerable to targeted attacks from a few systemically important entities (nodes). This is consistent with the connectivity patterns commonly observed during periods of high systemic risk in real-world financial systems, such as the S&P100 system (e.g., Bianchi et al. (2019)), EuroStoxx-600 realized volatility system (e.g., Ahelegbey et al. (2016b)) and the intersectoral system of the US economy (e.g., Acemoglu et al. (2012)).

Subsequently, states identification is based on the restriction

$$q^{\text{out}}(\tilde{A}(1)) < ... < q^{\text{out}}(\tilde{A}(K)). \tag{2.14}$$

This restriction directly separates the states according to the level of risk propagation for each state-specific graph, which allows us to give a clear economic interpretation to the states: the first state is characterized by the lowest WOutPR score of the graph and hence corresponds to the lowest level of connectedness, the second state corresponds to the next lowest level of connectedness, and so on, with the last state associated with the highest level of connectedness.

2.4 Model inference

The MS-GSVAR model is estimated with a Markov Chain Monte Carlo (MCMC) procedure. Specifically, we approximate the posterior distribution of the model by implementing a multi-move Gibbs sampling algorithm (see, e.g., Roberts and Sahu, 1997; Casella and Robert, 2004), where the graphs, the hidden states, and the parameters are all sampled in blocks. Denote with $\mathcal{T}_k = \{t : s_t = k\}$ the index set for observations in state k. At each iteration, the Gibbs sampler iterates over the following steps

Step 1: sample $S_{1:T}$ given $Y_{1:T}, \Theta, A$.
Step 2: sample $A_{0:p}(k)$ given $Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}$, for $k = 1, ..., K$.
Step 3: sample $B^*_{1:p}(k)$ given $Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, A_{0:p}(k), \Omega_\mu(k)$, for $k = 1, ..., K$.
Step 4: sample $B_0(k)$ given $Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, A_0(k), \Omega_\varepsilon(k), B^*_{1:p}(k)$, for $k = 1, ..., K$.
Step 5: sample $\Omega_\varepsilon(k)$ given $Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, B_0(k), B^*_{1:p}(k)$, for $k = 1, ..., K$.

Step 6: sample $\Omega_\mu(k)$ given $Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, B^*_{1:p}(k)$, for $k = 1, ..., K$.
Step 7: sample P given $S_{1:T}$.

We use 1,000 iterations of the Gibbs sampler and discard the first 100 iterations as burn-in for all of our simulation and empirical exercises. The simulation studies in Appendix 2.C illustrate the effectiveness of our Gibbs sampler, and suggest that the number of iterations and burn-in we choose is sufficient to have MCMC samples close to i.i.d. draws. Also note that from Step 2 to 6, the sampling under regimes $k \in \{1, ..., K\}$ is parallelized and implemented in MATLAB on one node at the High-performance computing (HPC) cluster provided by City University of London; each node has 2 CPUs (Intel Xeon Gold 6248R) with 24 cores and 384 GB RAM. In the following we describe the different steps of the Gibbs sampler.

2.4.1 Inference on graphs: the Hybrid-MCMC algorithm

The Hybrid-MCMC algorithm is a search algorithm used to build DAGs. In the Bayesian network structure learning literature [4], there are a few variants of this algorithm (see, e.g., Tsamardinos et al., 2006; Nandy et al., 2018; Kuipers et al., 2022). In this paper, we focus on the Hybrid-MCMC algorithm recently advanced by Kuipers et al. (2022) for learning the DAGs underlying high-dimensional Bayesian network models. The attractive property of their algorithm is a combination of two well-established graph search algorithms, namely the PC algorithm and the Order-MCMC algorithm. The PC algorithm (named after the authors Peter Spirtes and Clark Glymour from Spirtes et al. (2000)) is one of the most popular constraint-based structure learning methods. It learns the edges of the graph based on conditional independence (CI) tests (e.g. Fisher's z-statistic) given the property of DAGs that edges encode conditional dependences. The Order-MCMC algorithm, proposed by Friedman and Koller (2003), is a classical type of the score-and-search structure learning methods. It first samples a node order by running the Metropolis-Hastings algorithm on the order space, which are possible topological orderings (or permutations) of the node labels; it then draws a DAG compatible with a given sampled order. According to Kuipers et al. (2022), one can start with the PC algorithm to reduce the order space on which a subsequent Order-MCMC algorithm runs for posterior sampling of graphs. The resulting algorithm is well-suited for fast searching and sampling much larger DAGs (e.g., Kuipers et al. (2022) shows the feasibility of their hybrid algorithm when node dimension is up to 200).

We adapt the Hybrid-MCMC algorithm of Kuipers et al. (2022) to our model inference scheme. At each iteration, when the Gibbs sampler moves to Step 2, we generate a sample of $A_{0:p}(k)$ for a given $k \in \{1, ..., K\}$ based on a three-stage procedure as follows. Details of the algorithm at each stage are provided in Appendix 2.A.

[4] See Appendix 2.A for a brief review of current state-of-the-art methods for structure learning of Bayesian networks.

Input: an $n \times (np+n)$ all-ones matrix $A_{0:p}(k)$.
Stage 1 (Initialization): run the PC algorithm to recursively remove (or not) edges between nodes based on CI tests. This step initializes a configuration of the binary entries of $A_{0:p}(k)$.
Stage 2 (Structure learning): run the Order-MCMC algorithm sequentially for each lag-specific adjacency matrix stacked in the initialized $A_{0:p}(k)$ (i.e., $A_0(k)$,$A_1(k)$,...,$A_p(k)$) through a number of iterations. This step yields a number of posterior samples of $A_{0:p}(k)$.
Stage 3 (Bayesian model averaging): follow Ahelegbey et al. (2016a) to employ Bayesian model averaging for building a single sample of $A_{0:p}(k)$ from the set of posterior samples generated as before.
Output: a sample of $A_{0:p}(k)$ for a given $k \in \{1,...,K\}$, which is subsequently used to determine which entries of $B^*_{1:p}(k)$ and $B_0(k)$ to be estimated in Steps 3-4 of the Gibbs sampler.

Standard graph inference methods adopted in the literature on Bayesian Graphical model for multivariate systems normally rely on the Structure-MCMC algorithm (see,e.g., Ahelegbey et al., 2016a,b; Giudici and Spelta, 2016; Bianchi et al., 2019; Billio et al., 2019; Casarin et al., 2020; Agudze et al., 2022). The Structure-MCMC algorithm (proposed by Madigan et al. (1995), later refined by Giudici and Castelo (2003)) is also a classical type of the score-and-search structure learning methods. It samples a graph by running the Metropolis-Hastings algorithm on the graph space, which are possible configurations of edges (i.e., addition, deletion, or reversion of edges) between nodes. It is worth noting that due to the use of the Hybrid-MCMC algorithm, our graph inference method exhibits several advantages over the Structure-MCMC algorithm. The first advantage relates to the computational efficiency, which lies in two aspects. Compared with Structure-MCMC, our method offers considerable runtime saving, particularly for larger graphs, due to an initialization stage where the search space is reduced by the PC algorithm, and to a subsequent structure learning stage based on Order-MCMC that operates in the much smaller space of node orders.[5] Besides, since node order imposes acyclicity, our method relying on Order-MCMC does not require an additional stage to validate graph acyclicity, which however is required for Structure-MCMC as shown in previous implementations (see, e.g., Ahelegbey et al., 2016a; Casarin et al., 2020). The second advantage is associated with the improved accuracy in the graph inference, achieved by means of the inclusion of Order-MCMC which is known to have better mixing and convergence than Structure MCMC (Friedman and Koller, 2003). A formal comparison of our method and the Structure-MCMC algorithm is provided in Appendix 2.C.

2.4.2 Inference on parameters

Except for Step 2 (described in the previous section), the remaining steps of the Gibbs sampler are executed by sampling from the full conditional distributions of the parameters.

[5] Consider a graph with n nodes, order space has complexity $2^{O(n\log(n)}$ which is considerably smaller than $2^{\Omega(n^2)}$ for graph space (Friedman and Koller, 2003).

These steps are standard and we provide exact forms of the full conditional distributions as well as their sampling methods in Appendix 2.B. Below we discuss our choice of the prior distributions that are used to derive the full conditional distributions. We assume independent priors for the parameters and use a standard set of prior distributions as follows

$$\begin{aligned} &B_0(k), B^*_{1:p}(k)|A_{0:p}(k) \sim \mathcal{N}(m(k), M(k)), \\ &\Omega_\varepsilon(k) \sim \mathcal{W}_G(\nu_\varepsilon(k), R_\varepsilon(k)), \\ &\Omega_\mu(k) \sim \mathcal{W}(\nu_\mu(k), R_\mu(k)), \\ &p(k) \sim \mathcal{D}ir(c_{k1}, ..., c_{kK}). \end{aligned} \tag{2.15}$$

We assume the prior distributions for $B_0(k)$ and $B^*_{1:p}(k)$ conditional on $A_{0:p}(k)$ are Gaussian. We set the hyperparameters to be rather uninformative and common across states, i.e., $m(k) = \mathbf{0}$ and $M(k) = 100I_n$ for each $k = 1, ..., K$. We assume the prior for $\Omega_\varepsilon(k)$ is a G-Wishart distribution with prior expectation $\nu_\varepsilon(k)^{-1}R_\varepsilon(k)$ and $\nu_\varepsilon(k) > n$. The G-Wishart distribution is the conjugate prior for the precision matrix over the set of all symmetric, positive definite matrices with zeros in the off-diagonal elements that correspond to missing edges in the graph associated with ε_t. Since by assumption, the elements of ε_t are mutually independent, its associated graph is empty, which implies that $\Omega_\varepsilon(k)$ is a diagonal positive random matrix. We set the hyperparameters to be fairly vague, i.e., $\nu_\varepsilon(k) = n + 2$ and $R_\varepsilon(k) = \nu_\varepsilon(k)I_n$ for each $k = 1, ..., K$. We assume the prior for $\Omega_\mu(k)$ is a Wishart distribution with prior expectation $\nu_\mu(k)^{-1}R_\mu(k)$ and $\nu_\mu(k) > n$. We set the hyperparameters to be fairly vague, i.e., $\nu_\mu(k) = n+2$ and $R_\mu(k) = \nu_\mu(k)I_n$ for each $k = 1, ..., K$. The prior distribution for the kth row of the transition matrix P, i.e., $p(k) = (p_{k1}, ..., p_{kK})'$, are set independently and are given by K-dimensional Dirichlet distributions. The hyperparameters of these distributions, c_{km}, for $m \in \{1, ..., K\}$, are all set to 1 except those corresponding to the diagonal entries of the matrix P, denoted by c_{kk}, which are set to $0.98(K-1)/(1-0.98)$. This choice similar to that in Sims et al. (2008) expresses a prior belief about the persistence of states.

2.5 Empirical illustration

This section provides an empirical illustration of the proposed MS-GSVAR model. In the spirit of Demirer et al. (2018), we consider a connectedness analysis on a high-dimensional global banking dataset. The dataset consists of weekly logarithmic stock return volatilities of 96 banks, and the sample period is from September 12, 2003 to December 9, 2022 (1,005 observations). See Appendix 2.D for data details.

We start by searching for the model specification that best fits the empirical dataset. The proposed model inference procedure and graph inference method described in Section 2.4 allow us to undertake an extensive model estimation and selection in a reasonable time. Specifically, we estimate different model specifications in (2.6) based on a range of combinations of the

number of states and lag orders: $K = 1, 2, 3, 4$ and $p = 1, 2, 3$. The in-sample fits of each model specification are assessed and compared using the Deviance Information Criterion (DIC).[6] The DIC results in Table 2.1 are clearly in favor of MS(3)-GSVAR(2), a model specification with three states $K = 3$ and two lag orders $p = 2$. Given the states identification restriction in (2.14), we label the three estimated states as the low, high, and extreme connectedness states, respectively.

Table 2.1 In-sample fits of MS(K)-GSVAR(p) models

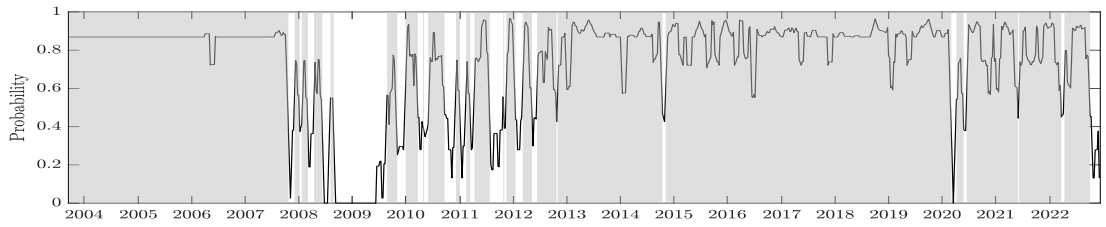
	$K = 1$	$K = 2$	$K = 3$	$K = 4$
$p = 1$	177155	172143	165092	166480
$p = 2$	175663	169259	**161210**	163687
$p = 3$	178633	167676	165877	162574

NOTES: This table presents the Deviance Information Criterion (DIC) of the competing MS(K)-GSVAR(p) models. The best fitting model which is highlighted in bold is the one with the smallest DIC.

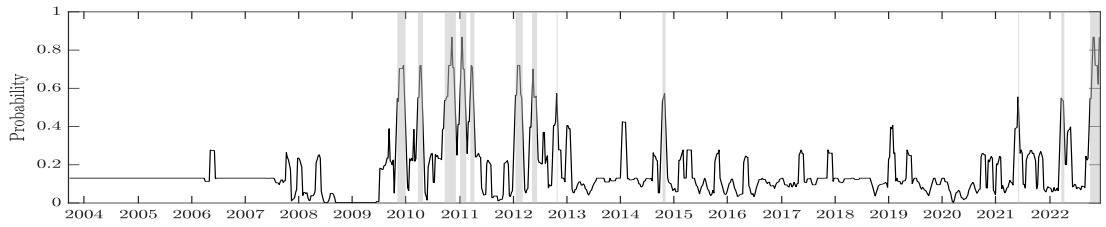
We proceed by investigating how the global banking system historically evolved across the estimated states. Such historical pattern identified by our model reflects the systemic risk developments and may behave as an early warning signal for how contagion-prone the system will be. The estimated states (gray bars) and their smoothed probabilities (black line) are given in Figure 2.1. Our model suggests that the evolution of the global banking system experiences long periods of low volatility connectedness (state 1), alternated by sporadic events of high and extreme volatility connectedness (states 2-3). A clear trend emerges: the states of high and extreme connectedness become less active and more short-lived in the second half of our sample, indicating that volatility connectedness and systemic risk of global banks to a certain extent were alleviated over the last decade. Model-implied turning points of the high and extreme connectedness states line up consistently with many financial, economic and political events that previously hit the banking system at different scale and magnitude. For instance, the events with known local-scale impact (e.g., a set of sovereign debt crises in European countries) are properly identified as the high connectedness state, while the events covering a much larger global scale (e.g., the global financial crisis, and the COVID-19) are correctly classified to the extreme connectedness state. Further discussions of these examples and other relevant events are provided in Appendix 2.D.

Under the states identification scheme outlined in Section 2.3, it is also possible to compute the state-specific WOutPR (WInPR) scores for each entity in the system based on (2.9). These scores can be used to explore individuals' systemic risk profiles, featured by their risk transmission and reception capacity. Table 2.2 shows the top 15 risk-transmitting (risk-

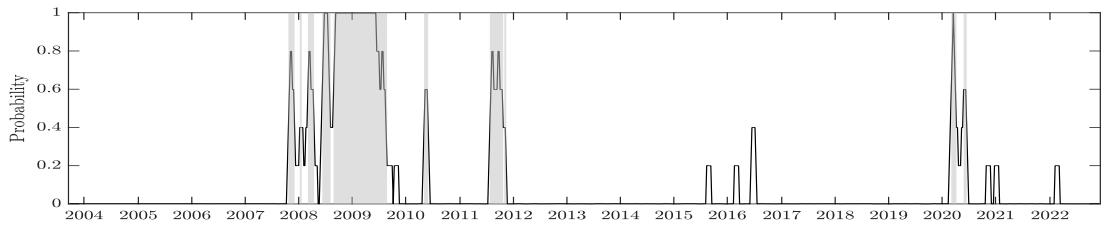
[6]The DIC has been successfully applied to discriminate between competing Markov-switching model specifications (see, e.g., Ardia, 2009; Raggi and Bordignon, 2012; Hur, 2017; Huber and Fischer, 2018; Billio et al., 2022). Since our model includes latent state variables, we adopt the modifications of the DIC, the DIC_3, introduced in Celeux et al. (2006, sec. 3.1).



(a) State 1: low connectedness state



(b) State 2: high connectedness state



(c) State 3: extreme connectedness state

Figure 2.1 Estimated states. We present the mean posterior smoothed probability of the low (panel (a)), high (panel (b)) and extreme (panel (c)) connectedness states. Gray bars in each plot correspond to the periods where the probability of the corresponding state is the maximum. Model-implied turning points of the high and extreme connectedness states: **1)** the high connectedness state (month/year): 11/09-12/09; 04/10; 09/10-11/10; 01/11; 03/11; 01/12-02/12; 05/12; 10/12; 10/14; 06/21; 03/22-04/22; 10/22-12/22. **2)** the extreme connectedness state (month/year): 10/07-11/07; 01/08; 03/08-04/08; 06/08-07/08; 09/08-08/09; 05/10; 08/11-11/11; 03/20; 06/20. See the Internet Appendix D for discussions of the turning points.

receiving) banks, ranked according to the posterior median of the WOutPR (WInPR) scores given in parenthesis. First, the North American banks, particularly the US banks, are major risk transmitters while the European and Asian banks are major risk receivers, as evidenced by the rankings in state 1 and state 3 which in total cover most of our sample. Despite that, the European and Asian banks occasionally replace the North American banks as the leading risk transmitters while remaining as the primary risk receiver, according to the rankings in state 2. This finding is not surprising since state 2 (the high connectedness state), as shown in Figure 2.1, is active during periods of local-scale events when the volatility shocks were expected to be absorbed within the region. Second, the three states are distinguishable in terms of banks' risk transmission capacity rather than their risk reception capacity. Clearly, the WOutPR scores of high-ranked banks (e.g., see banks in 1-5 positions) increase largely from

Table 2.2 Centrality ranking of individual banks

Rank	Risk transmitters			Risk receivers		
	State 1	State 2	State 3	State 1	State 2	State 3
1	rf.us	kbc.be	hban.us	seba.se	axis.in	inga.nl
	(0.0255)	(0.0280)	(0.0670)	(0.0188)	(0.0300)	(0.0272)
2	usb.us	shgp.cn	wfc.us	inga.nl	gle.fr	lloy.gb
	(0.0185)	(0.0235)	(0.0387)	(0.0175)	(0.0217)	(0.0200)
3	hban.us	pab.cn	rf.us	d05.sg	rbs.gb	dan.dk
	(0.0166)	(0.0201)	(0.0284)	(0.0158)	(0.0172)	(0.0183)
4	gle.fr	gle.fr	bk.us	bnp.fr	seba.se	ry.ca
	(0.0156)	(0.0187)	(0.0209)	(0.0157)	(0.0161)	(0.0166)
5	bac.us	mzh.jp	inga.nl	hdfc.in	sbk.kr	shf.kr
	(0.0152)	(0.0186)	(0.0192)	(0.0146)	(0.0159)	(0.0165)
6	wfc.us	key.us	ms.us	ubsn.ch	ucg.it	ffg.jp
	(0.0151)	(0.0179)	(0.0164)	(0.0143)	(0.0158)	(0.0159)
7	d05.sg	bp.it	stt.us	jpm.us	icici.in	uob.sg
	(0.0149)	(0.0152)	(0.0154)	(0.0136)	(0.0153)	(0.0154)
8	td.ca	ucg.it	ubsn.ch	cbb.jp	shgp.cn	mtbh.jp
	(0.0139)	(0.0151)	(0.0142)	(0.0133)	(0.0145)	(0.0146)
9	bbt.us	bmps.it	gs.us	ucg.it	pab.cn	seba.se
	(0.0137)	(0.0139)	(0.0130)	(0.0132)	(0.0143)	(0.0139)
10	ry.ca	bac.us	pnc.us	sbin.in	sab.es	kbc.be
	(0.0129)	(0.0138)	(0.0123)	(0.0128)	(0.0131)	(0.0135)
11	aib.ie	rbs.gb	c.us	anz.au	ubsn.ch	na.ca
	(0.0127)	(0.0136)	(0.0117)	(0.0126)	(0.0125)	(0.0134)
12	ubsn.ch	cmsb.cn	gle.fr	ry.ca	smt.jp	nor.fi
	(0.0125)	(0.0134)	(0.0115)	(0.0123)	(0.0122)	(0.0126)
13	c.us	ms.us	ucg.it	aca.fr	sbin.in	barc.gb
	(0.0123)	(0.0124)	(0.0113)	(0.0122)	(0.0120)	(0.0125)
14	icici.in	smt.jp	cmsb.cn	swe.se	puj.in	kb.kr
	(0.0120)	(0.0120)	(0.0111)	(0.0121)	(0.0118)	(0.0123)
15	fitb.us	mtbh.jp	smf.jp	key.us	cbb.jp	cba.au
	(0.0115)	(0.0116)	(0.0105)	(0.0120)	(0.0117)	(0.0122)

NOTES: This table presents the top 15 risk transmitters (receivers) in each state. The rankings are based on the posterior median of the WOutPR (WInPR) scores given in parenthesis.

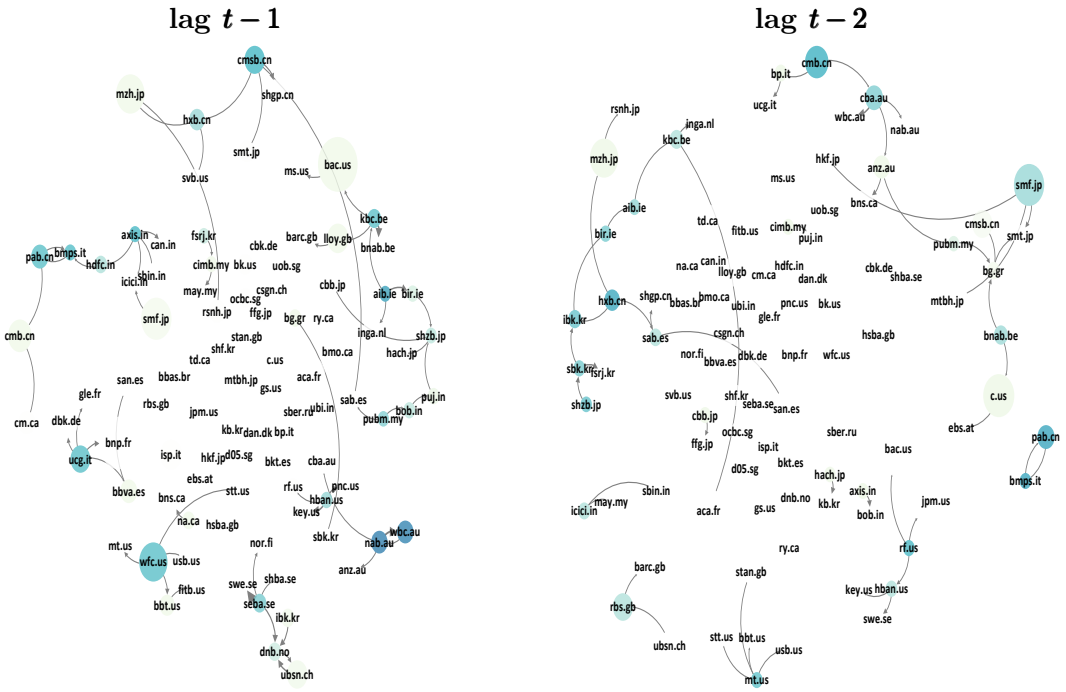
state 1 to state 3 while the WInPR scores of those banks are rather difficult to differentiate across states. This confirms that systemic risk is more associated with risk propagation than with absorption and thus validates our use of WOutPR for states identification. Third, the higher the connectedness level of the state, the larger contribution of risk transmission is concentrated on few top-ranked banks. For instance, comparing the total WOutPR scores of 1-5 and 10-15 ranked banks, the difference rises from 1.5 times to more than 2.5 times from state 2 to state 3. This confirms that the presence of a small number of entities, known as hubs or systemically important nodes, characterizes periods of high systemic risk, thus empirically supporting our intuition behind (2.13).

The just-completed analysis is based on the overall connectedness built in (2.8). Apparently, it suggests a picture that reveals how systemic risk changed over the sample and which individual is the main driver of systemic risk movements. In addition to learning the aggregate information embedded in the overall connectedness, a different perspective of (2.8) is to instead inspect the frequency-specific components of overall connectedness, which offer deeper insights explaining the build-up of systemic risk.

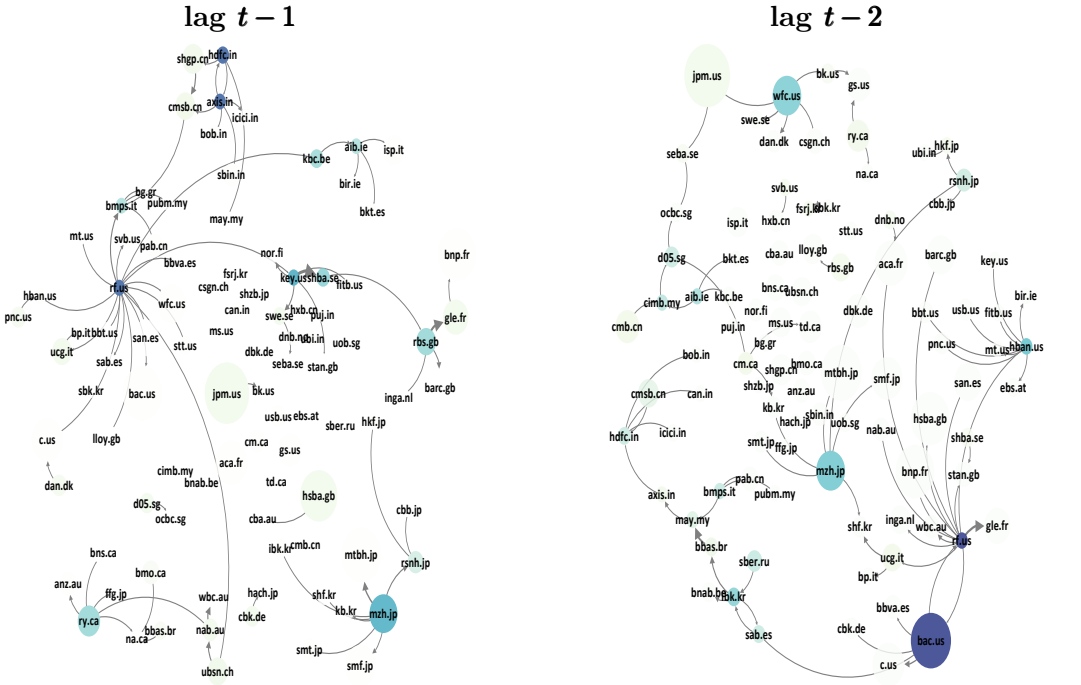
Information about the frequency-specific source of connectedness is encoded by the state-specific structural and autoregressive matrices of our model. The selected MS(3)-GSVAR(2) model specification implies that the global banking overall volatility graph at time t is created by contemporaneous graph (time t), short-term temporal graph (lag $t-1$), and long-term temporal graph (lag $t-2$). From the estimated entries of structural and autoregressive matrices, we visualize the underlying temporal and contemporaneous graphs in Figures 2.2-2.3, respectively:

Temporal graph. In the low connectedness state, both lag $t-1$ and lag $t-2$ graphs are formed by many small, separate clusters with relatively loose intra-cluster connections. At lag $t-1$, clusters are observed for banks in the same nations, such as the one related to the US and Australia, and for those in the same regions, such as the one located in Southern Europe and Northern Europe. Banks in Asian countries, such as those from India, Japan, and China, form three main groups, loosely connected with few US and European banks. At lag $t-2$, national clusters are only clearly observed for the US banks, and regional clusters appear less visible. Several banks in Asia Pacific and Europe merge together into larger clusters. In the high connectedness state, both lag $t-1$ and lag $t-2$ graphs are formed by fewer large, linked clusters with relatively dense intra-cluster connections. At lag $t-1$, banks from the North America, Europe, the Asian developing countries and Australia form one connected component, with the US banks (e.g., rf.us, key.us) creating a bridge transmitting volatility shocks between different geographical banking clusters. Banks in Japan and Korea form another independent cluster, with the Japanese bank mzh.jp being the center. Indian banks axis.in and hdfc.in, whose nodes are dark blue colored, generate strong volatility spillover to other domestic banks. At lag $t-2$, major Japanese and Korean banks (e.g., shf.kr, kb.kr, mzh.jp) attach to the US/European banking cluster. US banks bac.us and rf.us, as indicated by the dark blue colored nodes, generate very high volatility connectedness to both the domestic banks and their European counterparts. In the extreme connectedness state, both lag $t-1$ and lag $t-2$ are formed by one or two US-bank-dominated clusters. At lag $t-1$, two main clusters are led respectively by the US banks hban.us and rf.us. At lag $t-2$, a large cluster is clearly visible, with the US banks bbt.us, fitb.us and wfc.us being central.

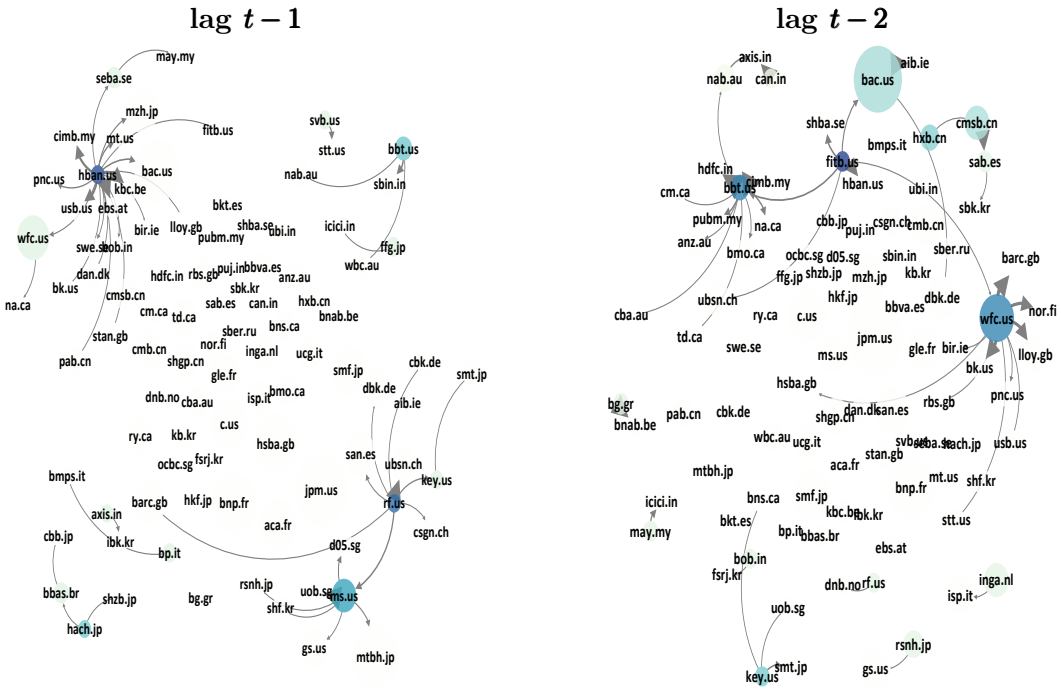
Contemporaneous graph. In the low connectedness state, the contemporaneous graph exhibits clear clustering at both national and regional levels. The national banking cluster is ubiquitous, ranging from countries with many banks in our sample (e.g., USA, Canada, Japan, India) to those with only two or three (e.g., Ireland, Singapore, Malaysia). The



(a) State 1: low connectedness state



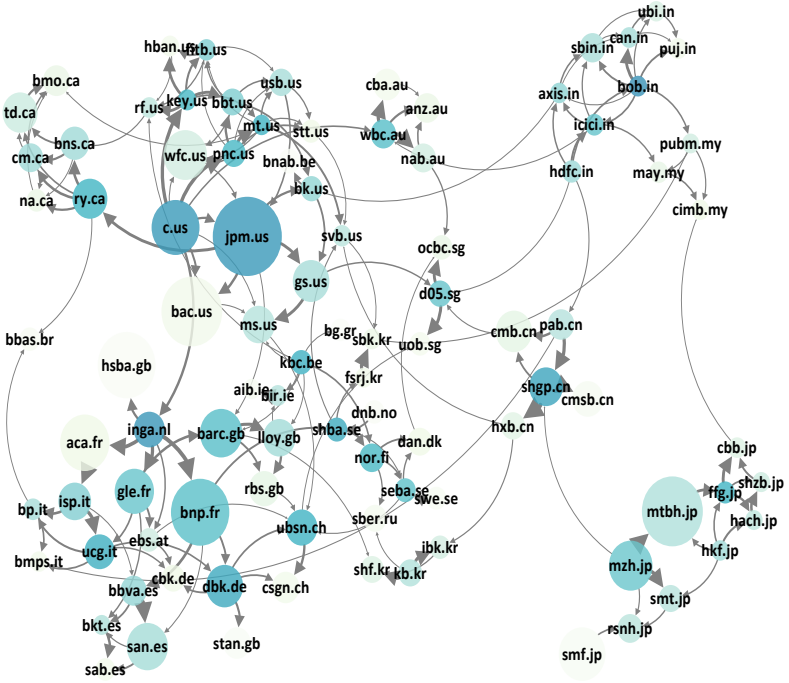
(b) State 2: high connectedness state



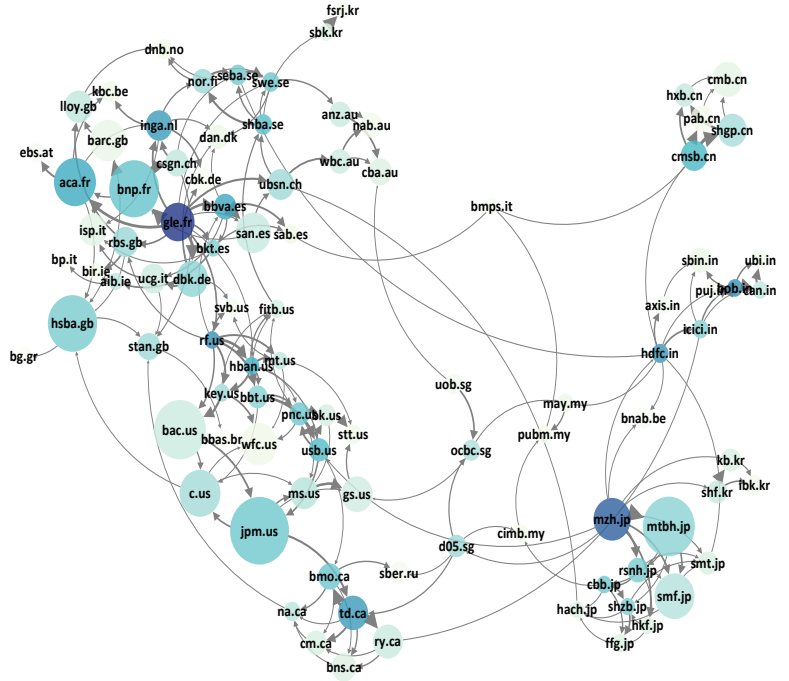
(c) State 3: extreme connectedness state

Figure 2.2 Temporal graphs. We present the maximum a posterior (MAP) estimates of the short-term (left column) and long-term (right column) temporal graphs in the low (panel (a)), high (panel (b)) and extreme (panel (c)) connectedness states. Gephi, an open-source software, is used for graph visualization. In each plot: node naming is the bank code given in the Internet Appendix D; node size indicates bank asset size; node color indicates its WOutPR score; node location, as suggested by Demirer et al. (2018), is determined by the ForceAtlas2 algorithm in Gephi; clockwise-oriented curves indicate directed edges between pairs of nodes; both curve thickness and edge arrow size indicate edge strength. The darker the color of the node, the higher its contribution to risk propagation.

