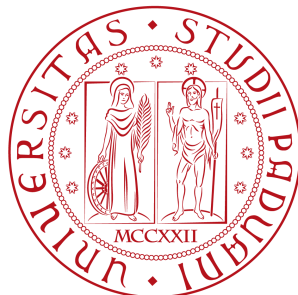


UNIVERSITÀ DEGLI STUDI DI PADOVA

Corso di Laurea Magistrale in SCIENZE STATISTICHE



Locally Adaptive Bayesian Covariance Regression.

Analysis of 2007-2012 global financial crisis.

Relatore:
Chiar.mo Prof. Bruno Scarpa
Dipartimento di Scienze Statistiche

Candidato:
Daniele Durante

Anno Accademico 2011/2012 - Sessione III

A mio padre e mia madre

Contents

Introduction	7
1 The Global Financial Crisis	9
1.1 What went wrong?	9
1.2 The burst of the bubble	11
1.3 A prolonged crisis	13
1.4 Policy responses	17
2 National Stock Market Indices	19
2.1 National Stock Market Indices (NSI)	19
2.2 Literature Overview	21
3 Locally Adaptive Bayesian Covariance Regression	27
3.1 Nested Gaussian Process	27
3.1.1 Simulation Smoother	29
3.2 Locally Adaptive Covariance Regression	31
3.2.1 Notation and Motivation	31
3.2.2 Latent Factor Model	31
3.2.3 Prior Specification	33
3.2.4 Hyperparameters interpretation	35
3.3 Posterior Computation	36
3.3.1 Main Steps	36
3.3.2 Detailed Algorithm	37
3.4 Online Updating	40
3.4.1 Online Updating Algorithm	41
4 Simulation Studies	43
4.1 Estimation Performance	43
4.2 Online Updating Performance	48

5	Application to National Stock Market Indices	51
5.1	Heteroscedastic Modeling of National Stock Market Indices . . .	52
5.2	Posterior results and economic facts	53
5.3	Updating and Predicting	58
	Discussion	61
	Bibliography	63

Introduction

This work arises from the combination of practical problems and theoretical interests. The attempt to provide a quantitative view on the evolution of the temporal and geo-economic relations between the major financial markets before and during the global financial crisis of 2007-2012 through the analysis of the historical quotes of the main National Stock Market Indices, motivates the search for statistical methodologies able to accommodate flexible dynamic structure of dependency and to answer the main issues and aims of multivariate financial time series analysis.

Large datasets and high frequency data, typical of this field, motivate the search for a formulation able to handle high-dimensional data through tractable computations and simple online updating and prediction procedures. Besides these issues, it is important for the model to allow the presence of missing values and to take also into account the possibility that covariances and variances change rapidly in particular financial scenarios. Finally, the model should consider that the marginal distributions of the returns are characterized by heavy tails. With this goals in mind, we develop a novel covariance stochastic process on continuous time with locally-varying smoothness to accommodate locally adaptive smoothing for the time-varying mean and covariance functions. This is accomplished by modifying recent method for Bayesian Covariance Regression to incorporate dictionary functions that are assigned nested Gaussian Process priors.

The use of latent factors in the formulation of the model and the possibility of representing the nested Gaussian Process priors through stochastic differential equations, allows us to answer the issue of high dimensionality and to develop a computationally tractable and efficient approach through MCMC, allowing also the implementation of an efficient online updating algorithm, particularly worthy in financial time series with data collected at high frequencies. Working with continuous time processes through a Bayesian

approach we can also easily handle missing values in the model estimation without the need for any imputation. Finally, multivariate analysis allows for significant improvements in terms of interpretation of the results and forecasts.

In Chapter 1, we provide a summarizing overview of the main events of the world financial crisis, which represents the main motivating problem for our proposed model. The focus is on the causes, the effects on the world finance scenario and policy responses, from the boom and burst of the U.S. housing bubble in 2006 until the recent happenings in the global economy. Chapter 2 motivates the use of National Stock Market Indices for the analysis of the dynamic dependence structure between financial markets before and during the crisis, defining the main features and problems that arise in the context of multivariate financial time series analysis and providing a summary of the literature on the main methods that address these issues. In Chapter 3 we provide a detailed description of the proposed model, with particular attention to (i) basic model structure, (ii) prior specification, (iii) posterior computation via MCMC and (iv) online updating algorithm. A couple of simulation studies implemented to assess the performance of our model and to compare the results to the main competing approach are reported in Chapter 4. Finally Chapter 5 shows the results of the application to our motivating example, highlighting the improvements provided by our approach in the analysis of the dependence structure of financial markets in relation to theory and economic facts, and considering also predictive performance.

All the following analysis has been performed using the free statistical software R.

Chapter 1

The Global Financial Crisis

The 2007-2012 global financial crisis has been the dominant theme in the recent history of the world economy and finance, and is often regarded as the worst crisis since the Great Depression of the thirties.

The direct consequences of this period of strong financial instability were the fall of the world stock markets and the collapse of large financial institutions or their bailout by national governments which had to come up with rescue packages to save their financial systems, even in the wealthier countries. At the outbreak of the U.S. housing bubble between 2006 and 2007 that resulted in the suffering of the real estate market, foreclosure and evictions, followed the 2008-2012 global recession which affected the entire world economy and was manifested through persistent high unemployment rates, declines in consumer confidence and wealth as well as a downturn in economic activity, leading to a slowdown in the growth of GDP in many countries as shown in Figure 1.1 and contributing to the European sovereign-debt crisis.

1.1 What went wrong?

Financial Crisis Inquiry Commission (2011) reported its findings about the causes of the U.S. financial crisis concluding that: *“the crisis was avoidable and was caused by: widespread failures in financial regulation, including the Federal Reserve (Fed) failure to stem the tide of toxic mortgages; dramatic breakdowns in corporate governance including too many financial firms acting recklessly and taking on too much risk; an explosive mix of excessive borrowing and risk by households and Wall Street that put the financial system on a collision course with crisis; key policy makers ill prepared for the crisis, lack-*

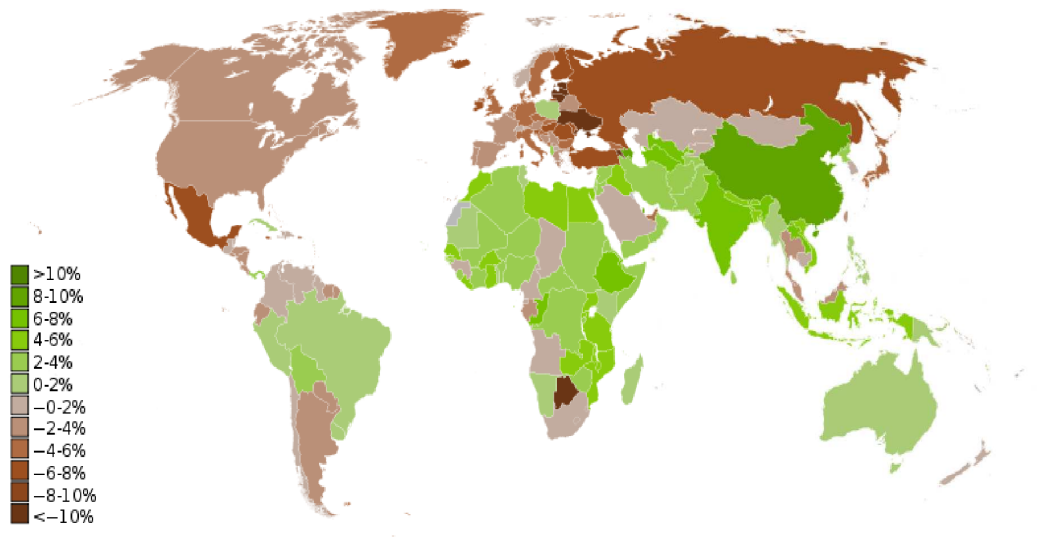


Figure 1.1: World map of the real GDP growth rates for 2009. Source: CIA world factbook.

ing a full understanding of the financial system they oversaw; and systemic breaches in accountability and ethics at all levels". Undoubtedly, the crisis was triggered by the coexistence of a complex system of causes such as easy credit conditions, the lack of proper regulation able to keep pace with the increasing importance of investment banks and hedge funds (shadow banking system), as well as the affirmation of new financial instruments which derived from the housing market.

Referring to Taylor (2009) the key factor behind the boom and burst of the housing bubble, that caused the outbreak of the recent financial crisis, relates to the monetary excesses stimulated by the unusually low interest rates decisions (around 1%) of the Federal Reserve to soften the effects of the 2000 Dot-com Bubble as well as to face the risk of deflation. Krugman (2002) argues that Fed *"needs to create a housing bubble to replace the Nasdaq bubble"*, to emphasize how the actual result of this loose fitting monetary policy was to fuel housing market instead of business investment.

A further incentive to the growth of the real estate market was represented by the growing demand for financial assets by foreign countries with high savings rates. This additional influx of "saving glut" (Bernanke, 2007), together with an underestimation of the risk of mortgage caused by optimistic forecasts on the expansion of the real estate market via backward-looking models, and, finally, the increasing competition between mortgage lenders

for revenue and market share, stimulated the decline of mortgage standards and risky loans proliferated between 2004 and 2007. Even the Government Sponsored Enterprises (GSE¹): Fannie Mae and Freddie Mac, which prior to 2003 maintained conservative underwriting standards, relaxed them in order to compete with the private banks. The direct result of this predatory lending was the growth of new risky loans with higher interest rates and less favorable terms to people who may have difficulty maintaining the repayment schedule (subprime mortgage), which rose from the historical 8% to approximately 20% from 2004 to 2006.

A further complication of this “originate and distribute” banking model, in which mortgage were pooled, tranced, and then resold via securitization (Brunnermeier, 2009), was the creation of securities of great complexity that stimulated large capital inflows from abroad, facilitating the creation of a wide network of dependencies between financial operators worldwide. Figure 1.2 shows an example of the complexity and financial innovation of two types of securities which derived their value from mortgage payments. More specifically, residential mortgage-backed securities (RMBS) and collateralized debt obligations (CDO) can be seen as tranches, characterized by different risk and returns, of diversified portfolios composed of mortgages and other loans, that the banks sold to investors with different “appetites”. The possibility to insure these obligations through credit default swaps (CDS) together with the complexity of these securities, contributed to an underestimation of risk by rating agencies. As a result the CDO issuance rose from an estimated 20 billion dollars in the first quarter of 2004 to its peak of over 180 billion dollars in the same period of 2007, increasing significantly the leverage of many banks (including Lehman Brothers, Bear Stearns, Merrill Lynch, Goldman Sachs and Morgan Stanley) and subjecting them to a real risk of liquidity in the event of a fall of the housing market. Specifically, the vulnerability of these entities were linked to the maturity mismatch, as they were borrowing short-term in liquid markets to purchase long-term, illiquid and risky assets.

1.2 The burst of the bubble

Witter, in August 2006, wrote in Barron’s magazine that “*a housing crisis*

¹Private enterprises with government support, operating in financial services.

Residential Mortgage-Backed Securities

Financial institutions packaged subprime, Alt-A and other mortgages into securities. As long as the housing market continued to boom, these securities would perform. But when the economy faltered and the mortgages defaulted, lower-rated tranches were left worthless.

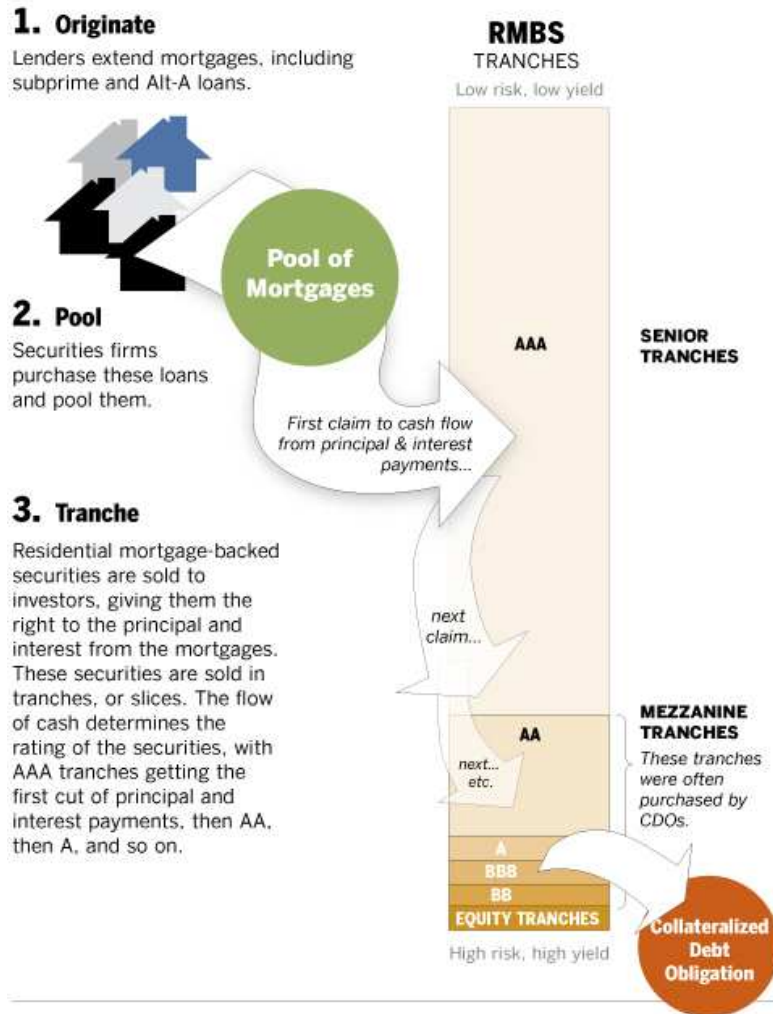


Figure 1.2: Diagram of residential mortgage-backed securities (RMBS) or collateralized debt obligations (CDO). Source: The Financial Crisis Inquiry Report.

approaches”, and noted that the median price of new homes in the U.S. had dropped almost 3% since January 2006. Indeed, once the initial grace period ended, subprime borrowers proved unable to pay their mortgage payments (this is an on-going crisis) increasing foreclosures and the supply of homes for sale nearly on 1.3 million properties in 2007. This placed downward pressure on housing prices, resulting in many owners holding negative equity: a mort-

gage debt higher than the value of the property. Borrowers in this situation had an incentive to default on their mortgages as a mortgage is typically non recourse debt secured against the property, triggering a vicious cycle at the base of the housing bubble burst between 2006 and 2007. As direct result, the declining mortgage payments caused the values of securities tied to U.S. real estate pricing to plummet, eroding the net worth and financial health of banks globally.

In an interview with the Financial Times, in July 2007, Citigroup CEO Chuck Prince, said about the subprime mortgage crisis: *“When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you’ve got to get up and dance. We are still dancing.”* If we can have some doubts about the fact that in early summer of 2007 finance was still dancing, certainly the active phase of the crisis manifested from August 2007 as a liquidity crisis due to the over-leveraged world financial institutions which were incurring in significant losses from their risky investments on the mortgage supply chain. The decline in the value of mortgage-backed securities held by these companies together with the inability to secure new funding in the credit markets, led to investor panic and bank run. One of the first victims was the highly leveraged British bank Northern Rock in mid-September 2007 whose problems proved to be an early indication of the incoming crisis. Subsequently over 100 mortgage lenders failed, were acquired under duress, or were subject to government takeover during 2007 and 2008. The five largest U.S. investment banks, either went bankrupt (Lehman Brothers), were taken over by other companies (Bear Stearns and Merrill Lynch), or were bailed-out by the U.S. government (Goldman Sachs and Morgan Stanley) during 2008. Even the GSE: Fannie Mae and Freddie Mac, were placed into receivership in September 2008. The results were the downturns in stock markets around the world, a worldwide slowdown in the economic growth and the rapid contagion of foreign banking and financial systems (especially in Europe).

1.3 A prolonged crisis

The systemic imbalances that followed the subprime mortgage crisis led to the 2008-2012 global recession which affected the entire world economy, with higher detriment in some countries than others. In most cases the recession was manifested through a sharp drop in international trade, increasing un-

employment rates, low consumer confidence, escalating sovereign-debt crisis (in particular in Europe), inflation, and rising petroleum and food prices.

In the U.S., beside the Dow Jones average's fall of more than 50% over a period of 17 months between October 2007 and March 2009, the crisis struck heavily wealth, consumptions and business investments. Figure 1.3 shows how since 2007 real median household income has declined for all race and Hispanic-origin groups. The decline was 5.4 percent for Non-Hispanic-White household income, 10.1 percent for Black household income, 7.5 percent for Asian household and 7.2 percent for Hispanic. The peak in the U.S. unemployment rate from 2008 in Figure 1.4 represents a further evident effect of the crisis that expands through a year-on-year decline in capital investment since the final quarter of 2006, and a plunge in the volumes of international trade in the second half of 2008, resulting in the contraction of the Real Gross Domestic Product (GDP) in the third quarter of 2008, and a sharp drop in early 2009.

The increasing interconnection between world financial markets and institutions, generated a contagion effect that took shape through the rapid development and spread of the crisis into a global economic shock. Referring to these issue Baily and Elliott (2009) reported that: *"The U.S. economy has been spending too much and borrowing too much for years and the rest of the world depended on the U.S. consumer as a source of global demand."* As a result, U.S. recession and the increased savings rate of U.S. consumers were accompanied by dramatic declines in various stock indices and in economic growth elsewhere. Declines in GDP at annual rates for the first quarter of 2009 were -14.4% for Germany, -15.2% for Japan, -7.4% for the UK, -9.8% for the Euro area and -21.5% for Mexico. In addition to these, also different states of Southeast Asia such as Taiwan, Singapore, Hong Kong and India suffered the contagion effect during the third and fourth quarter of 2008. Finally, of the largest economies in the world by GDP, China avoided the recession in 2008 experiencing a growth between 5% and 8% which, however, represents a slowdown compared to the 10% growth rates of the past five years.

A further downside of the crisis in the Eurozone is manifested through the ongoing European sovereign debt crisis. The transfer of private debts arising from the burst of the housing bubble to the already high sovereign debt as a result of banking system bailouts, together with the structural problem of Eurozone system based on monetary union without fiscal union, caused

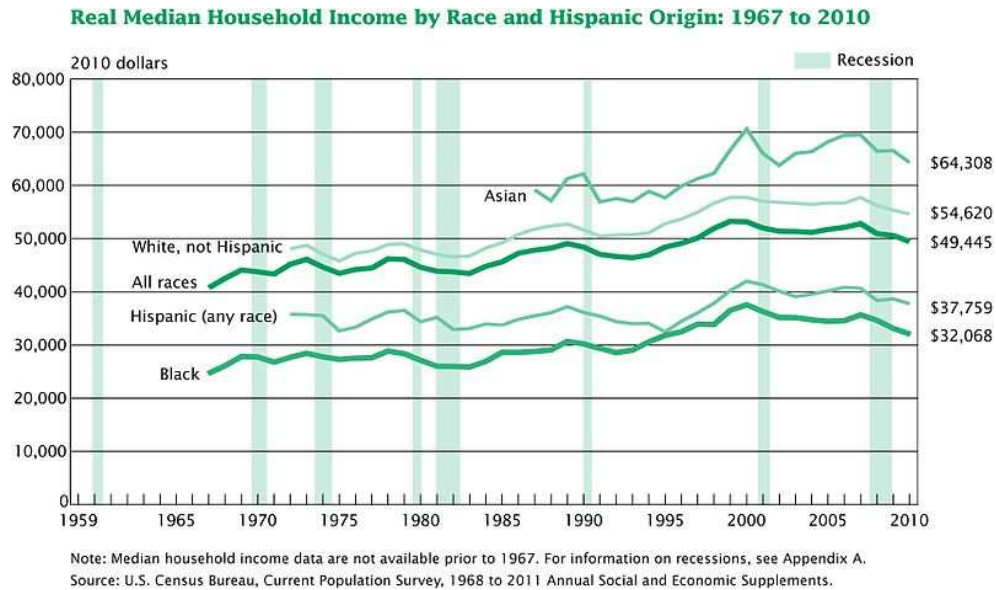


Figure 1.3: Real median Household income by Race from 1967 to 2010, in 2010 dollars. Source: U.S. Census Bureau, Current Population Survey, 1968 to 2011 Annual Social and Economic Supplements.

the impossibility for some European countries to re-finance their government debt without the assistance of third parties. The results were an increasing downgrade of the sovereign debt of many European countries by the credit rating agencies since early 2010, and repeated speculative attacks against euro by financial speculators and hedge funds that contributed to worsen the crisis. The countries most affected by the sovereign debt crisis were Greece, Portugal, Ireland and in June 2012 also Spain became a matter of concern because of its difficulties to access capital markets with rising interest rates.

The greek debt was the first to raise concerns at the beginning of 2010. Unsustainable public sector wages, pension commitments and high percentage of debt in the hands of foreign creditors generated a structural deficit that made it necessary continuous requests of loans from the EU and International Monetary Fund (IMF) to cover its financial needs since the early 2010. The result was the downgrade of Greece's sovereign debt from Standard & Poor's to BB+ or "junk", leading to the decline of Stock markets worldwide. Austerity measures that followed from mid-2010, through an increase in taxation were met with great anger by the Greek public and contributed to a worsening of the Greek recession, resulting in a decline of the Greek GDP in 2011 of -6.9% together with a growth of the unemployment rate.

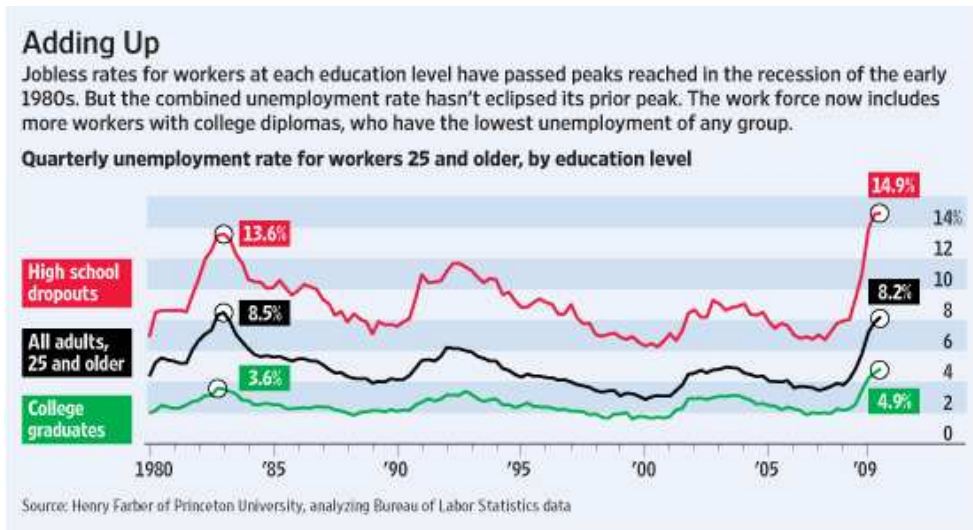


Figure 1.4: U.S. Quarterly unemployment rate for workers 25 and older, from 1980 and 2009, by education. Source: Henry Farber of Princeton University, analyzing Bureau of Labor Statistics data.

Unlike Greece, the Irish debt crisis was mainly based on the state guaranteeing the six main Irish-based banks that had invested heavily in real estate during the housing bubble, instead of on government over-spending. The burst of the bubble between 2006 and 2007 and the subsequent economic collapse in 2008, led the federal budget to a deficit of 32% GDP in 2010, the highest in the history of the Eurozone. The acute phase of the Irish debt crisis came in November 2010 when Ireland called for the intervention of the EU, the IMF and bilateral loans with non-euro countries through a "bailout" agreement. However, despite these measures, In April 2011 Moody's downgraded the banks' debt to junk status.

On May 2011 the Portugal became the third European country to receive emergency fund through a bailout package equally split between the European Financial Stabilization Mechanism, the European Financial Stability Facility, and the International Monetary Fund. The underlying causes of the Portuguese debt crisis are related to decades-long governmental overspending in unnecessary external consultancy, top management and head officer bonuses and wages, redundant public servants together with risky investments in housing bubble that led the country close to bankruptcy by 2011 and caused the cutting of Portugal's credit rating to junk status by Moody's on July 2011.

More recently, attention has focused on Spanish debt, whose crisis is

linked to high investments of Spanish banks in long-term mortgage during Spanish Real Estate boom. The subsequent building market crash weakened private banks, requiring government bailouts. As a result in May 2012, Bankia received a 19 billion euro bailout, and unemployment rates grew dramatically from March 2012.

The possibility of spread of the debt crisis to other states remains concrete. Among these it is useful to include Italy with a debt of almost 120% of GDP as well as lower economic growth with respect to EU average, and United Kingdom with its highly leveraged financial industry.

1.4 Policy responses

Since the beginning of the crisis several measures have been launched by central banks and governments to cover the risk of bankruptcy of their financial systems, prevent the growth of the debt crisis and address the problem of recession. To answer these goals, strong fiscal stimulus, monetary policy expansion and bank rescue packages together with austerity measures of spending cuts and taxation, have been carried out by different countries worldwide.

Beside the rescue of several financial institutions during the acute phase of the crisis, the U.S. responded through a continuous sharp reduction in the federal funds rate from the 5.25% in August 2007 to the 0% - 0.25% since December 2008; with the introduction of the term auction facility (TAF) in December 2007 to allow banks to borrow directly from the Fed and, finally, with Temporary Cash Infusions through 100 billion dollars defined in the Economic Stimulus Act of February 2008. Recalling Taylor (2008) the actual results of these policies were not in line with expectations as the three month LIBOR-OIS spread² remained substantially unchanged showing that a less expensive liquidity does not mean more confidence; in addition the attempt to encourage consumption through the Economic Stimulus Act resulted in an increase in savings instead of aggregate consumption, highlighting a growing pessimism.

Additional measures were taken later, with the commitment of the government to buy the toxic assets through the Troubled Asset Relief Program

²Difference between the 3-month London Interbank Offered Rate (LIBOR) and the 3-month Overnight Index Swap (OIS) rate.

(TARP) with a significant impact on the spread LIBOR-OIS, especially after the second announcement on October 13 of 2008. In addition to these, a more structured response to the problems of the recession through a program of rescue and creation of jobs, relief for those most affected by the crisis and investments in education, research, health, infrastructure and new energies came from the American Recovery and Reinvestment Act (ARRA) signed into law on February 17, 2009, by President Barack Obama.

Similarly to the U.S., the other countries responded to the financial crisis and the subsequent recession through low-rates monetary policy and stimulus plans aimed at economic recovery. In autumn 2008 states in Asia and Pacific reacted to the outbreak of the crisis through a cut in interest rates (in China for the first time since 2002) and stimulus packages particularly important in India, Japan, Australia, Indonesia, Taiwan and finally, China which announced a plan accounting for 16% of GDP, much higher than that of other countries that were equivalent to about 3%.

In addition to the monetary policies of the European Central Bank to address the growing financial and economic crisis, the Eurozone had to implement additional measures to prevent the sovereign debt crisis in their own countries. This led to policies by individual countries through austerity measures to cope with the increasing amount of sovereign debt, but also to agreement between Eurozone leaders to prevent the collapse of member economies. The results were an increasing taxation and spending cuts (particularly relevant in Greece, Ireland, Portugal, Spain and Italy) with side-effect on economic recession, and the introduction of rescue packages from the EU since early 2010 to ensure financial stability across Europe. The European Financial Stability Facility (EFSF) on 9 May 2010 and the European Financial Stabilization Mechanism (EFSM) on 5 January 2011 represent the specific temporary legal European Union funding vehicle (to be succeeded in July 2012 by the permanent European Stability Mechanism (ESM)) to provide financial assistance to Eurozone states in difficulty. Further measures to restore confidence in Europe, boost the economy and prevent other crisis in the Eurozone have been launched through the agreement to introduce a European Fiscal Compact including the Stability and Growth Pact aiming at straightening the debt rules with penalties in case of breaches.

Chapter 2

National Stock Market Indices

Spurred by the increasing growth of interest in the causes and consequences of the financial crisis on the worldwide financial and economic systems, we aim to provide a quantitative analysis of the dynamic dependence structure between financial markets in the main countries, and in its features during the crisis that have followed in recent years, through the joint study of the main National Stock Market Indices (NSI) between 2004 and 2012. Although only a window into a much larger problem with complex roots and ongoing effects, the multivariate analysis of these technical indicators designed through the synthesis of numerous data on the evolution of the various stocks, certainly represents an informative overview on the temporal and geo-economic changes in world financial market during the recent years.