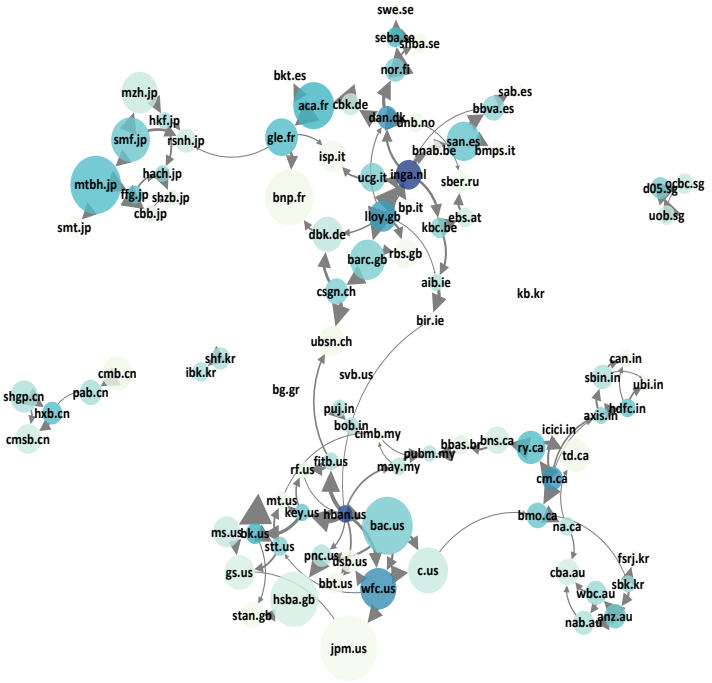
regional cluster is obvious throughout the graph, whose left side tends to contain banks of western countries, and whose middle and right side tend to contain banks of eastern countries. The left side breaks into a large European banking cluster and a smaller US/Canadian cluster, both of which are closely located with each other. The middle is segmented by the Singaporean/Australian, Korean, Indian/Malaysian and Chinese banking clusters, with the first two locating near the western banking groups. The right side is dominated by a separate Japanese cluster, scattering farther away from its Asian neighbors. In the high connectedness state, national and regional banking clusters remain but the relative location of clusters change. The Japanese cluster moves closer to banks of other eastern countries, while the Chinese cluster almost detaches from its regional group. The US/Canadian cluster seems to serve as a bridge connecting the European banks to the eastern banking groups. Aside from these changes, several clusters appear to have central nodes, or hubs (e.g., gle.fr, mzh.jp), the ones that display the stronger volatility spillover to others as indicated by their darker



(a) State 1: low connectedness state



(b) State 2: high connectedness state



(c) State 3: extreme connectedness state

Figure 2.3 Contemporaneous graphs. We present the maximum a posterior (MAP) estimates of the contemporaneous graphs in the low (panel (a)), high (panel (b)) and extreme (panel (c)) connectedness states. Gephi, an open-source software, is used for graph visualization. In each plot: node naming is the bank code given in the Internet Appendix D; node size indicates bank asset size; node color indicates its WOutPR score; node location, as suggested by Demirer et al. (2018), is determined by the ForceAtlas2 algorithm in Gephi; clockwise-oriented curves indicate directed edges between pairs of nodes; both curve thickness and edge arrow size indicate edge strength. The darker the color of the node, the higher its contribution to risk propagation.

colored nodes. In the extreme connectedness state, the relative position of clusters varies once again. Many Asian-Pacific banks (e.g., Malaysia, India, Australia) are now part of the US/Canadian cluster. Japanese banks, although geographically close to the Asia-Pacific region, move towards the European cluster. Banks from China and Singapore are separately grouped in two clusters, not connected at all with other banks. The most central bank nodes in terms of risk-spreading capability are inga.nl and hban.us. It is also noticeable that both intra- and inter-cluster connections become much stronger, as indicated by the thicker edges.

Overall, both temporal and contemporaneous graphs show some similarities that the state with higher level of connectedness is characterized by larger cross-geography banking clusters, stronger system-wide connections, and a few central banks dominating the volatility propagation. Distinctions between both graphs are also visible. In all three states, contemporaneous graphs are noticeably denser than temporal graphs, indicating that the dynamics of the global

banking overall volatility connectedness is mainly driven by changes in its contemporaneous component.

2.6 Conclusions

We proposed a Markov-Switching Graphical Structural Vector Autoregressive (MS-GSVAR) model, which facilitates an integrated analysis of some salient empirical features of financial connectedness, such as the mixture of contemporaneous and temporal dependences, high-dimensionality, and abrupt structural changes. Methodologically, our model synthesized a structural vector autoregression (SVAR) to jointly capture the contemporaneous and temporal dependence structures of a multivariate system, a graphical representation to generate parsimonious parameterization, and a hidden Markov chain process to account for potential structural changes in coefficients and dependence (graph) structures underlying the model. Building on this modeling framework, an identification scheme for different states of system connectivity was proposed by exploiting the centrality properties of a weighted directed graph that unionizes the contemporaneous and temporal graph structures in each state. Model inference was performed following a Markov Chain Monte Carlo (MCMC) procedure. An efficient graph inference method was developed to address the computational complexities arising from inference on graph structures in the context of high model dimension, numerous lags, and multiple states. Simulation studies validated the effectiveness of the proposed framework in recovering many empirically relevant graph structures, and in handling large datasets with changing graph structures. The MS-GSVAR model applied to the volatility series of 96 global banks detected different connectedness states, identified systemically important individuals, and uncovered the frequency-specific source of connectedness, which are relevant to systemic risk management.

Appendix 2.A

This appendix provides technical details on the graph inference method.

2.A.1 Graph theory concepts

Suppose that we have a set of random variables, $\{1,...,n\}$, and the number of observations is T. We denote the $T \times n$ data matrix by $\mathcal{D}$. Variables and conditional dependences among these variables can be represented by nodes and edges in a graph $\mathcal{G}$ with the following definition:

Definition 2.A.1 (Graphical model)**.** A graphical model $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a system of nodes (or vertices) $\mathcal{V} = \{1,...,n\}$ and connecting edges (or lines) $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, where n is the number of nodes in the system.

A graph can be classified as either undirected, i.e., when the edges have no orientation, or directed, when the edges are arrows indicating the direction of the relation. Based on these notions, two main types of graphs can be defined; namely, undirected and directed graphs. A special case of directed graphs corresponds to the so called Directed Acyclic Graph (DAG), which has received attention in the SVAR literature since the graphical representation of a recursive system is given by a DAG. For this reason, in this paper, we particularly focus on DAGs. We further adopt the following convention from DAGs regarding edge definitions. An edge $(r,l) \in \mathcal{E}$ is called directed if $(r,l) \in \mathcal{E}$ but $(l,r) \notin \mathcal{E}$, for some $r,l \in \mathcal{V}$. If both $(r,l) \in \mathcal{E}$ and $(l,r) \in \mathcal{E}$, the edge is called undirected. In the former case, the directed edge is denoted as $l \to r$, while in the latter case the undirected edge is denoted as $l - r$.

Formal definitions of both undirected and DAG graphs are given as follows. An undirected graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of the set $\mathcal{V}$ of nodes and $\mathcal{E}$ of edges, which are unordered pairs of elements of $\mathcal{V}$. Formally,

Definition 2.A.2 (Undirected graph)**.** A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is called undirected if $\forall r,l \in \mathcal{V}, (r,l) \in \mathcal{E}$ if and only if $(l,r) \in \mathcal{E}$.

The definition of Directed Acyclic Graph is based on the following two concepts:

Definition 2.A.3 (Directed cycle)**.** A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ has a directed cycle, if exists a sequence of directed edges $\{(r_1,r_2),(r_2,r_3),...,(r_{n-1},r_n)\}$, such that $(r_j,r_{j+1}) \in \mathcal{E}$ for each $j \in \{1,...,n-1\}$, with $r_1 = r_n$.

A directed graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of the set $\mathcal{V}$ of nodes and $\mathcal{E}$ of edges, which are ordered pairs of elements of $\mathcal{V}$. Formally,

Definition 2.A.4 (Directed graph)**.** A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is directed if $\forall r,l \in \mathcal{V}$ if $(r,l) \in \mathcal{E}$ then $(l,r) \notin \mathcal{E}$.

Now we are in condition to define a DAG:

Definition 2.A.5 (Directed Acyclic Graph)**.** A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is called a *Directed Acyclic Graph* (DAG) if all edges are directed and there are no directed cycles.

Whenever we want to focus on the structure of a DAG regardless of the direction of connectivity, the following concept is used:

Definition 2.A.6 (Skeleton of a DAG)**.** The graph generated by replacing all directed edges of a DAG with undirected edges is called a skeleton.

Besides, a related concept of DAG is the notion of Bayesian networks (or graphs):

Definition 2.A.7 (Bayesian networks (or graphs))**.** A class of graphical model with DAGs as their underlying structure is called Bayesian networks (or graphs).

We also introduce the following notations, which are invoked by the graph inference method developed in this paper.

If nodes $r, l \in \mathcal{V}$ and there is a directed edge $l \to r$, node l is said to be a *parent* of node r. For each $r \in \mathcal{V}$, denote Pa_r the set of parents of r.

The adjacency set of node r in graph $\mathcal{G}$, denoted by $Adj(\mathcal{G}, r)$, are all nodes l which are connected to r by an edge (either directed or undirected). The elements of $Adj(\mathcal{G}, r)$ are also called *neighbors* of r.

Graph $\mathcal{G}$ with n nodes can be represented through an n-dimensional binary matrix A, called adjacency matrix[7]. Each entry A_{rl} in A is equal to 1 if $l \to r$ with $r, l \in \mathcal{V}$, and 0 otherwise; where $r \neq l$, since self-loops are not allowed.

Any DAG has at least one topological order, denoted by $\prec$, which is simply a permutation of the n node labels. We associate a permutation $\pi_\prec$ with each order. A DAG is said to be compatible with an order, denoted by $\mathcal{G} \in \prec$, if the parents of each node have a higher index in the permutation. Mathematically,

$$\mathcal{G} \in \prec \overset{\text{def}}{\iff} \forall r, \forall \{l : l \in \text{Pa}_r\}, \pi_\prec[r] < \pi_\prec[l]. \tag{2.A.1}$$

Visually, when we place n nodes in a linear chain from left to right according to $\pi_\prec$, the parents of r can only come from $r+1, r+2, ..., n$; that is parents are chosen only from nodes that follow r in the order. For example, if for $n = 3$, we choose the order 3, 1, 2 then node 3 could have either of the others (or both or none) as parents. Node 1 may only have node 2 (or none) as parents while node 2 is forced to have no parents. The possible choices of parents can be represented as an adjacency matrix where the rows and columns are labeled

[7] In this paper, we use the terms *graph* and *adjacency matrix* interchangeably.

according to the order,

$$A = \begin{pmatrix} A_{1,1}=0 & A_{1,2}=\{0,1\} & A_{1,3}=0 \\ A_{2,1}=0 & A_{2,2}=0 & A_{2,3}=0 \\ A_{3,1}=\{0,1\} & A_{3,2}=\{0,1\} & A_{3,3}=0 \end{pmatrix}, \tag{2.A.2}$$

so that each choice for the entries in A is a different DAG and there are therefore eight different DAGs compatible with this order of three nodes.

Finally, we use $|\cdot|$ for the set cardinality and $\cdot\backslash\{\cdot\}$ for the set difference.

2.A.2 State-of-the-Art structure learning methods

Structure learning in graphical models refers to inference on their underlying graphs based on an observed data sample. Existing methods to structure learning fall into three categories:

1. Constraint-based methods use conditional independence (CI) tests to learn the edges of the graph. Examples are the PC (Spirtes et al., 2000; Kalisch and Bühlman, 2007), and RFCI algorithms (Colombo et al., 2012).

2. Score-and-search methods rely on an efficient search algorithm in the space of graphs (or node orders) and a score function to find the graph which best explains the data. The most basic method is Structure-MCMC (proposed by Madigan et al. (1995), later refined by Giudici and Castelo (2003)), where each step involves adding, deleting or reversing an edge and accepting the move according to a Metropolis–Hastings probability. Order-MCMC (Friedman and Koller, 2003) spearheads a path to reduce the search space by combining large collections of DAGs together, namely all DAGs compatible with the same topological ordering of the nodes. Other approaches to score-and-search include for example the greedy equivalence search (Chickering, 2002), dynamic or integer linear programming approaches (Koivisto and Sood, 2004), and Partition MCMC (Kuipers and Moffa, 2017).

3. Hybrid methods bring together the ease of constraint-based methods and the performance of score-and-search methods, to benefit from their individual advantages (Tsamardinos et al., 2006; Nandy et al., 2018; Kuipers et al., 2022).

For further material on the up-to-date Bayesian network structure learning methods, we refer the reader to Kitson et al. (2023).

2.A.3 Sample $A_{0:p}(k)$

From definitions 2.2.1-2.2.2 of the paper, $A_0(k)$ and $A_i(k)$ underlying the MS-GSVAR model takes the following form

$$
\begin{aligned}
A_0(k) &= \begin{pmatrix} 0 & A_{1,2}(k) & \cdots & A_{1,n}(k) \\ A_{2,1}(k) & 0 & \cdots & A_{2,n}(k) \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1}(k) & A_{n,2}(k) & \cdots & 0 \end{pmatrix}, \\
A_i(k) &= \begin{pmatrix} A_{1,ni+1}(k) & A_{1,ni+2}(k) & \cdots & A_{1,ni+n}(k) \\ A_{2,ni+1}(k) & A_{2,ni+2}(k) & \cdots & A_{2,ni+n}(k) \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,ni+1}(k) & A_{n,ni+2}(k) & \cdots & A_{n,ni+n}(k) \end{pmatrix}, \text{ for } 1 \leq i \leq p.
\end{aligned}
\tag{2.A.3}
$$

Given a data matrix $\mathcal{D}$ of dimension $|\mathcal{T}_k| \times (np+n)$ (note that $\mathcal{T}_k$ has been defined in Section 2.4 of the paper), our goal is to learn the configuration of the binary entries of each lag-specific adjacency matrix $A_0(k), ..., A_p(k)$ from the observed data. To sample these adjacency matrices, we implement a three-stage procedure based on the Hybrid-MCMC algorithm proposed by Kuipers et al. (2022) (as discussed in Section 2.4.1 of the paper). Details of each stage are provided in the following subsections.

2.A.3.1 Stage 1 (Initialization)

In Stage 1, we initialize a configuration of the binary entries of each adjacency matrix through the PC algorithm described in Spirtes et al. (2000, p.117-119).[8] This leads to a massive reduction in the space of graphs to be searched in Stage 2. The pseudo-code of Stage 1 is given in Algorithm 1. The implementation details are illustrated as follows:

The starting point is to form a complete undirected graph, which assumes all nodes of the system are connected. Particularly, we set those matrices in (2.A.3) to

$$
A_0(k) = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \; A_i(k) = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \text{ for } 1 \leq i \leq p. \tag{2.A.4}
$$

and stack them into a matrix $A_{0:p}(k)$ with dimension $n \times (np+n)$.

[8] The PC algorithm (Spirtes et al., 2000) has two main steps. In the first step, it learns from data a skeleton of DAG, which contains only undirected edges. In the second step, it orients the undirected edges to form a class of equivalent DAGs. Consistent with Kuipers et al. (2022), our Stage 1 simply involves the first step to avoid mistakes in orienting edges.

Using a $|\mathcal{T}_k| \times (np+n)$ data matrix $\mathcal{D}$ as an input, edges between nodes in $A_{0:p}(k)$ are then removed or retained based on the conditional independence (CI) tests. For each edge, the CI tests decide if the pair of nodes connected by the edge, r and l, are independent conditioning on a subset s of all neighbors of r and l:

- In the case of continuous multivariate Gaussian data, the CI tests use Fisher's z statistic (Fisher, 1915) to test for significance from zero

$$z(r,l|s) = \frac{1}{2}\log(\frac{1+\hat{\rho}_{r,l|s}}{1-\hat{\rho}_{r,l|s}}), \tag{2.A.5}$$

  where $\hat{\rho}_{r,l|s}$ is the sample partial correlation between r and l given s. To conduct a two-sided hypothesis test at significance level α, one may test if

$$\sqrt{T-|s|-3}\, z(r,l|s) \leq \Phi^{-1}(1-\alpha/2), \tag{2.A.6}$$

  where $T = |\mathcal{T}_k|$ is the sample size of data in state k; $|s|$ denotes the cardinality of s; Φ^{-1} is the inverse CDF of $\mathcal{N}(0,1)$. As with Kuipers et al. (2022), we use a conservative significance level $\alpha = \min(0.2, 10/n)$ throughout the paper to avoid Stage 1 initializes too many false positive edges for large-scale graphs.

The CI tests are organized by levels (based on the size of the conditioning sets, e.g. the depth d). At the first level ($d=0$), all pairs of nodes are tested conditioning on the empty set. Some of the edges would be deleted and the algorithm only tests the remaining edges in the next level ($d=1$). The size of the conditioning set, d, is progressively increased (by one) at each new level until d is greater than the size of the adjacent sets of the testing nodes. The output graph is a skeleton of $A_{0:p}(k)$ ($A^{\text{PC}}_{0:p}(k)$ hereafter).

2.A.3.2 Stage 2 (Structure learning)

In Stage 2, we run the Order-MCMC algorithm sequentially for each lag-specific adjacency matrix stacked in the initialized $A^{\text{PC}}_{0:p}(k)$ (i.e., $A^{\text{PC}}_0(k), A^{\text{PC}}_1(k), ..., A^{\text{PC}}_p(k)$) through M iterations. For expositional convenience, each lag-specific adjacency matrix is labeled as $A(k)$ hereafter in this section and Section 2.A.3.3. In all of our simulation and empirical exercises, we set $M = 40{,}000$ with 20% burn-in samples. Thus, a total of 32,000 iterations are used as the basis for graph inference, which is a reasonable number for the MCMC chain to converge[9] while still controlling the runtime.

The pseudo-code of Stage 2 is given in Algorithm 2. The implementation details are illustrated as follows:

[9]Similar to Ahelegbey et al. (2016a), the convergence of the MCMC chain with the selected number of iterations is monitored by the potential scale reduction factor (PSRF) and the multivariate PSRF (MPSRF). The MCMC chain is said to have properly converged if PSRF(MPSRF) ≤ 1.2. In all of our simulation and empirical exercises, PSRF and MPSRF are generally less than 1.1 in the case of 32,000 iterations.

Algorithm 1 Initializing the skeleton via the PC algorithm

Input: Data sample $\mathcal{D}$, CI test $Test(r,l|s,\mathcal{D})$ with significance level α.
Output: a skeleton $A^{\mathrm{PC}}_{0:p}(k)$.

```
Start with a complete undirected graph A_{0:p}(k).
Let depth d = 0
repeat
    for each ordered pair of adjacent nodes r and l in A_{0:p}(k) do
        if |Adj(A_{0:p}(k), r)\{l}| ≥ d then
            for each subset s ⊆ Adj(A_{0:p}(k), r)\{l} and |s| = d do
                Test(r, l|s, D)
                if Test(r, l|s, D) accepts then
                    Update A_{0:p}(k) by removing l − r from A_{0:p}(k)
                    break
                end if
            end for
        end if
    end for
    Let d = d + 1
until |Adj(A_{0:p}(k), r)\{l}| < d for every pair of adjacent nodes in A_{0:p}(k)
```

For $m = 1, 2, ..., M$

Step (a). Use the Metropolis–Hastings algorithm to sample the order according to its posterior probability defined in Kuipers et al. (2022, eq. (18))

$$p(\prec|\mathcal{D}) \propto \prod_{r=1}^{n} \sum_{\substack{\mathrm{Pa}_r \subseteq h^r \\ \mathrm{Pa}_r \in \prec}} \left[S(r, \mathrm{Pa}_r|\mathcal{D}) + \sum_{\substack{l \notin h^r \\ \pi_\prec[r] < \pi_\prec[l]}} S(r, \{\mathrm{Pa}_r, l\}|\mathcal{D}) \right], \tag{2.A.7}$$

where

- the first sum is restricted to parent subsets compatible with the order

$$\mathrm{Pa}_r \in \prec \overset{\text{def}}{\Longleftrightarrow} \forall \{l : l \in \mathrm{Pa}_r\}, \pi_\prec[r] < \pi_\prec[l], \tag{2.A.8}$$

  together with the additional constraint that all elements in the parent sets must belong to a set of possible parents h^r initialized through Stage 1

$$\mathrm{Pa}_r \subseteq h^r \text{ with } h^r = \{l : A^{\mathrm{PC}}_{r,l}(k) = 1\} \tag{2.A.9}$$

- the second sum is used for adding edges beyond a predefined skeleton, allowing for the correction of errors where edges may be missed in Stage 1. In particular, the possible parent sets of every node also include one additional parent from among the nodes outside its predefined parent sets according to the order, that is $l \notin h^r$, $\pi_\prec[r] < \pi_\prec[l]$.

- $S(\cdot|\mathcal{D})$ is a score function evaluating on each node and its parents. Since this paper assumes continuous multivariate Gaussian data, $S(\cdot|\mathcal{D})$ is formulated based on the Bayesian Gaussian equivalent (BGe) score function, as suggested by Kuipers et al. (2014).
- For the efficient computation of the sum, we follow Kuipers et al. (2022, sec. 2.7) to pre-compute and tabulate every score quantity $S(\cdot|\mathcal{D})$ needed for (2.A.7). During each Order-MCMC iteration, we then simply need to look up the relevant scores providing a very efficient sampler.

From the order $\prec^{[m-1]}$ at the previous iteration $[m-1]$ (The initial order $\prec^{[0]}$ is generated by randomly permuting the n nodes), a new order $\prec'$ at the current iteration $[m]$ is proposed based on three types of permutation (Kuipers et al., 2022):

- global swap: swapping two random nodes in $\prec^{[m-1]}$.
- local transposition: swapping two adjacent nodes in $\prec^{[m-1]}$.
- node relocation: sequentially transpose a random node with its adjacent nodes. Further details are provided in Kuipers et al. (2022, sec. 2.4).

To reduce computational complexity, the three types of permutation are mixed into a single scheme by selecting each of them at each iteration with a probability of $(\frac{6}{n+7}, \frac{n}{n+7}, \frac{1}{n+7})$, respectively.
If the new order is generated from the first two permutations, the new order is accepted, that is, $\prec^{[m]}=\prec'$, with the Metropolis-Hastings probability

$$\min\left\{1, \frac{p(\prec'|\mathcal{D})}{p(\prec^{[m-1]}|\mathcal{D})}\right\}. \tag{2.A.10}$$

where $p(\prec'|\mathcal{D})$ has been defined in (2.A.7).
If the new order is generated from the last permutation, the new order is always accepted.

Step (b). Sampling a graph structure $A(k)^{[m]}$ at the current iteration $[m]$ by sampling the parents of each node independently according to the scores of its admissible parent sets in $\prec^{[m]}$, namely $S(r, \mathrm{Pa}_r|\mathcal{D})$ where $\mathrm{Pa}_r \in \prec^{[m]}$.

2.A.3.3 Stage 3 (Bayesian model averaging)

In Stage 3, we follow Ahelegbey et al. (2016a) to employ Bayesian model averaging for building a single sample of every lag-specific adjacency matrix from their own posterior samples generated in Stage 2.

Algorithm 2 Sampling the graphs via the Order-MCMC algorithm

Input: Data sample $\mathcal{D}$, initial skeleton $A^{\text{PC}}_{0:p}(k)$, initial order $\prec^{[0]}$, precomputed scores $S(\cdot|\mathcal{D})$.
Output: posterior samples $A(k)^{[1]}, ..., A(k)^{[M]}$

For $m = 1, 2, ..., M$:

(a) Generate a new order $\prec'$ by randomly permuting the nodes in $\prec^{[m-1]}$ based on three types of permutation: global swap, local transposition, and node relocation;

(i) If the new order is generated from the first two permutations, the new order is accepted, that is, $\prec^{[m]}=\prec'$, with the Metropolis-Hastings probability

$$\min\left\{1, \frac{p(\prec'|\mathcal{D})}{p(\prec^{[m-1]}|\mathcal{D})}\right\}.$$

(ii) If the new order is generated from the last permutation, the new order is always accepted.

(b) Sample a graph $A(k)^{[m]}$ from the space of DAGs compatible with the order $\prec^{[m]}$ generated as before.

For each lag-specific adjacency matrix, the posterior estimate for an edge from node l to node r in state k, i.e., $A_{r,l}(k)$, is defined as

$$A_{r,l}(k) = \begin{cases} 1 & \text{if } q_{r,l,1-\alpha}(k) > \rho \\ 0 & \text{otherwise} \end{cases} \tag{2.A.11}$$

with

$$q_{r,l,1-\alpha}(k) = \bar{A}_{r,l}(k) - z_{1-\alpha}\sqrt{\frac{\bar{A}_{r,l}(k)(1-\bar{A}_{r,l}(k))}{n_{\text{eff}}}} \tag{2.A.12}$$

where $\bar{A}_{r,l}(k)$ is the posterior probability of the edge $A_{r,l}(k)$ and is approximated based on the average of posterior samples obtained in Stage 2; ρ is the posterior probability of edge inclusion. In simulation of Section 2.C.1, we examine $\rho \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$ following Kuipers et al. (2022). In simulation of Section 2.C.2 and the empirical illustration, we use $\rho = 0.5$ since this values balance well the true positives and false positives according to our examination; $z_{1-\alpha}$ is the z-score of the standard normal distribution at significance level $1-\alpha$ where $\alpha = 0.05$; n_{eff} is the effective sample size (see Casella and Robert, 2004, pp. 499–500) representing the number of independent posterior samples.

Appendix 2.B

This appendix introduces the posterior computation of model parameters.

2.B.1 Sample $S_{1:T}$

We sample the whole path $S_{1:T}$ from their joint full conditional distribution based on a forward-filtering backward-sampling (FFBS) algorithm (see Frühwirth-Schnatter, 2006, chap. 13). The FFBS algorithm consists of two stages. In the first stage, we carry out a forward recursion to obtain the filtered probability. In the second stage, we compute the joint full conditional density of $S_{1:T}$ using the filtered probabilities, and randomly draw $S_{1:T}$ from its full conditional distribution through a backward recursion. The stages of the FFBS algorithm are described in the following. First, the filtered probability at time t, $t = 1, ..., T$, is determined by implementing the forward recursion

$$p(s_t = j|Y_{1:t-1}, \Theta, A) = \sum_{i=1}^{K} p_{ij} p(s_{t-1} = i|Y_{1:t-1}, \Theta, A), \tag{2.B.1}$$

for $j = 1, ..., K$, where p_{ij} is the conditional probability of switching from state i at time $t-1$ to state j at time t. We initialize for $t = 1$, $P(s_0 = j|\Theta)$ to be equal to the ergodic probabilities. The filtered probability for all $j = 1, ..., K$ is computed as

$$p(s_t = j|Y_{1:t}, \Theta, A) = \frac{p(Y_t|s_t = j, Y_{1:t-1}, \Theta) p(s_t = j|Y_{1:t-1}, \Theta, A)}{p(Y_t|Y_{1:t-1}, \Theta)}, \tag{2.B.2}$$

where the denominator is the marginal predictive likelihood defined as

$$p(Y_t|Y_{1:t-1}, \Theta) = \sum_{i=1}^{K} p(Y_t|s_t = i, Y_{1:t-1}, \Theta) p(s_t = i|Y_{1:t-1}, \Theta, A).$$

The smoothing probabilities are obtained recursively and backward in time, once all the filtered probabilities $p(s_t = j|Y_{1:t}, \Theta, A)$ for $t = 1, ..., T$ are calculated. If $t = T$, smoothing probability and filtered probability are equal. For $t = T-1, T-2, ..., 1$ and for all $j = 1, ..., K$ the backward recursion proceeds as follows

$$\begin{aligned} p(s_t = j|Y_{1:T}, \Theta, A) &= \sum_{i=1}^{K} p(s_t = j, s_{t+1} = i|Y_{1:T}, \Theta, A) \\ &= \sum_{i=1}^{K} p(s_t = j|s_{t+1} = i, Y_{1:t}, \Theta, A) p(s_{t+1} = i|Y_{1:T}, \Theta, A) \end{aligned} \tag{2.B.3}$$

where

$$p(s_t = j|s_{t+1} = i, Y_{1:t}, \Theta, A) = \frac{p_{ji} p(s_t = j|Y_{1:t}, \Theta, A) p(s_t = j|Y_{1:t}, \Theta, A)}{p(s_{t+1} = i|Y_{1:T}, \Theta, A)}.$$

2.B.2 Sample $B^*_{1:p}(k)$

From (2.7) of the paper, the adjacency matrices associated with $B^*_{1:p}(k)$ can be expressed as $A^*_{1:p}(k) = (I_n - A_0(k))^{-1}A_{1:p}(k)$. Let matrix $A^*(k)$ be a stacked form of $A^*_1(k), ..., A^*_p(k)$ such that $A^*(k)$ is of dimension $n \times np$. Denote with $\pi_r(k) = \{l = 1, ..., np : A^*_{rl}(k) = 1\}$ the index set of non-zero entries in the rth row of $A^*(k)$, and with $|\pi_r(k)|$ its cardinality. We define the selection matrix $E_r(k) = (e_{j_1}, ..., e_{j_{|\pi_r(k)|}})$, where $E_r(k)$ is of dimension $np \times |\pi_r(k)|$, j_i is an element of the index set $\pi_r(k)$, i.e., $j_i \in \pi_r(k)$, and e_i is an $np \times 1$ vector of zeros with a one in the ith position. Let $Z_t = (Y'_{t-1}, ..., Y'_{t-p})$ be a $1 \times np$ vector of explanatory variables. The set of selected explanatory variables that predict the rth element of Y_t (i.e., $Y_{r,t}$) can be represented by $Z_{r,t} = Z_t E_r(k)$. We sample $B^*_{1:p}(k)$ from its full conditional distribution, which is defined as

$$p(B^*_{1:p}(k)|Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, A_{0:p}(k), \Omega_\mu(k)) \propto \mathcal{N}(M(k)^*(\sum_{t\in\mathcal{T}_k} Z^*_t \Omega_\mu(k) Y_t), M(k)^*), \tag{2.B.4}$$

where $Z^*_t = \text{diag}(Z_{1,t}, ..., Z_{n,t})$ is an $n \times \sum_{r=1}^n |\pi_r(k)|$ matrix in block diagonal form, and $M(k)^* = (\sum_{t\in\mathcal{T}_k} Z^*_t \Omega_\mu(k)(Z^*_t)' + M(k)^{-1})^{-1}$ with $M(k) = 100 I_{\sum_{r=1}^n |\pi_r(k)|}$ given the prior specified in Section 2.4.2 of the paper.

2.B.3 Sample $B_0(k)$

Given the sampled adjacency matrix $A_0(k)$, we denote with $\pi_r(k) = \{l = 1, ..., n : A_{rl,0}(k) = 1\}$ the index set of non-zero entries in the rth row of $A_0(k)$ and with $|\pi_r(k)|$ its cardinality. We define the selection matrix $E_r(k) = (e_{j_1}, ..., e_{j_{|\pi_r(k)|}})$, where $E_r(k)$ is of dimension $n \times |\pi_r(k)|$, j_i is an element of the index set $\pi_r(k)$, i.e., $j_i \in \pi_r(k)$, and e_i is an $n \times 1$ vector of zeros with a one in the ith position. It can be seen from (2.7) of the paper that $\mu_t = B_0(k)\mu_t + \varepsilon_t$, where $\mu_t = Y_t - Z^*_t B^*_{1:p}(k)$. The set of selected explanatory variables that predict the rth element of μ_t (i.e., $\mu_{r,t}$) can be represented by $U_{r,t} = \mu_t E_r(k)$. We sample $B_0(k)$ from its full conditional distribution, which is defined as

$$p(B_0(k)|Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, A_0(k), \Omega_\varepsilon(k), B^*_{1:p}(k)) \propto \mathcal{N}(M(k)^*(\sum_{t\in\mathcal{T}_k} U^*_t \Omega_\varepsilon(k) \mu_t), M(k)^*), \tag{2.B.5}$$

where $U^*_t = \text{diag}(U_{1,t}, ..., U_{n,t})$ is an $n \times \sum_{r=1}^n |\pi_r(k)|$ matrix in block diagonal form, and $M(k)^* = (\sum_{t\in\mathcal{T}_k} U^*_t \Omega_\varepsilon(k)(U^*_t)' + M(k)^{-1})^{-1}$ with $M(k) = 100 I_{\sum_{r=1}^n |\pi_r(k)|}$ given the prior specified in Section 2.4.2 of the paper.

2.B.4 Sample $\Omega_\varepsilon(k)$

Denote with $|\mathcal{T}_k|$ its cardinality. We sample $\Omega_\varepsilon(k)$ from its full conditional distribution, which is defined as

$$p(\Omega_\varepsilon(k)|Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, B_0(k), B^*_{1:p}(k)) \propto \mathcal{W}_G(\nu_\varepsilon(k)+|\mathcal{T}_k|, R_\varepsilon(k)+\sum_{t\in\mathcal{T}_k}\varepsilon_t\varepsilon_t'), \tag{2.B.6}$$

with $\varepsilon_t = \mu_t - U_t^* B_0(k)$.

2.B.5 Sample $\Omega_\mu(k)$

We sample $\Omega_\mu(k)$ from its full conditional distribution, which is defined as

$$p(\Omega_\mu(k)|Y_{\mathcal{T}_k}, S_{\mathcal{T}_k}, B^*_{1:p}(k)) \propto \mathcal{W}(\nu_\mu(k)+|\mathcal{T}_k|, R_\mu(k)+\sum_{t\in\mathcal{T}_k}\mu_t\mu_t'). \tag{2.B.7}$$

2.B.6 Sample P

We sample $p(k)$, the kth row of the transition matrix P, from its full conditional distribution, which is defined as

$$p(p(k)|S_{1:T}) \propto \mathcal{D}ir(c_{k1}+N_{k1}, ..., c_{kK}+N_{kK}), \tag{2.B.8}$$

where N_{km}, $m \in \{1, ..., K\}$, is the empirical distribution which counts the transition between the kth and the mth latent discrete states, i.e., $N_{km} = \sum_{t=1}^{T} \mathbb{I}(s_t = m)\mathbb{I}(s_{t-1} = k)$, where $\mathbb{I}(\cdot)$ is an indicator function.

Appendix 2.C

This appendix includes two simulation studies that highlight key features of our framework. The first simulation study aims to illustrate the validity of the proposed graph inference method in recovering empirically relevant graphs under different settings of data dimensions and sample sizes. The second aims to illustrate the performance of the proposed MS-GSVAR model and the effectiveness of its estimation procedure in handling large datasets with changing dependence structures.

2.C.1 Simulation study to compare graph inference methods

In this simulation, we evaluate the ability of our method to recover the true graphs underlying the SVAR(p) process in (2.1) of the paper. For simplicity, we simulate data from an SVAR(1) process with a unit diagonal covariance matrix. We consider four empirically-relevant scenarios of the true graph structures, each of which corresponds to a different form of the coefficient matrices B_i, $0 \leq i \leq 1$, of the SVAR(1). The first three are the *sparse*, *dense*, and *weak edge strength* scenarios, similar to the set-ups in Kuipers et al. (2022), and the fourth one is the *community* (or *block*) scenario, which has been considered in Billio et al. (2019)

1. *Sparse*: we assume the true graph structure underlying each of the coefficient matrices to be a random DAG with relatively sparse pattern. The setting described corresponds to a random matrix B_i in which the non-zero entries are drawn from the Erdős-Rényi model with an average of two parents per node and their values are sampled uniformly in the interval $[-0.7, -0.3] \cup [0.3, 0.7]$.

2. *Dense*: we assume the true graph structure underlying each of the coefficient matrices to be a random DAG with relatively dense pattern. The setting described corresponds to a random matrix B_i in which the non-zero entries are drawn from the Erdős-Rényi model with an average of three parents per node and their values are sampled uniformly in the same interval as the scenario *Sparse*.

3. *Weak edge strength*: we assume the true graph structure underlying each of the coefficient matrices to be a random DAG with relatively weak edge strengths. The setting described corresponds to a random matrix B_i in which the non-zero entries are drawn from the Erdős-Rényi model again with an average of two parents per node while their values are sampled uniformly in the interval $[-0.3, -0.1] \cup [0.1, 0.3]$.

4. *Community*: we assume the true graph structure underlying each of the coefficient matrices to be a random DAG with block pattern. The setting described corresponds to a block-diagonal matrix B_i in which the number of blocks is $n/4$ and each block has the same structure as the scenario *Sparse*.

For each scenario, we consider a total of six simulation settings based on different combinations of node dimensions and sample sizes: $n \in \{20 \text{ (small)}, 40 \text{ (medium)}, 80 \text{ (large)}\}$ and $T \in \{5n, 10n\}$. For each setting, we replicate the simulation and estimation exercises 50 times, with different randomly sampled graphs and data. All the results reported are aggregated based on 50 replications.

We consider the Structure-MCMC algorithm as a benchmark for comparison.[10] Both Structure-MCMC algorithm and our own algorithm are run for 8,000 iterations of burn-in, followed by 32,000 iterations as the basis for inference. To evaluate the performance of both algorithms, we use the true positive rate (TPR) and false positive rate (FPRp) as in Kuipers et al. (2022). Plotting TPR versus FPRp over varying thresholds on the posterior probability of edge inclusion produces a ROC curve, with curves close to the top left corner indicating better performance at returning the true graph structure. Figures 2.C.1-2.C.4 display respectively the ROC curves under the four scenarios. Within each scenario, our method yields better performance for all settings considered. It achieves a similarly high level of TPR as Structure-MCMC but notably lower level of FPRp across varying posterior probabilities of edge inclusion. This finding is particularly noticeable in the large-dimensional setting when $n = 80$. Moreover, the small interquartile ranges for TPR suggest that our method is stable across simulation replications. Figures 2.C.5-2.C.8 plot respectively the average runtimes for both algorithms under the four scenarios. It indicates the computational gain of using our method in the large-dimensional setting: when $n = 80$, our algorithm is more than 18 times faster than Structure-MCMC. Overall, our method relying on Hybrid-MCMC provides more reliable graph inference in much less time than Structure-MCMC in all cases we consider. The improved accuracy and the efficiency gain achieved by our method are substantial particularly in the higher dimension up to the case of 80 series.

[10] For the implementation of the Structure-MCMC algorithm, we use the code provided in Ahelegbey et al. (2016a), which is available in the *Journal of Applied Econometrics Data Archive*.

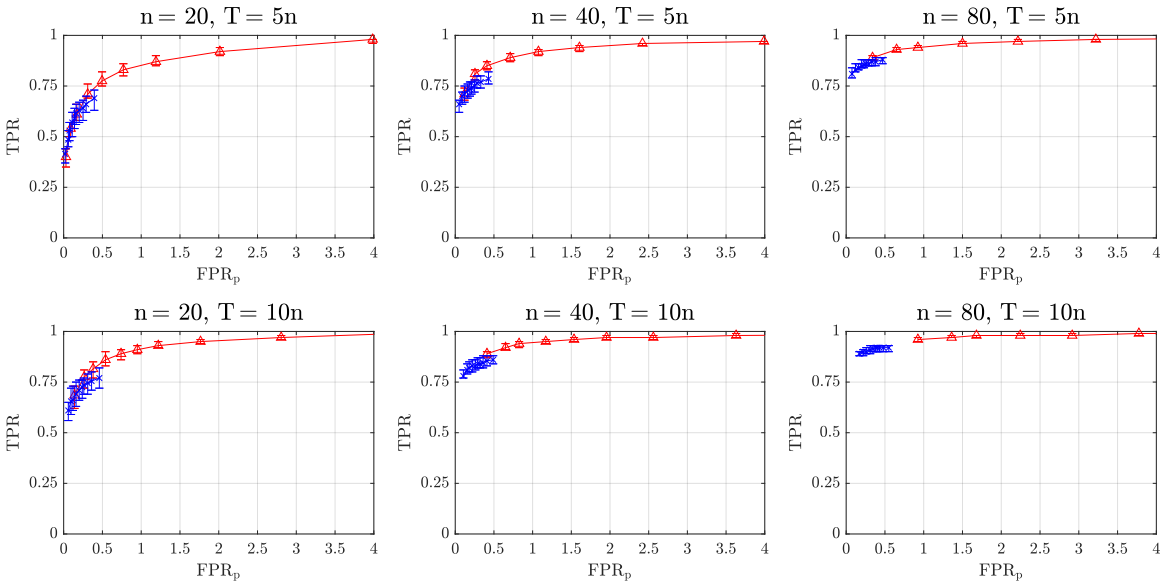


Figure 2.C.1 ROC curves under the scenario *Sparse*. We compare the performance of our method relying on Hybrid-MCMC (blue line with cross markers) to Structure-MCMC (red line with triangle markers). Under different settings of node dimensions and sample sizes, we compute the median TPR along with the first and third quartiles and plot versus the median FPRp over 50 replications for each algorithm under comparison. The markers on each graph indicate the following thresholds on the posterior probability of edge inclusion: $\rho \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$.

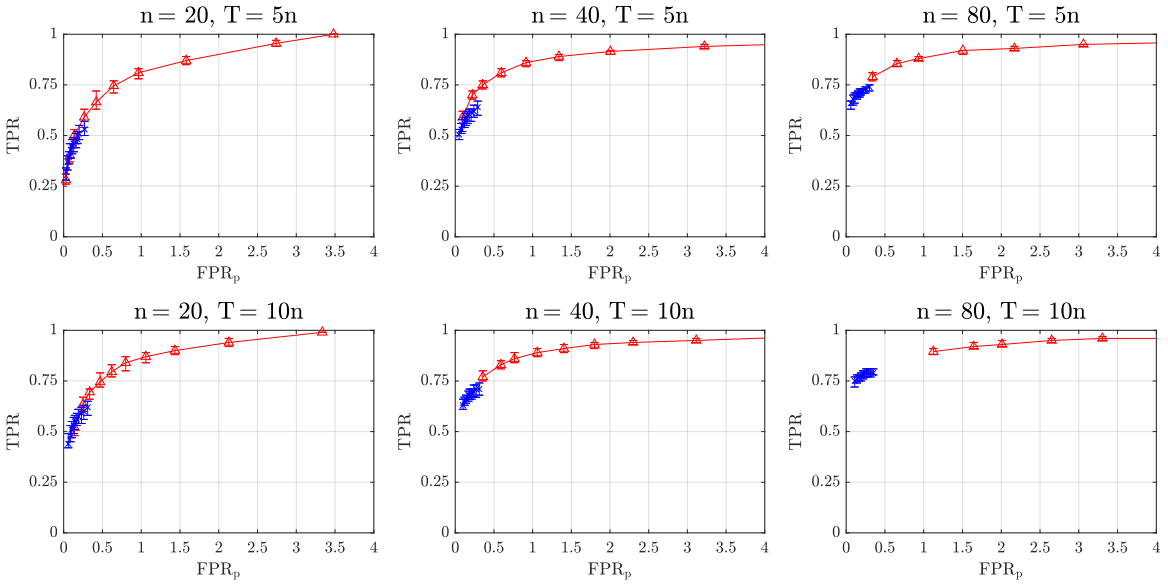


Figure 2.C.2 ROC curves under the scenario *Dense*. We compare the performance of our method relying on Hybrid-MCMC (blue line with cross markers) to Structure-MCMC (red line with triangle markers). Under different settings of node dimensions and sample sizes, we compute the median TPR along with the first and third quartiles and plot versus the median FPRp over 50 replications for each algorithm under comparison. The markers on each graph indicate the following thresholds on the posterior probability of edge inclusion: $\rho \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$.

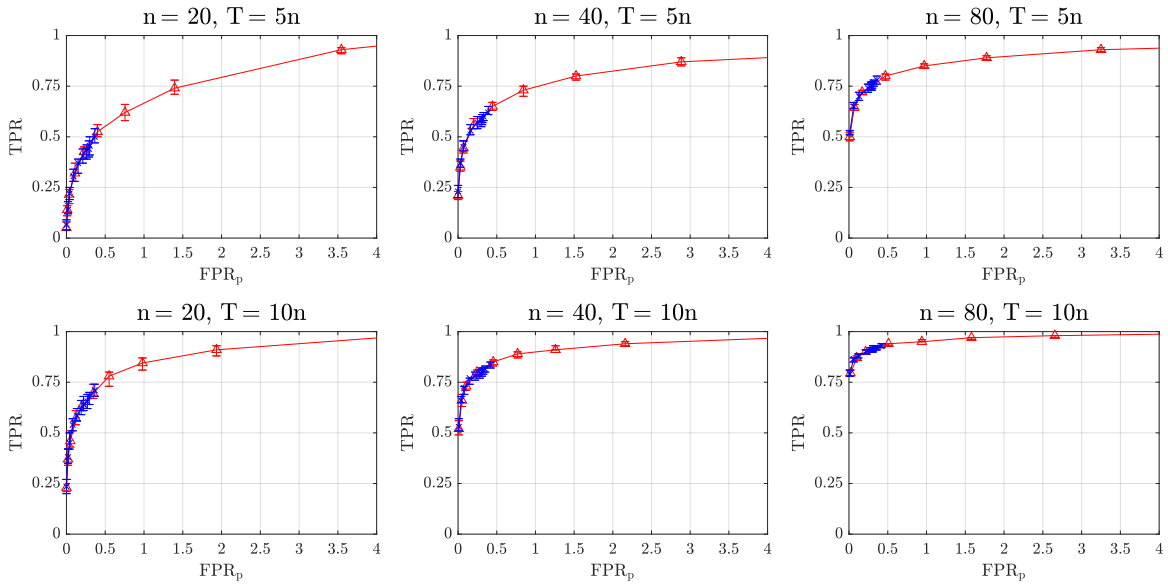


Figure 2.C.3 ROC curves under the scenario *Weak edge strength*. We compare the performance of our method relying on Hybrid-MCMC (blue line with cross markers) to Structure-MCMC (red line with triangle markers). Under different settings of node dimensions and sample sizes, we compute the median TPR along with the first and third quartiles and plot versus the median FPRp over 50 replications for each algorithm under comparison. The markers on each graph indicate the following thresholds on the posterior probability of edge inclusion: $\rho \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$.

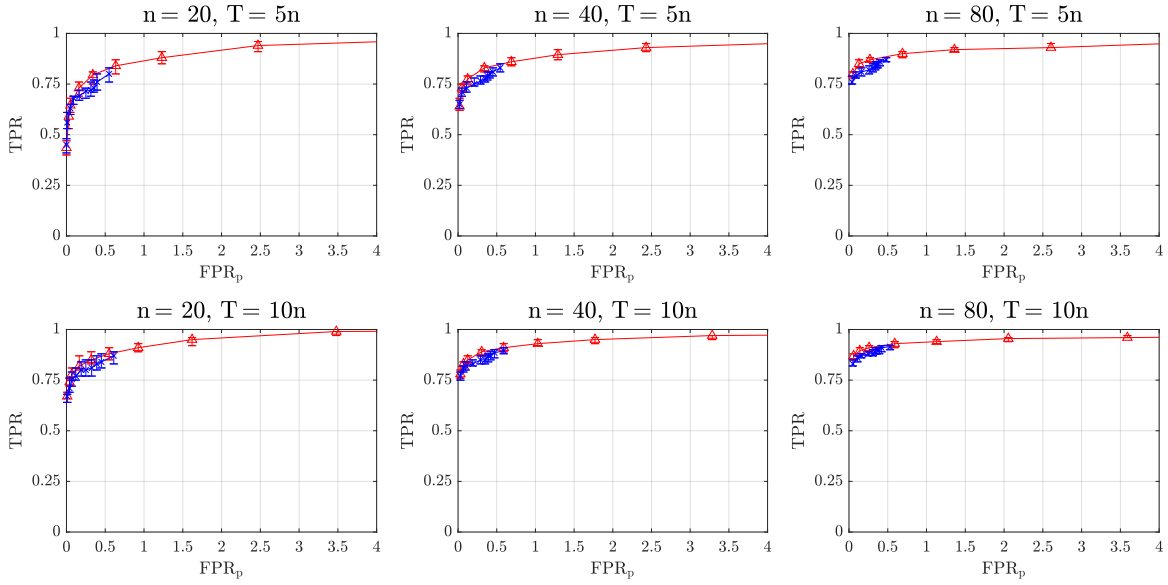


Figure 2.C.4 ROC curves under the scenario *Community*. We compare the performance of our method relying on Hybrid-MCMC (blue line with cross markers) to Structure-MCMC (red line with triangle markers). Under different settings of node dimensions and sample sizes, we compute the median TPR along with the first and third quartiles and plot versus the median FPRp over 50 replications for each algorithm under comparison. The markers on each graph indicate the following thresholds on the posterior probability of edge inclusion: $\rho \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$.

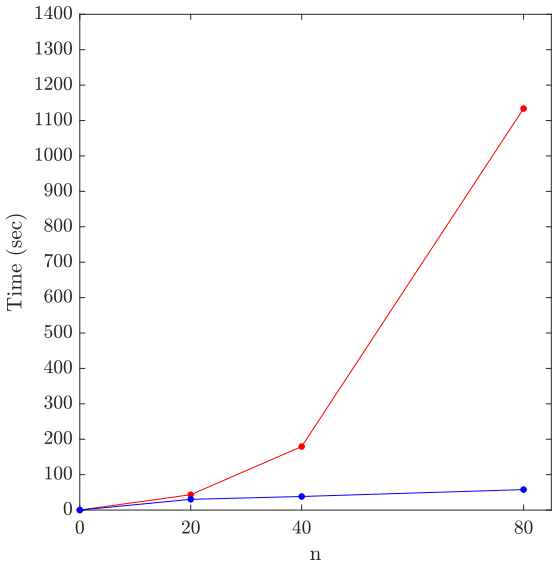


Figure 2.C.5 The runtimes under the scenario *Sparse*. We compare the performance of our method relying on Hybrid-MCMC (blue line) to Structure-MCMC (red line). Under different settings of node dimensions and sample sizes, we average the runtimes (in seconds) over 50 replications for each algorithm under comparison. These results are then average over the two sample sizes $T \in \{5n, 10n\}$ for each of the node dimensions listed.

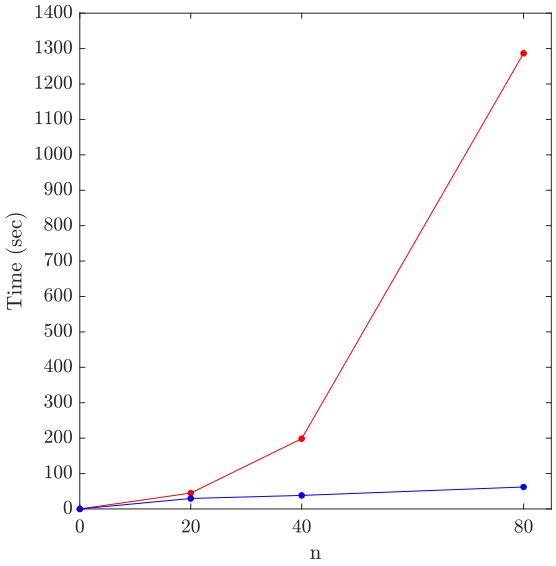


Figure 2.C.6 The runtimes under the scenario *Dense*. We compare the performance of our method relying on Hybrid-MCMC (blue line) to Structure-MCMC (red line). Under different settings of node dimensions and sample sizes, we average the runtimes (in seconds) over 50 replications for each algorithm under comparison. These results are then average over the two sample sizes $T \in \{5n, 10n\}$ for each of the node dimensions listed.

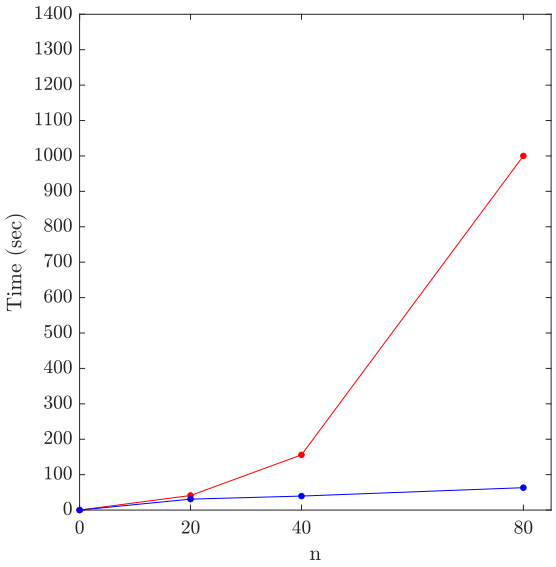


Figure 2.C.7 The runtimes under the scenario *Weak edge strength*. We compare the performance of our method relying on Hybrid-MCMC (blue line) to Structure-MCMC (red line). Under different settings of node dimensions and sample sizes, we average the runtimes (in seconds) over 50 replications for each algorithm under comparison. These results are then average over the two sample sizes $T \in \{5n, 10n\}$ for each of the node dimensions listed.

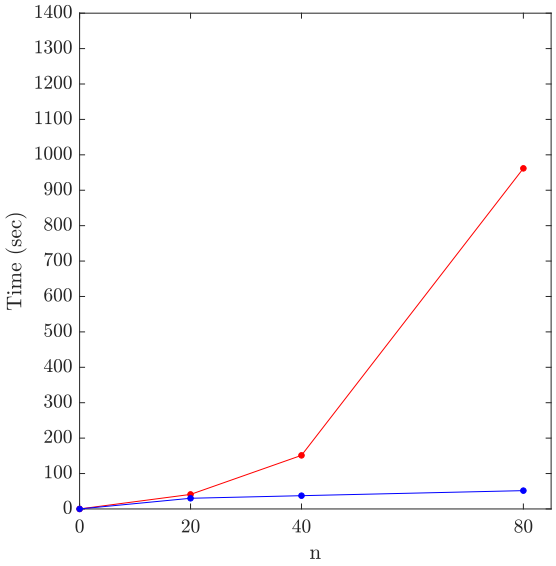


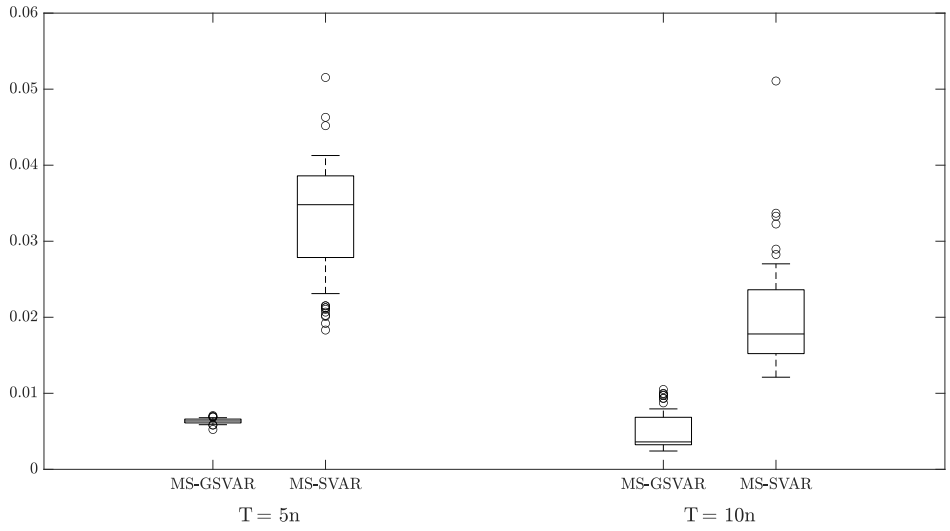
Figure 2.C.8 The runtimes under the scenario *Community*. We compare the performance of our method relying on Hybrid-MCMC (blue line) to Structure-MCMC (red line). Under different settings of node dimensions and sample sizes, we average the runtimes (in seconds) over 50 replications for each algorithm under comparison. These results are then average over the two sample sizes $T \in \{5n, 10n\}$ for each of the node dimensions listed.

2.C.2 Simulation study to assess model performance

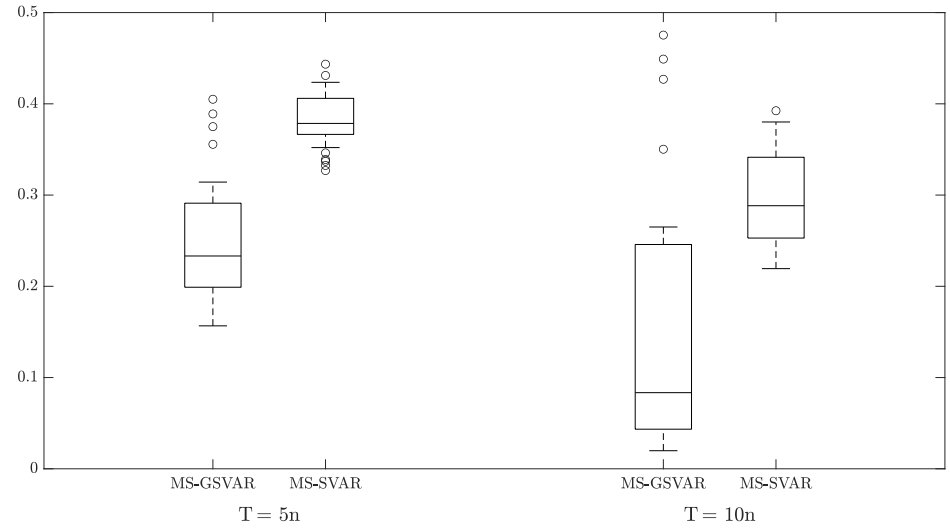
In this simulation, we illustrate the performance of the proposed MS-GSVAR model and the effectiveness of its estimation procedure in handling large datasets with changing dependence structures. We simulate data from an SVAR(1) process where the true graph structures are assumed to change across two states. In particular, we assume the true graph structure underlying each of the coefficient matrices $B_{0:p}(k)$ follows the structure of scenario *Sparse* in state $k=1$ and of scenario *Dense* in state $k=2$. Moreover, the edge values are sampled uniformly in the interval $[-0.5,-0.3]\cup[0.3,0.5]$ in state $k=1$ and $[-0.7,-0.3]\cup[0.3,0.7]$ in state $k=2$. Due to its relatively dense pattern as well as strong edge strengths, state $k=2$ corresponds to a state with the higher level of connectedness. The true covariance matrix $\Sigma_\varepsilon(k)$ is set to $0.2I_n$ in state $k=1$ and I_n in state $k=2$, which are roughly the same as our estimates for the global banking dataset analyzed in Section 2.5 of the paper. We generate the trajectory of the Markov process assuming a transition matrix with rows $p(k)=(0.95,0.05)$ for $k=1$ and $p(k)=(0.05,0.95)$ for $k=2$. We consider two different high-dimensional settings described in Section 2.C.1: $n=80$ and $T\in\{5n,10n\}$. For each setting, we replicate the simulation and estimation exercises 50 times, with different randomly sampled graphs and data. All the results reported are aggregated based on 50 replications. To illustrate the gain from our graphical representation, we compare the proposed MS-GSVAR model to a standard MS-SVAR model of Sims et al. (2008) in which each of the coefficient matrices is full.

We begin with the accuracy of the parameter estimates. The accuracy is evaluated using the Mean Square Deviation (MSD), which is defined as the average squared difference between the true value of the parameter and its posterior mean. Figure 2.C.9 shows the quartiles and median of the MSD statistics based on 50 replications by means of box plots. In all settings, it is clear that MS-GSVAR offers large gains over MS-SVAR in recovering the true model parameters since the interquartile ranges (i.e., the boxes) do not overlap. The gains are particularly significant when the ratio between the node dimension and the sample size (n/T) is relatively large.

To investigate the model fitness and detect potential misspecification, we perform a posterior predictive checking procedure suggested in Billio et al. (2022). Posterior predictive checking is based on the idea that, if the model fits, then data simulated from the model should look similar to observed data. Therefore, this procedure consists in (i) simulating data from the model, then (ii) measuring the discrepancy between these samples to the observed data via suitable test quantities. Denote with $T(Y)$ a test quantity that depends only on the data. We consider the following test quantities: (i) average density of Y_t over time, i.e., $T(Y)=1/(nT)\sum_r\sum_t Y_{r,t}$, (ii) cross-sectional average of the temporal standard deviation of Y_t, i.e., $T(Y)=1/n\sum_r\sqrt{\frac{1}{T}\sum_t(Y_{r,t}-\bar{Y}_r)^2}$, (iii) cross-sectional average of the autocorrelation at lag $i=1,...,p$, denoted by $\gamma_r(i)$, of the entries in Y_t, i.e., $T(Y)=1/(np)\sum_r\sum_{i=1}^p\gamma_r(i)$,



(a) MSD for $B_{0:p}(k)$



(b) MSD for $\Sigma_\varepsilon(k)$

Figure 2.C.9 Mean Square Deviation (MSD) of the estimated parameters from their true values. Shown are boxplots of MSD based on 50 replications, for two estimators: MS-GSVAR and MS-SVAR, two different sample sizes: $T = 5n, 10n$, and two sets of model parameters: $B_{0:p}(k)$ (panel (a)) and $\Sigma_\varepsilon(k)$ (panel (b)), for $k = 1, 2$. On each box, the central mark indicates the median, while the bottom and top edges of the box are the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme points and the outliers are plotted using $\circ$.

and (iv) cross-sectional average of the pairwise cross-correlation at lag 0, denoted by $\varsigma_{rl}(0)$, of the entries in Y_t, i.e., $T(Y) = 2/(n(n-1))\sum_l \sum_{r=l+1} \varsigma_{rl}(0)$. We visualize the discrepancy

between the test quantity computed on the simulated and the observed data by means of histogram. Under correct model specification, the histogram should be centered around 0. The results for our MS-GSVAR model and the MS-SVAR model are shown in Figure 2.C.10. We find that MS-GSVAR performs well in all the settings, being able to replicate all of the test quantities of the observed data. Conversely, the MS-SVAR model yields substantial bias in recovering the standard deviation of the observed data, and this performance deteriorates when n/T is relatively large.

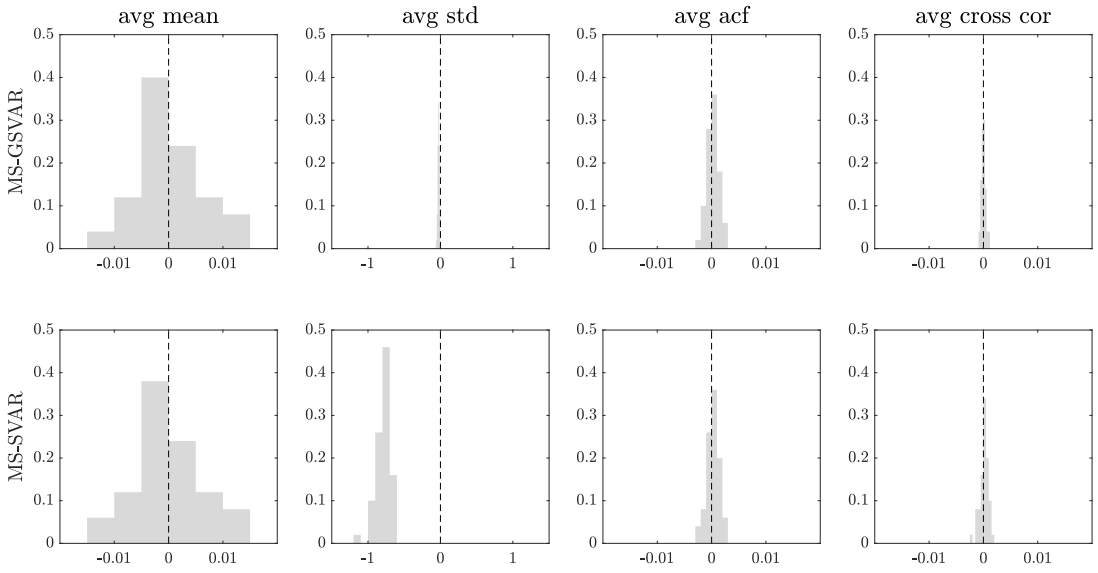
To assess the effectiveness of our model estimation procedure, we use the **coda** functions provided in the Econometrics toolbox[11] of LeSage (1999). Specifically, we rely on the diagnostic criteria including the autocorrelation function (ACF) at different lags, Geweke's test of equal mean, and Geweke's diagnostic based on the relative numerical efficiency (RNE). All criteria are first computed for each parameter of the model individually, and then averaged across all parameters of the model. Table 2.C.1 reports these diagnostic statistics, averaged over 50 replications and over the two simulation settings considered. The ACF values are small enough to be indicative of the i.i.d. nature of the parameter estimates, as suggested for example by Casarin et al. (2018). The Geweke's p-values are greater than 0.05 and the RNE values are close to 1, which are good signs of efficiency for the MCMC procedure according to previous studies (see, e.g., Billio et al., 2022). Overall, these criteria indicate well-mixing and convergence properties of our model estimation procedure.

Table 2.C.1 Diagnostic criteria

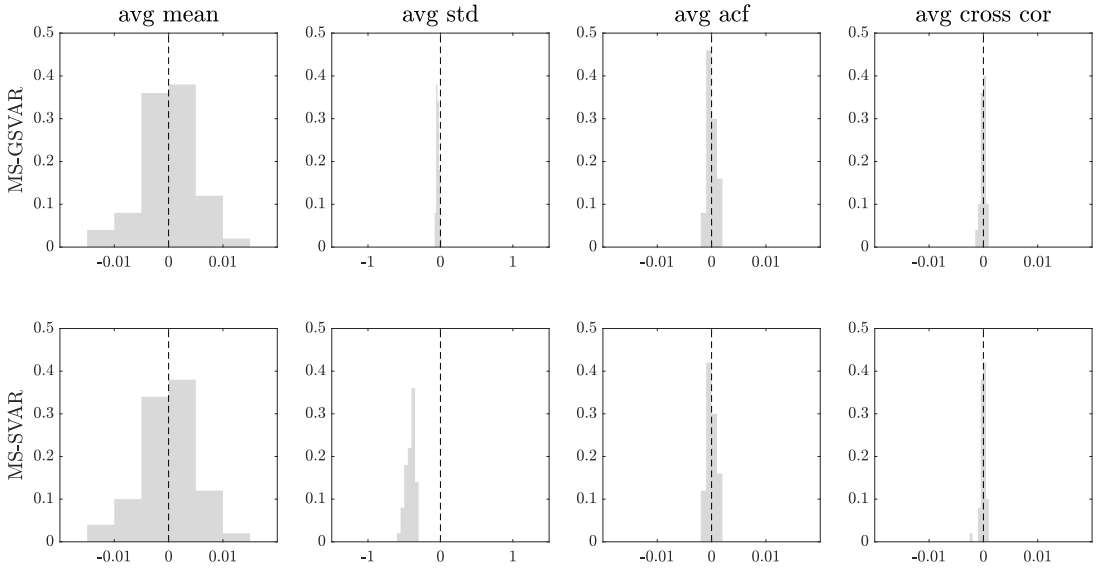
average ACF				average p-value Geweke test			average RNE		
lag 1	lag 5	lag 10	lag 50	4% taper	8% taper	15% taper	4% taper	8% taper	15% taper
0.056	0.023	0.021	0.013	0.471	0.468	0.460	0.957	0.998	1.108

NOTES: This table presents the autocorrelation function (ACF) at different lags, p-value of the Geweke test, and relative numerical efficiency (RNE), averaged over 50 replications and over the two simulation settings considered. Each statistic is an average across all parameters of the proposed model.

[11]The code is available in https://www.spatial-econometrics.com.



(a) T = 5n



(b) T = 10n

Figure 2.C.10 Posterior predictive checking based on 50 replications for the MS-GSVAR model (first row) and the MS-SVAR model (second row). Performance on a simulated dataset of dimension $n = 80$, with two different sample sizes $T = 5n$ (panel (a)), $= 10n$ (panel (b)). Each plot shows the posterior distribution of the discrepancy between the test quantity computed on the simulated and the observed data, averaged over all MCMC iterations for each replication.

Appendix 2.D

This appendix contains a description of the data used in the empirical illustration and further discussions of the estimation results.

2.D.1 Data details

Our dataset consists of weekly bank stock return volatilities, constructed using daily high–low-open-close prices of 96 banks obtained from *Datastream*. The dataset spans from September 12, 2003 to December 9, 2022 (1,005 observations). Our banks are those in the world's top 150 by assets that were publicly traded throughout the sample. They include all those designated as "globally systemically important banks" ("G-SIBs," as designated by the Basel Committee on Banking Supervision as of November 2022), except for four Chinese banks (Bank of China, Industrial and Commercial Bank of China, Agricultural Bank of China and China Construction Bank) and one French bank (Groupe BPCE), which we exclude because they were not publicly traded as of September 2003. Details on our sample banks regarding market capitalization (as of November 15, 2022 in billions of US dollars), bank code[12], and Reuters ticker, by bank assets (as of December 31, 2021 in billions of US dollars) are provided in Table 2.D.1 and by country are provided in Table 2.D.2.

Raw daily stock prices (high–low-open-close prices) for bank i are used to compute its weekly range-based realized volatility as proposed in Garman and Klass (1980)

$$\begin{aligned}\hat{\sigma}_{i,t}^2 =& 0.511(H_{i,t}-L_{i,t})^2 - 0.383(C_{i,t}-O_{i,t})^2 \\ & - 0.019[(C_{i,t}-O_{i,t})(H_{i,t}+L_{i,t}-2O_{i,t}) - 2(H_{i,t}-O_{i,t})(L_{i,t}-O_{i,t})],\end{aligned} \tag{2.D.1}$$

where $H_{i,t}$ is the weekly logarithmic high price, $L_{i,t}$ is the weekly logarithmic low price, $O_{i,t}$ is the weekly logarithmic opening price, and $C_{i,t}$ is the logarithmic closing price. The weekly prices are obtained by taking in a given week the maximum among the daily high prices (weekly High Price), the minimum among the daily low prices (weekly Low Price), the opening price of the first available day in a week (weekly Opening Price), and the closing price of the last available day in a week (weekly Closing Price). Besides, volatilities tend to be distributed asymmetrically with a right skew, and approximate normality is often obtained by taking natural logarithms (Andersen et al., 2003). Hence we use weekly logarithmic volatilities as the final datasets.

[12]Our bank codes are easier to interpret than the Reuters tickers, particularly as regards identifying banks' countries, so we use them in our empirical illustration.

Table 2.D.1 Global Bank Detail (Ordered by Assets)

Bank Name	Country	Mcap	Asset	Bank Code	Reuters Ticker
JP MORGAN CHASE & COMPANY	US	389.9	3,744	jpm.us	JPM
MITSUBISHI UFJ FINANCIAL GROUP	Japan	68.9	3,175	mtbh.jp	8306.T
BANK OF AMERICA	US	302.4	3,160	bac.us	BAC
BNP PARIBAS	France	67.1	2,991	bnp.fr	BNPP.PA
HSBC HOLDINGS	UK	114.0	2,956	hsba.gb	HSBA.L
CREDIT AGRICOLE	France	30.4	2,354	aca.fr	CAGR.PA
CITIGROUP	US	95.0	2,262	c.us	C
SUMITOMO MITSUI FINANCIAL GROUP	Japan	44.1	2,176	smf.jp	8316.T
MIZUHO FINANCIAL GROUP	Japan	29.5	1,957	mzh.jp	8411.T
WELLS FARGO & CO	US	178.0	1,948	wfc.us	WFC
BARCLAYS	UK	29.7	1,869	barc.gb	BARC.L
BANCO SANTANDER	Spain	45.4	1,793	san.es	SAN.MC
SOCIETE GENERALE	France	21.5	1,661	gle.fr	SOGN.PA
DEUTSCHE BANK	Germany	22.1	1,499	dbk.de	DBKGn.DE
GOLDMAN SACHS GROUP	US	129.7	1,464	gs.us	GS
CHINA MERCHANTS BANK	China	97.6	1,438	cmb.cn	600036.SS
TORONTO-DOMINION BANK	Canada	119.7	1,367	td.ca	TD.TO
ROYAL BANK OF CANADA	Canada	137.1	1,349	ry.ca	RY.TO
SHANGHAI PUDONG DEVELOPMENT BANK	China	29.5	1,267	shgp.cn	600000.SS
INTESA SANPAOLO	Italy	43.1	1,198	isp.it	ISP.MI
LLOYDS BANKING GROUP	UK	35.0	1,197	lloy.gb	LLOY.L
MORGAN STANLEY	US	153.0	1,188	ms.us	MS
UBS GROUP	Switzerland	65.3	1,109	ubsn.ch	UBSG.S
CHINA MINSHENG BANKING	China	17.5	1,083	cmsb.cn	600016.SS
ING GROEP	Netherlands	44.1	1,081	inga.nl	INGA.AS
NATWEST GROUP	UK	28.8	1,058	rbs.gb	NWG.L
UNICREDIT	Italy	27.4	1,029	ucg.it	CRDI.MI
BANK OF NOVA SCOTIA	Canada	62.3	936	bns.ca	BNS.TO
COMMONWEALTH BANK OF AUSTRALIA	Australia	122.3	834	cba.au	CBA.AX
STANDARD CHARTERED	UK	19.9	828	stan.gb	STAN.L
CREDIT SUISSE GROUP	Switzerland	12.0	825	csgn.ch	CSGN.S
BANK OF MONTREAL	Canada	67.4	781	bmo.ca	BMO.TO
PING AN BANK	China	33.1	766	pab.cn	000001.SZ
BANCO BILBAO VIZCAYA ARGENTARIA	Spain	33.8	737	bbva.es	BBVA.MC
AUSTRALIA AND NEW ZEALAND BANKING GROUP	Australia	49.1	710	anz.au	ANZ.AX
STATE BANK OF INDIA	India	66.1	697	sbin.in	SBI.NS
WESTPAC BANKING	Australia	56.8	679	wbc.au	WBC.AX
NATIONAL AUSTRALIA BANK	Australia	65.2	671	nab.au	NAB.AX
CANADIAN IMPERIAL BANK COMMERCE	Canada	43.0	663	cm.ca	CM.TO
RESONA HOLDINGS	Japan	10.3	662	rsnh.jp	8308.T
NORDEA BANK	Finland	37.7	647	nor.fi	NDASE.ST
DANSKE BANK	Denmark	14.3	602	dan.dk	DANSKE.CO
HUAXIA BANK	China	11.4	575	hxb.cn	600015.SS
US BANCORP	US	65.0	573	usb.us	USB
KB FINANCIAL GROUP	Korea	15.4	558	kb.kr	105560.KS
PNC FINANCIAL SERVICES GROUP	US	66.2	557	pnc.us	PNC
SUMITOMO MITSUI TRUST HOLDINGS	Japan	11.2	554	smt.jp	8309.T
SBERBANK OF RUSSIA	Russia	48.2	548	sber.ru	SBER.MM
SHINHAN FINANCIAL GROUP	Korea	14.3	545	shf.kr	055550.KS
TRUIST FINANCIAL	US	61.5	541	bbt.us	TFC
COMMERZBANK	Germany	10.4	534	cbk.de	CBKG.DE
DBS GROUP HOLDINGS	Singapore	66.5	508	d05.sg	DBSM.SI
BANK OF NEW YORK MELLON	US	35.4	444	bk.us	BK
BANQUE NATIONALE DE BELGIQUE	Belgium	0.4	409	bnab.be	BNAB.BR
OVERSEA-CHINESE BANKING	Singapore	41.1	402	ocbc.sg	OCBC.SI
KBC GROUP	Belgium	22.4	386	kbc.be	KBC.BR
SVENSKA HANDELSBANKEN	Sweden	19.7	370	shba.se	SHBa.ST
SKANDINAVISKA ENSKILDA BANKEN	Sweden	24.5	365	seba.se	SEBa.ST
ERSTE GROUP BANK	Austria	12.9	349	ebs.at	ERST.VI
UNITED OVERSEAS BANK	Singapore	36.9	340	uob.sg	UOBH.SI
BANCO DO BRASIL	Brazil	19.6	337	bbas.br	BBAS3.SA
INDUSTRIAL BANK OF KOREA	Korea	6.0	335	ibk.kr	024110.KS
DNB BANK	Norway	28.5	331	dnb.no	DNB.OL