2.1 National Stock Market Indices (NSI)

A National Market Index is a synthetic data calculated using statistical techniques for the construction of composite weighted price indices, which represents the performance of the stock market of a given nation. More specifically, referring to Gallo and Pacini (2002), most of Stock Market Indices represent composite price indices calculated through a weighted average of simple price indices relating a baskets of stocks, with reference to a base of time:

$$I_t = \frac{1}{f_t} \frac{\sum_{i=1}^n \frac{p_{i,t}}{p_{i,0}} W_{i,0}}{\sum_{i=1}^n W_{i,0}},$$

where n is the number of stocks in the basket on which the index is calculated, $p_{i,0}$ is the price of the i -th stock measured at the base time, $p_{i,t}$ is the price

of the i -th stock at time t , $W_{i,0}$ are the weights used, while f_t is a correction factor which takes into account changes over time in the life of the stocks like splits, recapitalizations and the change in the composition of the basket, ensuring the continuity of the index.

Typically the weighting is based on the value of capitalization of the business companies included in the calculation of the index at the base time, leading to $W_{i,0} = p_{i,0}q_{i,0}$, where $q_{i,0}$ represents the amount of stock i exchanged on the market at time 0. As a result

$$I_t = \frac{1}{f_t} \frac{\sum_{i=1}^n p_{i,t}q_{i,0}}{\sum_{i=1}^n p_{i,0}q_{i,0}},$$

showing a similar structure to the index of Laspeyres (see e.g., Predetti, 2006). Example of value weighted National Stock Market Indices are S&P 500, NASDAQ Composite, FTSE 100, CAC 40, DAX 30 and FTSE MIB. A special case is that of DOW JONES and NIKKEI 225 indices which represent the most important price-weighted indices of the few remaining. More specifically, price-weighted indices are obtained as the average of prices of n stocks, possibly corrected through a correction factor f_t :

$$I_t = \frac{1}{f_t} \sum_{i=1}^n \frac{p_{i,t}}{n}.$$

Note that, unlike the previous indices, these depend on the currency and do not take into account the importance of the stocks in terms of market capitalization and average trading volume.

The construction and the joint analysis of such indices can be traced back to the field of multivariate financial time series, with the specific features of financial data. The non-stationary behavior of the series of indices typically assumed as a *random walk process*, together with the different bases of currency for price weighted indices, and time for value weighted indices, motivates the use of logarithmic returns

$$y_{j,t} = \log(I_{j,t}/I_{j,t-1}), \quad (2.1)$$

where $I_{j,t}$ is the value of the j -th National Stock Market Index at time t . Beside this, large datasets and high frequency data, typical of this field, require models able to handle high-dimensional data through tractable computations and simple online updating and prediction procedures. Referring to Fama (1965), another aspect to consider is the empirical evidence related

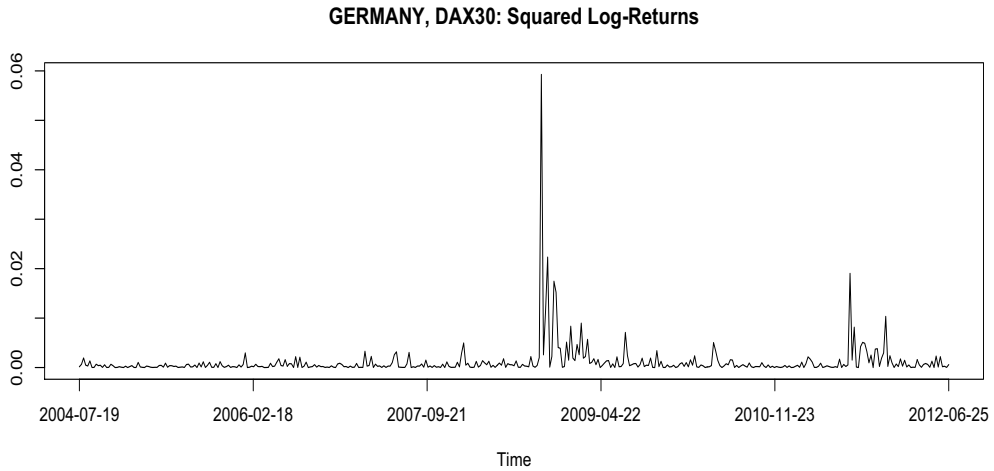


Figure 2.1: Squared Log>Returns of DAX30, using weekly data from 2004/07/19, to 2012/06/25.

to variance and covariance non-stationarity, manifesting through a strong heteroscedasticity characterized by the possibility of rapid changes in the dynamic evolution of volatilities and co-volatilities. Figure 2.1 shows an example where volatilities, described by the squared log-returns of the German stock market index (DAX30), change dramatically, motivating the search for fully flexible models able to capture such rapid changes which may occur during financial crisis. Finally, the possibility to easily handle missing data in fitting the model, together with the ability to accommodate heavy tails showed by the marginal distribution of logarithmic returns, represent additional key aspects to further improve the performance of the models in terms of accuracy of analysis and computational tractability.

2.2 Literature Overview

There is a rich literature on univariate stochastic volatility modeling, with an increasing emphasis on multivariate generalizations. One popular approach estimates the $p \times p$ time-varying covariance matrix Σ_t via an exponentially weighted moving average (EWMA; see, e.g., Tsay, 2005). Specifically, given a set of zero mean observations $\{y_1, \dots, y_T\}$ where $y_t \in \mathbb{R}^p$, the covariance matrix Σ_t can be recursively estimated from the equation

$$\hat{\Sigma}_t = (1 - \lambda)y_{t-1}y_{t-1}^T + \lambda\hat{\Sigma}_{t-1},$$

where $0 < \lambda < 1$ represents a decay factor and T sufficiently large such that $\lambda^{T-1} \approx 0$. In particular, the higher λ , the more weight is given to the observations that are more distant, leading to quite different final results. Besides its simplicity, the model has, however, some key limitations: for the estimation of λ , EWMA assumes the conditional normal distribution for the observations without considering the possibility of heavy tails (Fama, 1965) when financial time series are analyzed; to overcome this problem Guermat and Harris (2002) suggest a general power EWMA model based on the Generalized Error Distribution (GED). Another important limitation is referred to the use of a single time-constant smoothing parameter λ , with extensions to accommodate locally-varying smoothness λ_t not straightforward due to the need to maintain positive semidefinite Σ_t at every time. This issue leads to restrictive dynamics induced by the assumption of equal reaction of volatility to different economic events and the persistence in volatility for all assets considered in the model.

A generalization of the exponentially weighted moving-average approach is represented by the Diagonal VEC Model (Bollerslev, Engle and Wooldrige, 1988) in which the conditional variances and covariances are written as a linear combination of lagged conditional variances and covariances and lagged squared observations and their cross-product. More specifically a DVEC(m, s) follows the equation

$$\Sigma_t = A_0 + \sum_{j=1}^m A_j \odot (y_{t-j} y_{t-j}^T) + \sum_{r=1}^s B_r \odot \Sigma_{t-r},$$

where A_j and B_r are $p \times p$ symmetric parameter matrices and \odot denote the element-by-element multiplication (Hadamard product). The generality and flexibility of this formulation encounters a series of theoretical and computational disadvantages: conditions for Σ_t to be positive definite for all t are only sufficient and rather restrictive, estimation is computationally demanding involving a large number of parameters $(m + s + 1)p(p + 1)/2$ and models do not allow for interaction between different variances and covariances.

BEKK model (Engle and Kronos, 1995) answers the issues of positive definiteness and dynamic dependence between the volatility series, suggesting a model for the conditional covariance matrix in which

$$\Sigma_t = A_0 A_0^T + \sum_{j=1}^m A_j (y_{t-j} y_{t-j}^T) A_j^T + \sum_{r=1}^s B_r \Sigma_{t-r} B_r^T,$$

where A_0 is a lower triangular matrix and A_j and B_r are $p \times p$ matrices. The decomposition of the constant term into a product of two triangular matrices ensures positive definiteness of Σ_t by construction; however as for DVEC, also BEKK suffers from the curse of dimensionality limiting its applicability in large p settings. To address this problem Ding (1994) introduces the principal component GARCH (PC-GARCH), later depth by Alexander (2001) under the name O-GARCH. The dimensionality reduction is obtained by assuming a latent factor model for the observed variables $y_t = \Gamma f_t$, where Γ is a $p \times m$ matrix with orthogonal columns, and the latent factors in the $m \times 1$ vector process f_t are conditionally uncorrelated with GARCH conditional volatilities. The resulting time varying covariance matrix is

$$\Sigma_t = \Gamma D_t \Gamma^T,$$

where D_t represents the $m \times m$ diagonal matrix with conditional factor variances on the diagonal. This model shows good performance in highly correlated systems where the volatility and co-volatility of the observed variables can be explained by a few latent factors, however if the data are weakly correlated identification problems may arise. Another important limitation refers to the assumption of orthogonality of Γ . To overcome these issues van der Wiede (2002) introduces GO-GARCH, where the number of factors equals the number of series and the transformation matrix Γ is assumed to be invertible instead of orthogonal; however the formulation increases dramatically the number of parameters. Key assumptions of factor volatility models are that Γ is time-constant and the latent factors are uncorrelated. This ensure simplicity of analysis but permit only limited evolution of the the time-varying covariance matrix Σ_t .

To complete the introduction on multivariate GARCH it is important to highlight that such models fall far short of our goal of allowing Σ_t to be fully flexible with the dependence between Σ_t and $\Sigma_{t+\Delta}$ varying with not just the time-lag Δ but also time. In addition, these models do not handle missing data easily and tend to require long series for accurate estimation (Burns, 2005).

An alternative to multivariate GARCH is represented by stochastic volatility (SV) models (Harvey et al., 1994) where the volatility process is random rather than being a deterministic function of past returns. In general, SV assume

$$\Sigma_t = A \Gamma_t A^T,$$

with A real and $\Gamma_t = \text{diag}(\exp(h_{1t}), \dots, \exp(h_{pt}))$, where h_{it} , for $i = 1, \dots, p$, are independent autoregressive processes. The equation for Σ_t results from a state space formulation of the model which allows the implementation of Kalman filter and smoother for the Quasi Maximum Likelihood (QML) estimation (Ruiz, 1994) and enables to handle missing values. See Chib et al. (2009) for major details on such approaches. The main disadvantages of these models are that time variation of covariances are restricted by sole variation in variances, together with the fact that the conditional volatility of an asset depends only on its past variances and not on covariances with other assets.

To address the SV lack of flexibility, Philipov and Glickman (2006a) develop a multivariate stochastic volatility model in which the time-varying covariance structure follows a stochastic process based on Wishart distribution. More formally they consider

$$\begin{aligned}\Sigma_t^{-1} | \Sigma_{t-1} &\sim W(n, S_{t-1}), \\ S_{t-1} &= 1/n(A^{1/2})(\Sigma_{t-1}^{-1})^\nu(A^{1/2})^T.\end{aligned}$$

Greater flexibility has as its counterpart a challenging posterior computation and lack of simplicity in description of marginal distributions. Moreover in the case of large p setting, working with $p \times p$ covariance matrix could generate computational problems (an issue faced by Philipov and Glickman (2006b) through the application of the approach to the matrix of variance and covariance of a vector of latent factors in a standard model factor analysis). Finally it could not be optimal to control intra and inter-temporal covariance relationships through single parameters (n and ν). More recently Prado and West (2012) address the problem of posterior computation for dynamic covariance matrices via discounting methods that maintain simple update equations as new observations are added. In particular they assume

$$\begin{aligned}\Sigma_{t-1}^{-1} | y_{1:t-1}, \beta &\sim W(h_{t-1}, D_{t-1}^{-1}), \\ D_t &= \beta D_{t-1} + y_t y_t^T, \\ h_t &= \beta h_{t-1} + 1.\end{aligned}$$

This implies $\Sigma_t^{-1} | y_{1:t-1}, \beta \sim W(\beta h_{t-1}, (\beta D_{t-1})^{-1})$. The advantage of this formulation is that the update with a new observation y_t is conjugate, maintaining Wishart posterior; however the model constrains $h_t > p - 1$ and, therefore, $\beta > (p - 2)/(p - 1)$. Moreover the model restricts the evolution of

the covariance to be stationary and slowly-changing, this could be a key limitation in financial applications where covariances could change dramatically in particular economic scenarios.

Accommodating changes in continuous time is also important to avoid having the model being critically dependent on the time scale, with inconsistent models obtained as time units are varied. With reference to this, it is important to stress how the above-mentioned models assume that the observations evolve in discrete time on a regular grid, leading to discrete-time covariance dynamics. Our emphasis is instead on developing continuous time stochastic processes for time-varying covariance matrices, which accommodate locally-varying smoothness.

In this regard, a highly relevant development refers to Bayesian Nonparametric Covariance Regression (BNCR) model of Fox and Dunson (2011), which defines the covariance matrix as a regularized quadratic function of time-varying loadings in a latent factor model, characterizing the latter as a sparse combination of a collection of unknown Gaussian process (GP) dictionary functions. More specifically they assume

$$\text{cov}(y_i|t_i = t) = \Sigma(t) = \Theta\xi(t)\xi(t)^T\Theta^T + \Sigma_0, \quad t \in \mathcal{T} \subset \mathbb{R}^+,$$

where Θ is a $p \times L$ matrix of coefficients, $\xi(t)$ is a time-varying $L \times K$ matrix with unknown continuous dictionary functions entries $\xi_{lk} : \mathcal{T} \rightarrow \mathbb{R}$ that are modeled through independent Gaussian Process (GP) random functions, and finally Σ_0 is a positive definite diagonal matrix. The previous equation for $\Sigma(t)$ results from the marginalization of η_i in the latent factor model

$$y_i = \Theta\xi(t_i)\eta_i + \epsilon_i, \quad (2.2)$$

with the latent factor $\eta_i \sim N_K(0, I_K)$ and $\epsilon_i \sim N_p(0, \Sigma_0)$. A further generalization of the model allows also the possibility of including the nonparametric mean regression by assuming

$$\eta_i = \psi(t_i) + \nu_i, \quad (2.3)$$

where $\nu_i \sim N_K(0, I_K)$ and $\psi(t)$ is a $K \times 1$ matrix with unknown continuous entries $\psi_k : \mathcal{T} \rightarrow \mathbb{R}$ that can be modeled in a related manner to the dictionary elements in $\xi(t)$. The induced mean of y_i conditionally on $t_i = t$, and marginalizing out ν_i is then:

$$\mu(t) = \Theta\xi(t)\psi(t). \quad (2.4)$$

Although their approach provides a continuous time and highly flexible model that accommodates missing data and scales to large p , there are two limitations motivating this work. Firstly, their proposed covariance stochastic process assumes a stationary dependence structure, and hence tends to under-smooth during periods of stability and over-smooth during periods of dramatic change. Secondly, the well known computational problems with usual GP regression are inherited, leading to difficulties in scaling to long series and issues in mixing of MCMC algorithms for posterior computation.