Continued on next page

Table 2.D.1 – continued from previous page

Bank Name	Country	Mcap	Asset	Bank Code	Reuters Ticker
STATE STREET	US	29.2	314	stt.us	STT
SWEDBANK	Sweden	17.9	304	swe.se	SWEDa.ST
NATIONAL BANK OF CANADA	Canada	24.4	281	na.ca	NA.TO
BANCO DE SABADELL	Spain	4.9	279	sab.es	SABE.MC
BANK OF GREECE	Greece	0.3	270	bg.gr	BOGr.AT
HDFC BANK	India	111.3	268	hdfc.in	HDBK.NS
FUKUOKA FINANCIAL GROUP	Japan	3.4	252	ffg.jp	8354.T
ICICI BANK	India	78.4	226	icici.in	ICBK.NS
BANCO BPM	Italy	5.0	223	bp.it	BAMI.MI
MALAYAN BANKING	Malaysia	22.5	213	may.my	MBBM.KL
SVB FINANCIAL GROUP	US	14.2	211	svb.us	SIVB.O
FIFTH THIRD BANCORP	US	25.0	211	fitb.us	FITB.O
KEYCORP	US	17.9	186	key.us	KEY
PUNJAB NATIONAL BANK	India	6.0	179	puj.in	PNBK.NS
BANK OF IRELAND GROUP	Ireland	8.4	175	bir.ie	BIRG.I
HUNTINGTON BANCSHARES	US	21.8	174	hban.us	HBAN.O
STANDARD BANK GROUP	Korea	17.0	170	sbk.kr	SBKJ.J
BANK OF BARODA	India	10.4	168	bob.in	BOB.NS
CANARA BANK	India	6.9	165	can.in	CNBK.NS
REGIONS FINANCIAL NEW	US	21.6	163	rf.us	RF
CHIBA BANK	Japan	4.6	161	cbb.jp	8331.T
BANCA MONTE DEI PASCHI	Italy	2.4	156	bmps.it	BMPS.MI
M&T BANK	US	29.3	155	mt.us	MTB
AXIS BANK	India	32.5	152	axis.in	AXBK.NS
HOKUHOKU FINANCIAL GROUP	Japan	0.8	150	hkf.jp	8377.T
UNION BANK OF INDIA	India	5.3	149	ubi.in	UNBK.NS
CIMB GROUP HOLDINGS	Malaysia	12.7	149	cimb.my	CIMB.KL
AIB GROUP	Ireland	8.0	142	aib.ie	AIBG.I
SHIZUOKA FINANCIAL GROUP	Japan	3.9	127	shzb.jp	5831.T
FIRSTRAND	Korea	20.9	124	fsrj.kr	FSRJ.J
BANKINTER	Spain	5.3	122	bkt.es	BKT.MC
HACHIJUNI BANK	Japan	1.7	121	hach.jp	8359.T
PUBLIC BANK	Malaysia	18.7	111	pubm.my	PUBM.KL

Table 2.D.2 Global Bank Detail (Ordered by Country)

Bank Name	Country	Mcap	Asset	Bank Code	Reuters Ticker
JP MORGAN CHASE & COMPANY	US	389.9	3,744	jpm.us	JPM
BANK OF AMERICA	US	302.4	3,160	bac.us	BAC
CITIGROUP	US	95.0	2,262	c.us	C
WELLS FARGO & CO	US	178.0	1,948	wfc.us	WFC
GOLDMAN SACHS GROUP	US	129.7	1,464	gs.us	GS
MORGAN STANLEY	US	153.0	1,188	ms.us	MS
US BANCORP	US	65.0	573	usb.us	USB
PNC FINANCIAL SERVICES GROUP	US	66.2	557	pnc.us	PNC
TRUIST FINANCIAL	US	61.5	541	bbt.us	TFC
BANK OF NEW YORK MELLON	US	35.4	444	bk.us	BK
STATE STREET	US	29.2	314	stt.us	STT
SVB FINANCIAL GROUP	US	14.2	211	svb.us	SIVB.O
FIFTH THIRD BANCORP	US	25.0	211	fitb.us	FITB.O
KEYCORP	US	17.9	186	key.us	KEY
HUNTINGTON BANCSHARES	US	21.8	174	hban.us	HBAN.O
REGIONS FINANCIAL NEW	US	21.6	163	rf.us	RF
M&T BANK	US	29.3	155	mt.us	MTB
MITSUBISHI UFJ FINANCIAL GROUP	Japan	68.9	3,175	mtbh.jp	8306.T
SUMITOMO MITSUI FINANCIAL GROUP	Japan	44.1	2,176	smf.jp	8316.T
MIZUHO FINANCIAL GROUP	Japan	29.5	1,957	mzh.jp	8411.T
RESONA HOLDINGS	Japan	10.3	662	rsnh.jp	8308.T
SUMITOMO MITSUI TRUST HOLDINGS	Japan	11.2	554	smt.jp	8309.T
FUKUOKA FINANCIAL GROUP	Japan	3.4	252	ffg.jp	8354.T
CHIBA BANK	Japan	4.6	161	cbb.jp	8331.T
HOKUHOKU FINANCIAL GROUP	Japan	0.8	150	hkf.jp	8377.T
SHIZUOKA FINANCIAL GROUP	Japan	3.9	127	shzb.jp	5831.T
HACHIJUNI BANK	Japan	1.7	121	hach.jp	8359.T
STATE BANK OF INDIA	India	66.1	697	sbin.in	SBI.NS
HDFC BANK	India	111.3	268	hdfc.in	HDBK.NS
ICICI BANK	India	78.4	226	icici.in	ICBK.NS
PUNJAB NATIONAL BANK	India	6.0	179	puj.in	PNBK.NS
BANK OF BARODA	India	10.4	168	bob.in	BOB.NS
CANARA BANK	India	6.9	165	can.in	CNBK.NS
AXIS BANK	India	32.5	152	axis.in	AXBK.NS
UNION BANK OF INDIA	India	5.3	149	ubi.in	UNBK.NS
TORONTO-DOMINION BANK	Canada	119.7	1,367	td.ca	TD.TO
ROYAL BANK OF CANADA	Canada	137.1	1,349	ry.ca	RY.TO
BANK OF NOVA SCOTIA	Canada	62.3	936	bns.ca	BNS.TO
BANK OF MONTREAL	Canada	67.4	781	bmo.ca	BMO.TO
CANADIAN IMPERIAL BANK COMMERCE	Canada	43.0	663	cm.ca	CM.TO
NATIONAL BANK OF CANADA	Canada	24.4	281	na.ca	NA.TO
HSBC HOLDINGS	UK	114.0	2,956	hsba.gb	HSBA.L
BARCLAYS	UK	29.7	1,869	barc.gb	BARC.L
LLOYDS BANKING GROUP	UK	35.0	1,197	lloy.gb	LLOY.L
NATWEST GROUP	UK	28.8	1,058	rbs.gb	NWG.L
STANDARD CHARTERED	UK	19.9	828	stan.gb	STAN.L
KB FINANCIAL GROUP	Korea	15.4	558	kb.kr	105560.KS
SHINHAN FINANCIAL GROUP	Korea	14.3	545	shf.kr	055550.KS
INDUSTRIAL BANK OF KOREA	Korea	6.0	335	ibk.kr	024110.KS
STANDARD BANK GROUP	Korea	17.0	170	sbk.kr	SBKJ.J
FIRSTRAND	Korea	20.9	124	fsrj.kr	FSRJ.J
CHINA MERCHANTS BANK	China	97.6	1,438	cmb.cn	600036.SS
SHANGHAI PUDONG DEVELOPMENT BANK	China	29.5	1,267	shgp.cn	600000.SS
CHINA MINSHENG BANKING	China	17.5	1,083	cmsb.cn	600016.SS
PING AN BANK	China	33.1	766	pab.cn	000001.SZ
HUAXIA BANK	China	11.4	575	hxb.cn	600015.SS
BANCO SANTANDER	Spain	45.4	1,793	san.es	SAN.MC
BANCO BILBAO VIZCAYA ARGENTARIA	Spain	33.8	737	bbva.es	BBVA.MC
BANCO DE SABADELL	Spain	4.9	279	sab.es	SABE.MC
BANKINTER	Spain	5.3	122	bkt.es	BKT.MC
INTESA SANPAOLO	Italy	43.1	1,198	isp.it	ISP.MI
UNICREDIT	Italy	27.4	1,029	ucg.it	CRDI.MI
BANCO BPM	Italy	5.0	223	bp.it	BAMI.MI
BANCA MONTE DEI PASCHI	Italy	2.4	156	bmps.it	BMPS.MI

Continued on next page

Table 2.D.2 – continued from previous page

Bank Name	Country	Mcap	Asset	Bank Code	Reuters Ticker
COMMONWEALTH BANK OF AUSTRALIA	Australia	122.3	834	cba.au	CBA.AX
AUSTRALIA AND NEW ZEALAND BANKING GROUP	Australia	49.1	710	anz.au	ANZ.AX
WESTPAC BANKING	Australia	56.8	679	wbc.au	WBC.AX
NATIONAL AUSTRALIA BANK	Australia	65.2	671	nab.au	NAB.AX
SVENSKA HANDELSBANKEN	Sweden	19.7	370	shba.se	SHBa.ST
SKANDINAVISKA ENSKILDA BANKEN	Sweden	24.5	365	seba.se	SEBa.ST
SWEDBANK	Sweden	17.9	304	swe.se	SWEDa.ST
DBS GROUP HOLDINGS	Singapore	66.5	508	d05.sg	DBSM.SI
OVERSEA-CHINESE BANKING	Singapore	41.1	402	ocbc.sg	OCBC.SI
UNITED OVERSEAS BANK	Singapore	36.9	340	uob.sg	UOBH.SI
MALAYAN BANKING	Malaysia	22.5	213	may.my	MBBM.KL
CIMB GROUP HOLDINGS	Malaysia	12.7	149	cimb.my	CIMB.KL
PUBLIC BANK	Malaysia	18.7	111	pubm.my	PUBM.KL
BNP PARIBAS	France	67.1	2,991	bnp.fr	BNPP.PA
CREDIT AGRICOLE	France	30.4	2,354	aca.fr	CAGR.PA
SOCIETE GENERALE	France	21.5	1,661	gle.fr	SOGN.PA
UBS GROUP	Switzerland	65.3	1,109	ubsn.ch	UBSG.S
CREDIT SUISSE GROUP	Switzerland	12.0	825	csgn.ch	CSGN.S
BANK OF IRELAND GROUP	Ireland	8.4	175	bir.ie	BIRG.I
AIB GROUP	Ireland	8.0	142	aib.ie	AIBG.I
DEUTSCHE BANK	Germany	22.1	1,499	dbk.de	DBKGn.DE
COMMERZBANK	Germany	10.4	534	cbk.de	CBKG.DE
BANQUE NATIONALE DE BELGIQUE	Belgium	0.4	409	bnab.be	BNAB.BR
KBC GROUP	Belgium	22.4	386	kbc.be	KBC.BR
SBERBANK OF RUSSIA	Russia	48.2	548	sber.ru	SBER.MM
DNB BANK	Norway	28.5	331	dnb.no	DNB.OL
ING GROEP	Netherlands	44.1	1,081	inga.nl	INGA.AS
BANK OF GREECE	Greece	0.3	270	bg.gr	BOGr.AT
NORDEA BANK	Finland	37.7	647	nor.fi	NDASE.ST
DANSKE BANK	Denmark	14.3	602	dan.dk	DANSKE.CO
BANCO DO BRASIL	Brazil	19.6	337	bbas.br	BBAS3.SA
ERSTE GROUP BANK	Austria	12.9	349	ebs.at	ERST.VI

2.D.2 Discussions of the states turning points

This section provides further details on the states turning points shown in Figure 2.1 of the paper.

From the beginning of the sample period to the late 2007, the global banking system stayed in the state of low connectedness with no sign of volatility contagion. The substantial increase in the system-wide connectedness was observed in late October 2007 when the model generates the first sign of an extreme connectedness state. This state transition reflected the heightened risk perceived by investors at the time, in part due to the the worsening in the third quarter financial statements of major US banks and the sudden resignation of CEOs from Merrill Lynch and Citigroup.

Since then until late August 2008, the global banking system lingered in the extreme connectedness state for most of the periods but transited back to the low connectedness state at times roughly corresponding to the main policy actions. The first of these transitions coincided with the initiation of the *Term Auction Facility* (TAF) program, which was instituted by the Federal Reserve on December 12, 2007 to alleviate financial strains in the inter-bank money market. The second transition took place around February 2008, when the *Economic Stimulus Act of 2008* was enacted with aim to boost consumer spending in US.

The next transition was partly attributable to a $75 billion expansion of the TAF program on May 2, 2008. The last transition followed the enactment of the *Housing and Economic Recovery Act of 2008* on July 30, 2008, which was designed primarily to support Fannie Mae and Freddie Mac, the two large US suppliers of mortgage funding.

From September 2008 until August 2009, the global banking system entered a long period of extreme connectedness state without temporarily returning to the low connectedness state. During this time, the subprime mortgage crisis expanded into a global one with the failures of Lehman Brothers, AIG, Fannie Mae, Freddie Mac, Merrill Lynch and many others. It was then clear that the global economy was in its worst shape since the Second World War as evidenced by the announcement by the World Bank in mid-2009 that global production would decline for the first time since that war. Consequently, pessimism seems to continuously grip the bank stock market, leading to the period in which the extreme connectedness state persisted.

In the last quarter of 2009 and throughout 2010, the European sovereign debt crisis drove the global banking system continually out of the low connectedness state. Since this crisis is incomparable to the previous global financial crisis regarding the scale and magnitude of its effects, the banking system usually ended up at the high connectedness state, with two exceptions that it went into the extreme connectedness state. The high connectedness state was firstly observed during November-December 2009, when concerns about some EU member states' debts started to grow following the Dubai sovereign debt crisis. In the first quarter of 2010, the system briefly turned back to the low connectedness state in response to the launch of Greece's *Stability and Growth Program* and its first austerity plan. However, as the Greek debt crisis intensified and concerns started to build about other heavily indebted countries in Europe - Portugal and Spain, the system transited again to the high connectedness state in April and then rushed to the extreme connectedness state in May. Over the subsequent three months, austerity measures put in effect by the Italian government as well as the IMF and ECB's endorsement of an additional €9 billion payment to Greece calmed the markets down, prompting the banking system back to the low connectedness state. The next high-connectedness state appeared in late September due to the worsening of Ireland's fiscal conditions. After two months of negotiation, the EU and IMF approved a €85 billion bailout package for Ireland, inducing the banking system to re-entering the low connectedness state in December.

Sovereign debt troubles of Greece and Portugal continued in early 2011, when the Greek sovereign debt was downgraded to junk status in January and Portuguese government collapsed in a fight over austerity measures in March. These events contributed to two high-connectedness states emerging in the corresponding month. After that, some good news about the approvement of €78 billion bailout package for Portugal and the launch of the permanent rescue fund the *European Stability Mechanism* relieved the markets, bringing on a low connectedness state until the end of July. From August onwards, unimpressed

with EU countries' recovery in the wake of the previous bailout package, rating agencies further downgraded the sovereign debt ratings of Italy, Spain, Belgium and Portugal. An international alarm over a Eurozone crisis grew, when the US president Barack Obama stated on 26 September the debt crisis in Europe was "scaring the world". As a result, global investors became increasingly skittish about Europe's prospects for resolving its crisis, contributing to an extreme connectedness state occurring around August-November. In the last month of 2011, the system returned to the low connectedness state, which may be associated with the announcement of the *Long-Term Refinancing Operation* through which the European Central Bank offered three-year loans to more than 500 Euro Area banks.

The first half of 2012 was marked by two high-connectedness states, taking place around February and May respectively. The former was largely driven by Moody's cut of the debt ratings of six European countries, along with a pessimistic outlook of the EU economy according to the European Commission. The latter was partly attributable to an escalated political crisis in Greece. In the summer of 2012, European Central Bank president Mario Draghi delivered the famous "Whatever it takes" speech, which marked the turnaround of the euro crisis. Since then, investor sentiment has gradually recovered and the bank stock market has steadily returned to stability. Therefore, the global banking system tended to stay in the low connectedness state toward the end of the year, with one exception that a short-lived high-connectedness state occurred in October following the announcement of a near junk status of Spain's sovereign credit rating.

The subsequent period 2013-2019 was primarily characterized by the low connectedness state. In mid-October 2014, the global banking system transited to the high connectedness state for a short while due to the flash crash in the US Treasury bill market.

The global banking system was once in the extreme connectedness state in March 2020 and the first half of June of 2020, coinciding with the first major wave of COVID-19 around the world. At this time, banks were exposed to a large panel of issues related to the financing of the real economy, a decrease in assets due to the repayment delay of SMEs, volatility of assets under management, a reduced amount of capital exchanged because of the lockdown, and volatility of reserves resulting from exchange rate volatility. The year following 2020 was characterized by one fleeting high-connectedness state in early June, possibly due to the advance of the delta variant of the coronavirus in Europe and the potential drag this could have on the normalization of activity (e.g., for tourism). With the combination of geopolitical disruption and soaring inflation in 2022, the banking system transited to the high connectedness state twice during March-April and October-December. The former transition was driven by the heightened Russia-Ukraine conflict and the continued energy market pressures. The latter was associated with the announcement of rising interest rates in most developed economies.

Chapter3

Determination of Regime Dimension in Markov-Switching Panel Models

3.1 Introduction

In Markov-switching (MS) panel models, several latent regimes can summarize the dynamics of data from observational units (such as workers, firms, and countries). This is achieved by allowing individuals within a given regime to share similar distributions (e.g., the means and the variances), while distributions can differ across the regimes. From the identified regimes, such models are useful to glean a range of economic insights, which may relate to the business cycle synchronization among countries or regions in the economy (Agudze et al., 2022), the propagation of macroeconomic shocks on a panel of equity returns (Zhu et al., 2017), the dynamic transitions across business model groups for a collection of financial institutions (Joao et al., 2023), and many others. Despite the usefulness of the MS panel models, one major challenge exists when applying them in practice, which is determining how many regimes are necessary to adequately characterize the observed data.

Existing solutions for the determination of the number of regimes in MS panel models typically employ the assumption that the regime dimension is homogeneous in the cross-section. In other words, the same set of regimes characterizes all individual units. With this assumption, the problem of determining the regime dimension for an MS panel model is treated as the problem of finding out the value of a variable that represents the regime dimension for the whole panel. Sometimes, the value of such a single variable can be dictated by the actual applications. For example, two regimes are often considered for macroeconomic panel data when investigating the business cycle phases (see, e.g., Hamilton and Owyang, 2012; Francis et al., 2017; Casarin et al., 2018; Owyang et al., 2022; Agudze et al., 2022). For

more general applications where there is no external information on the number of regimes, a common approach to determine the regime dimension follows a model selection procedure. This generally consists in estimating the MS panel model for increasing values of the regime dimension variable up to an upper bound and then choosing the optimal value of the variable by using a model selection criterion (see, e.g., Billio et al., 2016). Apparently, estimation of the model needs to be repeated several times for different values of regime dimension, which makes the overall procedure time-consuming and computationally demanding. Moreover, the choice of both the upper bound of the regime dimension and the model selection criterion suffers from a degree of subjectivity. These limitations motivate an alternative approach, which is based on Bayesian nonparametric (BNP) inference and aims to estimate the unknown regime dimension variable in MS panel models (see, e.g., Casarin et al., 2024). The idea of this BNP inference approach is to extend the finite number of regimes of the MS panel model to the case in which this number goes to infinity, thus transforming the parametric MS panel framework into a nonparametric structure. A direct consequence is that the regime transition matrix has an infinite dimension. Then with a proper nonparametric prior (e.g, the Dirichlet process or its variant, the Pitman-Yor process) imposed on this infinite-dimensional matrix, the superfluous matrix dimension can be emptied during estimation and the non-empty dimension finally indicates the finite regime dimension of the panel.

Assuming homogeneous regime dimension across the cross-section, however, may be restrictive as individuals are likely to be characterized only by one or a subset of regimes identified from the panel. For example, in spite of the wide-spreading effects of the 1997 Asian financial crisis, both China and India were not victims as opposed to most Asian economies (Park et al., 2010). Similarly, for the recent Global Financial Crisis after 2007, some US states were heavily affected, such as Michigan, while other states were essentially unaffected, such as Texas, which benefited from high oil prices in 2009 and 2010. Together, these observations imply that the regimes available in the panel do not necessarily affect all units and the set of regimes owned by individuals might be of heterogeneous dimension.

This paper proposes a general framework to estimate the number of regimes in MS panel models, allowing possible heterogeneity in cross-sectional regime dimension. We model individual heterogeneity via a binary matrix, where each row corresponds to a panel unit, each column corresponds to a regime, and an entry of one indicates a given unit owns a particular regime available in the panel. Hence, the column dimension of this binary matrix indicates the total number of regimes for the panel, and the number of nonzero entries in a particular row of this binary matrix indicates the number of regimes for the individual units. Importantly, our framework includes the MS panel models with cross-sectional homogeneous regime dimension as a special case where the binary matrix has all one entries. To maintain an unbounded number of regimes in the panel, we then model the binary matrix nonparametrically by letting its column dimension go to infinity. As a result, the problem of determining the number of regimes for an MS panel model is recast as the problem of estimating the unknown

column dimension and configuration of a binary matrix with a finite number of rows and an infinite number of columns.

Methodologically, we develop new BNP inference to jointly estimate the latent binary matrix and the other parameters in MS panel models. In particular, we specify the binary matrix using a nonparametric prior based on an Indian buffet process (Griffiths and Ghahramani, 2005, 2011), which allows an infinite total number of regimes in the panel while encouraging the sharing of regimes across units. We formulate the unit-specific regime transition matrix via the usual Beta-to-Dirichlet transformation and a prior based on the zero-inflated Beta distribution, which constraints individuals to transit among its set of present regimes informed by the binary matrix. We place a hierarchical conjugate prior over the unit-specific regime-switching parameters, which enables units within the same regime to share similar dynamics and helps deliver a closed-form data marginal likelihood that improves posterior computation efficiency. To accomplish the estimation, we present a six-step Markov chain Monte Carlo procedure that exploits a combination of multi-move Gibbs, Metropolis-Hasting and reversible jump samplers.

Our inference approach makes contributions that relate to two branches of literature. First, it contributes to the expanding literature on Bayesian nonparametric Markov-switching models concerning an infinite number of regimes. Existing studies routinely use the Dirichlet process or its variants to infer the regime dimension from the finite dimension of one single common transition matrix. For this reason, previous BNP solutions to the regime dimension determination problem are limited to univariate models (see, e.g., Otranto and Gallo, 2002; Song, 2014; Dufays, 2016; Bauwens et al., 2017; Jin et al., 2019; Luo et al., 2022), and multivariate models assuming common regimes and transitions for all individuals (see, e.g., Hou, 2017; Casarin et al., 2024). Our approach complements these BNP solutions by utilizing the Indian buffet process as an alternative to the Dirichlet process, and estimating the regime dimension for multivariate models in the presence of possibly heterogeneous regimes occurring to some individuals. Second, the proposed inference approach contributes to the Bayesian nonparametrics literature on applications of the Indian buffet process (IBP). The IBP so far has received considerable attention in latent feature modeling (see, e.g., Broderick et al., 2013; Ročková and George, 2016; James, 2017; Ni et al., 2020; Lui et al., 2023), yet only a few studies have considered the IBP in MS modeling. A closely related work is Fox et al. (2014), which employs the IBP as a prior on the infinite-dimensional binary representation of the unknown regime pattern underlying multiple time series. Our approach differs in two main aspects. On the one hand, Fox et al. (2014) model the unit-specific regime transition matrix via a Gamma-to-Dirichlet transformation, which requires sampling an extra scaling factor for normalizing the transition probabilities. We bypass the additional sampling and normalization steps by using a more straightforward Beta-to-Dirichlet transformation. On the other hand, Fox et al. (2014) assume the regime-switching parameters are identical between individuals. Instead, we explicitly model the unit-specific regime-switching parameters via

a hierarchical conjugate prior, which allows individuals' dynamics within a given regime to be quite similar but not the same. Therefore, our approach, benefiting from more suitable priors on the transition matrix and regime-switching parameters, improves the flexibility and applicability of the IBP in the field of MS multivariate modeling.

Through simulations, we examine the effectiveness of our framework under a range of panel settings, differing in the cross-section dimensions, the time series dimensions, the degrees of cross-sectional heterogeneity in the regime dimension, and the degrees of individual dynamics across regimes. For panels with homogeneous regime dimension in the cross-section, our approach can efficiently recover the true regime dimension of the whole panel, no matter the size of cross-section, the length of time series, and the stability of individual regime-switching dynamics. For panels with weak and strong levels of cross-sectional heterogeneity in the regime dimension, regardless of how unstable the individual regime-switching dynamics are, our approach continues to perform well in estimating the total number of regimes so long as the sample size of the panel is not too large. In every case where the total number of regimes for the panel is correctly estimated, our framework can further precisely estimate the regime dimension and the regime-switching dynamics for each unit.

An application to 50 US state-level macroeconomic indices over the period 1979-2023 demonstrates the empirical gains of relaxing the assumption of cross-sectional homogeneous regime dimension in MS panel models. Our framework that assumes cross-sectional heterogeneous regime dimension identifies seven regimes of the whole panel over the 40-year sample. About 25% of US states possess five of the identified seven regimes whereas the remaining states possess all seven regimes, indicating our assumption is plausible empirically. From the turning points of the identified regimes, it appears that our framework produces accurate timing of recessions that agrees with the NBER-designated cycle dates and is fully flexible in capturing the noncommon and non-synchronized dynamics underlying individual US states. A recursive out-of-sample forecasting exercise shows that the MS panel (or multivariate) framework assuming cross-sectional homogeneous regime dimension as well as the MS univariate framework are unable to rival the forecast accuracy generated from our MS panel framework allowing for likely heterogeneous regime dimension in the cross-section.

The remainder of this paper is organized as follows. Section 3.2 introduces the methodology. Section 3.3 conducts Monte Carlo simulations and Section 3.4 provides an empirical application to US state-level macroeconomic indices. Section 3.5 concludes. Technical details, further simulation results and additional material on the empirical illustration are provided in Appendix 3.

3.2 Methodology

3.2.1 Model

We consider a panel with $i = 1, ..., N$ units and $t = 1, ..., T$ time periods that is subject to an unknown number of regimes, denoted by K. We consider an unbounded number of regimes by letting K go to infinity. The i-th unit may own just a single one or a subset or all of the K regimes over the T time periods. Let f_{ik} denote a unit-specific binary indicator which equals one if the i-th unit owns the k-th regime and zero otherwise. The k-th regime contains $N_k = \sum_{i=1}^{N} f_{ik}$ number of units, subject to $1 \leq N_k \leq N$. The set $\mathcal{N}_k = \{i : f_{ik} = 1\}$ of dimension N_k thus collects those units possessing the k-th regime. The i-th unit possesses $K_i = \sum_{k=1}^{K} f_{ik}$ number of regimes, subject to $1 \leq K_i \leq K$. The set $\mathcal{K}_i = \{k : f_{ik} = 1\}$ of dimension K_i thus collects those regimes possessed by the i-th unit. Let $\boldsymbol{f}_i = \{f_{i1}, ..., f_{iK}\}$ denote a K-dimensional vector collecting all the unit-specific binary indicators. Given $\boldsymbol{f}_i$, the set of regimes possessed by a panel can be further expressed in the form of a binary matrix $F = \{\boldsymbol{f}_1, ..., \boldsymbol{f}_N\}'$ with a finite number of rows (N) and an infinite number of columns (K). Conditional on this infinite-dimensional binary matrix F, for units $i = 1, ..., N$, a simple nonparametric MS panel model with possibly cross-sectional heterogeneous regime dimension takes the following form[1]

$$\begin{aligned} &y_{it} = \mu_{i,s_{it}} + \sigma_{i,s_{it}} \varepsilon_{it}, \quad \varepsilon_{it} \sim \text{ i.i.d. } \mathcal{N}(0,1) \\ &s_{it} | s_{it-1} = k, P_i \sim \boldsymbol{p}_{i,k}, \end{aligned} \tag{3.1}$$

where y_{it} is the time series from the i-th observational units; $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2; s_{it} is a hidden Markov chain process with values in the set $\mathcal{K}_i$; P_i is a $K_i \times K_i$ regime transition matrix; $\boldsymbol{p}_{i,k} = (p_{i,k1}, ..., p_{i,kl}, ...)$ is the k-th row of P_i which collects the transition probabilities out of regime k, i.e., $p_{i,kl} = \text{Prob}(s_{it} = l | s_{it-1} = k)$, for $l, k \in \mathcal{K}_i$. The functional form for the regime-switching parameters is

$$\mu_{i,s_{it}} = \sum_{k \in \mathcal{K}_i} \mu_{i,k} \mathbb{I}(s_{it} = k), \qquad \sigma^2_{i,s_{it}} = \sum_{k \in \mathcal{K}_i} \sigma^2_{i,k} \mathbb{I}(s_{it} = k). \tag{3.2}$$

where $\mathbb{I}(\cdot)$ is the indicator function which takes value one if the chain s_{it} in the regime k and zero otherwise.

[1] Similar to several studies in the literature (see, e.g., Hamilton and Owyang, 2012; Agudze et al., 2022; Casarin et al., 2024), we regard the parsimonious MS panel framework in (3.1) to illustrate the proposed methodology. There is no conceptual problem with adding autoregressive terms, exogenous predictors, or cross-section error correlations to this equation, though that would greatly increase the number of parameters for which one needs to draw an inference.

3.2.2 Model intuition and comparison with similar approaches

To illustrate the importance of assuming cross-sectional heterogeneity in the regime dimension, we display in Figure 3.1 three hypothetical regime patterns the panel may exhibit. In the left and middle plots, the number of regimes is homogeneous in the cross-section (all four units own over time the three regimes in the panel). The left and middle plots differ in the time at which the three regimes hit the units, with the former considering synchronized regimes (hitting individuals simultaneously) while the latter permitting non-synchronized regimes (hitting individuals with different delays). In the right plot, the number of regimes is heterogeneous in the cross-section (unit owns over time a single one or two or all of the three regimes in the panel) and non-synchronized regimes are permitted. The regime pattern

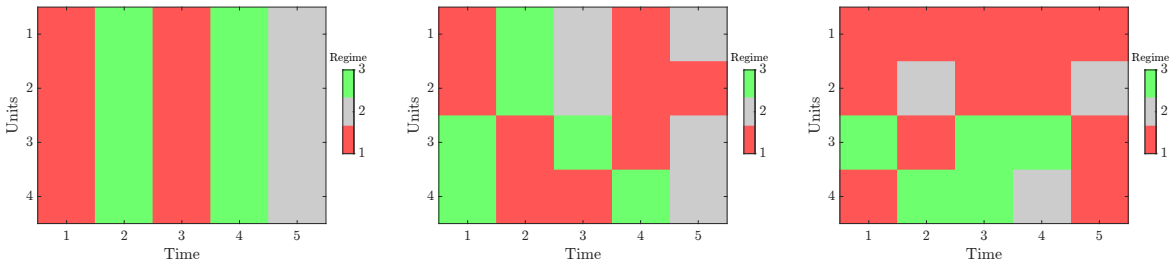


Figure 3.1 Illustration of three hypothetical regime patterns the panel may exhibits.

in the left plot is the result of applying the existing BNP MS multivariate frameworks to the panel considered (see, e.g., Hou, 2017; Casarin et al., 2024). The next one in the middle plot can be identified by a standard MS panel model, with the number of regimes $K = 3$ being set through prior belief (see, e.g., Hamilton and Owyang, 2012; Francis et al., 2017; Casarin et al., 2018; Owyang et al., 2022; Agudze et al., 2022) or a model selection procedure (see, e.g., Billio et al., 2016). The last one in the right plot cannot be captured by any of the above frameworks but could be captured to some extent by the existing BNP MS univariate frameworks, when applied separately to each unit in the cross-section (see, e.g., Otranto and Gallo, 2002; Song, 2014; Dufays, 2016; Bauwens et al., 2017; Jin et al., 2019; Luo et al., 2022). The identified regimes from such univariate approaches, however, are specific to individual units and thus overlook information in the cross-section. By allowing for cross-sectional heterogeneous regime dimension in MS panel models, our methodology therefore has the benefit of accommodating the last regime pattern as the BNP MS univariate framework while still preserving much of the increased power derived from the cross-section. Our approach is also fully flexible in its ability to encompass the regime patterns in the left and middle plots since after estimation the binary matrix may have all-one entries and the unit-specific Markov chain process may (or not) be synchronized with each other.

3.2.3 Prior distributions

We now specify the prior distributions over the latent binary matrix and the other parameters in the proposed nonparametric MS panel framework.

3.2.3.1 Prior on infinite-dimensional binary matrix

For a prior on the infinite-dimensional binary matrix F, we adopt the Indian buffet process (IBP) (Griffiths and Ghahramani, 2005, 2011). The IBP is based on a culinary metaphor in which customers arrive at an infinitely long buffet line of dishes. By viewing customers as panel units and dishes as regimes, we show our proposed infinite-dimensional binary matrix can be specified through a procedure by which customers (units) choose dishes (regimes):

1. Imagine a sequence of N customers entering an Indian buffet restaurant. Each customer selects a finite number of dishes (i.e., K_i), chosen from a limitless supply of potential dishes to taste (i.e., K).

2. The first customer starts at the left of the buffet and takes a serving from each dish, stopping after a Pois(β) number of dishes. The dishes are labeled $1,...,K_1$. This gives

$$\begin{aligned} K_1 &\sim \text{Pois}(\beta), \\ \mathcal{K}_1 &= \{1,...,K_1\}, \end{aligned} \tag{3.3}$$

   where Pois($\cdot$) denotes the Poisson distribution; $\beta > 0$ is called the *mass parameter*, which controls the total number of dishes tried by N customers.

3. For every $i > 1$, the i-th customer chooses which dishes to taste in two parts. First, for each dish k that has previously been tasted by any customer in $1,...,i-1$, the i-th customer tastes disk k or not (i.e., f_{ik}) according to its popularity which is specified as the probability $\frac{f_{1k}+f_{2k}+...+f_{i-1k}}{c+i-1}$. Having reached the end of all previously tasted dishes, the i-th customer then chooses a Pois($\frac{\beta c}{c+i-1}$) number of new dishes to try. The new dishes receive unique labels $M_{i-1}+1, M_{i-1}+2, ..., M_{i-1}+K_i^+$. Here M_i represents the number of tasted dishes after i customers and K_i^+ denotes the number of new dishes.

   Let $\mathcal{M}_i = \mathcal{K}_1 \cup \mathcal{K}_2 ... \cup \mathcal{K}_i$ be the set of previously tasted dishes after i customers. The cardinality of $\mathcal{M}_i$ is M_i. Let $N_{i,k} = f_{1k} + f_{2k} + ... + f_{ik}$ be the number of customers indexed $1,...,i$ who have tasted dish k. The above behavior of customer i gives

$$\begin{aligned} K_i &= \sum_{k \in \mathcal{M}_{i-1}} f_{ik} + K_i^+, \text{ with } f_{ik} \sim \text{Bern}(\frac{N_{i-1,k}}{c+i-1}), K_i^+ \sim \text{Pois}(\frac{\beta c}{c+i-1}), \\ \mathcal{K}_i &= \{k : f_{ik} = 1, k \in \mathcal{M}_{i-1}\} \cup \{M_{i-1}+1, M_{i-1}+2, ..., M_{i-1}+K_i^+\}, \end{aligned} \tag{3.4}$$

where Bern$(\cdot)$ denotes the Bernoulli distribution; $c > 0$ is called the *concentration parameter*, which controls the average number of dishes per customer.

4. After N customers, the total number of tasted dishes is $M_N = M_{N-1} + K_N^+$.

An example of the infinite-dimensional binary matrix produced by the above procedure is shown in Figure 3.2. The first customer tasted four dishes: $K_1 = 4, \mathcal{K}_1 = \{1,2,3,4\}, \mathcal{M}_1 = \mathcal{K}_1$ and $M_1 = 4$. The second customer tasted two of those dishes, and then tried two new dishes: $K_2 = 4, \mathcal{K}_2 = \{2,4\} \cup \{5,6\}, \mathcal{M}_2 = \mathcal{K}_1 \cup \mathcal{K}_2 = \{1,2,3,4,5,6\}, M_2 = 6$ and $K_2^+ = 2$. Vertically concatenating the choices of the $N = 5$ customers produces the binary matrix shown in the figure.

Through the illustration of the Indian buffet metaphor, the probability of any particular infinite-dimensional binary matrix $\hat{F}$ is the result of the Poisson distribution for the number of new dishes times the Bernoulli distributions for the choice of existing dishes. Formally,

$$p(\hat{F}) = \prod_{i=1}^{N} f^{\text{Pois}}(K_i^+; \frac{\beta c}{c+i-1}) \times \prod_{k=1}^{M_N} \prod_{i=1}^{N} f^{\text{Bern}}(f_{ik}; \frac{N_{i-1,k}}{c+i-1}), \tag{3.5}$$

where f^A denotes the probability mass function of a distribution A.

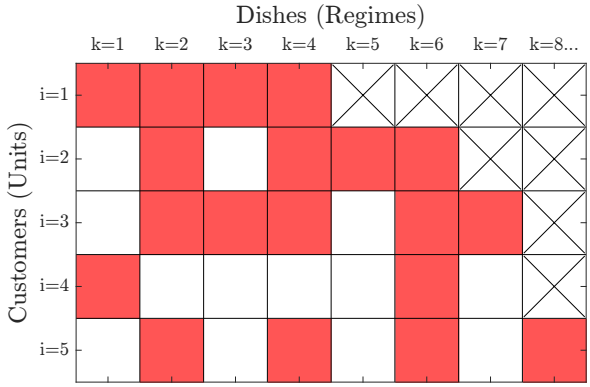


Figure 3.2 Illustration of an infinite-dimensional binary matrix produced by IBP. The buffet consists of an infinitely long vector of dishes, corresponding to regimes. Each customer—corresponding to a panel unit—who enters first decides whether or not to eat dishes that the other customers have already tasted and then tries a random number of new dishes, not previously tasted by any customer. A red box in position (i,k) indicates dish k has been chosen by customer i, a white box indicates the dish has not been chosen by the customer, and a box marked with $\times$ indicates dish k has not yet been chosen by anyone.

From Figure 3.2, we also see that two dishes $k = 2$ and $k = 4$ are tasted by exactly the same set of customers; i.e., $\mathcal{N}_2 = \{1,2,3,5\}$ and $\mathcal{N}_4 = \{1,2,3,5\}$. Therefore, random permutations of columns $k = 2$ and $k = 4$ will yield the same infinite-dimensional binary matrix $\hat{F}$ shown in the figure. Taking this into account, a random infinite-dimensional binary matrix F will refer to the class of all particular matrices $\hat{F}$ that exhibit the same left-orderings (Griffiths and Ghahramani, 2005, 2011). Hence, a combinatoric term is augmented to the probability mass

functions in (3.5) such that the probability of random matrix F is unaffected by changing the order of its columns

$$\begin{aligned} p(F=\hat{F}) &= \frac{\prod_{i=1}^{N} K_i^+!}{\prod_{h=1}^{2^N-1} \tilde{K}_h!} \times p(\hat{F}), \\ &= (\prod_{h=1}^{2^N-1} \tilde{K}_h!)^{-1} (\beta c)^{M_N} \exp(-\beta c \sum_{i=1}^{N} (c+i-1)^{-1}) \prod_{k=1}^{M_N} \frac{\Gamma(N_k)\Gamma(N-N_k+c)}{\Gamma(N+c)}, \end{aligned} \tag{3.6}$$

where $\tilde{K}_h$ represents the number of columns in F for the h-th possible configuration. The cardinality of all possible non-zero binary columns in F of length N is 2^N-1 so $\prod_{h=1}^{2^N-1} \tilde{K}_h!$ iterates over the sample space of all distinct non-zero columns in F. It is also clear that the probability mass function specified in (3.6) has no dependence on the ordering of the indices in N. Hence, the distribution of F is invariant under permutations of the indexes of columns and rows in F. Overall, the prior on our infinite-dimensional binary matrix F follows an IBP(β, c) distribution[2] with the probability mass function specified in (3.6). To infer the values of the IBP parameters β and c, we additionally place gamma priors so that $\beta \sim \mathcal{G}(a_\beta, b_\beta)$ and $c \sim \mathcal{G}(a_c, b_c)$, where $\mathcal{G}(a,b)$ denotes the gamma distribution with shape and scale parameters a and b.

3.2.3.2 Prior on zero-embedded regime transition matrix

Previous studies (see, e.g., Casarin et al., 2018; Agudze et al., 2022; Casarin et al., 2024) show that the unit-specific regime transition matrices P_i within the panel are usually specified by a common Dirichlet distribution as follows

$$\boldsymbol{p}_{i,k} = (p_{i,k1}, ..., p_{i,kK}) \sim \text{ i.i.d. Dirichlet}(\alpha_1, ..., \alpha_K), \quad i = 1, ..., N,\ k = 1, ..., K \tag{3.7}$$

where $\boldsymbol{p}_{i,k}$ is the k-th row of P_i; $\alpha_1, ..., \alpha_K$ are Dirichlet distribution parameters.

In our context assuming cross-sectional heterogeneous regime dimension, each unit transits only between its set of present regimes informed by the identified binary vector $\boldsymbol{f}_i$ (i.e., the i-th row of the infinite-dimensional binary matrix F) with dimension up to $\max(\mathcal{K}_i)$. This means the unit-specific probability vector $\boldsymbol{p}_{i,k}$ is now defined over the finite set of regimes $l \in \{1, ..., \max(\mathcal{K}_i)\}$, but assigns nonzero probabilities only at regimes l where $f_{il} = 1$. In

[2] In this paper, we use the two-parameter IBP, i.e., IBP(β, c), under which the asymptotic behavior of M_N (i.e., the total number of regimes) is $\mathcal{O}(\log N)$ (Griffiths and Ghahramani, 2005) (for fixed β and c) and it is clear that M_N grows slowly with N. Hence, the two-parameter IBP is suitable for many MS applications in macroeconomics and finance where a small total number of regimes is usually observed. For applications where one expects the distribution of the total number of regimes to exhibit a heavy tail with N, a three-parameter IBP (Teh and Gorur, 2009; Broderick et al., 2012) can be used since asymptotically M_N is $\mathcal{O}(N^d)$ $(0 < d < 1)$ and grows with N faster than the two-parameter case.

other words, the entry $p_{i,kl}$ in $\boldsymbol{p}_{i,k}$ now takes the form

$$\begin{cases} p_{i,kl} = 0 & \text{if} \quad f_{il} = 0 \\ p_{i,kl} > 0 & \text{if} \quad f_{il} = 1 \end{cases} \tag{3.8}$$

for $l \in \{1, 2, ..., \max(\mathcal{K}_i)\}$, and satisfies $\sum_l p_{i,kl} = 1$. As a result, (3.8) leads to a zero-embedded regime transition matrix which is different from the standard one in (3.7) where all entries are nonzero.

To specify a prior on our zero-embedded regime transition matrix, we start by following Connor and Mosimann (1969) and Ishwaran and James (2001) to rewrite (3.7) based on a set of mutually independent Beta variables $\boldsymbol{z}_{i,k} = (z_{i,k1}, ..., z_{i,kK-1})$

$$\begin{aligned} &p_{i,k1} = z_{i,k1}, \quad p_{i,kl} = z_{i,kl} \prod_{m=1}^{l-1} (1 - z_{i,km}) \text{ for } l = 2, ..., K-1, \\ &\text{and } p_{i,kK} = 1 - \sum_{m=1}^{K-1} p_{i,km}, \end{aligned} \tag{3.9}$$

where $z_{i,kl} \sim$ i.i.d. Beta$(\alpha_l, \sum_{m=l+1}^{K} \alpha_m)$ for $l = 1, ..., K-1$ and $K = \max(\mathcal{K}_i)$. To make transition matrix conditional on the identified binary vector $\boldsymbol{f}_i$, we instead employ a zero-inflated Beta (ZIB) distribution (see, e.g., Tang and Chen, 2019) for the Beta variables $\boldsymbol{z}_{i,k}$, that is

$$\begin{cases} z_{i,kl} = 0 & \text{if} \quad f_{il} = 0 \\ z_{i,kl} | f_{il} = 1 \sim \text{ i.i.d. Beta}(\alpha_l, \sum_{m=l+1}^{K} \alpha_m) & \text{if} \quad f_{il} = 1 \end{cases} \tag{3.10}$$

Consequently, transforming those zero-inflated Beta variables to $\boldsymbol{p}_{i,k}$ through (3.9) gives a zero-embedded regime transition matrix. As many macroeconomic and financial time series evolve with high persistence, we further follow the sticky version of Fox et al. (2011) to define $\alpha_k = \gamma + \kappa\delta_k$ in (3.10), where δ_k is the measure of a point mass of one at k. The term $\kappa\delta_k$ indicates that to element α_k (i.e., the k-th parameter in the Dirichlet distribution) is added $\kappa \geq 0$. Through κ, regime self-transition can be reinforced. To infer the values of the ZIB parameters γ and κ, we additionally place gamma priors so that $\gamma \sim \mathcal{G}(a_\gamma, b_\gamma)$ and $\kappa \sim \mathcal{G}(a_\kappa, b_\kappa)$.

3.2.3.3 Prior on regime-switching parameters

For priors on the regime-switching parameters $(\mu_{i,k}, \sigma_{i,k})$, we specify a hierarchical conjugate normal–gamma distribution

$$\begin{aligned} &\mu_{i,k} | \sigma_{i,k}^{-2} \sim \mathcal{N}(\mu_k, (\sigma_{i,k}^{-2})^{-1} h_k^{-1}), \\ &\sigma_{i,k}^{-2} \sim \mathcal{G}(\frac{\nu_k}{2}, 2V_k^{-1}), \end{aligned} \tag{3.11}$$

with

$$\begin{aligned}
\mu_k &\sim \mathcal{N}(\underline{\mu}, (\underline{h}^{-2})^{-1}), \\
h_k &\sim \mathcal{G}(\frac{\underline{d}}{2}, 2\underline{c}^{-1}), \\
\nu_k &\sim \text{Exp}(\underline{\lambda}), \\
V_k &\sim \mathcal{G}(\frac{\underline{\nu}}{2}, 2\underline{V}^{-1}),
\end{aligned} \tag{3.12}$$

where $\text{Exp}(\lambda)$ is the exponential distribution with mean λ^{-1}. The hierarchical conjugate prior distribution for $(\mu_{i,k}, \sigma_{i,k})$ has several advantages. First, given (3.11), there is a degree of similarity in the regime-switching parameters of those units that share the same regime since parameters are shrunk towards the regime-specific common value (μ_k, V_k). Second, due to conjugacy, the regime-switching parameters can be integrated analytically to derive the marginal likelihood of y_{it}. This property provides a closed-form function of the joint posterior distribution of data observations, binary matrix and regime sequences, which can be used to simplify the posterior sampling of the infinite-dimensional binary matrix F as detailed in Appendix 3.A.

3.2.4 Posterior Approximation

Let $Y_{\tau:t} = (\boldsymbol{y}_\tau, ..., \boldsymbol{y}_t)$ and $S_{\tau:t} = (\boldsymbol{s}_\tau, ..., \boldsymbol{s}_t)$ be the collection between time τ and t of observation vectors $\boldsymbol{y}_t = (y_{1t}, ..., y_{Nt})$ and of hidden Markov chain vectors $\boldsymbol{s}_t = (s_{1t}, ..., s_{Nt})$, respectively. Let $\Theta = (\{\mu_{i,k}\}_{i=1}^N, \{\sigma_{i,k}^{-2}\}_{i=1}^N, \mu_k, h_k, \nu_k, V_k)$ be the collection of regime-specific parameters in the hierarchical conjugate normal–gamma distribution. Let $\Theta_{\text{IBP}} = \{\beta, c\}$ and $\Theta_{\text{ZIB}} = \{\gamma, \kappa\}$ be the collection of parameters in the IBP and ZIB distributions, respectively. Let $P = \{P_i\}_{i=1}^N$ be the collection of regime transition matrices.

Our framework is estimated using a Markov chain Monte Carlo (MCMC) procedure. Specifically, we approximate the posterior distribution of the proposed BNP MS panel model by implementing a multi-move Gibbs sampling algorithm, where the infinite-dimensional binary matrix, the hidden regime sequences, and the parameters are all sampled in blocks. At each iteration, the Gibbs sampler iterates over the following six steps. Computational details of each step are given in Appendix 3.A.

Step 1: Sampling the infinite-dimensional binary matrix F given $\Theta, \Theta_{\text{IBP}}, P, S_{1:T}, Y_{1:T}$.
Step 2: Sampling the hidden regime sequences $S_{1:T}$ given $F, \Theta, P, Y_{1:T}$.
Step 3: Sampling the regime transition matrices P given $F, \Theta_{\text{ZIB}}, S_{1:T}$.
Step 4: Sampling the regime-specific parameters Θ given $F, S_{1:T}, Y_{1:T}$.
Step 5: Sampling the IBP parameters Θ_{IBP} given F.
Step 6: Sampling the ZIB parameters Θ_{ZIB} given P.

In MS modeling, one has to deal with the nonidentifiability of the regimes, which is commonly called the label-switching problem. See, for example, Frühwirth-Schnatter (2001) for a discussion on the effects that the label switching and the lack of identification have on the results of MCMC-based Bayesian inference. To avoid the label-switching problem, in this paper we impose the regime identifying restriction $\mu_1 < \mu_2 < ... < \mu_K$, which implies that a number of K regimes are separated and identified according to the means the regime characterize. This restriction follows to a large extent the application of MS panel models in macroeconomics where data are growth rate of economic variables and is related to the natural interpretation of the different regimes as different phases (e.g., recession and expansion) of the business cycle (see, e.g., Billio et al., 2016; Casarin et al., 2018). Alternative identifying restrictions based on volatilities or other statistical properties of the regimes can also be considered in our framework for applications in which one wants to discover regimes with some specific economic interpretation.

3.3 Simulation study

This section illustrates the ability of our BNP MS panel framework to correctly estimate (i) K, namely the total number of regimes available in the panel; (ii) F, namely the regime pattern the panel exhibits (i.e., the regime dimension for the individuals); (iii) the individual regime-switching dynamics, namely its regime-switching parameters and hidden regime sequence.

3.3.1 Simulation design

We consider time series dimensions (T) of 150, 250, and 500, which correspond to approximately ten, twenty and forty years of monthly data. We consider cross-section dimensions (N) of 25, 50, and 100. We assume there are $K = 3$ regimes in the panel. We consider many empirically relevant panel settings, examining the sensitivity of our approach to panels with different degrees of cross-sectional heterogeneity in the regime dimension and with various degrees of individual dynamics across regimes. This is investigated via six different data generating processes (DGPs). The first two differ from the last four in the regime patterns of the panel. Among each two, they further differ from each other in the regime transition matrices and the degrees of separation in the regime-specific parameters. More specifically, we consider three scenarios for the regime patterns of the panel, which are reflected in the configuration of F

1. *Homogeneous F (HoF)*: this scenario specifies an $N \times K$ matrix F with all one entries, which implies that the regime dimension is homogeneous in the cross-section.

2. *Weak Heterogeneous F (WHeF)*: this scenario specifies an $N \times K$ matrix F with random binary entries for 25% of units in the panel and all one entries for the remaining units, which implies a weak level of cross-sectional heterogeneity in the regime dimension.

3. *Strong Heterogeneous F (SHeF)*: this scenario specifies an $N \times K$ matrix F with random binary entries for 50% of units in the panel and all one entries for the remaining units, which implies a strong level of cross-sectional heterogeneity in the regime dimension.

Further, we consider two settings of regime-specific parameters and regime transition matrices, where parameter values are set based on the in-sample estimation results obtained from the empirical section

1. *Weakly Separated Regimes (WS)*: this setting mimics individual regime-switching dynamics over tranquil periods.

$$\begin{aligned} &\mu_{i,1} = -0.2 + 0.01\eta_{i,1},\ \mu_{i,2} = 0 + 0.01\eta_{i,2},\ \mu_{i,3} = 0.2 + 0.01\eta_{i,3}. \\ &\sigma^2_{i,1} = 0.015 + 0.01^2\zeta_{i,1},\ \sigma^2_{i,2} = 0.003 + 0.01^2\zeta_{i,2},\ \sigma^2_{i,3} = 0.01 + 0.01^2\zeta_{i,3}. \end{aligned} \tag{3.13}$$

   where $\eta_{i,k} \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\zeta_{i,k} \sim$ i.i.d. $\mathcal{N}(0,1)$ for $i = 1,...,N$ and $k = 1,...,K$. When $1 < K_i \leq K$, each row in the transition matrix P_i is sampled from a K_i-dimensional Dirichlet distribution with the mean of the regime self-transition probabilities equaling to 0.95; i.e., $E(p_{i,kk}) = 0.95$ for $k \in \mathcal{K}_i$. This implies that an expected duration of each regime is equal to $20(=1/(1-p_{i,kk}))$ observations which represents nearly two years of monthly data. When $K_i = 1$, the transition matrix P_i is set equal to one since the unit stays in one regime all the time.

2. *Weakly Separated Regimes + Short-lived Strongly Separated Regimes (WSSS)*: this setting mimics individual regime-switching dynamics over tranquil and turbulent periods. The regime $k = 1$ corresponds to the turbulent periods.

$$\begin{aligned} &\mu_{i,1} = -0.6 + 0.01\eta_{i,1},\ \mu_{i,2} = 0 + 0.01\eta_{i,2},\ \mu_{i,3} = 0.2 + 0.01\eta_{i,3}. \\ &\sigma^2_{i,1} = 0.1 + 0.01^2\zeta_{i,1},\ \sigma^2_{i,2} = 0.003 + 0.01^2\zeta_{i,2},\ \sigma^2_{i,3} = 0.01 + 0.01^2\zeta_{i,3}. \end{aligned} \tag{3.14}$$

   where $\eta_{i,k} \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\zeta_{i,k} \sim$ i.i.d. $\mathcal{N}(0,1)$ for $i = 1,...,N$ and $k = 1,...,K$. When $1 < K_i \leq K$, each row in the transition matrix P_i is sampled from a K_i-dimensional Dirichlet distribution, with the mean of the regime self-transition probabilities equaling to 0.75 for the regime $k = 1$ and 0.95 for the remaining regimes; i.e., $E(p_{i,11}) = 0.75$, $E(p_{i,22}) = 0.95$ and $E(p_{i,33}) = 0.95$. This implies that the expected durations of the regime $k = 1$ and the other regimes are equal to $4(=1/(1-p_{i,11}))$ months and nearly two years of monthly data, respectively. When $K_i = 1$, the transition matrix P_i is set equal to one.

Overall, the six DGPs are labelled as *HoF + WS*, *HoF + WSSS*, *WHeF + WS*, *WHeF + WSSS*, *SHeF + WS*, and *SHeF + WSSS*. Each DGP is simulated from (3.1) with cross-section dimensions $N = 25, 50, 100$, time series dimensions $T = 150, 250, 500$ and the true number of regimes $K = 3$. For each DGP with a specific value of N and T, we conduct 20 independent experiments, simulating different panels. For each simulated panel, we estimate our framework by running the proposed MCMC procedure for 5,000 iterations. To reduce the serial correlation among the MCMC samples and to mitigate the dependence on the initial conditions, we discard the first 500 iterations as burn-in and thin the remaining 4,500 iterations at an interval of five. This leaves 900 MCMC samples for each simulated panel and a total of 18,000 MCMC samples across the 20 different simulated panels.

We specify the following values for the hyperparameters in the prior distributions described in Section 3.2.3. We set $a_\beta = 2$, $b_\beta = 1$, $a_c = 1$, and $b_c = 1$ in the gamma priors on the IBP parameters. We set $a_\gamma = 10$, $b_\gamma = 1$, $a_\kappa = 10$, and $b_\kappa = 1$ in the gamma priors on the ZIB parameters. Finally, we set the hyperparameters in the hierarchical conjugate distribution on the regime-switching parameters as follows[3]: $\underline{\mu} = 0$, $\underline{h} = 100$; $\underline{d} = (2/1) * (s/0.0001)^2$, $\underline{c} = (2/1) * (s/0.0001)$ with s denoting the average sample variance of $Y_{1:T}$; $\underline{\lambda} = 1/1000$; $\underline{\nu} = (2/100^2) * (s/\underline{\lambda})^2$, $\underline{V} = (2/100^2) * (s/\underline{\lambda})$.

3.3.2 Simulation results

Table 3.1 illustrates, for all DGPs and different values of N and T, the estimation accuracy of the main variables and parameters. The error e_K evaluates the frequency of correctly estimating the true total number of regimes in the panel across experiments. The estimate of the number of regimes K is approximated as the posterior mode of the nonempty column dimension that the estimator of F exhibits through MCMC iterations. The errors e_F and $e_{S_{1:T}}$ evaluate the Hamming distance between their corresponding estimates and the truths. The estimate of the infinite-dimensional binary matrix F is approximated based on the posterior probability of each entry therein, which is computed as the average of each entry's MCMC samples. We set an entry of F equal to 0 if its posterior probability is less than 0.5 and to 1 otherwise. The estimate of the hidden regime sequences $S_{1:T}$ is approximated as the maximum a posteriori (MAP) estimate of their MCMC samples. The

[3]Given the specified hyperparameters and the priors in (3.12), the mean of μ_k is equal to 0 and its variance is equal to 100^2, which are rather uninformative. The mean of h_k is equal to $s/0.0001$ and its variance is equal to 1, which according to (3.11) implies $\mu_{i,k}$ for those units i that share the same regime k is concentrated on μ_k with a small variance around 0.0001. The mean of ν_k is equal to 1000, which according to (3.11) implies $\sigma_{i,k}^{-2}/\sigma_{j,k}^{-2}$ follows an F-distribution (i.e., $\sigma_{i,k}^{-2}/\sigma_{j,k}^{-2} \sim \mathrm{F}(1000, 1000)$, $i \neq j$) and the probability that the ratio $\sigma_{i,k}^{-2}/\sigma_{j,k}^{-2}$ is less than 0.9 or greater than 1 is about 0.1. Therefore, the ratio $\sigma_{i,k}^{-2}/\sigma_{j,k}^{-2}$ is close to 1, suggesting $\sigma_{i,k}^{-2}$ for those units i that share the same regime k are quite similar. The mean of V_k is equal to $s/\underline{\lambda}$ (i.e., $s * \mathrm{E}(\nu_k)$) and its variance is equal to 100^2, which according to (3.11) implies $\sigma_{i,k}^2$ is concentrated on s (i.e., the average sample variance) but is rather uninformative for different regimes k given a large variance around 100^2.

error e_θ evaluates the average mean square error (MSE) between the estimates and the truths for all regime-switching parameters $(\mu_{i,k}, \sigma_{i,k})$, where $k \in \mathcal{K}_i$ and $i = 1, ..., N$. The estimate of $\theta = (\mu_{i,k}, \sigma_{i,k})$ is approximated as the posterior means of their MCMC samples. In the table, the results reported for e_K are in percentage, and for e_F, e_θ, and $e_{S_{1:T}}$ are averages over the experiments that detect the true regime dimension of the panel.

Table 3.1 Summary of estimation errors

	e_K (%)			e_F			e_θ			$e_{S_{1:T}}$		
	$N = 25$	50	100	25	50	100	25	50	100	25	50	100
DGP 1 (HoF + WS)												
$T = 150$	100	95	85	0	0	0	4×10^{-5}	4×10^{-5}	4×10^{-5}	0.02	0.02	0.02
250	100	95	100	0	0	0	4×10^{-5}	4×10^{-5}	4×10^{-5}	0.02	0.02	0.02
500	100	95	100	0	0	0	3×10^{-5}	3×10^{-5}	3×10^{-5}	0.02	0.02	0.02
DGP 2 (HoF + WSSS)												
150	95	95	85	1×10^{-3}	0	0	1×10^{-4}	9×10^{-5}	8×10^{-5}	0.02	0.02	0.02
250	90	100	90	1×10^{-3}	3×10^{-4}	1×10^{-3}	9×10^{-5}	9×10^{-5}	5×10^{-5}	0.02	0.02	0.02
500	100	95	85	0	7×10^{-4}	2×10^{-4}	7×10^{-5}	5×10^{-5}	4×10^{-5}	0.02	0.02	0.02
DGP 3 (WHeF + WS)												
150	95	95	85	2×10^{-2}	1×10^{-2}	2×10^{-2}	3×10^{-4}	5×10^{-4}	2×10^{-4}	0.03	0.05	0.03
250	95	95	90	2×10^{-3}	2×10^{-3}	2×10^{-3}	4×10^{-5}	4×10^{-5}	3×10^{-5}	0.02	0.01	0.01
500	100	70	80	0	0	2×10^{-2}	3×10^{-5}	3×10^{-5}	2×10^{-4}	0.01	0.01	0.03
DGP 4 (WHeF + WSSS)												
150	100	95	100	8×10^{-3}	9×10^{-3}	9×10^{-3}	2×10^{-4}	1×10^{-4}	8×10^{-5}	0.01	0.02	0.01
250	90	100	90	7×10^{-4}	3×10^{-3}	3×10^{-3}	1×10^{-4}	6×10^{-5}	6×10^{-5}	0.01	0.01	0.01
500	75	55	30	0	0	0	8×10^{-5}	6×10^{-5}	4×10^{-5}	0.01	0.01	0.01
DGP 5 (SHeF + WS)												
150	95	90	85	2×10^{-2}	9×10^{-3}	7×10^{-3}	6×10^{-4}	4×10^{-5}	4×10^{-5}	0.05	0.01	0.01
250	95	90	85	0	4×10^{-4}	1×10^{-3}	4×10^{-5}	3×10^{-5}	3×10^{-5}	0.01	0.01	0.01
500	95	85	40	0	0	0	2×10^{-5}	2×10^{-5}	2×10^{-5}	0.01	0.01	0.01
DGP 6 (SHeF + WSSS)												
150	95	95	100	4×10^{-3}	7×10^{-3}	4×10^{-3}	1×10^{-4}	6×10^{-5}	6×10^{-5}	0.01	0.01	0.01
250	80	70	70	3×10^{-3}	1×10^{-3}	1×10^{-3}	2×10^{-4}	7×10^{-5}	5×10^{-5}	0.01	0.01	0.01
500	60	20	30	0	0	0	7×10^{-5}	4×10^{-5}	4×10^{-5}	0.01	0.01	0.01

NOTES: The table reports, for all DGPs and different values of N and T, the estimation accuracy of the main variables and parameters. The results reported for e_K are in percentage, and for e_F, e_θ, and $e_{S_{1:T}}$ are averages over the experiments that detect the true number of regimes.

We first examine how well the proposed framework performs in estimating the total number of regimes in the panel. For DGP 1 (HoF + WS) and DGP 2 (HeF + WS), our approach manages to detect the correct total number of regimes in more than 85% case, even with a relatively short time series ($T = 150$) and large panel ($N = 100$). This shows that when individuals are quite alike (i.e., panel exhibits homogeneous regime dimension in the cross-section), our method can efficiently estimate the regime dimension of the whole panel, no matter the size of cross-section, the length of times series, and the stability of individual regime-switching dynamics. In DGP 3 (WHeF + WS) and DGP 4 (WHeF + WSSS), the

correct detection rate is also high in most cases, although it drops quickly in large sample ($T = 500$) as N increases. Allowing a stronger level of cross-sectional regime heterogeneity in DGP 5 (SHeF + WS) leads to less accurate estimation of the overall regime dimension for large panel and sample ($N = 100$ and $T = 500$). Nevertheless, our approach still manages to maintain high correct detection rates (with correct detection rates of more than 85%) in other cases of N and T. Further, introducing more instability to individual regime-switching dynamics as in DGP 6 (SHeF + WSSS) makes the estimation accuracy more sensitive to N in large sample ($T = 500$). These results indicate that as long as the sample size is not too large, our approach continues to perform well in cases with slightly or very distinct individuals (i.e., panel exhibits weak or strong levels of cross-sectional heterogeneity in the regime dimension), regardless of how notable individual dynamics differ across regimes. Also in such cases, once the sample size grows, our approach suffers from an increasing loss of estimation accuracy[4] in larger panels, especially with unstable individual regime-switching dynamics. This is possibly because at the same level of cross-sectional heterogeneity in the regime dimension, the larger panels result in more individuals subject to likely heterogeneous regimes, the more unstable individual regime-switching dynamics lead those heterogeneous regimes to be more distinct from each other, and the longer sample means the heterogeneous regimes may appear more often and visible. These reasons jointly contribute to the difficulty in overlapping many individual-specific heterogeneous regimes, with the aim of finally delivering a sparse set of regimes that can be shared across the panel. For applications involving such panel data sets (i.e., large panel and sample, very different and unstable individual dynamics), two additional implementations based on our framework may be preferable: the user could either (i) specify the hyperparameters a_β, b_β, a_c, and b_c in the gamma priors on the IBP distribution at small values to favor more overlapping (or sharing) of regimes in the cross-section, as suggested in Griffiths and Ghahramani (2011), or (ii) augment our MCMC procedure with the split-merge Metropolis-Hasting algorithm (see, e.g., Smith, 2023) that allows large changes in the regime assignment across individuals at each iteration.

We now examine the accuracy of the parameter estimates obtained from our framework. When the total number of regimes in the panel is correctly estimated, we observe that the errors e_F, e_θ, and $e_{S_{1:T}}$, in every case of N and T, are equal or close to zero under all DGPs considered. This indicates that our framework is able to precisely estimate the regime pattern (i.e., the regime dimension for the individuals) and the unit-specific regime-switching dynamics.

In Appendix 3.B, we also evaluate the convergence and stability of the regime dimension estimators K and F for those successful experiments that detect the true regime dimension of the panel. We provide the MCMC trace plots of the estimated number of regimes K and of the estimation error e_F under all DGPs with different values of N and T. In general, the

[4]Unreported results suggest that when our approach fails to estimate the true number of regimes of the whole panel in DGPs 3-6, it typically overestimates them by one extra regime.

regime dimension estimators K and F are converged since the median values (red line) of K and e_F obtained from different experiments remain respectively at the true value $K = 3$ and $e_F = 0$ across MCMC iterations. Moreover, the estimators K and F are generally stable since the 10th and 90th quantile (grey line) values of K and e_F obtained from different experiments sweep through small regions[5] around the true value $K = 3$ and $e_F = 0$ across MCMC iterations.

3.4 Empirical illustration

MS panel models have been widely used in analyzing business cycles. Under an assumption of homogeneous regime dimension in the cross-section, normally two regimes with interpretations of expansion and recession are determined for business cycles at the region or global level (see, e.g., Casarin et al., 2018; Owyang et al., 2022; Agudze et al., 2022). In this section, we revisit the dataset analyzed in Agudze et al. (2022), which includes a large panel of US state-level macroeconomic indices. Our aim is to illustrate the incremental value added for business cycle analysis by relaxing the assumption of cross-sectional homogeneous regime dimension in the MS panel model. To this aim, some in-sample posterior estimates from our BNP MS panel framework with cross-sectional regime heterogeneity compared with those from the framework without this feature (see, e.g., Hou, 2017; Casarin et al., 2024) are first analyzed. We further compare out-of-sample forecasts produced by the above two frameworks along with those generated by a standard two-regime MS panel model and a BNP MS univariate framework applied separately to each unit in the cross-section. Benchmark model specifications are detailed in Appendix 3.C.

3.4.1 Data and implementation

Our sample consists of $N = 50$ states in the US. We download monthly state-level coincident indices from the Federal Reserve Bank of Philadelphia (FED)[6] between October 1979 and October 2023, implying $T = 529$. Following Agudze et al. (2022), we transform all indices into stationary growth rates by the log-first-difference. Data details regarding state name, abbreviation, and summary statistics are provided in Appendix 3.C. The proposed framework is implemented with the same prior hyperparameter values and MCMC samples specified in the simulation study.

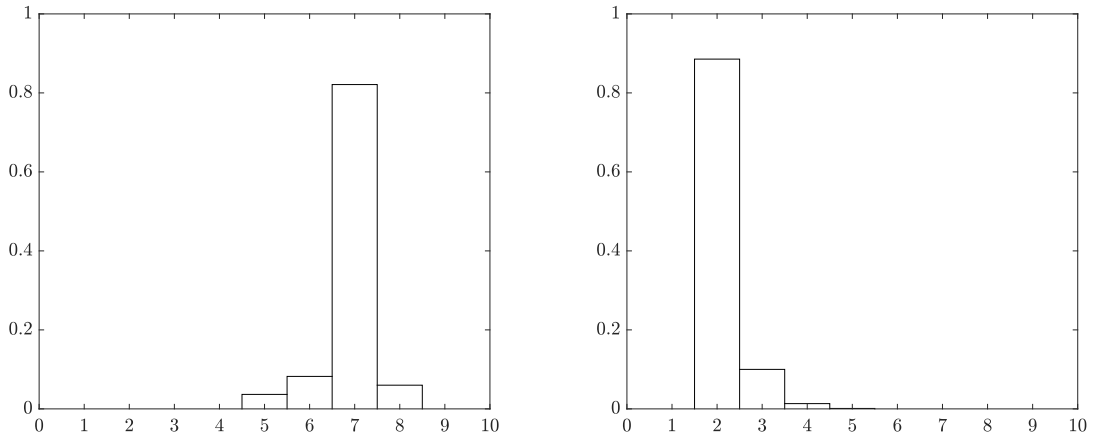
3.4.2 In-sample estimations

We start by addressing the problem of determining the number of regimes for the MS panel model. Under the assumption of cross-sectional homogeneous regime dimension, the

[5] We see that K is overestimated by only 1 or 2 and e_F deviates zero by no more than 0.4.

[6] Data are available at https://philadelphiafed.org/research-and-data/regional-economy/indexes/coincident.

problem of determining the number of regimes for an MS panel model is treated as the problem of knowing the nonempty dimension of an infinite-dimensional transition matrix P. Under our assumption of cross-sectional heterogeneous regime dimension, the problem of determining the number of regimes for an MS panel model is recast as the problem of knowing the nonempty column dimension and the configuration of an infinite-dimensional binary matrix F. Figure 3.3 plots the posterior distribution of the regime dimension for the whole panel (i.e., K) inferred respectively from the nonempty column dimension of F in our BNP MS panel framework (plot (a)) and the nonempty dimension of P in the BNP MS panel framework assuming cross-sectional homogeneous regime dimension (plot (b)). The figure shows that both methods take the uncertainty of regime dimension into account. Moreover, it is obvious that the shapes of these posterior distributions are quite different from each other. The posterior mode under our framework peaks at $K = 7$ regimes while that under the benchmark framework points to $K = 2$ regimes. As the BNP MS panel framework assuming cross-sectional homogeneous regime dimension treats the dynamics of all individuals equally, it is not surprising that it tends to use less regimes to characterize the panel dynamics. The potential harm of ignoring the heterogeneous regimes is that those regimes occurring to only a part of units may be averaged out, thus acting against the correct estimation of panel's regime dimension.



(a) BNP MS panel with regime heterogeneity **(b)** BNP MS panel with regime homogeneity

Figure 3.3 Posterior distribution of the regime dimension for the panel.

Given the identified seven regimes in the panel, the configuration of binary matrix F is shown as an $N \times K$ heatmap in Figure 3.4. The results imply that our assumption about heterogeneous regime dimension in the cross-section is plausible empirically. Specifically, about 25% of US states possess five of the identified seven regimes in the panel, supporting our assumption that the regime dimension for the individuals can be less than the regime dimension for the whole panel. Compared with those owning five regimes, the remaining US

states possess all seven regimes, corroborating our assumption that the regime dimension for the individuals may exhibit cross-sectional heterogeneity. Besides, the US states exhibiting

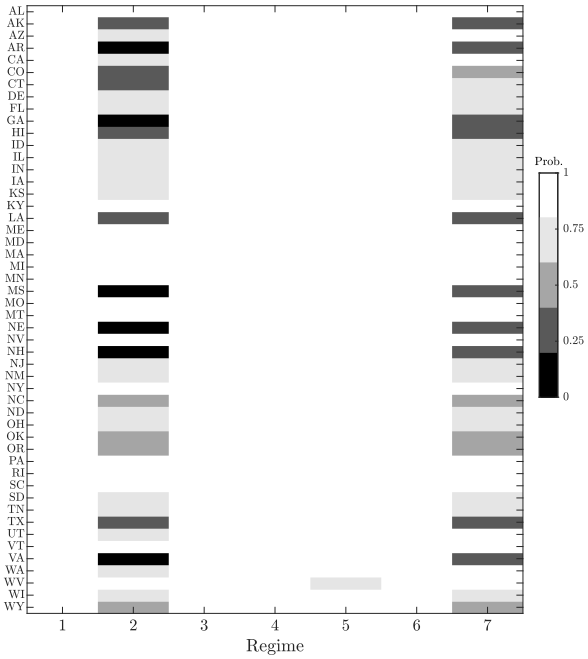
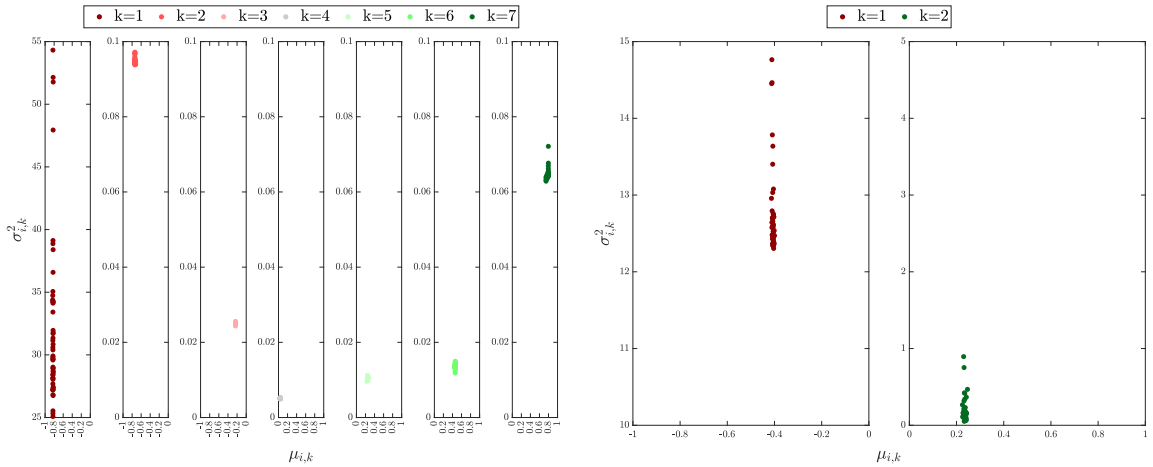


Figure 3.4 Heatmap of the estimated binary matrix. The columns are regimes in the panel and the rows are 50 US states. The value of each cell is the posterior probability, which is obtained by averaging the posterior estimates of each entry in F across MCMC samples. Dark and light cells represent 0 (absent regime) and 1 (present regime), respectively.

five regimes appear to own the same set of regimes, including all except regimes 2-7. Those US states therefore share the same set of dynamics over the entire sample period, and behave unlike others exhibiting seven regimes. The BNP MS panel framework assuming cross-sectional homogeneous regime dimension lacks the flexibility to capture these features.