Chapter 3

Locally Adaptive Bayesian Covariance Regression

Our focus is on developing a Locally Adaptive Bayesian Covariance Regression (LABNCR) model that allows covariance and mean to vary flexibly over time, and additionally accommodate locally adaptive smoothing. Locally adaptive smoothing to accommodate varying smoothness in a trajectory over time has been well studied, but such approaches have not yet been developed for time-varying covariance matrices and multivariate nonparametric mean regression, to our knowledge. To address this gap, we generalize recently developed methods for Bayesian covariance regression to incorporate random dictionary elements with locally varying smoothness. Using a differential equation representation, we additionally develop a fast computational approach via MCMC, with online algorithms also considered.

The detailed description of the proposed model is provided in Section 3.2, while Section 3.1 introduces the problem of locally adaptive modeling via nested Gaussian Process (nGP) prior, as it has been outlined by Zhu and Dunson (2012) with reference to nonparametric mean regression in univariate models. We consider a similar approach for time-varying covariances by also allowing for the multivariate case.

3.1 Nested Gaussian Process

Zhu and Dunson (2012) consider the problem of nonparametric mean regres-

sion in the univariate model

$$y_i = U(t_i) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_\epsilon^2), \quad (3.1)$$

where

$$U(t) = \mathbb{E}[y_i | t_i = t], \quad t \in \mathcal{T} \subset \mathbb{R}^+, \quad U : \mathcal{T} \rightarrow \mathbb{R},$$

is an unknown continuous mean function to be estimated at $\mathcal{T}_o = \{t_1, \dots, t_T\} \subset \mathcal{T}$. To allow the smoothness of U to vary locally as a function of t , a nested Gaussian Process prior is assumed. More specifically the nGP prior for U specifies a GP for U 's m th-order derivative $D^m U$, centered on a local instantaneous mean function $C : \mathcal{T} \rightarrow \mathbb{R}$ which represents an higher-level GP, that induces adaptivity to locally-varying smoothing. Formally, both GP are defined by the following stochastic differential equations (SDEs):

$$D^m U(t) = C(t) + \sigma_U W_U(t), \quad m \in \mathbb{N}, \quad m \geq 2, \quad \sigma_U \in \mathbb{R}^+, \quad (3.2)$$

$$D^n C(t) = \sigma_C W_C(t), \quad n \in \mathbb{N}, \quad n \geq 1, \quad \sigma_C \in \mathbb{R}^+, \quad (3.3)$$

where $W_U : \mathcal{T} \rightarrow \mathbb{R}$ and $W_C : \mathcal{T} \rightarrow \mathbb{R}$ are two independent Gaussian white noise processes with zero mean and covariance function defined by a delta function $\delta(t - t')$. The initial value of U and its derivatives up to order $m - 1$ at t_1 are assumed $[U(t_1), D^1 U(t_1), \dots, D^{m-1} U(t_1)]^T \sim N_m(0, \sigma_\mu^2 I_m)$; the same goes for the initial value $C(t_1)$ and its derivatives up to order $n - 1$ leading to $[C(t_1), D^1 C(t_1), \dots, D^{n-1} C(t_1)]^T \sim N_n(0, \sigma_\alpha^2 I_n)$. In addition the initial values, the corresponding derivatives and the white noise Gaussian Processes are assumed mutually independent. This formulation naturally induces a prior for U whose smoothness, measured by $D^m U$, is expected to be centered on a continuous time stochastic process C . Finally, to conclude prior specification, they assume independent Inverse Gamma priors for σ_ϵ^2 , σ_U^2 and σ_C^2 .

The markovian property implied by SDEs in equations (3.2) and (3.3) allows to obtain a simple state space formulation for the assumed nGP prior, particularly worthy for posterior computation. Specifically, for $m = 2$ and $n = 1$ (this can be easily extended for higher m and n), and for $\delta_i = t_{i+1} - t_i$ sufficiently small, the nGP for U along with its first order derivative U' and C , follow the approximated state equation

$$\begin{bmatrix} U(t_{i+1}) \\ U'(t_{i+1}) \\ C(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U(t_i) \\ U'(t_i) \\ C(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i,U} \\ \omega_{i,C} \end{bmatrix} \quad (3.4)$$

where $[\omega_{i,U}, \omega_{i,C}]^T \sim N_2(0, V_i)$, with $V_i = \text{diag}(\sigma_U^2 \delta_i, \sigma_C^2 \delta_i)$. Note that this formulation allows continuous time and irregular grid of observations over t by relating the latent states at $i + 1$ to those at i through the distance between t_{i+1} and t_i , where i represents a discrete order index and $t_i \in \mathcal{T}$ the time observation related to the i -th observation.

For posterior computation, note that, given σ_ϵ^2 , σ_U^2 and σ_C^2 , the observation equation (3.1) combined with the above state equation, forms a state space model for which the vector of latent states $[U(t_i), U'(t_i), C(t_i)]^T$ with $t_i \in \mathcal{T}$, can be efficiently updated using a simulation smoother algorithm (Durbin and Koopman, 2002). Conditionally to $[U(t_i), U'(t_i), C(t_i)]^T$, posterior samples of σ_ϵ^2 , σ_U^2 and σ_C^2 can be easily obtained by drawing from the standard conjugate Inverse Gamma distribution.

3.1.1 Simulation Smoother

Considering the importance of the simulation smoother for posterior computation under the assumption of nGP prior (which plays a key role also in our model), we summarize below the main features of the one proposed by Durbin and Koopman (2002) for drawing samples from the conditional distribution of latent state vectors in a state space model, given the observations. This simple and computationally efficient technique represent a key element for posterior computation in our model, allowing us to reduce the GP computational burden involving matrix inversions from $O(T^3)$ to $O(T)$, with T denoting the length of the time series.

To introduce the simulation smoother, consider the state space model

$$\begin{aligned} y_i &= Z_i \beta_i + \epsilon_i, & \epsilon_i &\sim N(0, H_i), \\ \beta_{i+1} &= T_i \beta_i + R_i \omega_i, & \omega_i &\sim N(0, Q_i), \\ \beta_1 &\sim N(b_1, P_1), & i &= 1, \dots, T, \end{aligned} \tag{3.5}$$

where y_i is a $p \times 1$ vector of observations, β_i an $m \times 1$ vector of latent states, and ϵ_i and ω_i are vectors of disturbances. Z_i , T_i , R_i , H_i and Q_i are matrices assumed to be known and b_1 and P_1 represent the initial conditions.

The algorithm for generating draws of the state vector $\beta = [\beta_1^T, \dots, \beta_T^T]^T$ from the conditional density $p(\beta|y)$ with $y = [y_1^T, \dots, y_T^T]$, iterates between standard Kalman filtering and state smoothing algorithms. The first outputs the one step ahead state prediction $b_{i+1} = E[\beta_{i+1}|y_1, \dots, y_i]$ and prediction variance $P_{i+1} = \text{var}(\beta_{i+1}|y_1, \dots, y_i)$ from the following recursive equation

based on the theoretical results of Multivariate Normal distribution:

$$\begin{aligned} v_i &= y_i - Z_i b_i, & F_i &= Z_i P_i Z_i^T + H_i, \\ K_i &= T_i P_i Z_i^T F_i^{-1}, & L_i &= T_i - K_i Z_i, \\ b_{i+1} &= T_i b_i + K_i v_i, & P_{i+1} &= T_i P_i L_i^T + R_i Q_i R_i^T, \end{aligned} \quad (3.6)$$

for $i = 1, \dots, T$ and with b_1 and P_1 respectively mean a variance of the initial state β_1 . Given the Kalman filter estimates and conditioning on the whole observation data y , the state smoother outputs $\hat{\beta}_i = \mathbb{E}[\beta_i|y]$ and $V_i = \text{var}(\beta_i|y)$ for $i = 1, \dots, T$ by backwards recursion through the following equation, again based on the theoretical results of Multivariate Normal distribution:

$$\begin{aligned} r_{i-1} &= Z_i^T F_i^{-1} v_i + L_i^T r_i, & N_{i-1} &= Z_i^T F_i^{-1} Z_i + L_i^T N_i L_i, \\ \hat{\beta}_i &= b_i + P_i r_{i-1}, & V_i &= P_i - P_i N_{i-1} P_i, \end{aligned} \quad (3.7)$$

where r_T and N_T are set to 0 to start the backwards recursion. Based on these results the simulation smoother draws random vector $\tilde{\beta}$ through the following step:

1. Draw a random vector $w^+ = [\epsilon_1^T, \omega_1^T, \dots, \epsilon_T^T, \omega_T^T]^T$ from the joint density $p(w) \sim N_{2 \times T}(0, \text{diag}(H_1, Q_1, \dots, H_T, Q_T))$ and use it to generate β^+ and y^+ by means of recursion from the state space model in (3.5) with w replaced by w^+ , where the recursion is initialized by the draw $\beta_1^+ \sim N(b_1, P_1)$.
2. Compute $\hat{\beta} = \mathbb{E}[\beta|y]$ and $\hat{\beta}^+ = \mathbb{E}[\beta^+|y^+]$ using (3.6) forward, (3.7) backward and finally forward the equation $\hat{\beta}_{i+1} = T_i \hat{\beta}_i + R_i Q_i R_i^T r_i$ for $i = 1, \dots, T$.
3. Take $\tilde{\beta} = \hat{\beta} + (\beta^+ - \hat{\beta}^+)$.

The entire approach is mainly based on the property of $V = \text{var}(\beta|y)$ not depending upon y and on the assumption of linear and Gaussian model. As a result, instead of drawing directly from the density of $\beta|y$ which is $N(\hat{\beta}, V)$, we can simulate samples from $N(0, V)$ and adding these to the known vector $\hat{\beta}$. The difference $\beta^+ - \hat{\beta}^+$ is the desired drawn from $N(0, V)$ since $\hat{\beta}^+ = \mathbb{E}[\beta^+|y^+]$, and recalling the independence between V and y , $\text{var}(\beta^+|y^+) = V$. Note also that the initialization at $i = 1$ for the filter, and the assumption $r_T = 0$ and $N_T = 0$ for the backward smoother, lead to larger conditional variances of β_i at the beginning and the end of the sample as discussed in Durbin and Koopman (2001).

3.2 Locally Adaptive Covariance Regression

In order to allow locally adaptive smoothing for the time-varying covariance and mean functions we develop a novel stochastic process with locally-varying smoothness. This is accomplished by modifying the model of Fox and Dunson (2011) to incorporate dictionary functions ξ_{lk} and ψ_k , that are assigned nested Gaussian process priors. Note also that compared to Zhu and Dunson (2012) our approach allows generalizations in: (i) extending the analysis to the multivariate case (i.e. y_i is a p -dimensional vector instead of a scalar) and (ii) accommodating locally adaptive smoothing not only for the mean but also for the time-varying variance and covariance functions.

3.2.1 Notation and Motivation

Working in a context of multivariate time series, let y_i represent a $p \times 1$ vector of observations, where $i = 1, \dots, T$ represents a discrete order index and assume

$$y_i \sim N_p(\mu(t_i), \Sigma(t_i)),$$

with $\mu(t_i)$ and $\Sigma(t_i)$ denoting respectively the $p \times 1$ mean vector and $p \times p$ covariance matrix at "location" $t_i \in \mathcal{T} \subset \mathbb{R}^+$, where t_i represents the time observation related to the i -th observation.

Our aim is to define a prior Π_Σ for $\Sigma_{\mathcal{T}} = \{\Sigma(t), t \in \mathcal{T}\}$ and, if $\mu(t)$ is not known, to also define a prior Π_μ for $\mu_{\mathcal{T}} = \{\mu(t), t \in \mathcal{T}\}$. We look for priors with the properties of large support as well as good performance in large- p samples, but differently from previous proposals, we require that these priors allows for locally varying smoothing. Note that even if we will focus on multivariate time series (i.e., t is a time observation), our formulation can be easily extended without any loss of generality to the case where t is an arbitrary predictor value.

3.2.2 Latent Factor Model

Following the approach of Fox and Dunson (2011) outlined in Section 2.2, we address the problem of dimensionality reduction by modeling a $p \times p$ covariance matrix $\Sigma(t)$ over an arbitrary predictor space \mathcal{T} (which represents an enormous dimensional regression problem in large- p settings), through a

lower dimensional $p \times K$, with $K \ll p$, factor loadings matrix $\Lambda(t)$ indexed by predictors t , by assuming the decomposition

$$\Sigma(t) = \Lambda(t)\Lambda(t)^T + \Sigma_0,$$

where $\Sigma_0 = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. Note that such a decomposition for $\Sigma(t)$ is naturally induced by the marginalization of the $K \times 1$ vector of latent factors η_i , in the latent factor model

$$y_i = \Lambda(t_i)\eta_i + \epsilon_i, \quad (3.8)$$

where $\eta_i \sim N_K(0, I_K)$ and $\epsilon_i \sim N_p(0, \Sigma_0)$. To further improve the tractability of the model, we assume that the time-varying factor loadings matrix $\Lambda(t)$ is a linear combination of a much smaller set of continuous dictionary functions $\xi_{lk} : \mathcal{T} \rightarrow \mathbb{R}$ comprising the $L \times K$, with $L \ll p$, matrix $\xi(t)$. As a result

$$\Lambda(t_i) = \Theta\xi(t_i),$$

where Θ is a $p \times L$ matrix of coefficients relating the matrix $\Lambda(t)$ to the time-varying dictionary elements in $\xi(t)$. Such a decomposition for $\Lambda(t_i)$ reproduces the equation (2.2) described in Section 2.2, and further reduces the number of continuous random function to be modeled from $p \times L$ to $L \times K$, leading to an higher flexible and computationally tractable formulation in which the induced covariance structure, after marginalizing out the latent factors, follows the equation

$$\text{cov}(y_i|t_i = t) = \Sigma(t) = \Theta\xi(t)\xi(t)^T\Theta^T + \Sigma_0. \quad (3.9)$$

As stated in Fox and Dunson (2011) the above decomposition for $\Sigma(t)$ is not unique as it is not guaranteed the identifiability of Θ and $\xi(t)$. A possibility to cope with this issue is to constrain the factor loading matrix (Geweke and Zhou, 1996), but this approach induces order dependence among responses (Aguilar and West, 2000, West, 2003, Lopes and West, 2004, Carvalho et al., 2008). However, we are focusing on inference and prediction on the covariance matrix $\Sigma(t)$, rather than on the structure of Θ and $\xi(t)$. It follows that the issue of identifiability is not troublesome, as it does not cause problems on the uniqueness of $\Lambda(t) = \Theta\xi(t)$, allowing us to avoid restrictions and define priors with good computational properties. The characterization of the class of time-varying covariance matrices $\Sigma(t)$ is proved by Lemma 2.1

of Fox and Dunson (2011), which states that for K and L sufficiently large, any covariance regression can be decomposed as in (3.9).