We proceed by evaluating the parameter estimates under different assumptions of regime dimension. Figure 3.5 displays the scatterplots of the posterior means of the regime-switching parameters (i.e., $\mu_{i,k}, \sigma^2_{i,k}$) for each regime identified from our framework (plot (a)) and the benchmark framework (plot (b)). Focusing on plot (a), it is evident that the seven regimes differ mainly in the regime-specific means whereas the regime-specific variances are quite similar for most regimes except regime 1. This indicates that the identified regimes are generally well separated for the mean dynamics they characterize and validates our regime identification restriction on means. Turning to plot(b), it is obvious that the BNP MS panel framework that ignores regime dimension heterogeneity results in parameters being pooled across all units in every identified regime. This inability to allow the degree of pooling to vary across regimes leads to imprecise parameter estimates if some units behave unlike any other.



(a) BNP MS panel with regime heterogeneity **(b)** BNP MS panel with regime homogeneity

Figure 3.5 Scatterplots of the posterior means of the regime-switching parameters for each of the identified regimes.

We further investigate how data are categorized into different regimes over time by plotting an $N \times T$ heatmap in Figure 3.6. For a comparative assessment, the figure also displays the NBER-referenced chronology, which serves as the standard for business cycle dates. It is worth noting that our framework locates those regimes with negative means (i.e., regimes 1-3) exactly around the recession periods according to the NBER dating. This illustrates the ability of our framework to provide accurate inferences about the recession turning points that agree with the NBER-designated cycle dates. In addition, Figure 3.6 confirms our intuition in Section 3.2.2 that the empirical gain brought by our framework comes from the identification of heterogeneity and non-synchronization in regimes. As for the heterogeneity in regimes, the figure suggests that many oil-producing states (e.g., Louisiana-LA, Oklahoma-OK, Texas-TX, and Wyoming-WY) have never entered regime 2 and regime 7, with the former exploding considerably during the crisis period (e.g., the 1980, 1981-1982 national recessions and the Global Financial Crisis) and the latter occurring mostly during the subsequent recovery period after crises. This finding, consistent with the evidence in Baumeister et al. (2024), points out the noncommon dynamics of the regional economy. Regarding the non-synchronization in regimes, the figure shows that most recessions correspond to non-synchronized regimes (i.e., regimes 2-3) that impact US states with different delays, with the COVID-19 pandemic being a notable exception where almost all states are hit around the same time. This result, in line with the observation in Owyang et al. (2005), distinguishes the synchronization level of regional economic downturn. Furthermore, the magnitude of the identified regimes (i.e., the regime-specific means) accurately reflects the severity of the different recessionary episodes, with the 1990-1991 and 2001 recessions being milder compared to the Global Financial Crisis and the COVID-19 recession. Relative to our framework, the BNP MS panel framework assuming cross-sectional homogeneous regime dimension treats the regime pattern of all units

no differently (see Figure 3.1 in Section 3.2.2) and thus lacks the flexibility to capture the noncommon and non-synchronized dynamics of individuals.

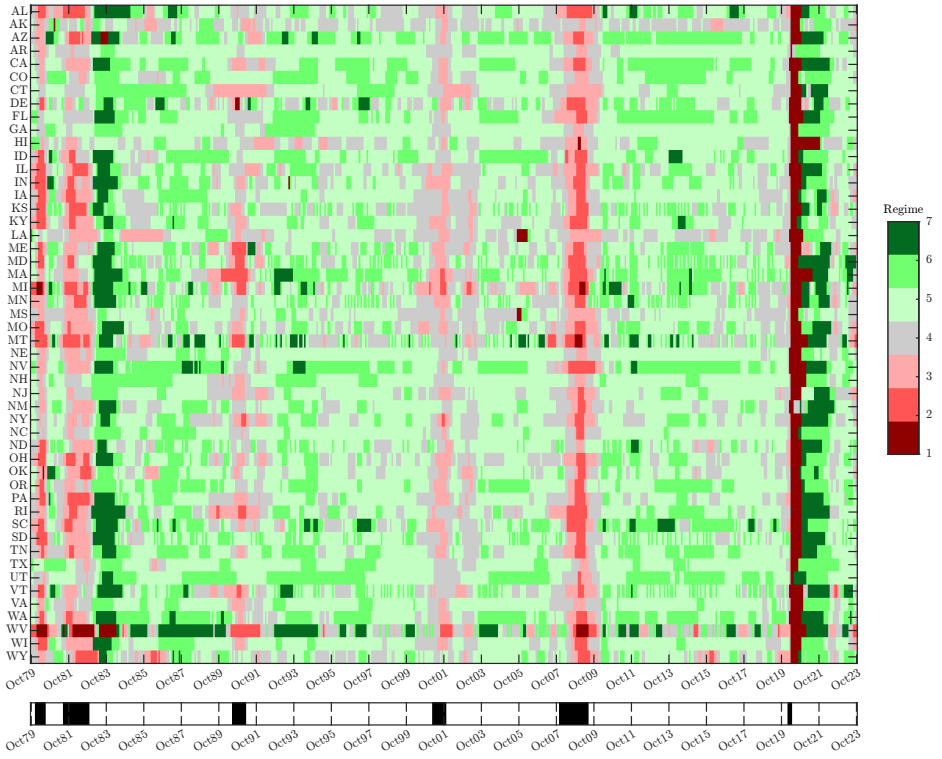


Figure 3.6 Estimated regimes over time. The top heatmap displays the maximum a posteriori (MAP) estimates of the regime indicators s_{it} of the 50 US states. Seven levels of colors indicate periods of the identified seven regimes, with red, green, and gray colors representing negative, positive, and zero valued means the regimes characterize, i.e., μ_k, for $k = 1, ..., 7$. The bottom heatmap displays the NBER-referenced business cycle dates, with black (white) color indicating periods of recession (expansion).

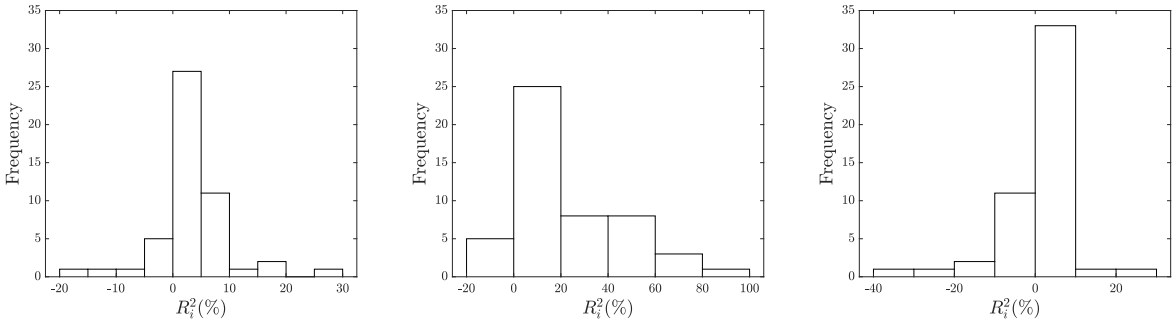
3.4.3 Out-of-sample evaluation

We consider a recursive forecasting exercise to remain close to the majority of the literature (see, e.g., Bauwens et al., 2017; Hou, 2017; Luo et al., 2022). About 2/3 sample which covers a total of 352 months is selected as the in-sample period and the remaining 1/3 sample which covers 177 months is employed for the out-of-sample evaluations. We recursively move through the sample one month at a time generating forecasts at each point using only the data available at the time the forecast is made.

The comparative forecasting performance of our framework relative to each of the benchmarks is measured for the i-th US state by the popular out-of-sample R_i^2 value

$$R_i^2 = 1 - \text{MSE}_{i,HeF}/\text{MSE}_{i,Bmk}, \tag{3.15}$$

in which $\text{MSE}_{i,HeF}$ and $\text{MSE}_{i,Bmk}$ denote the mean squared forecast errors for the i-th US state obtained from our BNP MS panel framework assuming cross-sectional heterogeneous regime dimension and the given benchmark framework. Positive R_i^2 values indicate our framework is outperforming the benchmark, while negative values reflect underperformance. For ease of illustration, all reported R_i^2 values are multiplied by 100 as in Smith (2023). Each plot in Figure 3.7 corresponds to a different benchmark and displays the R_i^2 relative to that benchmark across the 50 US states. Our framework delivers positive R_i^2 values for 42 of the 50 cases relative to the BNP MS panel framework assuming cross-sectional homogeneous regime dimension (plot (a)). Compared with the two-regime MS panel model (plot (b)), equally strong outperformance is observed as 45 cases have positive R_i^2. Relative to the BNP MS univariate framework (plot (c)), our panel framework outperforms a majority of times, delivering positive R_i^2 values for 35 cases.



(a) BNP MS panel with regime homogeneity

(b) Two-regime MS panel

(c) BNP MS univariate

Figure 3.7 R_i^2 values. Each plot displays the R_i^2 values obtained from comparing the forecasting performance of our proposed framework (BNP MS panel with regime heterogeneity) with the benchmark labeled in the subcaption for each of the 50 US states.

To evaluate the statistical significance of any out- or under-performance delivered by our framework, we use the test statistic of Clark and West (2007) which adjusts for the Diebold–Mariano test-statistic (Diebold and Mariano, 1995) possibly having a nonstandard distribution as a result of nested frameworks. The results are displayed in Table 3.2. The left panel of the table displays the number of the US states for which our methodology significantly underperforms ($-*$), insignificantly underperforms ($-$), insignificantly outperforms ($+$), or significantly outperforms ($+*$) the benchmark according to the Clark-West (CW) test. Each row corresponds to a different benchmark. In general, our approach never significantly underperforms relative to all three benchmarks. Across all 150 cases (50 cases against each of the three benchmarks), our framework outperforms (both insignificantly and significantly) 145 times, and especially outperforms at a significant level for 97 out of the 150 cases.

Aside from the point forecasts, we also evaluate the performances of density forecasts using the continuously ranked probability score (CRPS) popularized by Gneiting and Raftery

Table 3.2 Statistical significance of pairwise forecast comparisons

CW				AG			
$-*$	$-$	$+$	$+*$	$-*$	$-$	$+$	$+*$
BNP MS panel with regime homogeneity							
0	0	10	40	0	6	14	30
Two-regime MS panel							
0	0	17	33	0	5	8	37
BNP MS univariate							
0	5	21	24	0	7	18	25

NOTES: This table reports the statistical significance from pairwise MSE (left panel) and CRPS (right panel) comparisons of forecasts from our proposed model (BNP MS panel with regime heterogeneity) versus forecasts from a range of benchmarks listed in each row of the table. Significance is evaluated using the Clark-West (CW) and Amisano-Giacomoni (AG) tests. For each test the table displays the number of the 50 US states for which our method produces significantly worse ($-*$), insignificantly worse ($-$), insignificantly better ($+$), and significantly better ($+*$) forecasts at the 5% level.

(2007). Let $\text{CRPS}_{it,Bmk}$ denote the corresponding score for the benchmark model, our approach yields better (worse) predictive accuracy for the i-th US states at each out-of-sample time point t relative to the benchmark model if $\Delta L_{it} = (\text{CRPS}_{it,Bmk} - \text{CRPS}_{it,HeF})$ is positive (negative). To test whether any improved CRPS performance ΔL_{it} generated by our methodology relative to the benchmark is significant, we use the test statistic[7] of Amisano and Giacomini (2007). The Amisano-Giacomoni (AG) statistic in the right panel of Table 3.2 shows that the statistical significance of outperformance in terms of the density forecasts is almost as strong relative to the results of point forecasts, with our methodology outperforming (both insignificantly and significantly) 132 times, significantly outperforming for 92 out of the 150 cases, and never significantly underperforming.

3.5 Conclusions

Markov-switching panel models face a major challenge in practical implementation, which is determining how many regimes are necessary to adequately characterize the observed data. Existing solutions typically rely on the assumption that the regime dimension is homogeneous in the cross-section. Such an assumption may be restrictive as individuals are likely to be characterized only by one or a subset of regimes identified from the panel. This paper proposed a general framework to estimate the number of regimes in Markov-switching panel models, allowing possible heterogeneity in cross-sectional regime dimension. Within the proposed framework, the unknown regime dimension for the whole panel and the units was inferred

[7] We compute the unweighted version of the Amisano-Giacomoni test. The test statistic is $W_i = (T-m)((T-m)^{-1}\sum_{t=m+1}^{T} h_t \Delta L_{it})' \Omega^{-1} ((T-m)^{-1}\sum_{t=m+1}^{T} h_t \Delta L_{it})$ in which $m = 352$ denotes the in-sample period, $h_t = 1$ for $t = m+1, ..., T$, Ω is an appropriate heteroscedasticity and autocorrelation consistent estimate of the covariance matrix. The null hypothesis is that the two approaches have equal predictive accuracy under which the distribution of W_i tends to χ^2_1 as $(T-m) \to \infty$.

respectively from the column dimension and the configuration of an infinite-dimensional binary matrix. Methodologically, we developed new Bayesian nonparametric inference to jointly estimate the latent binary matrix and the other parameters in Markov-switching panel models. A nonparametric prior based on an Indian buffet process was specified for the infinite-dimensional binary matrix. A set of mutually independent zero-inflated Beta variables was employed for constructing a prior on the regime transition matrices restricted by unit-specific present regimes. A hierarchical conjugate prior was placed over the regime-switching parameters, which allows units within the same regime to share similar dynamics and helps derive the data marginal likelihood that improves posterior computation efficiency. Estimation was accomplished by an efficient Markov chain Monte Carlo procedure, which exploits a combination of multi-move Gibbs, Metropolis-Hasting and reversible jump samplers. Simulation studies were conducted to validate the effectiveness of the proposed framework in recovering the total number of regimes, the regime patterns, and the regime-switching dynamics under many empirically relevant panel settings. An empirical application to 50 US state-level macroeconomic indices over the period 1979-2023 was offered for in-sample evaluation and forecasting purposes. The results showed that assuming likely cross-sectional heterogeneous regime dimension in Markov-switching panel models is helpful in revealing the noncommon and non-synchronized dynamics of panel units, and in generating significantly more accurate forecasts than a range of popular benchmarks.

Appendix 3.A

This appendix provides the computational details of the Bayesian nonparametric inference procedure. Our BNP MS panel framework is estimated using a Markov chain Monte Carlo (MCMC) procedure. There are six estimation steps. In Step 1, the infinite-dimensional binary matrix F is estimated using the Metropolis-Hasting and reversible jump methods. In Step 2, the hidden regime sequences $S_{1:T}$ are estimated using a forward-filtering backward-sampling (FFBS) algorithm. In Steps 3-4, the regime transition matrices P and the regime-specific parameters Θ are estimated from their full conditional distributions. In Steps 5-6, the IBP parameters Θ_{IBP} and the ZIB parameters Θ_{ZIB} are estimated using a Metropolis-Hasting method. We now explain each step in detail.

3.A.1 Sampling F

We sample the infinite-dimensional binary matrix F row-by-row. That is, we sample the binary vector $\boldsymbol{f}_i$ for $i = 1, ..., N$. For $\boldsymbol{f}_i$, we sample the entry f_{ik} separately for those regime k unique to unit i and those regime k owned by any other units in the panel. We define the set of unique regimes for unit i to be the regimes that only i possesses; we use the notation $\mathcal{U}_i$. In other words, $\mathcal{U}_i$ is the column indices for those columns in F with all entries of zero but a entry of one on the i-th row. We define the set of shared regimes for unit i to be the regimes that are owned by any other units in the panel; we use the notation $\mathcal{S}_i$. In other words, $\mathcal{S}_i$ is the column indices obtained by excluding $\mathcal{U}_i$ from $\mathcal{M}_N$, which is defined in Section 3.2.3.1 of the paper and corresponds to the indices for those columns in F with at least one nonzero entry. Taking Figure 3.2 in the paper as an example, we can see $M_5 = 8$, $\mathcal{M}_5 = \{1,2,3,4,5,6,7,8\}$, and for $i = 2$: $\mathcal{U}_2 = \{5\}$, $\mathcal{S}_2 = \{1,2,3,4,6,7,8\}$.

Shared regimes. We sample the binary indicator f_{ik} for each shared regimes $k \in \mathcal{S}_i$ from its full conditional distribution, which is defined as

$$p(f_{ik}|F_{-ik}, \Theta, \Theta_{\text{IBP}}, P, Y_{1:T}) \propto p(f_{ik}|F_{-ik}, \Theta_{\text{IBP}}) p(y_{i,1:T}|\boldsymbol{f}_i, \Theta, P_i), \tag{3.A.1}$$

where F_{-ik} denotes all the entries of F excluding f_{ik}. According to the IBP prior which indicates $f_{ik} \sim \text{Bern}(\frac{N_{i-1,k}}{c+i-1})$, the first term on the right hand side of (3.A.1) is calculated as

$$p(f_{ik}|F_{-ik}, \Theta_{\text{IBP}}) = f_{ik}\frac{N_{k,-i}}{c+N-1} + (1-f_{ik})(1-\frac{N_{k,-i}}{c+N-1}), \tag{3.A.2}$$

where $N_{k,-i}$ denotes the number of units other than i possessing regime k, which is equal to $N_k - 1$ if $f_{ik} = 1$ and N_k if $f_{ik} = 0$.

The second term on the right hand side of (3.A.1) is derived from (3.A.11)

$$p(y_{i,1:T}|\boldsymbol{f}_i,\Theta,P_i)=\prod_{t=1}^{T}p(y_{it}|y_{i,1:t-1},\boldsymbol{f}_i,\Theta,P_i), \tag{3.A.3}$$

As (3.A.1) is not conditionally conjugate, we adopt a Metropolis–Hastings method to evaluate a new proposal which flips the current indicator value f_{ik} to its binary complement $\tilde{f}_{ik}=1-f_{ik}$. We accept the proposal $\tilde{f}_{ik}$ with probability

$$\min\left\{1,\frac{p(\tilde{f}_{ik}|F_{-ik},\Theta,\Theta_{\text{IBP}},P,Y_{1:T})}{p(f_{ik}|F_{-ik},\Theta,\Theta_{\text{IBP}},P,Y_{1:T})}\right\}. \tag{3.A.4}$$

Unique regimes. We sample the binary indicator f_{ik} for each unique regimes $k\in\mathcal{U}_i$ using a birth and death reversible jump MCMC sampler (Green, 1995), where the birth proposal means that we introduce a new unique regime, and the death proposal means that we delete an existing unique regime. We first make a random choice between birth and death. Let $\text{card}(\mathcal{U}_i)$ denote the number of unique features for unit i. If $\text{card}(\mathcal{U}_i)$ is zero (i.e., no unique regimes), we always choose a birth and thus the function that determines the probability of choosing a birth conditional on the set $\mathcal{U}_i$ is $p_{\text{birth}}(\mathcal{U}_i)=1$. Otherwise, we either choose a birth with probability $p_{\text{birth}}(\mathcal{U}_i)=0.5$, or a death of a member of $\mathcal{U}_i$ with equal probability $(1-p_{\text{birth}}(\mathcal{U}_i))/\text{card}(\mathcal{U}_i)$.

If a birth is chosen, a new unique regime $\tilde{k}=M_N+1$ is introduced for unit i. This leads to new proposals for the binary vector and the regime sequence: for $\boldsymbol{f}_i$, the newly proposed one is $\tilde{\boldsymbol{f}}_i=\{f_{i1},f_{i2},...,f_{iM_N},f_{iM_N+1}=1\}$. We denote this proposal density by $q(\tilde{\boldsymbol{f}}_i,\boldsymbol{f}_i)$; for $s_{i,1:T}$, the newly proposed one is $\tilde{s}_{i,1:T}$ which needs to evolve across the regimes indicated by the newly proposed $\tilde{\boldsymbol{f}}_i$. We denote this proposal density by $q(\tilde{s}_{i,1:T},s_{i,1:T})$. To construct the proposed new regime sequence $\tilde{s}_{i,1:T}$, we first define the new regime transition matrix using the Beta variables with values setting to their prior means. Besides, for existing regimes, we set the regime-specific means and the variances to their posterior means. For the new regime $\tilde{k}$, we define its associated mean $\mu_{i,\tilde{k}}$ and variance $\sigma^2_{i,\tilde{k}}$ based on a data-driven approach similar as in Fox et al. (2014). This requires us to choose a random subwindow W of the current unit i. Next, given the chosen window, we set $\mu_{i,\tilde{k}}$ and $\sigma^2_{i,\tilde{k}}$ to their posterior means given the data in the window $\{y_{it}:t\in W\}$. Finally, given the new regime transition matrix and regime-specific parameters, we sample the proposal $\tilde{s}_{i,1:T}$ based on Section 3.A.2. We adopt a Metropolis–Hastings method to decide whether the proposed configuration $\tilde{\boldsymbol{f}}_i$ and $\tilde{s}_{i,1:T}$ is accepted or rejected. We accept the proposal with probability

$$\min\left\{1,\frac{p(Y_{1:T},\tilde{F},\tilde{S}_{1:T})q(s_{i,1:T},\tilde{s}_{i,1:T})q(\boldsymbol{f}_i,\tilde{\boldsymbol{f}}_i)}{p(Y_{1:T},F,S_{1:T})q(\tilde{s}_{i,1:T},s_{i,1:T})q(\tilde{\boldsymbol{f}}_i,\boldsymbol{f}_i)}\right\}. \tag{3.A.5}$$

where the proposal density $q(\tilde{s}_{i,1:T}, s_{i,1:T})$ and the corresponding reverse proposal density $q(s_{i,1:T}, \tilde{s}_{i,1:T})$ is obtained by summing the quantity in (3.A.12) for $\tilde{s}_{it}$ and s_{it} respectively over T; the proposal density is $q(\tilde{\boldsymbol{f}}_i, \boldsymbol{f}_i) = p_{\text{birth}}(\mathcal{U}_i)$ and the reverse proposal density (i.e., death at the new regime $\tilde{k}$) is $q(\boldsymbol{f}_i, \tilde{\boldsymbol{f}}_i) = (1 - p_{\text{birth}}(\tilde{\mathcal{U}}_i))/\text{card}(\tilde{\mathcal{U}}_i)$ for $\tilde{\mathcal{U}} = \mathcal{U}_i \cup \{\tilde{k}\}$; the joint posterior distribution of data observations, binary matrix and regime sequences is given by

$$\begin{aligned} p(Y_{1:T}, F, S_{1:T}) &= p(Y_{1:T}|S_{1:T})p(F)p(S_{1:T}|F), \\ &= \int_{\Theta} p(Y_{1:T}|\Theta)p(\Theta)d\Theta \, p(F) \int_{P} p(S_{1:T}|P)p(P|F)dP \end{aligned} \tag{3.A.6}$$

Since the prior distributions of Θ enjoy conjugacy, the first integral term on the right hand side of (3.A.6) reduces to a closed-form function

$$\begin{aligned} \int_{\Theta} p(Y_{1:T}|\Theta)p(\Theta)d\Theta &= \prod_{i=1}^{N} \prod_{k \in \mathcal{K}_i} \int \int p(y_{i,k}|\mu_{i,k}, \sigma_{i,k}^{-2}) p(\mu_{i,k}) p(\sigma_{i,k}^{-2}) d\mu_{i,k} d\sigma_{i,k}^{-2} \\ &= \prod_{i=1}^{N} \prod_{k \in \mathcal{K}_i} \frac{\Gamma(\bar{\nu}_k/2)}{\Gamma(\nu_k/2)} \frac{(V_k/2)^{\nu_k/2}}{(\bar{V}_k/2)^{\bar{\nu}_k/2}} (\frac{h_k}{\bar{h}_{i,k}})^{1/2} (2\pi)^{-T_{i,k}/2}, \end{aligned} \tag{3.A.7}$$

where parameters are defined in Section 3.A.4. The second term on the right hand side of (3.A.6) is the probability mass function of IBP defined in Section 3.2.3.1 of the paper. The last integral term on the right hand side of (3.A.6) due to the conjugacy of the Dirichlet-multinomial distribution reduces to a closed-form function

$$\begin{aligned} \int_{P} p(S_{1:T}|P)p(P|F)dP &= \prod_{i=1}^{N} \int p(s_{i,1:T}|P_i)p(P_i|f_i)dP_i \\ &= \prod_{i=1}^{N} \frac{\Gamma(\sum_k \alpha_k)}{\Gamma(\sum_k (\sum_l n_{i,lk} + \alpha_k))} \prod_{k \in \mathcal{K}_i} \frac{\Gamma(\sum_l n_{i,lk} + \alpha_k)}{\Gamma(\alpha_k)}, \end{aligned} \tag{3.A.8}$$

where parameters are defined in Section 3.A.3.

If a death is chosen, a randomly chosen regime $k \in \mathcal{U}_i$ is deleted for unit i. This leads to new proposals for the binary vector and the regime sequence: for $\boldsymbol{f}_i$, the newly proposed one is $\tilde{\boldsymbol{f}}_i = \{f_{i1}, f_{i2}, ..., f_{ik} = 0, ..., f_{iM_N}\}$. We denote this proposal density by $q(\tilde{\boldsymbol{f}}_i, \boldsymbol{f}_i)$; for $s_{i,1:T}$, the newly proposed one is $\tilde{s}_{i,1:T}$ which needs to evolve across the regimes indicated by the newly proposed $\tilde{\boldsymbol{f}}_i$. We denote this proposal density by $q(\tilde{s}_{i,1:T}, s_{i,1:T})$. The proposed new regime sequence $\tilde{s}_{i,1:T}$ is constructed as in the birth proposal but here only for the reduced set of exsiting regimes indicated by $\tilde{\boldsymbol{f}}_i$. We adopt a Metropolis–Hastings method to decide whether the proposed configuration $\tilde{\boldsymbol{f}}_i$ and $\tilde{s}_{i,1:T}$ is accepted or rejected. We accept the proposal with probability given in (3.A.5) but here the proposal density is $q(\tilde{\boldsymbol{f}}_i, \boldsymbol{f}_i) = (1 - p_{\text{birth}}(\mathcal{U}_i))/\text{card}(\mathcal{U}_i)$ and the reverse proposal density (i.e., birth at the deleted regime k) is $q(\boldsymbol{f}_i, \tilde{\boldsymbol{f}}_i) = p_{\text{birth}}(\tilde{\mathcal{U}}_i)$ for $\tilde{\mathcal{U}} = \mathcal{U}_i - \{k\}$.

3.A.2 Sampling $S_{1:T}$

For each unit-specific regime sequence $s_{i,1:T}$ contained in $S_{1:T}$, we either set $s_{i,1:T} = \{k : f_{ik} = 1\}$ when $K_i = 1$ or sample its whole path from the full conditional distribution based on a forward-filtering backward-sampling (FFBS) algorithm (see Frühwirth-Schnatter, 2006, chap. 11.2) when $1 < K_i \leq K$. The FFBS algorithm consists of two stages. In the first stage, we carry out a forward recursion to obtain the filtered probability. In the second stage, we compute the full conditional density of $s_{i,1:T}$ using the filtered probabilities, and randomly draw $s_{i,1:T}$ from its full conditional distribution through a backward recursion. The stages of the FFBS algorithm are described in the following. First, the filtered probability at time t, $t = 1,...,T$, is determined by implementing the forward recursion

$$p(s_{it} = l|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i) = \sum_{k \in \mathcal{K}_i} p_{i,kl} p(s_{it-1} = k|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i), \tag{3.A.9}$$

for $l \in \mathcal{K}_i$, where $p_{i,kl}$ is the regime transition probability of unit i from regime k at time $t-1$ to regime l at time t. We initialize for $t = 1$, $P(s_{i0} = k|y_{i,0}, \boldsymbol{f}_i, \Theta, P_i)$ to be equal to the ergodic probabilities. The filtered probability for all $l \in \mathcal{K}_i$ is computed as

$$\begin{aligned} p(s_{it} = l|y_{i,1:t}, \boldsymbol{f}_i, \Theta, P_i) &= \frac{p(y_{it}|s_{it} = l, y_{i,1:t-1}, \Theta) p(s_{it} = l|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i)}{p(y_{it}|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i)}, \\ &= \frac{\mathcal{N}(\mu_{i,l}, \sigma^2_{i,l}) p(s_{it} = l|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i)}{p(y_{it}|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i)} \end{aligned} \tag{3.A.10}$$

where the denominator is the marginal predictive likelihood defined as

$$\begin{aligned} p(y_{it}|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i) &= \sum_{l \in \mathcal{K}_i} p(y_{it}|s_{it} = l, y_{i,1:t-1}, \Theta) p(s_{it} = l|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i) \\ &= \sum_{l \in \mathcal{K}_i} \mathcal{N}(\mu_{i,l}, \sigma^2_{i,l}) p(s_{it} = l|y_{i,1:t-1}, \boldsymbol{f}_i, \Theta, P_i) \end{aligned} \tag{3.A.11}$$

The smoothing probabilities are obtained recursively and backward in time, once all the filtered probabilities $p(s_{it} = l|y_{i,1:t}, \boldsymbol{f}_i, \Theta, P_i)$ for $t = 1,...,T$ are calculated. If $t = T$, smoothing probability and filtered probability are equal. For $t = T-1, T-2, ..., 1$ and for all $k \in \mathcal{K}_i$ the backward recursion proceeds as follows

$$\begin{aligned} p(s_{it} = k|y_{i,1:T}, \boldsymbol{f}_i, \Theta, P_i) &= \sum_{l \in \mathcal{K}_i} p(s_{it} = k, s_{it+1} = l|y_{i,1:T}, \boldsymbol{f}_i, \Theta, P_i) \\ &= \sum_{l \in \mathcal{K}_i} p(s_{it} = k|s_{it+1} = l, y_{i,1:t}, \boldsymbol{f}_i, \Theta, P_i) p(s_{it+1} = l|y_{i,1:T}, \boldsymbol{f}_i, \Theta, P_i) \end{aligned} \tag{3.A.12}$$

where

$$p(s_{it}=k|s_{it+1}=l,y_{i,1:t},\boldsymbol{f}_i,\Theta,P_i)=\frac{p_{i,kl}p(s_{it}=k|y_{i,1:t},\boldsymbol{f}_i,\Theta,P_i)}{\sum_{j\in\mathcal{K}_i}p_{i,jl}p(s_{it}=j|y_{i,1:t},\boldsymbol{f}_i,\Theta,P_i)}.$$

3.A.3 Sampling P

For each unit-specific regime transition matrix P_i contained in P, we either set P_i equal to one when $K_i=1$ or sample P_i row-by-row based on a set of Beta variables $\boldsymbol{z}_{i,k}=(z_{i,k1},....,z_{i,kK-1})$ for $k\in\mathcal{K}_i$ when $1<K_i\leq K$. Because regimes not present in $\boldsymbol{f}_i$ have zero probability of occurring, we only need to sample the entries of $\boldsymbol{z}_{i,k}$ that correspond to regimes l satisfying $f_{il}=1$. The full conditional distribution of those entries according to the prior in Section 3.2.3.2 of the paper is defined as

$$p(z_{i,kl}|f_{il}=1,s_{i,1:T})\propto\text{Beta}(\alpha_l+n_{i,kl},\sum_{m=l+1}^{\max(\mathcal{K}_i)}\alpha_m+\sum_{m=l+1}^{\max(\mathcal{K}_i)}n_{i,km}),\tag{3.A.13}$$

where $\alpha_k=\gamma+\kappa\delta_k$ defined in Section 3.2.3.2 of the paper; $n_{i,kl}=\sum_{t=1}^{T}\mathbb{I}(s_{it-1}=k)\mathbb{I}(s_{it}=l)$ counts the transitions from regime k to l. Finally, we transform $\boldsymbol{z}_{i,k}$ to $\boldsymbol{p}_{i,k}$ using (3.9) in the paper.

3.A.4 Sampling Θ

We sample $\sigma_{i,k}^{-2}$ for $i=1.,,,.N$ and $k\in\mathcal{K}_i$ from its full conditional distribution, which according to the prior in Section 3.2.3.3 of the paper is defined as

$$p(\sigma_{i,k}^{-2}|\Theta_{-\sigma_{i,k}^{-2}},s_{i,1:T},y_{i,1:T})\propto\mathcal{G}(\frac{\bar{\nu}_k}{2},2\bar{V}_k^{-1}),\tag{3.A.14}$$

with $\bar{\nu}_k=\nu_k+T_{i,k}$ and $\bar{V}_k=V_k+(y_{i,k}-\iota_{T_{i,k}}\bar{y}_{i,k})'(y_{i,k}-\iota_{T_{i,k}}\bar{y}_{i,k})+\frac{T_{i,k}h_k}{h_k+T_{i,k}}(\mu_k-\bar{y}_{i,k})^2$. We define $\mathcal{T}_{i,k}=\{t:s_{it}=k\}$ the index set for observations in regime k, $T_{i,k}=\text{card}(\mathcal{T}_{i,k})$ its cardinality, $y_{i,k}=\{y_{it}:t\in\mathcal{T}_{i,k}\}$ the vector containing the $T_{i,k}$ observations in regime k, $\bar{y}_{i,k}=(T_{i,k})^{-1}\iota'_{T_{i,k}}y_{i,k}$, and ι_n denotes an $n\times1$ all-ones vector. The notation Θ_{-r} indicates that element r is excluded from the vector Θ.

We sample $\mu_{i,k}$ for $i=1.,,,.N$ and $k\in\mathcal{K}_i$ from its full conditional distribution, which according to the prior in Section 3.2.3.3 of the paper is defined as

$$p(\mu_{i,k}|\Theta_{-\mu_{i,k}},s_{i,1:T},y_{i,1:T})\propto\mathcal{N}(\bar{\mu}_{i,k},(\sigma_{i,k}^{-2})^{-1}\bar{h}_{i,k}^{-1}),\tag{3.A.15}$$

with $\bar{\mu}_{i,k}=(h_k\mu_k+T_{i,k}\bar{y}_{i,k})/(h_k+T_{i,k})$ and $\bar{h}_{i,k}=h_k+T_{i,k}$.

We sample h_k for $k = 1, ..., K$ from its full conditional distribution, which according to the prior in Section 3.2.3.3 of the paper is defined as

$$\begin{aligned} p(h_k|\Theta_{-h_k}, F) &\propto \mathcal{G}(\frac{\underline{d}}{2}, 2\underline{c}^{-1}) \prod_{i \in \mathcal{N}_k} \mathcal{N}(\mu_{i,k}|\mu_k, (\sigma_{i,k}^{-2})^{-1} h_k^{-1}) \\ &\propto \mathcal{G}(\frac{\underline{d} + N_k + 1}{2}, 2(\underline{c} + \sum_{i \in \mathcal{N}_k} \sigma_{i,k}^{-2}(\mu_{i,k} - \mu_k)^2)^{-1}). \end{aligned} \tag{3.A.16}$$

We sample ν_k for $k = 1, ..., K$ from its full conditional distribution, which according to the prior in Section 3.2.3.3 of the paper is defined as

$$\begin{aligned} p(\nu_k|\Theta_{-\nu_k}, F) &\propto \text{Exp}(\underline{\lambda}) \prod_{i \in \mathcal{N}_k} \mathcal{G}(\sigma_{i,k}^{-2}|\frac{\nu_k}{2}, 2V_k^{-1}) \\ &\propto \exp(-\underline{\lambda}\nu_k) \prod_{i \in \mathcal{N}_k} (\Gamma(\frac{\nu_k}{2})^{-1} (2V_k^{-1})^{-\frac{\nu_k}{2}} (\sigma_{i,k}^{-2})^{\frac{\nu_k}{2}}). \end{aligned} \tag{3.A.17}$$

The above distribution is not conditionally conjugate. Therefore, a Metropolis–Hastings method is applied to sample ν_k. We use a Gamma random walk proposal distribution $q(\cdot)$ with mean equal to the current iteration value and variance equal to a fixed value $\sigma_{\nu_k}^2 = 1$. This is parameterized as $q(\tilde{\nu}_k|\nu_k) \sim \mathcal{G}(\frac{\nu_k^2}{\sigma_{\nu_k}^2}, \frac{\sigma_{\nu_k}^2}{\nu_k})$, where $\tilde{\nu}_k$ is the proposal value and ν_k is the current iteration value. We accept the proposal with probability

$$\min\left\{1, \frac{p(\tilde{\nu}_k|\Theta_{-\nu_k}, F) q(\nu_k|\tilde{\nu}_k)}{p(\nu_k|\Theta_{-\nu_k}, F) q(\tilde{\nu}_k|\nu_k)}\right\}. \tag{3.A.18}$$

We sample μ_k for $k = 1, ..., K$ from its full conditional distribution, which according to the prior in Section 3.2.3.3 of the paper is defined as

$$\begin{aligned} p(\mu_k|\Theta_{-\mu_k}, F) &\propto \mathcal{N}(\underline{\mu}, (\underline{h}^{-2})^{-1}) \prod_{i \in \mathcal{N}_k} \mathcal{N}(\mu_{i,k}|\mu_k, (\sigma_{i,k}^{-2})^{-1} h_k^{-1}) \\ &\propto \mathcal{N}(\bar{\mu}_k, (\bar{h}_k^{-2})^{-1}), \end{aligned} \tag{3.A.19}$$

with $\bar{h}_k^2 = (\underline{h}^{-2} + h_k \sum_{i \in \mathcal{N}_k} \sigma_{i,k}^{-2})^{-1}$ and $\bar{\mu}_k = \bar{h}_k^2 (\underline{h}^{-2}\underline{\mu} + h_k \sum_{i \in \mathcal{N}_k} \sigma_{i,k}^{-2} \mu_{i,k})$.

We sample V_k for $k = 1, ..., K$ from its full conditional distribution, which according to the prior in Section 3.2.3.3 of the paper is defined as

$$\begin{aligned} p(V_k|\Theta_{-V_k}, F) &\propto \mathcal{G}(\frac{\underline{\nu}}{2}, 2\underline{V}^{-1}) \prod_{i \in \mathcal{N}_k} \mathcal{G}(\sigma_{i,k}^{-2}|\frac{\nu_k}{2}, 2V_k^{-1}) \\ &\propto \mathcal{G}(\frac{\underline{\nu} + \nu_k N_k}{2}, 2(\underline{V} + \sum_{i \in \mathcal{N}_k} \sigma_{i,k}^{-2})^{-1}). \end{aligned} \tag{3.A.20}$$

3.A.5 Sampling $\Theta_{\mathbf{IBP}}$

We sample the IBP parameter β from its full conditional distribution, which according to the probability mass function and the prior in Section 3.2.3.1 of the paper is defined as

$$\begin{aligned} p(\beta|c,F) &\propto \mathcal{G}(a_\beta, b_\beta)p(F) \\ &\propto \beta^{a_\beta-1}\exp(-\beta b_\beta^{-1})\beta^{M_N}\exp(-\beta\sum_{i=1}^{N}\frac{c}{c+i-1}) \\ &\propto \mathcal{G}(a_\beta+M_N, (b_\beta^{-1}+\sum_{i=1}^{N}\frac{c}{c+i-1})^{-1}), \end{aligned} \tag{3.A.21}$$

We sample the IBP parameter c from its full conditional distribution, which according to the probability mass function and the prior in Section 3.2.3.1 of the paper is defined as

$$\begin{aligned} &p(c|\beta,F) \propto \mathcal{G}(a_c, b_c)p(F) \\ &\propto c^{a_c-1}\exp(-cb_c^{-1})c^{M_N}\exp(-\beta\sum_{i=1}^{N}\frac{c}{c+i-1})\prod_{k=1}^{M_N}\frac{\Gamma(N_k)\Gamma(N-N_k+c)}{\Gamma(N+c)}. \end{aligned} \tag{3.A.22}$$

The above distribution is not conditionally conjugate. Therefore, a Metropolis–Hastings method is applied to sample c. We use a Gamma random walk proposal distribution $q(\cdot)$ with mean equal to the current iteration value and variance equal to a fixed value $\sigma_c^2 = 1$. This is parameterized as $q(\tilde{c}|c) \sim \mathcal{G}(\frac{c^2}{\sigma_c^2}, \frac{\sigma_c^2}{c})$, where $\tilde{c}$ is the proposal value and c is the current iteration value. We accept the proposal with probability

$$\min\left\{1, \frac{p(\tilde{c}|\beta,F)q(c|\tilde{c})}{p(c|\beta,F)q(\tilde{c}|c)}\right\}. \tag{3.A.23}$$

3.A.6 Sampling $\Theta_{\mathbf{ZIB}}$

We sample the ZIB parameter γ from its full conditional distribution, which according to the prior in Section 3.2.3.2 of the paper and the probability density function of the Dirichlet distribution is defined as

$$\begin{aligned} &p(\gamma|\kappa,P) \propto \mathcal{G}(a_\gamma, b_\gamma)p(P|F) \\ &\propto \gamma^{a_\gamma-1}\exp\{-\gamma b_\gamma^{-1}\}\prod_{i=1}^{N}\prod_{k\in\mathcal{K}_i}(\frac{\Gamma(\gamma K_i+\kappa)}{\Gamma(\gamma)^{(K_i-1)}\Gamma(\gamma+\kappa)}\prod_{l\in\mathcal{K}_i}p_{i,kl}^{\gamma+\kappa\delta_k-1}). \end{aligned} \tag{3.A.24}$$

Because of non-conjugacy we rely on a Metropolis–Hastings method to sample γ. We use a Gamma random walk proposal distribution $q(\cdot)$ with mean equal to the current iteration value and variance equal to a fixed value $\sigma_\gamma^2 = 100$. This is parameterized as $q(\tilde{\gamma}|\gamma) \sim \mathcal{G}(\frac{\gamma^2}{\sigma_\gamma^2}, \frac{\sigma_\gamma^2}{\gamma})$, where $\tilde{\gamma}$ is the proposal value and γ is the current iteration value. We accept the proposal

with probability

$$\min\left\{1, \frac{p(\tilde{\gamma}|\kappa, P)q(\gamma|\tilde{\gamma})}{p(\gamma|\kappa, P)q(\tilde{\gamma}|\gamma)}\right\}. \tag{3.A.25}$$

We sample the ZIB parameter κ from its full conditional distribution, which according to the prior in Section 3.2.3.2 of the paper and the probability density function of the Dirichlet distribution is defined as

$$\begin{aligned} p(\kappa|\gamma, P) &\propto \mathcal{G}(a_\kappa, b_\kappa)p(P|F) \\ &\propto \kappa^{a_\kappa - 1}\exp\{-\kappa b_\kappa^{-1}\}\prod_{i=1}^{N}\frac{\Gamma(\gamma K_i + \kappa)^{K_i}}{\Gamma(\gamma+\kappa)^{K_i}}\prod_{l\in\mathcal{K}_i} p_{i,ll}^{\gamma+\kappa-1}. \end{aligned} \tag{3.A.26}$$

Because of non-conjugacy we rely on a Metropolis–Hastings method to sample κ. We use a Gamma random walk proposal distribution $q(\cdot)$ with mean equal to the current iteration value and variance equal to a fixed value $\sigma_\kappa^2 = 100$. This is parameterized as $q(\tilde{\kappa}|\kappa) \sim \mathcal{G}(\frac{\kappa^2}{\sigma_\kappa^2}, \frac{\sigma_\kappa^2}{\kappa})$, where $\tilde{\kappa}$ is the proposal value and κ is the current iteration value. We accept the proposal with probability

$$\min\left\{1, \frac{p(\tilde{\kappa}|\gamma, P)q(\kappa|\tilde{\kappa})}{p(\kappa|\gamma, P)q(\tilde{\kappa}|\kappa)}\right\}. \tag{3.A.27}$$

Appendix 3.B

This appendix presents additional simulation results on convergence and efficiency.

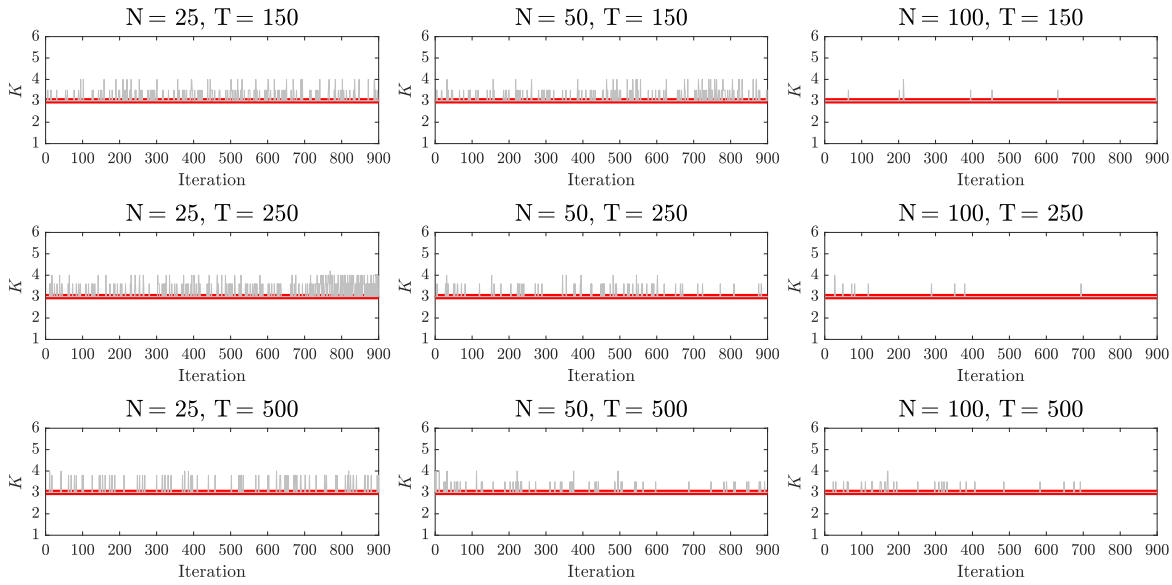


Figure 3.B.1 DGP 1 (HoF + WS). The median (red line) and 10th and 90th quantile (grey line) values of the regime dimension K computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

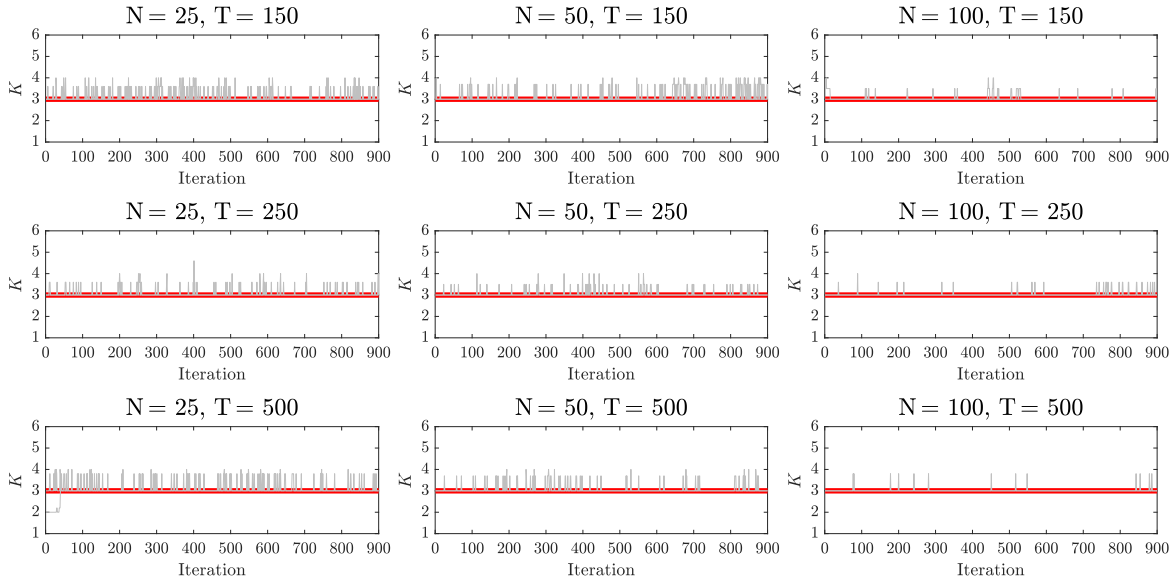


Figure 3.B.2 DGP 2 (HoF + WSSS). The median (red line) and 10th and 90th quantile (grey line) values of the regime dimension K computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

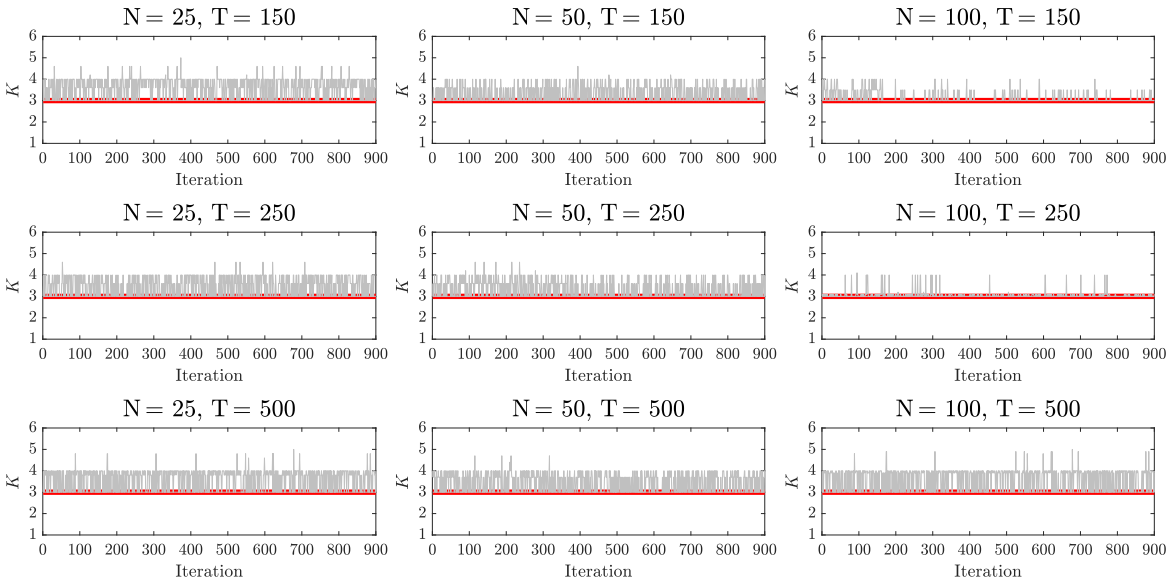


Figure 3.B.3 DGP 3 (WHeF + WS). The median (red line) and 10th and 90th quantile (grey line) values of the regime dimension K computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

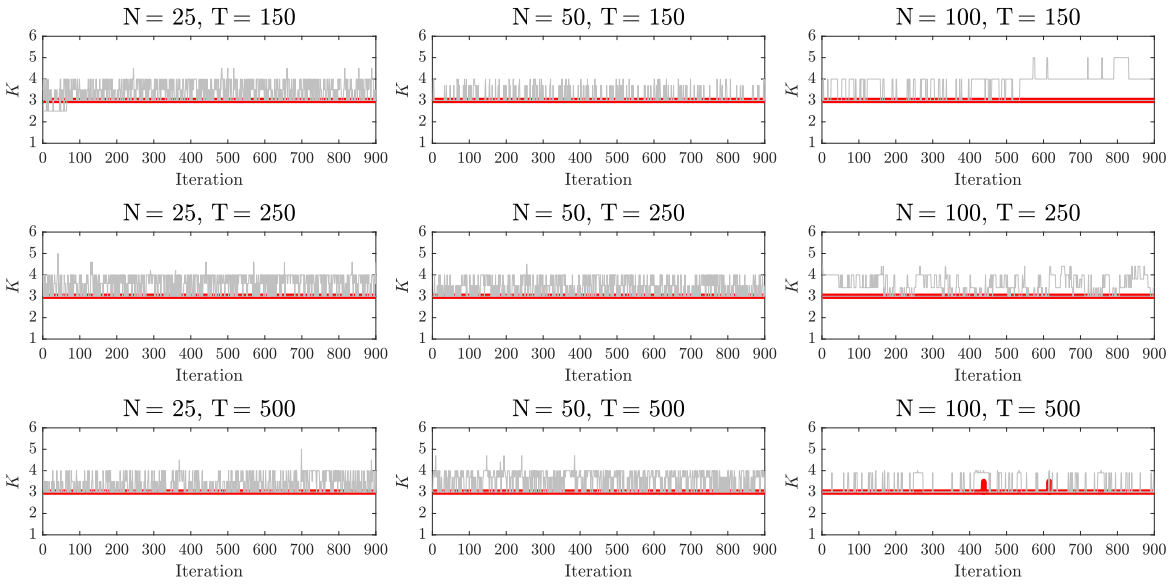


Figure 3.B.4 DGP 4 (WHeF + WSSS). The median (red line) and 10th and 90th quantile (grey line) values of the regime dimension K computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

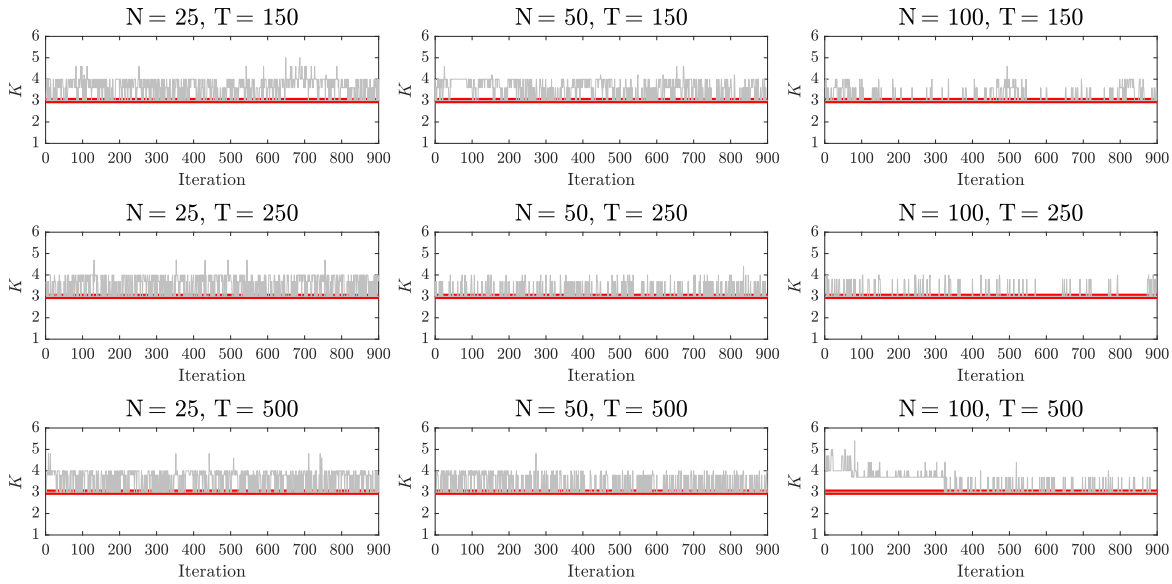


Figure 3.B.5 DGP 5 (SHeF + WS). The median (red line) and 10th and 90th quantile (grey line) values of the regime dimension K computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

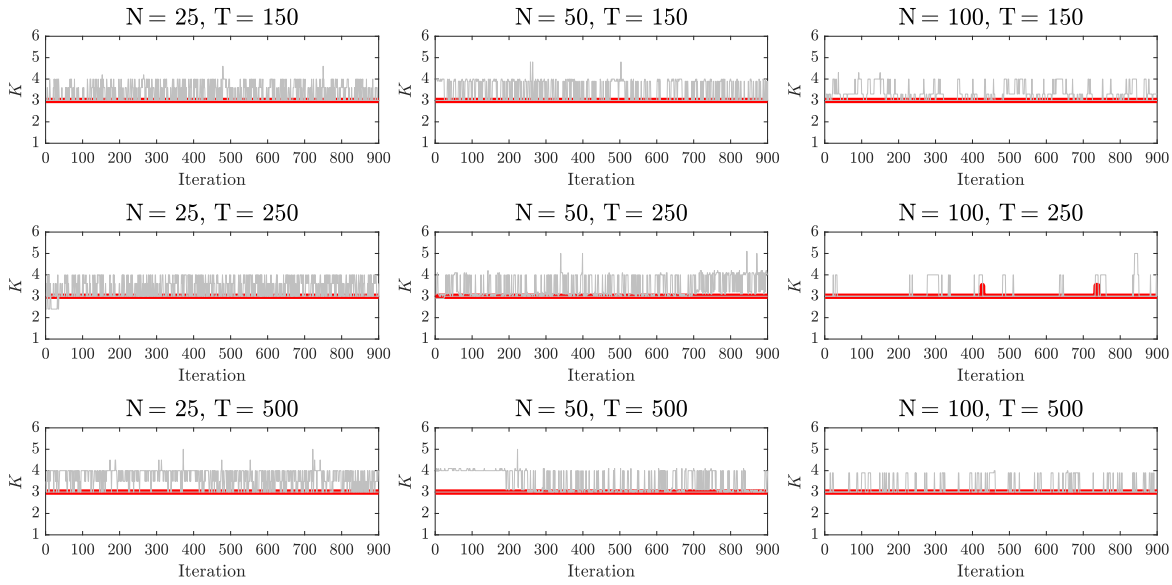


Figure 3.B.6 DGP 6 (SHeF + WSSS). The median (red line) and 10th and 90th quantile (grey line) values of the regime dimension K computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

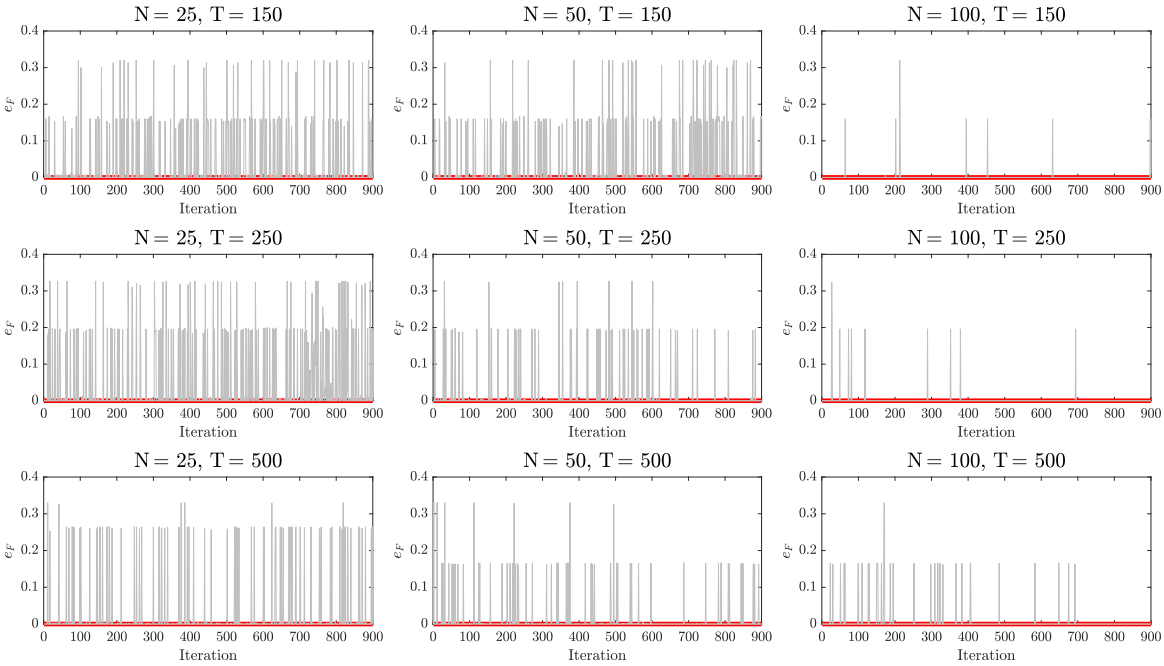


Figure 3.B.7 DGP 1 (HoF + WS). The median (red line) and 10th and 90th quantile (grey line) values of the estimation error e_F computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

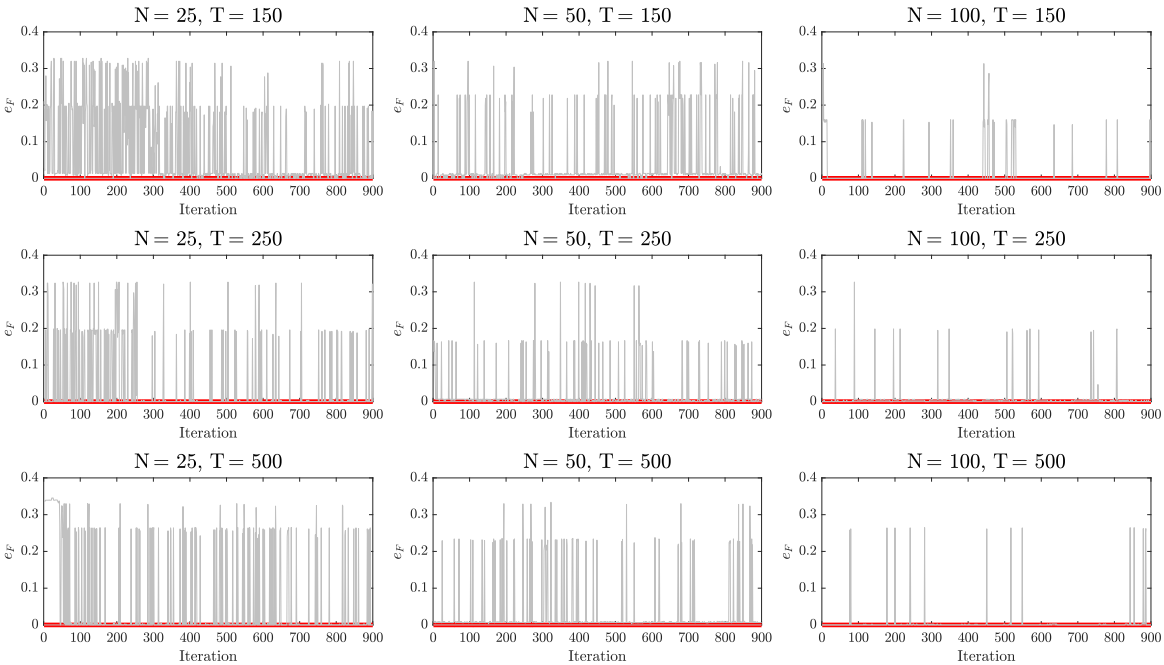


Figure 3.B.8 DGP 2 (HoF + WSSS). The median (red line) and 10th and 90th quantile (grey line) values of the estimation error e_F computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

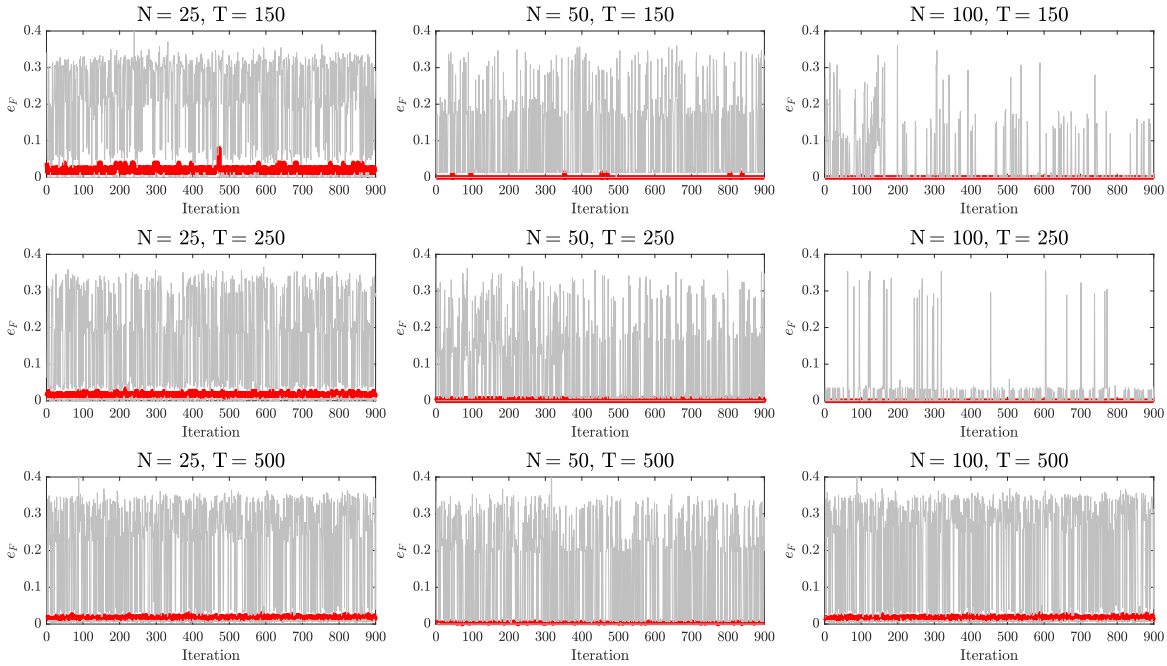


Figure 3.B.9 DGP 3 (WHeF + WS). The median (red line) and 10th and 90th quantile (grey line) values of the estimation error e_F computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

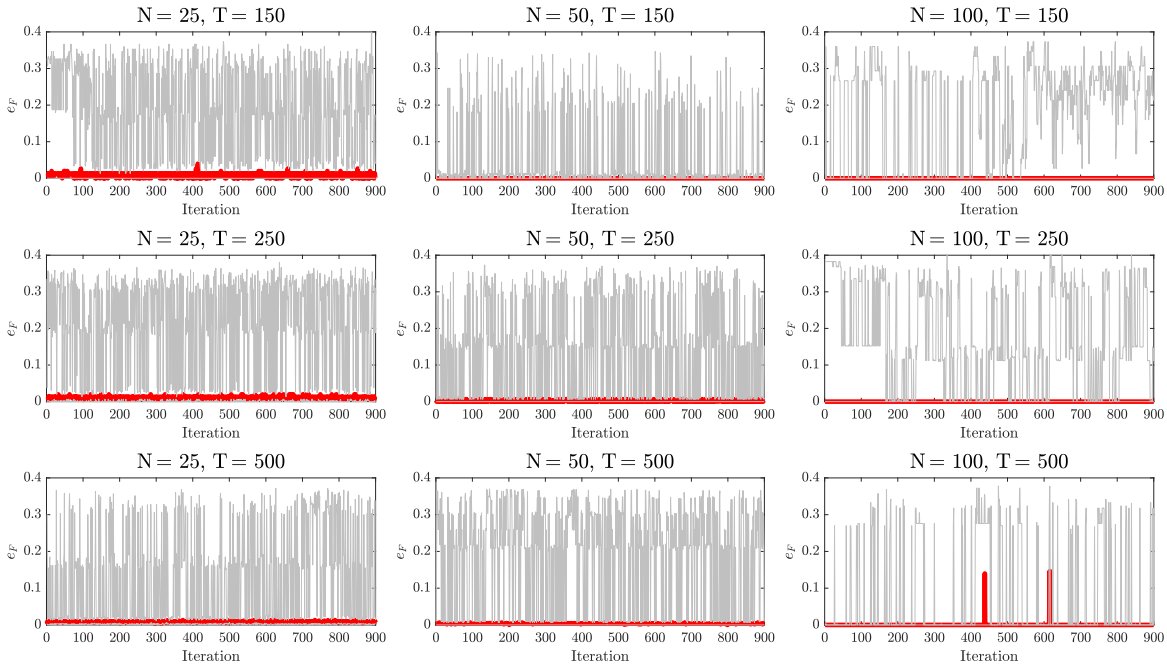


Figure 3.B.10 DGP 4 (WHeF + WSSS). The median (red line) and 10th and 90th quantile (grey line) values of the estimation error e_F computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

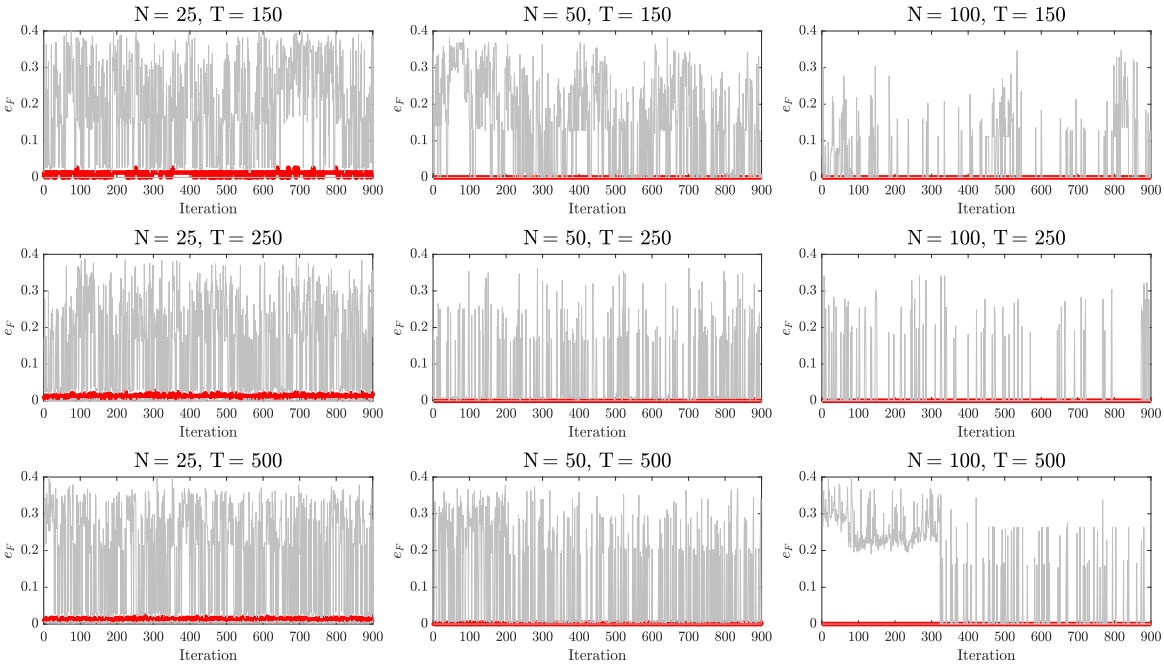


Figure 3.B.11 DGP 5 (SHeF + WS). The median (red line) and 10th and 90th quantile (grey line) values of the estimation error e_F computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

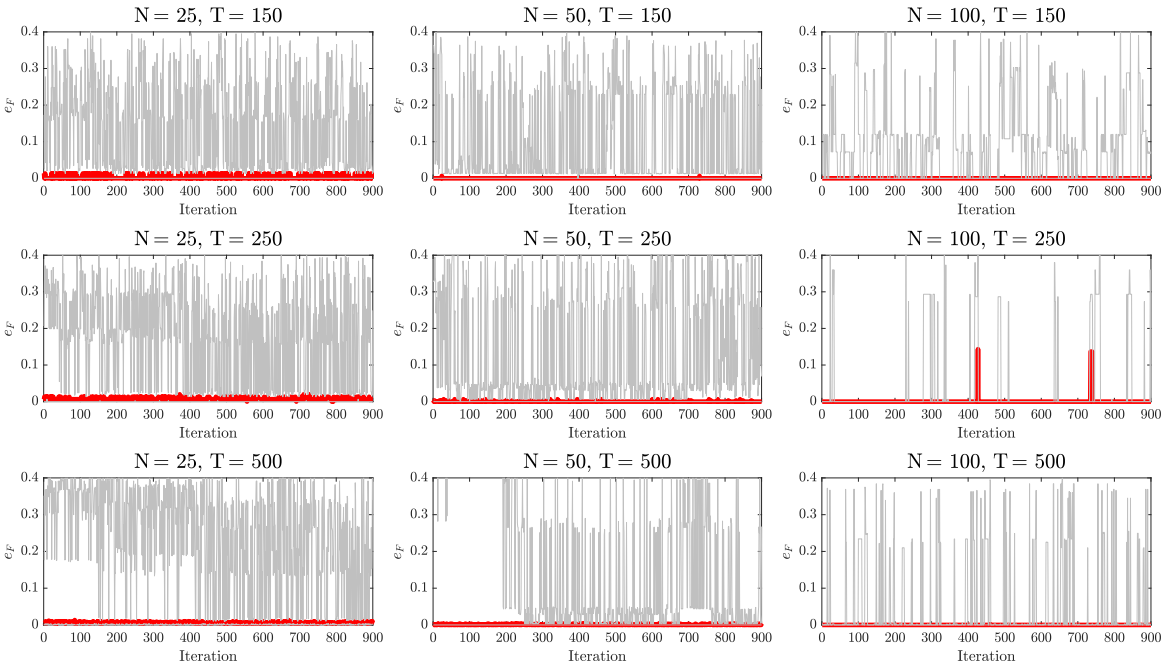


Figure 3.B.12 DGP 6 (SHeF + WSSS). The median (red line) and 10th and 90th quantile (grey line) values of the estimation error e_F computed through MCMC iterations from successful experiments that detect the true regime dimension of the panel.

Appendix 3.C

This appendix presents benchmark model specifications and contains a description of the data used in the empirical illustration.

3.C.1 Benchmark model specifications

The first benchmark framework is a BNP MS panel framework assuming cross-sectional homogeneous regime dimension. For $i = 1, ..., N$ and $t = 1, ..., T$, it is specified as

$$\begin{aligned} &y_{it} = \mu_{i,s_t} + \sigma_{i,s_t}\varepsilon_{it}, \quad \varepsilon_{it} \sim \text{ i.i.d. } \mathcal{N}(0,1) \\ &s_t | s_{t-1} = k, P \sim \boldsymbol{p}_k \\ &\boldsymbol{p}_k | \pi \sim \text{DP}(\gamma + \kappa, \frac{\gamma\pi + \kappa\delta_k}{\gamma + \kappa}) \\ &\pi \sim \text{SBP}(\eta), \end{aligned} \tag{3.C.1}$$

where s_t is a hidden Markov chain process with values in an infinite set $\{1, 2, ...\}$; P is an infinite-dimensional matrix of regime transition probabilities; $\boldsymbol{p}_k = (p_{k1}, ..., p_{kl}, ...)$ is the k-th row of P which collects the transition probabilities out of regime k, i.e., $p_{kl} = \text{Prob}(s_t = l | s_{t-1} = k)$, for $l, k \in \{1, 2, ...\}$; $\text{DP}(\cdot, \cdot)$ denotes the Dirichlet process[8]; $\text{SBP}(\cdot)$ stands for the stick-breaking process, which allows the sequence $\pi = \{\pi_j\}_{j=1}^{\infty}$ to satisfy $\sum_{j=1}^{\infty} \pi_j = 1$ with probability one and therefore defines a distribution over the positive integer. To infer the values of the DP and SBP parameters γ, κ and η, we place the following priors, $\gamma + \kappa \sim \mathcal{G}(1, 10)$, $\rho = \frac{\kappa}{\gamma+\kappa} \sim \text{Beta}(10, 1)$ and $\eta \sim \mathcal{G}(1, 1)$. The rest of the parameters in (3.C.1) are given the same priors as our proposed framework. Posterior sampling is based on the beam sampler; for implementation details, see Dufays (2016).