Recalling equations (2.3) and (2.4) in Section 2.2, when $\mu(t)$ is not known, we can incorporate the nonparametric mean regression in the model formulation by assuming

$$\eta_i = \psi(t_i) + \nu_i, \quad \nu_i \sim N_K(0, I_K), \quad (3.10)$$

where $\psi(t_i) = [\psi_1(t_i), \dots, \psi_K(t_i)]^T$ is a vector of continuous functions $\psi_j : \mathcal{T} \rightarrow \mathbb{R}$ that can be modeled similarly to the dictionary elements functions ξ_{lk} . As a result, marginalizing out the latent factors, the induced mean of y_i conditionally on $t_i = t$, turns to be

$$\mu(t) = \Theta \xi(t) \psi(t). \quad (3.11)$$

3.2.3 Prior Specification

The key point is to identify independent priors Π_ξ , Π_Θ , Π_{Σ_0} and Π_ψ for $\xi_{\mathcal{T}} = \{\xi(t), t \in \mathcal{T}\}$, Θ , Σ_0 and $\psi_{\mathcal{T}} = \{\psi(t), t \in \mathcal{T}\}$ respectively, to induce priors Π_Σ and Π_μ on $\Sigma_{\mathcal{T}}$ and $\mu_{\mathcal{T}}$, through equations (3.9) and (3.11), with the goal of maintaining simple computation and allowing both covariances and means to vary flexibly over time. Fox and Dunson (2011) address this issue by considering dictionary function as

$$\xi_{lk} \sim GP(0, c),$$

independently for all l, k , with c squared exponential correlation function having $c(\xi, \xi') = \exp(-\kappa \|\xi - \xi'\|_2^2)$, but proves unable to accommodate locally adaptive smoothing. We, instead, specify the dictionary functions as independent nGP prior to explicitly model the expectation of the derivative of ξ_{lk} , for each (l, k) with $l = 1, \dots, L$ and $k = 1, \dots, K$, as a function of t through the specification of a GP prior for ξ_{lk} 's m th-order derivative $D^m \xi_{lk}$, centered on a higher-level GP. As a result we allow the smoothness of the induced GP prior on ξ_{lk} to be measured by a time-varying set of derivatives, centered on a GP instantaneous mean function $A_{lk} : \mathcal{T} \rightarrow \mathbb{R}$ to accommodate nonparametric locally-adaptive smoothing.

Recalling equations (3.2) and (3.3) from Zhu and Dunson (2012), outlined in Section 3.1, we use the nGP by defining a GP prior for the random dictionary function ξ_{lk} and the local instantaneous mean A_{lk} , through the following

stochastic differential equations with parameters $\sigma_{\xi_{lk}} \in \mathbb{R}^+$ and $\sigma_{A_{lk}} \in \mathbb{R}^+$:

$$D^m \xi_{lk}(t) = A_{lk}(t) + \sigma_{\xi_{lk}} W_{\xi_{lk}}(t), \quad m \in \mathbb{N}, \quad m \geq 2, \quad (3.12)$$

$$D^n A_{lk}(t) = \sigma_{A_{lk}} W_{A_{lk}}(t), \quad n \in \mathbb{N}, \quad n \geq 1, \quad (3.13)$$

where $W_{\xi_{lk}} : \mathcal{T} \rightarrow \mathbb{R}$ and $W_{A_{lk}} : \mathcal{T} \rightarrow \mathbb{R}$ are two independent Gaussian white noise processes with mean function $\mathbb{E}[W_{\xi_{lk}}(t)] = \mathbb{E}[W_{A_{lk}}(t)] = 0, \forall t \in \mathcal{T}$; and covariance function $\mathbb{E}[W_{\xi_{lk}}(t)W_{\xi_{lk}}(t')] = \mathbb{E}[W_{A_{lk}}(t)W_{A_{lk}}(t')] = \delta(t - t')$ a delta function. This formulation naturally induces a prior for ξ_{lk} with varying smoothness, where $\mathbb{E}[D^m \xi_{lk}(t)|A_{lk}(t)] = A_{lk}(t)$ and initialization at t_1 based on the assumption

$$(\xi_{lk}(t_1), D^1 \xi_{lk}(t_1), \dots, D^{m-1} \xi_{lk}(t_1)) \sim N_m(0, \sigma_{\mu_{lk}}^2 I_m).$$

The same goes for the initial value $A_{lk}(t_1)$ and its derivatives up to order $n - 1$ leading to the prior

$$(A_{lk}(t_1), D^1 A_{lk}(t_1), \dots, D^{n-1} A_{lk}(t_1)) \sim N_n(0, \sigma_{\alpha_{lk}}^2 I_n).$$

The prior for the initial values and for $W_{\xi_{lk}}$ and $W_{A_{lk}}$ are assumed mutually independent. Finally, we assume

$$\begin{aligned} \sigma_{\xi_{lk}}^2 &\sim \text{invGa}(a_\xi, b_\xi), \\ \sigma_{A_{lk}}^2 &\sim \text{invGa}(a_A, b_A), \end{aligned}$$

independently for each (l, k) ; where $\text{invGa}(a, b)$ denote the Inverse Gamma distribution with shape a and scale b .

Recalling equation (3.4), for $m = 2$ and $n = 1$ and for δ_i sufficiently small, the approximated state equation induced by the nGP prior for ξ_{lk} , turns to be

$$\begin{bmatrix} \xi_{lk}(t_{i+1}) \\ \xi'_{lk}(t_{i+1}) \\ A_{lk}(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi_{lk}(t_i) \\ \xi'_{lk}(t_i) \\ A_{lk}(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i, \xi_{lk}} \\ \omega_{i, A_{lk}} \end{bmatrix} \quad (3.14)$$

where $[\omega_{i, \xi_{lk}}, \omega_{i, A_{lk}}]^T \sim N_2(0, V_{i, lk})$, with $V_{i, lk} = \text{diag}(\sigma_{\xi_{lk}}^2 \delta_i, \sigma_{A_{lk}}^2 \delta_i)$.

To address the issue related to the selection of the number of dictionary elements, a shrinkage prior Π_Θ is proposed for Θ . In particular, as proposed in Bhattacharya and Dunson (2011) we assume:

$$\begin{aligned} \theta_{jl} | \phi_{jl}, \tau_l &\sim N(0, \phi_{jl}^{-1} \tau_l^{-1}), \quad \phi_{jl} \sim \text{Ga}(3/2, 3/2), \\ \vartheta_1 &\sim \text{Ga}(a_1, 1), \quad \vartheta_h \sim \text{Ga}(a_2, 1), h \geq 2, \quad \tau_l = \prod_{h=1}^l \vartheta_h, \end{aligned} \quad (3.15)$$

Note that if $a_2 > 1$ the expected value for ϑ_h is greater than 1. As a result, as l goes to infinity, τ_l tends towards infinity shrinking θ_{jl} towards zero. This leads to a flexible prior for θ_{jl} with a local shrinkage parameter ϕ_{jl} and a global column-wise shrinkage factor τ_l , which allows many elements of Θ being close to zero as L increases.

Finally for the variances of the error terms in vector ϵ_i , we assume the usual Inverse Gamma prior distribution. Specifically Π_{Σ_0} is defined through

$$\sigma_j^{-2} \sim Ga(a_\sigma, b_\sigma),$$

independently for each $j = 1, \dots, p$. To conclude prior specification, if $\mu(t)$ is unknown, a similar approach based on dictionary elements, can be considered in the definition of the prior Π_ψ . In particular, recalling the previous results of the prior for ξ_{lk} , we can represent the nGP for ψ_k with the following state equation:

$$\begin{bmatrix} \psi_k(t_{i+1}) \\ \psi'_k(t_{i+1}) \\ B_k(t_{i+1}) \end{bmatrix} = \begin{bmatrix} 1 & \delta_i & 0 \\ 0 & 1 & \delta_i \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \psi_k(t_i) \\ \psi'_k(t_i) \\ B_k(t_i) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{i,\psi_k} \\ \omega_{i,B_k} \end{bmatrix} \quad (3.16)$$

independently $\forall k$, with $k = 1, \dots, K$. Where $[\omega_{i,\psi_k}, \omega_{i,B_k}]^T \sim N_2(0, S_{i,k})$, with $S_{i,k} = \text{diag}(\sigma_{\psi_k}^2 \delta_i, \sigma_{B_k}^2 \delta_i)$. Similarly to ξ_{lk} , the priors for the initial values are assumed

$$\begin{aligned} (\psi_k(t_1), D^1 \psi_k(t_1), \dots, D^{m-1} \psi_k(t_1)) &\sim N_m(0, \sigma_{\mu_k}^2 I_m), \\ (B_k(t_1), D^1 B_k(t_1), \dots, D^{n-1} B_k(t_1)) &\sim N_n(0, \sigma_{\alpha_k}^2 I_n). \end{aligned}$$

While those for the variances in the state equation follow

$$\begin{aligned} \sigma_{\psi_k}^2 &\sim \text{invGa}(a_\psi, b_\psi), \\ \sigma_{B_k}^2 &\sim \text{invGa}(a_B, b_B). \end{aligned}$$

3.2.4 Hyperparameters interpretation

We now focus our attention on the hyperparameters of the priors for $\sigma_{\xi_{lk}}^2$, $\sigma_{A_k}^2$, $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$. Several simulation studies have shown that the higher the variances in the latent state equations, the better our formulation accommodates locally adaptive smoothing for sudden changes in covariances and means. A theoretical support for this data-driven consideration can be identified in the connection between the nGP prior and nested smoothing splines.

It has been shown (Zhu and Dunson, 2012) that the posterior mean of U , under nGP prior can be related to the minimizer of the equation

$$\frac{1}{T} \sum_{i=1}^T (y_i - U(t_i))^2 + \lambda_U \int_{\mathcal{T}} (D^m U(t) - C(t))^2 dt + \lambda_C \int_{\mathcal{T}} (D^n C(t))^2 dt,$$

where $\lambda_U \in \mathbb{R}^+$ and $\lambda_C \in \mathbb{R}^+$ regulate the smoothness of unknown functions U and A respectively, leading to less smoothed patterns when fixed at low values. The resulting inverse relationship between these smoothing parameters and the variances in the state equation, together with the results in the simulation studies, suggest to fix the hyperparameters in the Inverse Gamma priors for $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$, $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$, so as to allow high variances in the case in which the time series analyzed are expected to have strong changes in their covariance (or mean) dynamic.

In practical applications, it may be useful to obtain a first estimate of the covariance matrix $\tilde{\Sigma}(t)$ and the mean vector $\tilde{\mu}(t)$, to set the hyperparameters according to the results of the graphical analysis of the estimated values. More specifically, $\tilde{\mu}_j(t_i)$ can be the output of a standard moving average on each time series $y_j = [y_{j,1}, \dots, y_{j,T}]$, while $\tilde{\Sigma}(t_i)$ can be obtained by a simple estimator, such as the EWMA procedure. With these choices, the recursive equation

$$\tilde{\Sigma}(t_i) = (1 - \lambda)\{[y_{i-1} - \tilde{\mu}(t_{i-1})][y_{i-1} - \tilde{\mu}(t_{i-1})]^T\} + \lambda\tilde{\Sigma}(t_{i-1}),$$

become easy to implement.

3.3 Posterior Computation

Posterior computation can proceed via a straightforward modification of the Gibbs sampling algorithm proposed by Fox and Dunson (2011). The algorithm alternates between a simulation smoother step to update state space formulation of the nGP, and standard Gibbs sampling steps for updating the parametric component model from their conditional distributions. When also the mean process needs to be estimated, an additional step with a block sampling for $\{\psi(t_i)\}_{i=1}^T$ and $\{\nu_i\}_{i=1}^T$ is implemented.

3.3.1 Main Steps

We outline here the main features of the algorithm for posterior computation, based on observations (y_i, t_i) for $i = 1, \dots, T$.

- A. Given Θ and $\{\eta_i\}_{i=1}^T$, a multivariate version of the MCMC algorithm proposed by Zhu and Dunson (2012) draws posterior samples from each dictionary element's function $\{\xi_{lk}(t_i)\}_{i=1}^T$, its first order derivative $\{\xi'_{lk}(t_i)\}_{i=1}^T$, the corresponding instantaneous mean $\{A_{lk}(t_i)\}_{i=1}^T$, the variances in the state equations $\sigma_{\xi_{lk}}^2$, $\sigma_{A_{lk}}^2$ and the variances of the error terms in the observation equation σ_j^2 with $j = 1, \dots, p$.
- B. If the mean process needs not to be estimated, recalling the prior $\eta_i \sim N_{K^*}(0, I_{K^*})$ and model (3.8), the standard conjugate posterior distribution from which to sample the vector of latent factors for each i , given Θ , $\{\sigma_j^{-2}\}_{j=1}^p$, $\{y_i\}_{i=1}^T$ and $\{\xi(t_i)\}_{i=1}^T$ is Gaussian.

Otherwise, if we want to incorporate the mean regression, through model (3.10), we implement a block sampling of $\{\psi(t_i)\}_{i=1}^T$ and $\{\nu_i\}_{i=1}^T$ following a similar approach used for the dictionary elements process, based on the Durbin and Koopman (2002) simulation smoother and the implementation by Zhu and Dunson (2012).

- C. Finally, conditioned on $\{y_i\}_{i=1}^T$, $\{\eta_i\}_{i=1}^T$, $\{\sigma_j^{-2}\}_{j=1}^p$ and $\{\xi(t_i)\}_{i=1}^T$, and recalling the shrinkage prior for the elements of Θ in (3.15), we update Θ , each local shrinkage hyperparameter ϕ_{jl} and the global shrinkage hyperparameters τ_l , following the standard conjugate analysis as in Fox and Dunson (2011).

3.3.2 Detailed Algorithm

For a fixed truncation level L^* and a latent factor dimension K^* , the detailed steps of the Gibbs sampler for posterior computations, without nonparametric mean regression are:

1. Define the vector of the latent state and the error terms in the state space equation resulting from nGP prior for dictionary elements as

$$\begin{aligned}\Xi_i &= [\xi_{11}(t_i), \xi_{21}(t_i), \dots, \xi_{L^*K^*}(t_i), \xi'_{11}(t_i), \dots, \xi'_{L^*K^*}(t_i), A_{11}(t_i), \dots, A_{L^*K^*}(t_i)]^T, \\ \Omega_{i,\xi} &= [\omega_{i,\xi_{11}}, \omega_{i,\xi_{21}}, \dots, \omega_{i,\xi_{L^*K^*}}, \omega_{i,A_{11}}, \omega_{i,A_{21}}, \dots, \omega_{i,A_{L^*K^*}}]^T,\end{aligned}$$

Given Θ , $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, Σ_0 and the variances in latent state equations $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$; update $\{\Xi_i\}_{i=1}^T$ by using the simulation smoother in the following state space model:

$$y_i = [\eta_i^T \otimes \Theta, 0_{p \times (2 \times K^* \times L^*)}] \Xi_i + \epsilon_i, \quad (3.17)$$

$$\Xi_{i+1} = T_i \Xi_i + R_i \Omega_{i,\xi}, \quad (3.18)$$

where the observation equation in (3.17) results by applying the vec operator in the latent factor model $y_i = \Theta\xi(t_i)\eta_i + \epsilon_i$. More specifically recalling the property $vec(ABC) = (C^T \otimes A)vec(B)$ we obtain

$$\begin{aligned} y_i = vec(y_i) &= vec\{\Theta\xi(t_i)\eta_i + \epsilon_i\} \\ &= vec\{\Theta\xi(t_i)\eta_i\} + vec(\epsilon_i) \\ &= (\eta_i^T \otimes \Theta)vec\{\xi(t_i)\} + \epsilon_i. \end{aligned}$$

The state equation in (3.18) is a joint representation of the equations resulting from the nGP prior on each ξ_{lk} defined in (3.14). As a result, the $(3 \times L^* \times K^*) \times (3 \times L^* \times K^*)$ matrix T_i together with the $(3 \times L^* \times K^*) \times (2 \times L^* \times K^*)$ matrix R_i reproduce, for each dictionary element, the state equation in (3.14) by fixing to 0 the coefficients relating latent states with different (l, k) (from the assumption of independence between the dictionary elements). Finally, recalling the assumptions on $\omega_{i,\xi_{lk}}$ and $\omega_{i,A_{lk}}$, $\Omega_{i,\xi}$ is normally distributed with $E[\Omega_{i,\xi}] = 0$ and $E[\Omega_{i,\xi}\Omega_{i,\xi}^T] = \text{diag}(\sigma_{\xi_{11}}^2 \delta_i, \sigma_{\xi_{21}}^2 \delta_i, \dots, \sigma_{\xi_{L^*K^*}}^2 \delta_i, \sigma_{A_{11}}^2 \delta_i, \sigma_{A_{21}}^2 \delta_i, \dots, \sigma_{A_{L^*K^*}}^2 \delta_i)$.

2. Given $\{\Xi_i\}_{i=1}^T$ sample each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ respectively from

$$\begin{aligned} \sigma_{\xi_{lk}}^2 | \{\Xi_i\} &\sim invGa \left(a_\xi + \frac{T}{2}, b_\xi + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(\xi_{lk}(t_{i+1})' - \xi_{lk}(t_i)' - A_{lk}(t_i)\delta_i)^2}{\delta_i} \right), \\ \sigma_{A_{lk}}^2 | \{\Xi_i\} &\sim invGa \left(a_A + \frac{T}{2}, b_A + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(A_{lk}(t_{i+1}) - A_{lk}(t_i))^2}{\delta_i} \right). \end{aligned}$$

3. Conditioned on Θ , $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, and $\{\xi(t_i)\}_{i=1}^T$ (obtained from Ξ_i), the standard conjugate posterior from which to update σ_j^{-2} is

$$\sigma_j^{-2} | \Theta, \{\eta_i\}, \{y_i\}, \{\xi(t_i)\} \sim Ga \left(a_\sigma + \frac{T}{2}, b_\sigma + \frac{1}{2} \sum_{i=1}^T (y_{ji} - \theta_j \xi(t_i) \eta_i)^2 \right),$$

where $\theta_j = [\theta_{j1}, \dots, \theta_{jL^*}]$.

4. In the case nonparametric mean regression is not considered, given Θ , Σ_0 , y_i , and $\xi(t_i)$, the vector of latent factors at each time t_i can sampled from the Gaussian conditional distribution of $\eta_i | \Theta, \Sigma_0, y_i, \xi(t_i)$:

$$N_{K^*} \left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} y_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \right).$$

5. Given $\{\eta_i\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, $\{\xi(t_i)\}_{i=1}^T$ and the hyperparameters ϕ and τ , the shrinkage prior on Θ combined with the likelihood for the latent factor model lead to the Gaussian posterior

$$\theta_j | \{\eta_i\}, \{y_i\}, \{\xi(t_i)\}, \phi, \tau \sim N_{L^*} \left(\tilde{\Sigma}_\theta \tilde{\eta}^T \sigma_j^{-2} \begin{bmatrix} y_{j1} \\ \vdots \\ y_{jT} \end{bmatrix}, \tilde{\Sigma}_\theta \right),$$

where $\tilde{\eta}^T = [\xi(t_1)\eta_1, \xi(t_2)\eta_2, \dots, \xi(t_T)\eta_T]$ and

$$\tilde{\Sigma}_\theta^{-1} = \sigma_j^{-2} \tilde{\eta}^T \tilde{\eta} + \text{diag}(\phi_{j1}\tau_1, \dots, \phi_{jL^*}\tau_{L^*}).$$

6. The Gamma prior on the local shrinkage hyperparameter ϕ_{jl} implies the standard conjugate posterior given θ_{jl} and τ_l

$$\phi_{jl} | \theta_{jl}, \tau_l \sim Ga \left(2, \frac{3 + \tau_l \theta_{jl}^2}{2} \right).$$

7. Following Bhattacharya and Dunson (2011), conditioned on Θ and τ , sample the global shrinkage hyperparameters from

$$\vartheta_1 | \Theta, \tau^{(-1)} \sim Ga \left(a_1 + \frac{pL^*}{2}, 1 + \frac{1}{2} \sum_{l=1}^{L^*} \tau_l^{(-1)} \sum_{j=1}^p \phi_{jl} \theta_{jl}^2 \right),$$

$$\vartheta_h | \Theta, \tau^{(-h)} \sim Ga \left(a_2 + \frac{p(L^* - h + 1)}{2}, 1 + \frac{1}{2} \sum_{l=1}^{L^*} \tau_l^{(-h)} \sum_{j=1}^p \phi_{jl} \theta_{jl}^2 \right),$$

where $\tau_l^{(-h)} = \prod_{t=1, t \neq h}^l \vartheta_t$ for $h = 1, \dots, p$.

If one wishes to consider also the problem of mean regression as in (3.10), the Gibbs sampler mimics the previous algorithm except for the step 4 which is replaced by a block sampling of $\{\psi(t_i)\}_{i=1}^T$ and $\{\nu_i\}_{i=1}^T$. In particular

4.1. Similarly to Ξ_i and $\Omega_{i,\xi}$ let

$$\begin{aligned} \Psi_i &= [\psi_1(t_i), \psi_2(t_i), \dots, \psi_{K^*}(t_i), \psi'_1(t_i), \dots, \psi'_{K^*}(t_i), B_1(t_i), \dots, B_{K^*}(t_i)]^T, \\ \Omega_{i,\psi} &= [\omega_{i,\psi_1}, \omega_{i,\psi_2}, \dots, \omega_{i,\psi_{K^*}}, \omega_{i,B_1}, \omega_{i,B_2}, \dots, \omega_{i,B_{K^*}}]^T, \end{aligned}$$

be the vectors of the latent state and error terms in the state space equation resulting from nGP prior for ψ .

Conditioned to Θ , $\{\xi(t_i)\}_{i=1}^T$, $\{y_i\}_{i=1}^T$, Σ_0 , and the variances in latent state equations $\{\sigma_{\psi_k}^2\}$, $\{\sigma_{B_k}^2\}$, with $k = 1, \dots, K^*$; sample $\{\Psi_i\}_{i=1}^T$ from the simulation smoother in the following state space model:

$$y_i = [\Theta\xi(t_i), 0_{p \times (2 \times K^*)}] \Psi_i + \varpi_i, \quad (3.19)$$

$$\Psi_{i+1} = G_i \Psi_i + F_i \Omega_{i,\psi}, \quad (3.20)$$

with $\varpi_i \sim N_p(0, \Theta\xi(t_i)\xi(t_i)^T\Theta^T + \Sigma_0)$. The observation equation in (3.19) results by marginalizing out ν_i in the latent factor model with nonparametric mean regression $y_i = \Theta\xi(t_i)\psi(t_i) + \Theta\xi(t_i)\nu_i + \epsilon_i$. Analogously to Ξ_i , the state equation in (3.20) is a joint representation of the state equation induced by the nGP prior on each ψ_k defined in (3.16); where the $(3 \times K^*) \times (3 \times K^*)$ matrix G_i and the $(3 \times K^*) \times (2 \times K^*)$ matrix F_i are constructed with the same goal of the matrices T_i and R_i in the state space model for Ξ_i . Finally, $\Omega_{i,\psi} \sim N_{2 \times K^*}(0, \text{diag}(\sigma_{\psi_1}^2 \delta_i, \sigma_{\psi_2}^2 \delta_i, \dots, \sigma_{\psi_{K^*}}^2 \delta_i, \sigma_{B_1}^2 \delta_i, \sigma_{B_2}^2 \delta_i, \dots, \sigma_{B_{K^*}}^2 \delta_i))$.

4.2. Given $\{\Psi_i\}_{i=1}^T$ update each $\sigma_{\psi_k}^2$ and $\sigma_{B_k}^2$ respectively from

$$\begin{aligned} \sigma_{\psi_k}^2 | \{\Psi_i\} &\sim \text{invGa} \left(a_\psi + \frac{T}{2}, b_\psi + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(\psi_k(t_{i+1})' - \psi_k(t_i)' - B_k(t_i)\delta_i)^2}{\delta_i} \right), \\ \sigma_{B_k}^2 | \{\Psi_i\} &\sim \text{invGa} \left(a_B + \frac{T}{2}, b_B + \frac{1}{2} \sum_{i=1}^{T-1} \frac{(B_k(t_{i+1}) - B_k(t_i))^2}{\delta_i} \right). \end{aligned}$$

4.3. Conditioned on Θ , Σ_0 , y_i , $\xi(t_i)$ and $\psi(t_i)$, and recalling $\nu_i \sim N_{K^*}(0, I_{K^*})$; the standard conjugate posterior distribution $\nu_i | \Theta, \Sigma_0, \tilde{y}_i, \xi(t_i), \psi(t_i)$ is

$$N_{K^*} \left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} \tilde{y}_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \right),$$

with $\tilde{y}_i = y_i - \Theta\xi(t_i)\psi(t_i) = \Theta\xi(t_i)\nu_i + \epsilon_i$.

3.4 Online Updating

The problem of online updating represents a key point in multivariate time series with high frequency data. Referring to our formulation, we are interested in finding an update approximated posterior distribution for $\Sigma(t_{T+h})$ and $\mu(t_{T+h})$ with $h = 1, \dots, H$, once a new vector of observation $\{y_i\}_{i=T+1}^{T+H}$ is available, instead of rerunning posterior computation for the whole time series.

Conditionally to the posterior estimates of the Gibbs sampler based on observations available up to time T , $\{y_i\}_{i=1}^T$, it is easy to implement a highly computationally tractable online updating algorithm which alternates between steps 1 and 4 outlined in Section 3.3.2 for the new set of observations, and that can be initialized at $T + 1$ using the one step ahead predictive distribution for the latent state vector in the state space formulation.

3.4.1 Online Updating Algorithm

Consider $\Theta, \Sigma_0, \{\sigma_{\xi_{lk}}^2\}, \{\sigma_{A_{lk}}^2\}, \{\sigma_{\psi_k}^2\}$ and $\{\sigma_{B_k}^2\}$ fixed at their posterior mean $\hat{\Theta}, \hat{\Sigma}_0, \{\hat{\sigma}_{\xi_{lk}}^2\}, \{\hat{\sigma}_{A_{lk}}^2\}, \{\hat{\sigma}_{\psi_k}^2\}, \{\hat{\sigma}_{B_k}^2\}$ respectively, and let $\hat{\Xi}_T, \hat{\Sigma}_{\Xi_T}$ and $\hat{\Psi}_T, \hat{\Sigma}_{\Psi_T}$ be the sample mean and covariance matrix of the posterior distribution respectively for Ξ_T and Ψ_T , obtained from the posterior estimates of the Gibbs sampler conditioned on $\{y_i\}_{i=1}^T$.

1. Given $\hat{\Theta}, \hat{\Sigma}_0, \{\hat{\sigma}_{\xi_{lk}}^2\}, \{\hat{\sigma}_{A_{lk}}^2\}, \{\eta_i\}_{i=T+1}^{T+H}$ and $\{y_i\}_{i=T+1}^{T+H}$, update $\{\Xi_i\}_{i=T+1}^{T+H}$ by using the simulation smoother in the following state space model

$$\begin{aligned} y_i &= [\eta_i^T \otimes \hat{\Theta}, 0_{p \times (2 \times K^* \times L^*)}] \Xi_i + \epsilon_i, \\ \Xi_{i+1} &= T_i \Xi_i + R_i \Omega_{i,\xi}, \end{aligned}$$

where Ξ_{T+1} can be initialized from the standard one step ahead predictive distribution

$$\Xi_{T+1} \sim N(T_T \hat{\Xi}_T, T_T \hat{\Sigma}_{\Xi_T} T_T^T + R_T E[\Omega_{T,\xi} \Omega_{T,\xi}^T] R_T^T).$$

2. Conditioned on $\hat{\Theta}, \hat{\Sigma}_0, \{\hat{\sigma}_{\psi_k}^2\}, \{\hat{\sigma}_{B_k}^2\}, \{\xi(t_i)\}_{i=T+1}^{T+H}$ and $\{y_i\}_{i=T+1}^{T+H}$, sample $\{\Psi_i\}_{i=T+1}^{T+H}$ through the simulation smoother in the state space model

$$\begin{aligned} y_i &= [\hat{\Theta} \xi(t_i), 0_{p \times (2 \times K^*)}] \Psi_i + \varpi_i, \\ \Psi_{i+1} &= G_i \Psi_i + F_i \Omega_{i,\psi}, \end{aligned}$$

similarly to Ξ_{T+1} , $\Psi_{T+1} \sim N(G_T \hat{\Psi}_T, G_T \hat{\Sigma}_{\Psi_T} G_T^T + F_T E[\Omega_{T,\psi} \Omega_{T,\psi}^T] F_T^T)$.