The second benchmark is a standard MS panel model with the pre-specified number of regimes $K = 2$. For $i = 1, ..., N$ and $t = 1, ..., T$, it is specified as

$$\begin{aligned} y_{it} &= \begin{cases} \mu_{i,1} + \sigma_{i,1}\varepsilon_{it}, & \varepsilon_{it} \sim \text{ i.i.d. } \mathcal{N}(0,1) \quad \text{if} \quad s_{it} = 1 \\ \mu_{i,2} + \sigma_{i,2}\varepsilon_{it}, & \varepsilon_{it} \sim \text{ i.i.d. } \mathcal{N}(0,1) \quad \text{if} \quad s_{it} = 2 \end{cases} \\ s_{it} | s_{it-1} = k, P_i &\sim \boldsymbol{p}_{i,k}, \quad \text{for } k \in \{1, 2\}, \end{aligned} \tag{3.C.2}$$

where s_{it} is a hidden Markov chain process with values in a finite set $\{1, 2\}$; P is a 2×2 matrix of regime transition probabilities; $\boldsymbol{p}_k = (p_{k1}, p_{k2})$ is the k-th row of P which collects the transition probabilities out of regime k, i.e., $p_{kl} = \text{Prob}(s_t = l | s_{t-1} = k)$, for $l, k \in \{1, 2\}$. For the regime sequences and regime-switching parameters, posterior sampling can be performed in a similar fashion as in our proposed framework with $\mathcal{K}_i = \{1, 2\}, \forall i \in \{1, ..., N\}$. For the

[8] To facilitate comparison, we use the sticky version of the Dirichlet process (see, e.g., Dufays, 2016; Bauwens et al., 2017; Jin et al., 2022), which relies on parameters γ, κ, and the measure of a point mass of one at k, namely δ_k, so that it reinforces the regime self-transition in the similar way as our proposed framework.

regime transition matrix, posterior sampling relies on the full conditional distribution of a two-dimensional Dirichlet distribution.

The third is a BNP MS univariate framework applied separately to each unit in the cross-section. For $i = 1, ..., N$ and $t = 1, ..., T$, it is specified as

$$\begin{aligned} &y_{it} = \mu_{i,s_{it}} + \sigma_{i,s_{it}} \varepsilon_{it}, \quad \varepsilon_{it} \sim \text{ i.i.d. } \mathcal{N}(0,1) \\ &s_{it} | s_{it-1} = k, P_i \sim \boldsymbol{p}_{i,k} \\ &\boldsymbol{p}_{i,k} | \pi_i \sim \text{DP}(\gamma + \kappa, \frac{\gamma \pi_i + \kappa \delta_k}{\gamma + \kappa}) \\ &\pi_i \sim \text{SBP}(\eta). \end{aligned} \tag{3.C.3}$$

This framework is implemented independently for each unit with the same prior hyperparameter values and posterior sampler as in the first benchmark.

3.C.2 Data details

Table 3.C.1 Data description

State	Abbreviation	Mean	St.dev.	AR(1)	Min.	Max.
Alabama	AL	0.171	1.049	0.200	-20.902	5.561
Alaska	AK	0.088	0.545	0.252	-10.504	4.425
Arizona	AZ	0.308	0.713	0.263	-13.324	4.003
Arkansas	AR	0.192	0.216	0.280	-3.780	0.739
California	CA	0.255	1.015	0.204	-21.098	4.519
Colorado	CO	0.269	0.564	0.142	-10.847	4.977
Connecticut	CT	0.166	0.662	0.567	-9.920	4.049
Delaware	DE	0.195	1.191	0.227	-21.545	13.177
Florida	FL	0.272	0.907	0.171	-18.615	4.838
Georgia	GA	0.272	0.343	0.153	-6.478	2.822
Hawaii	HI	0.141	2.223	0.059	-48.708	8.142
Idaho	ID	0.277	0.485	0.314	-8.648	2.116
Illinois	IL	0.143	1.094	0.225	-22.622	5.251
Indiana	IN	0.169	0.860	0.220	-17.134	3.333
Iowa	IA	0.156	0.701	0.176	-14.238	2.399
Kansas	KS	0.160	1.034	0.147	-21.004	6.721
Kentucky	KY	0.150	1.054	0.133	-19.386	12.041
Louisiana	LA	0.071	0.783	0.300	-12.944	4.589
Maine	ME	0.173	0.858	0.204	-16.273	5.784
Maryland	MD	0.209	0.844	0.169	-17.198	2.310
Massachusetts	MA	0.226	2.105	0.087	-45.624	7.830
Michigan	MI	0.108	2.255	0.109	-46.366	16.941
Minnesota	MN	0.210	0.805	0.204	-16.137	3.249
Mississippi	MS	0.134	0.493	0.143	-9.057	4.772
Missouri	MO	0.142	0.864	0.180	-17.361	5.040
Montana	MT	0.143	1.615	0.151	-32.991	8.868
Nebraska	NE	0.200	0.305	0.487	-5.010	1.585
Nevada	NV	0.309	1.842	0.231	-38.150	11.674
New Hampshire	NH	0.264	0.571	0.158	-11.721	2.128
New Jersey	NJ	0.169	1.017	0.115	-20.245	9.938
New Mexico	NM	0.199	0.378	0.439	-6.670	1.004
New York	NY	0.160	1.310	0.104	-28.236	4.690
North Carolina	NC	0.245	0.390	0.244	-7.361	1.529
North Dakota	ND	0.210	0.913	0.120	-19.394	3.465
Ohio	OH	0.140	0.924	0.189	-18.698	4.935
Oklahoma	OK	0.160	0.761	0.169	-15.147	4.772
Oregon	OR	0.252	0.611	0.197	-12.205	2.055
Pennsylvania	PA	0.145	0.967	0.150	-20.505	2.757
Rhode Island	RI	0.154	1.237	0.114	-25.904	6.039
South Carolina	SC	0.233	0.855	0.240	-16.588	4.600
South Dakota	SD	0.230	0.692	0.190	-13.703	2.260
Tennessee	TN	0.230	0.727	0.195	-14.657	2.446
Texas	TX	0.266	0.334	0.291	-6.315	1.063
Utah	UT	0.309	0.302	0.411	-4.936	1.235
Vermont	VT	0.207	1.294	0.151	-27.061	6.818
Virginia	VA	0.223	0.639	0.159	-13.199	2.956
Washington	WA	0.263	0.857	0.273	-17.468	3.242
West Virginia	WV	0.096	1.927	0.282	-36.618	6.146
Wisconsin	WI	0.170	0.668	0.202	-13.383	2.446
Wyoming	WY	0.126	0.412	0.716	-4.441	1.519

NOTES: The table reports data details regarding state name, abbreviation, and summary statistics of the monthly coincident indices for 50 US states. The sample period is from October 1979 to October 2023.

Conclusive remarks and extensions

This dissertation centered on modeling and forecasting time series subject to regime shifts, with the emphasis on applications of Markov-switching (MS) models in economics and finance. Chapter 1 modeled the regime shifts in funds' timing behavior as well as the potential drivers behind such shifts using an empirical multi-factor model with endogenous MS factor loadings. Chapter 2 proposed a Bayesian, graph-based MS model for detecting the regime shifts in financial system connectivity, with methodological advances highlighted by an application in the global banking system. Chapter 3 proposed a general framework to estimate the regime dimension for Markov-switching panel models in the presence of possible heterogeneity in cross-sectional regime dimension, with practical usefulness demonstrated by a US business cycle analysis. In the following, we highlight the contributions and implications that we can draw from the results in this dissertation, and provide some discussions regarding future extensions.

Methodological implications

This dissertation has several methodological implications and contributions. Chapter 1 expanded upon the extensive research that assumed time-invariant funds' timing behavior and investigated this issue through a multi-factor model with constant factor loadings (see, e.g., Treynor and Mazuy, 1966; Cao et al., 2013; Bodnaruk et al., 2019; Bali et al., 2021, among many others). Specifically, our MS approach showed that even when evidence for funds' timing behavior was found, this behavior was not necessarily constant and tended to experience regime shifts across varying market conditions. This held important lessons for researchers and practitioners who ignored the situation where funds timed the market strategically and intermittently, rather than continuously over the time span of study. Moreover, although greater attention has been directed towards time-varying funds' timing abilities, some important economic insights were still missing in the literature, which included (i) the rich dynamics (e.g., changes in the existence, directions, and strengths) of timing behavior, and (ii) the drivers (i.e., potential economic mechanisms) underlying such dynamics. Compared to the existing methodological advances in this aspect (see, e.g., Cao et al., 2013; Siegmann and Stefanova, 2017; Li et al., 2020a), our MS approach allowed for multiple

regimes and endogeneity in the regime shifts, which facilitated further analysis on the rich dynamics and the potential drivers of funds' timing behavior.

The claimed methodological contribution of Chapter 2 was twofold. The first contribution of the chapter lay in the literature on financial connectedness analysis. Specifically, a novel MS-GSVAR model building on the framework in Ahelegbey et al. (2016a) was proposed to simultaneously capture several notable empirical features of financial connectedness, such as the mixture of contemporaneous and temporal dependences, its high-dimensionality, and the occurrence of abrupt structural changes. Although many earlier studies made methodological advances in capturing one (or two) of the above features, our model facilitated a "full sweep" of the list. The second contribution of the paper lay in the literature on using Bayesian Graphical model for multivariate systems. Specifically, an efficient graph inference method building on a Hybrid-MCMC algorithm of Kuipers et al. (2022) was developed to address the computational complexities arising from inference on graphical multivariate models in the context of high model dimension, numerous lags, and multiple states. Although the literature normally considered the Structure-MCMC algorithm for graph inference, the chapter confirmed in simulations that our method provided more reliable graph inference in much less time, particularly noticeable in the large-dimensional setting.

Chapter 3 addressed the long-standing issue emerging in modeling panel datasets with MS models, which was the determination of the number of regimes that adequately characterized the observed panel data. Existing solutions to this issue typically relied on the assumption that the number of regimes was homogeneous in the cross-section (i.e., every individual transited across all regimes available in the whole panel); examples of methods in this respect included the Dirichlet-process-based MS multivariate frameworks (Hou, 2017; Casarin et al., 2024), and the standard MS panel frameworks with the number of regimes being set through prior belief (Hamilton and Owyang, 2012; Francis et al., 2017; Casarin et al., 2018; Owyang et al., 2022; Agudze et al., 2022) or a model selection procedure (Billio et al., 2016). However, such an assumption may be restrictive as individuals are likely to be characterized only by one or a subset of regimes identified from the panel. The modeling framework proposed in this chapter can estimate the unknown regime dimension for both the whole panel and units in the presence of likely cross-sectional heterogeneous regime dimension. Our approach was also fully flexible in its ability to encompass the case of cross-sectional homogeneous regime dimension, as evidenced by a battery of simulations.

Implications in economics and finance

This dissertation has some implications in economics and finance. In Chapter 1, we found evidence of currency liquidity timing at the aggregate level for a sample of globally-diversified funds, which was not explained away by other reasons except currency liquidity timing (e.g., currency market return and volatility timing, liquidity reaction, illiquid holdings, funding

constraints, and fund flow). Interestingly, this timing behavior appeared to exhibit regime shifts across varying market conditions, in that funds timed currency liquidity negatively (adjusted their currency exposure in the opposite direction to the currency liquidity movement) during tranquil market periods but switched to aggressively positive timing (adjusted their currency exposure in the same direction to the liquidity movement with increasing aggressivity) during turbulent market periods. We provided a potential explanation of the observed dynamic pattern of funds' currency-liquidity-timing behavior based on their currency hedging practices. Besides, we documented that the dynamics of currency liquidity timing were possibly driven by currency liquidity deterioration and negative shocks to fund returns. The primary focus of previous studies relevant to timing was on certain timing skills in the market that was presumably viewed as funds' focus market, e.g., timing skills of equity funds in the equity market or that of bond funds in the bond market. Consequently, for funds exposed to various markets, their timing skills in the risk-exposed market, apart from what was conceptually thought as their traded market, tended to be neglected. An example was globally-diversified mutual funds, of whom the timing skills on the traded markets (e.g., global equity markets) have been explored in the literature (see, e.g., Glassman and Riddick, 2006; Rodríguez, 2008) but a clear understanding of the same skills on the trade-exposed markets (e.g., currency markets) was still missing. Hence, the findings in this chapter encouraged greater attention to be directed toward funds' timing skills on those markets central to fund risks.

In Chapter 2, we proposed an econometric model that has practical applications in finance. The chapter included a real-data illustration in which we applied the model to measure the volatility connectedness in a global banking system consisting of 96 banks between 2003-2022. First, our model provided a data-driven detection of different connectedness states that best describe the historical evolution of the global banking volatility network. It was found that the evolution of the global banking system experienced long periods of low volatility connectedness, alternated by sporadic events of high and extreme volatility connectedness. In addition, the systemic risk profiles of individual banks, featured by their risk transmission and reception capacity, were readily inferred from the model output. It was found that the North American banks, particularly the US banks, were major risk transmitters while the European and Asian banks were major risk receivers. The model also offered deeper insights about the frequency-specific source of volatility connectedness. It was found that for both temporal and contemporaneous connectedness the regime with higher level of connectedness was characterized by larger cross-geography banking clusters, stronger system-wide connections, and a few central banks dominating the volatility propagation. Interestingly, contemporaneous connectedness patterns were noticeably denser than temporal patterns, indicating that the dynamics of the global banking overall volatility connectedness was mainly driven by changes in its contemporaneous component. Overall, the proposed statistical tool in this chapter can assist policymakers to monitor how contagion-prone the system will be, identify risk transmitters as well as recipients in the financial system, and thus

to shape targeted policy to protect vulnerable risk recipient as individual or group whenever necessary.

In Chapter 3, we demonstrated the empirical gains of considering likely heterogeneous regime dimension in the cross-section for MS panel models. The chapter included a business cycle analysis on a panel of 50 US state-level macroeconomic indices over the period 1979-2023. The results implied that the assumption about heterogeneous regime dimension in the cross-section was plausible empirically: about 25% of US states possessed part of the identified regimes whereas the remaining states possessed all the identified regimes. From the turning points of the identified regimes, it appeared that our framework produced accurate timing of recessions that agrees with the NBER-designated cycle dates and was fully flexible in capturing the noncommon and non-synchronized dynamics underlying individual US states. In contrast, the framework assuming cross-sectional homogeneous regime dimension lacked the flexibility to capture these features. Additionally, an out-of-sample forecasting exercise showed that the MS panel framework assuming cross-sectional homogeneous regime dimension as well as the MS univariate framework were unable to rival the forecast accuracy generated from our MS panel framework allowing for cross-sectional heterogeneous regime dimension. In summary, these results should be valuable for state policymakers to take targeted actions earlier in an effort to counteract contractionary tendencies, especially in turbulent economic times. Understanding cross-state heterogeneity in economic conditions within one country is also useful for understanding aggregate dynamics and for informing federal policymaking.

Future extensions

Based on the results in this dissertation, several developments can be pursued in the near future. In Chapter 1, one of the research lines discussed is whether funds with distinct fund characteristics (e.g., age, size, expense ratio, turnover ratio) differ in their dynamic currency-liquidity-timing behavior. A possible extension is to implement the proposed modeling framework for funds grouped by characteristics. In Chapter 2, we pointed out that the proposed model is fast and easy to implement for multivariate time series with dimension up to around 100 series. Thus, the model may find application in a wide spectrum of research where connectedness analysis on large financial datasets is of interest (e.g., connectedness of the global commodity market or the foreign exchange market). In Chapter 3, the most parsimonious form of the MS panel models is considered, but there is no conceptual problem with adding autoregressive terms, exogenous predictors, or cross-section error correlations to this basic form. For instance, a further research can involve augmenting the proposed MS panel framework with some exogenous factors used in the asset pricing literature, and applying the model to a panel of stock returns. This may allow to identify the noncommon and non-synchronized regime shifts in the cross-sectional risk premia.

Since the nature of MS models is to deal with persistent and recurrent instability in not only the evolution of data series but also the relationships within empirical variables, their applications are certainly not limited to those presented in this dissertation. In this respect, a possible research line is instability in forecast combination. Forecast combination (or pooling) which combines individual forecasts rather than relying on a single forecast has attracted considerable attention in the forecasting literature because it offers diversification and hedging gains. Given that the relative performance of different forecasts is likely to change over time, it is natural to consider forecast pooling with time-varying combination weights. To accommodate persistent and recurrent changes in combination weights as a result of policy changes, economy fluctuations or institutional evolution, it is desirable to view the weights subject to regime shifts. Further research can consider methodological and/or empirical extensions of MS models with the applications in this field.

Bibliography

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Acharya, V., R. Engle, and M. Richardson (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review 102*(3), 59–64.

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. Richardson (2017). Measuring systemic risk. *Review of Financial Studies 30*(1), 2–47.

Adrian, T. and M. K. Brunnermeier (2016). CoVaR. *American Economic Review 106*(7), 1705.

Agudze, K. M., M. Billio, R. Casarin, and F. Ravazzolo (2022). Markov switching panel with endogenous synchronization effects. *Journal of Econometrics 230*(2), 281–298.

Ahelegbey, D. F., M. Billio, and R. Casarin (2016a). Bayesian graphical models for structural vector autoregressive processes. *Journal of Applied Econometrics 31*(2), 357–386.

Ahelegbey, D. F., M. Billio, and R. Casarin (2016b). Sparse graphical vector autoregression: A Bayesian approach. *Annals of Economics and Statistics 123*, 333–361.

Ahelegbey, D. F., M. Billio, and R. Casarin (2021). Modeling turning points in the global equity market. *Econometrics and Statistics* (forthcoming).

Ahelegbey, D. F. and P. Giudici (2022). NetVIX - A network volatility index of financial markets. *Physica A: Statistical Mechanics and its Applications 594*, 127017.

Alda, M., M. Vargas, and L. Ferruz (2015). Perverse timing or biased coefficients? *Quantitative Finance 15*(1), 171–183.

Almazan, A., K. C. Brown, M. Carlson, and D. A. Chapman (2004). Why constrain your mutual fund manager? *Journal of Financial Economics 73*(2), 289–321.

Amisano, G. and G. Fagan (2013). Money growth and inflation: A regime switching approach. *Journal of International Money and Finance 33*, 118–145.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*(2), 177–190.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Andersson, S. A., D. Madigan, and M. D. Perlman (1997). A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics 25*(2), 505–541.

Ang, A. and G. Bekaert (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics 20*(2), 163–182.

Ardia, D. (2009). Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with Student-t innovations. *The Econometrics Journal 12*(1), 105–126.

Avalos, F. H., R. Moreno, and T. Romero (2015). Leverage on the buy side. BIS Working Papers No 517. Available at: https://www.bis.org/publ/work517.pdf.

Bali, T. G., S. J. Brown, M. O. Caglayan, and U. Celiker (2021). Does industry timing ability of hedge funds predict their future performance, survival, and fund flows? *Journal of Financial and Quantitative Analysis 56*(6), 2136–2169.

Barigozzi, M. and C. Brownlees (2019). NETS: Network estimation for time series. *Journal of Applied Econometrics 34*(3), 347–364.

Baruník, J. and T. Křehlík (2018). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics 16*(2), 271–296.

Baumeister, C., D. Leiva-León, and E. Sims (2024). Tracking weekly state-level economic conditions. *Review of Economics and Statistics*, 1–22.

Bauwens, L., J.-F. Carpantier, and A. Dufays (2017). Autoregressive moving average infinite hidden Markov-switching models. *Journal of Business & Economic Statistics 35*(2), 162–182.

Benoit, S., J.-E. Colliard, C. Hurlin, and C. Pérignon (2017). Where the risks lie: A survey on systemic risk. *Review of Finance 21*(1), 109–152.

Bianchi, D., M. Billio, R. Casarin, and M. Guidolin (2019). Modeling systemic risk with Markov switching graphical SUR models. *Journal of Econometrics 210*(1), 58–74.

Billio, M., R. Casarin, M. Costola, and M. Iacopini (2021). COVID-19 spreading in financial networks: A semiparametric matrix regression model. *Econometrics and Statistics* (forthcoming).

Billio, M., R. Casarin, and M. Iacopini (2022). Bayesian Markov-switching tensor regression for time-varying networks. *Journal of the American Statistical Association*, 1–13.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. Van Dijk (2016). Interconnections between eurozone and US booms and busts using a Bayesian panel Markov-switching VAR model. *Journal of Applied Econometrics 31*(7), 1352–1370.

Billio, M., R. Casarin, and L. Rossini (2019). Bayesian nonparametric sparse VAR models. *Journal of Econometrics 212*(1), 97–115.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics 104*(3), 535–559.

Bodnaruk, A., B. Chokaev, and A. Simonov (2019). Downside risk timing by mutual funds. *The Review of Asset Pricing Studies 9*(1), 171–196.

Bodson, L., L. Cavenaile, and D. Sougné (2013). A global approach to mutual funds market timing ability. *Journal of Empirical Finance 20*, 96–101.

Boguth, O. and M. Simutin (2018). Leverage constraints and asset prices: Insights from mutual fund risk taking. *Journal of Financial Economics 127*(2), 325–341.

Boney, V., G. Comer, and L. Kelly (2009). Timing the investment grade securities market: Evidence from high quality bond funds. *Journal of Empirical Finance 16*(1), 55–69.

Broderick, T., M. I. Jordan, and J. Pitman (2012). Beta processes, stick-breaking and power laws. *Bayesian Analysis 7*(2), 439–476.

Broderick, T., J. Pitman, and M. I. Jordan (2013). Feature allocations, probability functions, and paintboxes. *Bayesian Analysis 8*(4), 801–836.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen (2008). Carry trades and currency crashes. *NBER Macroeconomics Annual 23*(1), 313–348.

Brusa, F., T. Ramadorai, and A. Verdelhan (2014). The international CAPM redux. Saïd Business School Working Paper. Available at: https://core.ac.uk/download/pdf/288287858.pdf.

Busse, J. A. (1999). Volatility timing in mutual funds: Evidence from daily returns. *Review of Financial Studies 12*(5), 1009–1041.

Cao, C., Y. Chen, B. Liang, and A. W. Lo (2013). Can hedge funds time market liquidity? *Journal of Financial Economics 109*(2), 493–516.

Cao, C., T. T. Simin, and Y. Wang (2013). Do mutual fund managers time market liquidity? *Journal of Financial Markets 16*(2), 279–307.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Casarin, R., M. Costantini, and A. Osuntuyi (2024). Bayesian nonparametric panel Markov-switching GARCH models. *Journal of Business & Economic Statistics 42*(1), 135–146.

Casarin, R., C. Foroni, M. Marcellino, and F. Ravazzolo (2018). Uncertainty through the lenses of a mixed-frequency Bayesian panel Markov-switching model. *Annals of Applied Statistics 12*(4), 2559–2586.

Casarin, R., M. Iacopini, G. Molina, E. Ter Horst, R. Espinasa, C. Sucre, and R. Rigobon (2020). Multilayer network analysis of oil linkages. *The Econometrics Journal 23*(2), 269–296.

Casarin, R., D. Sartore, and M. Tronzano (2018). A Bayesian Markov-switching correlation model for contagion analysis on exchange rate markets. *Journal of Business & Economic Statistics 36*(1), 101–114.

Casella, G. and C. P. Robert (2004). *Monte Carlo Statistical Methods*, Volume 2. New York: Springer.

Celeux, G., F. Forbes, C. P. Robert, and D. M. Titterington (2006). Deviance information criteria for missing data models. *Bayesian Analysis 1*(4), 651–673.

Chan, K. F., S. Treepongkaruna, R. Brooks, and S. Gray (2011). Asset market linkages: Evidence from financial, commodity and real estate assets. *Journal of Banking & Finance 35*(6), 1415–1426.

Chen, Y., W. Ferson, and H. Peters (2010). Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics 98*(1), 72–89.

Chen, Y. and B. Liang (2007). Do market timing hedge funds time the market? *Journal of Financial and Quantitative Analysis 42*(4), 827–856.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*(2), 221–241.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association 96*(453), 270–281.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research 3*, 507–554.

Choi, J., M. Kronlund, and J. Y. J. Oh (2022). Sitting bucks: Stale pricing in fixed income funds. *Journal of Financial Economics 145*(2), 296–317.

Ciccarelli, M. and A. Rebucci (2007). Measuring contagion and interdependence with a Bayesian time-varying coefficient model: An application to the Chilean FX market during the Argentine crisis. *Journal of Financial Econometrics 5*(2), 285–320.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291–311.

Colombo, D., M. H. Maathuis, M. Kalisch, and T. S. Richardson (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics 40*(1), 294–321.

Connor, R. J. and J. E. Mosimann (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. *Journal of the American Statistical Association 64*(325), 194–206.

Demirer, M., F. X. Diebold, L. Liu, and K. Yilmaz (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics 33*(1), 1–15.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

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

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

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics 7*(2), 197–226.

Dufays, A. (2016). Infinite-state Markov-switching for dynamic volatility. *Journal of Financial Econometrics 14*(2), 418–460.

Edelen, R. M. (1999). Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics 53*(3), 439–466.

Ferson, W. E. and R. W. Schadt (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance 51*(2), 425–461.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika 10*(4), 507–521.

Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance 57*(5), 2223–2261.

Fox, E. B., M. C. Hughes, E. B. Sudderth, and M. I. Jordan (2014). Joint modeling of multiple time series via the beta process with application to motion capture segmentation. *The Annals of Applied Statistics 8*(3), 1281–1313.

Fox, E. B., E. B. Sudderth, M. I. Jordan, and A. S. Willsky (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics 5*(2A), 1020–1056.

Francis, N., M. T. Owyang, and O. Savascin (2017). An endogenously clustered factor approach to international business cycles. *Journal of Applied Econometrics 32*(7), 1261–1276.

Fricke, D. and H. Wilke (2020). Connected funds. Deutsche Bundesbank Discussion Paper No 48. Available at: https://www.econstor.eu/bitstream/10419/223253/1/172798353X.pdf.

Friedman, N. and D. Koller (2003). Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. *Machine Learning 50*, 95–125.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*(453), 194–209.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. New York: Springer.

Garman, M. B. and M. J. Klass (1980). On the estimation of security price volatilities from historical data. *Journal of business 53*(1), 67–78.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian Statistics 4*, 641–649.

Giudici, P. and R. Castelo (2003). Improving Markov chain Monte Carlo model search for data mining. *Machine Learning 50*, 127–158.

Giudici, P. and A. Spelta (2016). Graphical network models for international financial flows. *Journal of Business & Economic Statistics 34*(1), 128–138.

Glassman, D. A. and L. A. Riddick (2006). Market timing by global fund managers. *Journal of International Money and Finance 25*(7), 1029–1050.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association 102*(477), 359–378.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika 82*(4), 711–732.

Griffiths, T. L. and Z. Ghahramani (2005). Infinite latent feature models and the Indian buffet process. In Y. Weiss, B. Schölkopf, and J. Platt (Eds.), *Advances in Neural Information Processing Systems*, Volume 18, pp. 475–482. Cambridge, MA: MIT Press.

Griffiths, T. L. and Z. Ghahramani (2011). The Indian Buffet Process: An introduction and review. *Journal of Machine Learning Research 12*, 1185–1224.

Guidolin, M., E. Hansen, and M. Pedio (2019). Cross-asset contagion in the financial crisis: A Bayesian time-varying parameter approach. *Journal of Financial Markets 45*, 83–114.

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

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hamilton, J. D. and M. T. Owyang (2012). The propagation of regional recessions. *Review of Economics and Statistics 94*(4), 935–947.

Horn, R. A. and C. R. Johnson (2013). *Matrix Analysis*. Cambridge: Cambridge University Press.

Hou, C. (2017). Infinite hidden Markov switching VARs with application to macroeconomic forecast. *International Journal of Forecasting 33*(4), 1025–1043.

Huber, F. and M. M. Fischer (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.

Hur, J. (2017). Monetary policy and asset prices: A Markov-switching DSGE approach. *Journal of Applied Econometrics 32*(5), 965–982.

Hwu, S.-T., C.-J. Kim, and J. Piger (2021). An N-state endogenous Markov-switching model with applications in macroeconomics and finance. *Macroeconomic Dynamics 25*(8), 1937–1965.

Ibert, M., R. Kaniel, S. Van Nieuwerburgh, and R. Vestman (2018). Are mutual fund managers paid for investment skill? *Review of Financial Studies 31*(2), 715–772.

Ishwaran, H. and L. F. James (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American statistical Association 96*(453), 161–173.

James, L. F. (2017). Bayesian Poisson calculus for latent feature modeling via generalized Indian buffet process priors. *The Annals of Statistics 45*(5), 2016–2045.

Jeffreys, H. (1961). Theory of Probability (3rd ed.). Oxford: Oxford University Press.

Jiang, X., Y. Yuan, S. Mahadevan, and X. Liu (2013). An investigation of Bayesian inference approach to model validation with non-normal data. *Journal of Statistical Computation and Simulation 83*(10), 1829–1851.

Jin, X., J. M. Maheu, and Q. Yang (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics 34*(5), 641–660.

Jin, X., J. M. Maheu, and Q. Yang (2022). Infinite Markov pooling of predictive distributions. *Journal of Econometrics 228*(2), 302–321.

Joao, I. C., A. Lucas, J. Schaumburg, and B. Schwaab (2023). Dynamic clustering of multivariate panel data. *Journal of Econometrics 237*(2), 105281.

Jutasompakorn, P., R. Brooks, C. Brown, and S. Treepongkaruna (2014). Banking crises: Identifying dates and determinants. *Journal of International Financial Markets, Institutions and Money 32*, 150–166.

Kalisch, M. and P. Bühlman (2007). Estimating high-dimensional directed acyclic graphs with the PC-algorithm. *Journal of Machine Learning Research 8*(3), 613–636.

Karnaukh, N., A. Ranaldo, and P. Söderlind (2015). Understanding FX liquidity. *Review of Financial Studies 28*(11), 3073–3108.

Karolyi, G. A. and Y. Wu (2021). Is currency risk priced in global equity markets? *Review of Finance 25*(3), 863–902.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kaufmann, S. (2015). K-state switching models with time-varying transition distributions-does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics 187*(1), 82–94.

Kim, Y. M. and K. H. Kang (2022). Bayesian inference of multivariate regression models with endogenous Markov regime-switching parameters. *Journal of Financial Econometrics 20*(3), 391–436.

Kitson, N. K., A. C. Constantinou, Z. Guo, Y. Liu, and K. Chobtham (2023). A survey of Bayesian Network structure learning. *Artificial Intelligence Review 56*, 1–94.

Koivisto, M. and K. Sood (2004). Exact Bayesian structure discovery in Bayesian networks. *Journal of Machine Learning Research 5*, 549–573.

Kuipers, J. and G. Moffa (2017). Partition MCMC for inference on acyclic digraphs. *Journal of the American Statistical Association 112*(517), 282–299.

Kuipers, J., G. Moffa, and D. Heckerman (2014). Addendum on the scoring of Gaussian directed acyclic graphical models. *The Annals of Statistics 42*(4), 1689–1691.

Kuipers, J., P. Suter, and G. Moffa (2022). Efficient sampling and structure learning of Bayesian networks. *Journal of Computational and Graphical Statistics 31*(3), 639–650.

LeSage, J. P. (1999). Applied econometrics using MATLAB. Manuscript, Dept. of Economics, University of Toledo.

Li, B., J. Luo, and K.-H. Tee (2017). The market liquidity timing skills of debt-oriented hedge funds. *European Financial Management 23*(1), 32–54.

Li, C., B. Li, and K.-H. Tee (2020a). Are hedge funds active market liquidity timers? *International Review of Financial Analysis 67*, 101415.

Li, C., B. Li, and K.-H. Tee (2020b). Measuring liquidity commonality in financial markets. *Quantitative Finance 20*(9), 1553–1566.

Lui, A., J. Lee, P. F. Thall, M. Daher, K. Rezvani, and R. Basar (2023). A Bayesian feature allocation model for identifying cell subpopulations using CyTOF data. *Journal of the Royal Statistical Society Series C: Applied Statistics 72*(3), 718–738.

Luo, J., T. Klein, Q. Ji, and C. Hou (2022). Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HAR models. *International Journal of Forecasting 38*(1), 51–73.

Luo, J., K.-H. Tee, and B. Li (2017). Timing liquidity in the foreign exchange market: Did hedge funds do it? *Journal of Multinational Financial Management 40*, 47–62.

Lustig, H., N. Roussanov, and A. Verdelhan (2011). Common risk factors in currency markets. *Review of Financial Studies 24*(11), 3731–3777.

Madigan, D., J. York, and D. Allard (1995). Bayesian graphical models for discrete data. *International Statistical Review 63*(2), 215–232.

Maheu, J. M., T. H. McCurdy, and Y. Song (2012). Components of bull and bear markets: Bull corrections and bear rallies. *Journal of Business & Economic Statistics 30*(3), 391–403.

Mancini, L., A. Ranaldo, and J. Wrampelmeyer (2013). Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *Journal of Finance 68*(5), 1805–1841.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics 197*(1), 101–129.

Massacci, D. (2023). Testing for regime changes in portfolios with a large number of assets: A robust approach to factor heteroskedasticity. *Journal of Financial Econometrics 21*(2), 316–367.

Melvin, M. and M. P. Taylor (2009). The crisis in the foreign exchange market. *Journal of International Money and Finance 28*(8), 1317–1330.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2012). Carry trades and global foreign exchange volatility. *Journal of Finance 67*(2), 681–718.

Nandy, P., A. Hauser, and M. H. Maathuis (2018). High-dimensional consistency in score-based and hybrid structure learning. *The Annals of Statistics 46*(6A), 3151–3183.

Newman, M. (2010). *Networks: An Introduction*. Oxford: Oxford University Press.

Ni, Y., P. Müller, and Y. Ji (2020). Bayesian double feature allocation for phenotyping with electronic health records. *Journal of the American Statistical Association 115*(532), 1620–1634.

Omori, Y. and T. Watanabe (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis 52*(6), 2892–2910.

Opie, W. and S. J. Riddiough (2020). Global currency hedging with common risk factors. *Journal of Financial Economics 136*(3), 780–805.

Osinga, A. J., M. B. Schauten, and R. C. Zwinkels (2021). Timing is money: The factor timing ability of hedge fund managers. *Journal of Empirical Finance 62*, 266–281.

Otranto, E. and G. M. Gallo (2002). A nonparametric Bayesian approach to detect the number of regimes in Markov switching models. *Econometric Reviews 21*(4), 477–496.

Owyang, M. T., J. Piger, and D. Soques (2022). Contagious switching. *Journal of Applied Econometrics 37*(2), 415–432.

Owyang, M. T., J. Piger, and H. J. Wall (2005). Business cycle phases in US states. *Review of Economics and Statistics 87*(4), 604–616.

Park, A., D. Yang, X. Shi, and Y. Jiang (2010). Exporting and firm performance: Chinese exporters and the Asian financial crisis. *The Review of Economics and Statistics 92*(4), 822–842.

Poon, A. and D. Zhu (2022). A new Bayesian model for contagion and interdependence. *Econometric Reviews 41*(7), 806–826.

Raggi, D. and S. Bordignon (2012). Long memory and nonlinearities in realized volatility: A Markov switching approach. *Computational Statistics & Data Analysis 56*(11), 3730–3742.

Ranaldo, A. and P. S. de Magistris (2022). Liquidity in the global currency market. *Journal of Financial Economics 146*(3), 859–883.

Roberts, G. O. and S. K. Sahu (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society 59*(2), 291–317.

Ročková, V. and E. I. George (2016). Fast Bayesian factor analysis via automatic rotations to sparsity. *Journal of the American Statistical Association 111*(516), 1608–1622.

Rodríguez, J. (2008). Market timing: A global endeavor. *Journal of International Financial Markets, Institutions and Money 18*(5), 545–556.

Scholes, M. and J. Williams (1977). Estimating betas from nonsynchronous data. *Journal of Financial Economics 5*(3), 309–327.

Sialm, C. and Q. Zhu (2022). Currency management by international fixed income mutual funds. Nanyang Business School Research Paper No. 20-29. Forthcoming: Journal of Finance. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3732286.

Siegmann, A. and D. Stefanova (2017). The evolving beta-liquidity relationship of hedge funds. *Journal of Empirical Finance 44*, 286–303.

Sims, C. A., D. F. Waggoner, and T. Zha (2008). Methods for inference in large multiple-equation Markov-switching models. *Journal of Econometrics 146*(2), 255–274.

Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *Journal of Finance 53*(5), 1589–1622.

Smith, S. C. (2023). Structural breaks in grouped heterogeneity. *Journal of Business & Economic Statistics 41*(3), 752–764.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics 29*(5), 825–842.

Spirtes, P., C. N. Glymour, R. Scheines, and D. Heckerman (2000). *Causation, Prediction, and Search*. Cambridge, Massachusetts: MIT Press.

Tang, Z.-Z. and G. Chen (2019). Zero-inflated generalized Dirichlet multinomial regression model for microbiome compositional data analysis. *Biostatistics 20*(4), 698–713.

Teh, Y. and D. Gorur (2009). Indian buffet processes with power-law behavior. In B. Yoshua, S. Dale, L. John, W. Chris, and C. Aron (Eds.), *Advances in Neural Information Processing Systems*, Volume 22, pp. 1838–1846. Cambridge, MA: MIT Press.

Treynor, J. and K. Mazuy (1966). Can mutual funds outguess the market. *Harvard Business Review 44*(4), 131–136.

Tsamardinos, I., L. E. Brown, and C. F. Aliferis (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning 65*, 31–78.

Vivar, L. M., M. Wedow, and C. Weistroffer (2023). Burned by leverage? Flows and fragility in bond mutual funds. *Journal of Empirical Finance 72*, 354–380.

Warther, V. A. (1995). Aggregate mutual fund flows and security returns. *Journal of Financial Economics 39*(2-3), 209–235.

Zhang, P., T. Wang, and J. Yan (2022). Pagerank centrality and algorithms for weighted, directed networks. *Physica A: Statistical Mechanics and its Applications 586*, 126438.

Zhu, H., X. Su, W. You, and Y. Ren (2017). Asymmetric effects of oil price shocks on stock returns: Evidence from a two-stage Markov regime-switching approach. *Applied Economics 49*(25), 2491–2507.