3. Given $\hat{\Theta}, \hat{\Sigma}_0, \{y_i\}, \xi(t_i)$ and $\psi(t_i)$, for $i = T+1, \dots, T+H$, sample ν_i from the standard conjugate posterior distribution for $\nu_i | \Theta, \Sigma_0, \tilde{y}_i, \xi(t_i), \psi(t_i)$:

$$N_{K^*} \left((I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \xi(t_i)^T \Theta^T \Sigma_0^{-1} \tilde{y}_i, (I + \xi(t_i)^T \Theta^T \Sigma_0^{-1} \Theta \xi(t_i))^{-1} \right),$$

with $\tilde{y}_i = y_i - \Theta \xi(t_i) \psi(t_i) = \Theta \xi(t_i) \nu_i + \epsilon_i$.

4. Compute the updated covariance $\{\Sigma(t_i)\}_{i=T+1}^{T+H}$ and mean $\{\mu(t_i)\}_{i=T+1}^{T+H}$ from the usual equations

$$\begin{aligned}\Sigma(t_i) &= \hat{\Theta}\xi(t_i)\xi(t_i)^T\hat{\Theta}^T + \hat{\Sigma}_0, \\ \mu(t_i) &= \hat{\Theta}\xi(t_i)\psi(t_i).\end{aligned}$$

Note that the initialization procedure for latent state vectors in the algorithm depends on the sample moments of posterior distribution for the latent states at T . As it is known for Kalman smoother (see, e.g., Durbin and Koopman, 2001), this could lead to computational problems in the online updating, due to the larger conditional variances of the latent states at the end of the sample (i.e., at T). To overcome this problem, we replace the previous assumptions for the initial values with a data-driven initialization scheme. In particular, instead of using only the new observations for the online updating, we run the algorithm starting from $\{y_i\}_{i=T-k}^{T+H}$, with k small enough, and choosing diffuse but proper priors for the initial states at $T - k$. As a result the distribution of the smoothed states at T is not anymore affected by the problem of large conditional variances, leading to better online updating performance.

Chapter 4

Simulation Studies

The aim of the following simulation studies is to assess whether and to what extent the assumption of nGP prior for the dictionary elements can accommodate, in practice, even dramatic changes in the time-varying covariances, and to compare the performance of our proposal with respect to BNCR proposed by Fox and Dunson (2011) with GP in the dictionary elements, which represents the main competing alternative. In the last subsection we also analyze the performance of the online updating algorithm proposed.

4.1 Estimation Performance

We generate a set of 5-dimensional observations y_i for each t_i in the discrete set $\mathcal{T}_o = \{1, 2, \dots, 100\}$, from the latent factor model in (3.8) with $\Lambda(t_i) = \Theta\xi(t_i)$ and η_i defined as in (3.10). To allow dramatic changes of the covariances in the generating mechanism, we consider a 2×2 (i.e. $L = K = 2$) matrix $\{\xi(t_i)\}_{i=1}^{100}$ from the time-varying functions adapted from Donoho and Johnstone (1994), with locally-varying smoothness (more specifically we choose functions bumps). The latent mean dictionary elements $\{\psi(t_i)\}_{i=1}^{100}$ are simulated from a Gaussian processes $\text{GP}(0, c)$ with $c(t, t') = \exp(-\kappa\|t - t'\|_2^2)$ and length scale $\kappa = 10$, while the elements in matrix Θ can be obtained from the shrinkage prior in (3.15) with $a_1 = a_2 = 10$. Finally the elements of the diagonal matrix Σ_0^{-1} are sampled independently from a $Ga(1, 0.1)$.

Posterior computation for our proposed approach is performed by using truncation levels $K^* = L^* = 2$; placing a $Ga(1, 0.1)$ prior on the precision parameters σ_j^{-2} , and choosing $a_1 = a_2 = 2$. As regards the nGP prior for

each dictionary element ξ_{lk} with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$, we choose diffuse but proper priors for the initial values by setting $\sigma_{\mu_{lk}}^2 = \sigma_{\alpha_{lk}}^2 = 100$ and place an $invGa(2, 5 \times 10^8)$ prior on each $\sigma_{\xi_{lk}}^2$ and $\sigma_{A_{lk}}^2$ in order to allow less smoothed behavior according to a previous graphical analysis of $\tilde{\Sigma}(t_i)$ estimated via EWMA. Similarly we set $\sigma_{\mu_k}^2 = \sigma_{\alpha_k}^2 = 100$ in the prior for the initial values of the latent state equations resulting from the nGP prior for ψ_k , and consider $a_\psi = a_B = b_\psi = b_B = 0.005$ to balance the rough behavior induced on the nonparametric mean functions by the settings of the nGP prior on ξ_{lk} , as suggested from previous graphical analysis. Note also that for posterior computation, we first scale the predictor space to $(0, 1]$, leading to $\delta_i = 1/100, \forall i = 1, \dots, 100$.

For inference in BNCR we consider the same previous hyperparameters setting for Θ and Σ_0 priors as well as the same truncation levels K^* and L^* , while the length scale κ in GP prior for ξ_{lk} and ψ_k has been set to 10 using the data-driven heuristic outlined by Fox and Dunson (2011). In both cases we run 50,000 Gibbs iterations discarding the first 20,000 as burn-in, and tinning the chain every 5 samples.

For monitoring convergence we analyzed several trace plot together with the Gelman-Rubin (see e.g. Gelman and Rubin, 1992) diagnostic which is based on the comparison of within-chain and between-chain variances (similar to a classical analysis of variance) in parallel chains after burn-in. Values of the potential scale reduction factor \hat{R} near 1 suggest converge. Having a single chain, we compared the within and between variances from samples obtained by splitting the chains in 6 pieces of same length. We consider this approach because of the substantial independence between samples after tinning the chain. Examination of trace plots for the elements of $\{\Sigma(t_i)\}_{i=1}^{100}$ and $\{\mu(t_i)\}_{i=1}^{100}$ in Figure 4.1 shows no evidence against convergence. Similar conclusions derive from the analysis of Gelman-Rubin's \hat{R} . In LABNCR the 95% of the chains have a potential reduction factor lower than 1.35, with a median equal to 1.11 and a maximum of 1.50. BNCR shows more problematic mixing, but not enough to reject the convergence, with an $\hat{R} < 1.44$ in the 95% of the chains and a median and a maximum of 1.18 and 1.67, respectively.

Figure 4.2 compares true mean and covariance functions to posterior mean respectively of our proposed approach and BNCR. A more detailed comparison between the true and posterior mean (for both approaches), for selected components of $\Sigma(t)$ and $\mu(t)$ over the predictor space \mathcal{T}_o together with the point-wise 95% high posterior density intervals is shown in Figure 4.3.

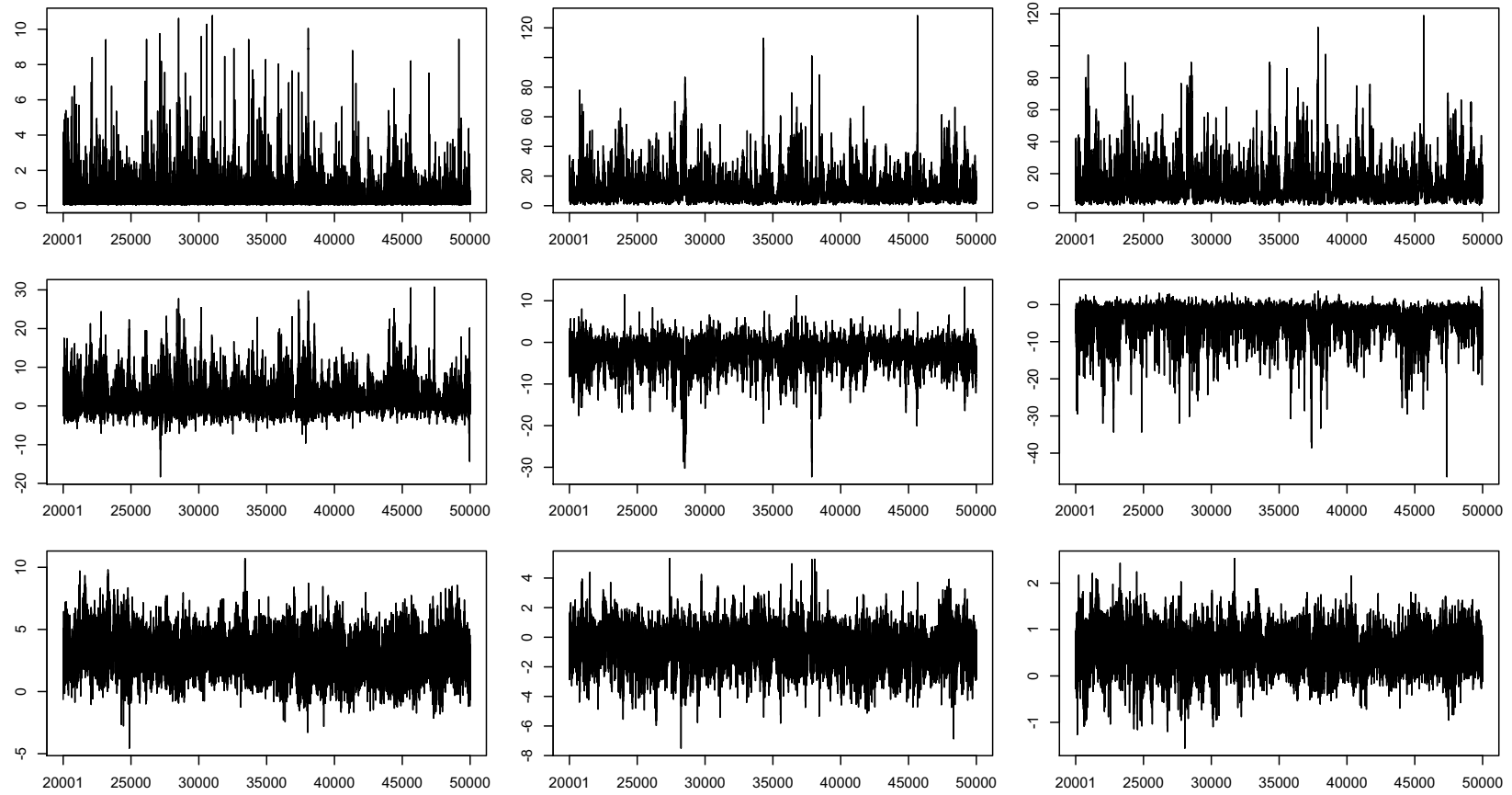


Figure 4.1: Some trace plot for posterior computation in LABNCR model after discarding the first 20,000 iterations as burn-in. Variances (Top), Covariances (Middle), Means (Bottom).

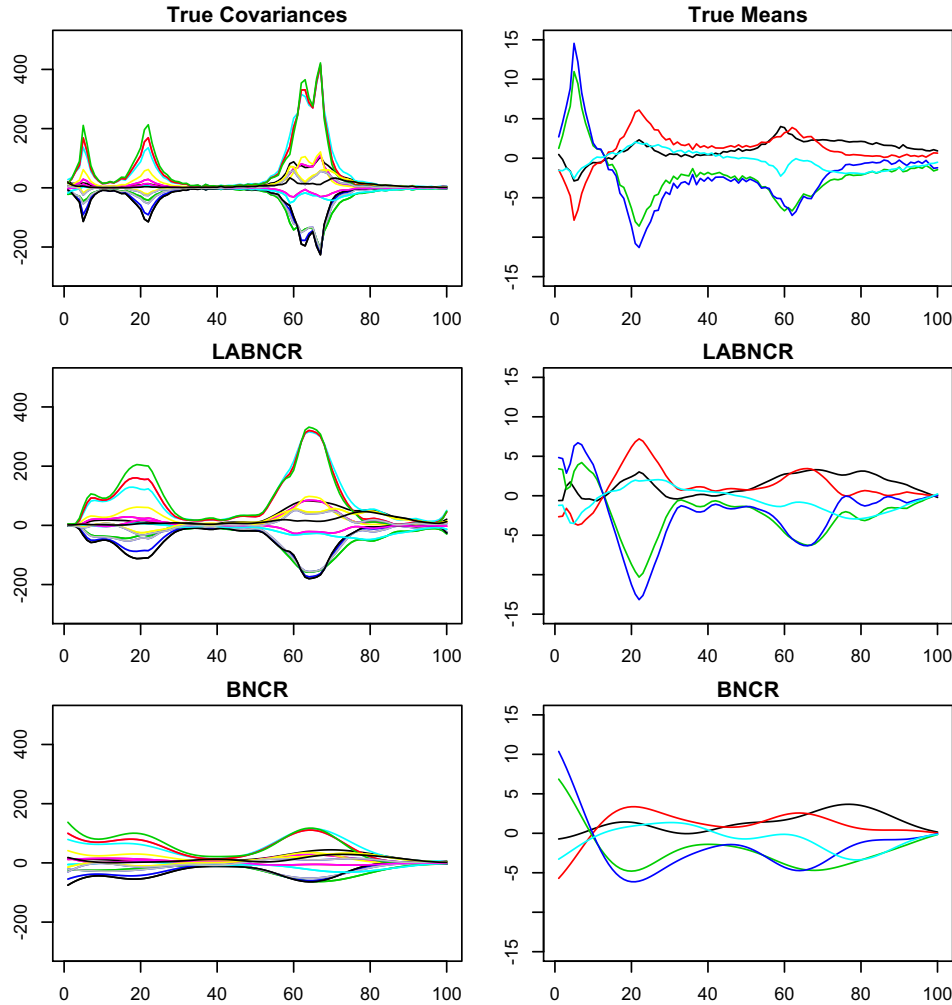


Figure 4.2: Left column: comparison between true covariance matrix functions $\Sigma(t)$ over the predictor space \mathcal{T}_o and the posterior mean respectively of LABNCR and BNCR. Right column: Similar comparison conducted on each component mean function

From these plots we can clearly note that our approach is able to capture conditional heteroscedasticity as well as mean patterns, also in correspondence of dramatic changes in the time-varying true functions. The major differences compared to the true values can be found at the beginning and at the end of the series and are likely to be related to the structure of the simulation smoother which also causes a widening of the credibility bands at the very end of the series. For references regarding this issue see Durbin and Koopman (2001). However, even in the most problematic cases, the true values are within the bands of the 95% high posterior density intervals. Much more problematic is the behavior of the posterior distributions for the com-

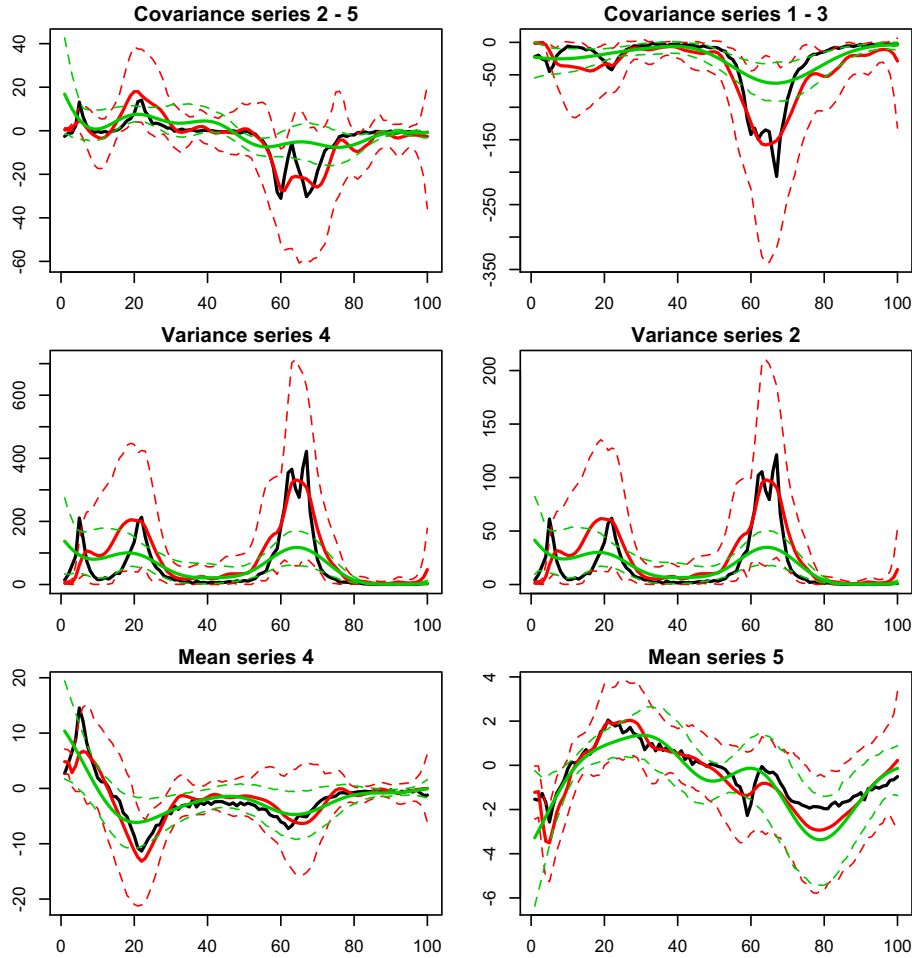


Figure 4.3: Plots of truth (black) and posterior mean respectively of LABNCR (solid red line) and BNCR (solid green line) for selected components of the covariance (top), variance (middle), mean (bottom). For both the approach the dotted lines represent the 95% high posterior density intervals.

peting alternative which badly over-smooth both covariance and mean functions leading also to many 95% high posterior density intervals not containing the true values. The comparison of the summaries of the squared errors between true values $\{\mu(t_i)\}_{i=1}^{100}$ and $\{\Sigma(t_i)\}_{i=1}^{100}$ and posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{100}$ and $\{\hat{\mu}(t_i)\}_{i=1}^{100}$ respectively for BNCR and LABNCR in Table 4.1, once again confirms the overall better performance of our approach.

To better understand the improvement of our approach in allowing locally varying smoothness and to evaluate the consequences of the over-smoothing induced by BNCR on the distribution of y_i with $i = 1, \dots, 100$, consider Figure 4.4 which shows, for some selected series $\{y_{ji}\}_{i=1}^{100}$, the time varying

	Mean $\{\mu(t_i)\}$		Covariance $\{\Sigma(t_i)\}$	
	BNCR	LABNCR	BNCR	LABNCR
Mean	2.24	1.64	1351.09	667.27
90th Quantile	4.86	3.22	1608.45	1556.38
95th Quantile	9.39	4.98	5140.102	3460.47
Max	74.57	67.87	96092.35	22373.37

Table 4.1: Summaries of the squared errors between true values $\{\mu(t_i)\}_{i=1}^{100}$ and $\{\Sigma(t_i)\}_{i=1}^{100}$ and posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{100}$ and $\{\hat{\mu}(t_i)\}_{i=1}^{100}$ obtained respectively with BNCR and our LABNCR.

mean together with the point-wise 2.5% and 97.5% quantiles of the marginal distribution of y_{ji} induced respectively by the true mean and true variance, the posterior mean of $\mu_j(t_i)$ and $\Sigma_{jj}(t_i)$ from our proposed approach, and the posterior mean of the same quantities from the competing alternative. We can clearly see that the marginal distribution of y_{ji} induced by BNCR is over-concentrated along the mean leading to incorrect inferences on the series analyzed. Note that our proposal is also able to accommodate for heavy tails, typical characteristic in financial series.

4.2 Online Updating Performance

To analyze the performance of the online updating algorithm proposed, we simulate 50 new observations $\{y_i\}_{i=101}^{150}$ with $t_i \in \mathcal{T}_o^* = \{101, \dots, 150\}$, considering the same Θ and Σ_0 used in the generating mechanism for the previous simulated data and taking the 50 subsequent observations from the bumps functions for the dictionary elements $\{\xi(t_i)\}_{i=101}^{150}$; finally the additional latent mean dictionary elements $\{\psi(t_i)\}_{i=101}^{150}$ are simulated as before, maintaining the continuity with the previously simulated functions $\{\psi(t_i)\}_{i=1}^{100}$.

According to the algorithm described in section 3.4, we fix Θ , Σ_0 , $\{\sigma_{\xi_{lk}}^2\}$, $\{\sigma_{A_{lk}}^2\}$, $\{\sigma_{\psi_k}^2\}$ and $\{\sigma_{B_k}^2\}$ at their posterior mean from the previous Gibbs sampler, and consider the last three observations y_{98} , y_{99} and y_{100} (i.e. $k = 3$) to initialize the simulation smoother in $i = 101$ through a data-driven approach. Posterior computation shows a good performance in terms of mixing, and convergence was assessed after 5,000 Gibbs iterations with a small burn-in of 500.

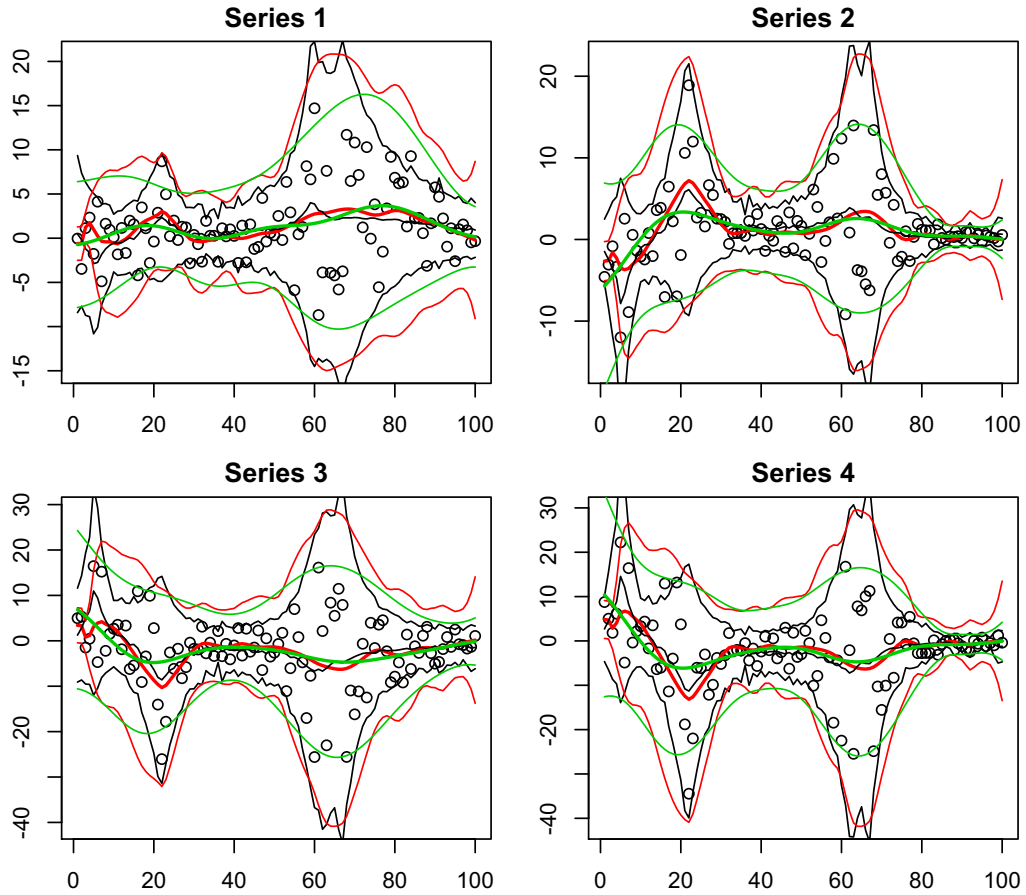


Figure 4.4: Plot for 4 selected simulated series of the time-varying mean $\mu_j(t_i)$ and the time-varying 2.5th and 97.5th quantiles of the marginal distribution of y_{ji} with true mean and variance (black), mean and variance from posterior mean of LABNCR (red), mean and variance from posterior mean BNCR (green). Black points represent the simulated data.

Figure 4.5 compares true mean and covariance to posterior mean of a selected set of components of $\{\mu(t_i)\}_{i=101}^{150}$ and $\{\Sigma(t_i)\}_{i=101}^{150}$, including also the 95% high posterior density intervals. The results clearly show that the online updating is characterized by a good performance, which allows to capture the behavior of new observations conditioning on the previous estimates.

Note that the posterior distribution of the approximated mean and covariance functions tends to slightly over-estimate the patterns of the functions at dramatic changes, however also in these cases the true values are within the bands of the credibility intervals. Finally, note that the data-driven initialization ensures a good behavior at the beginning of the series, while the results

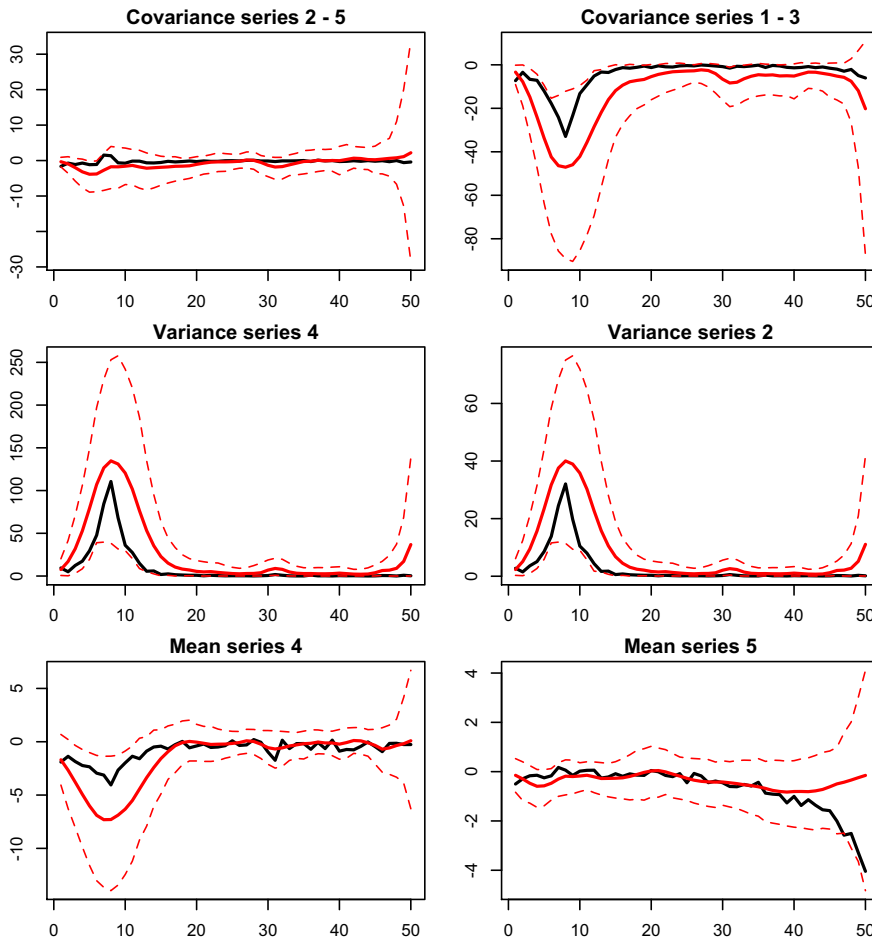


Figure 4.5: Plots of truth (black) and posterior mean of the online updating procedure (solid red line) for selected components of the covariance (top), variance (middle), mean (bottom). The dotted lines represent the 95% high posterior density intervals.

at the very end continue to remain troublesome because of the initialization scheme of the backward smoother at the end of the series.

Chapter 5

Application to National Stock Market Indices

In this application we focus our attention on the multivariate weekly time series of the main 33 (i.e. $p = 33$) National Stock Market Indices from 12/07/2004 to 25/06/2012 ($T = 416$ weeks). The dataset has been downloaded from *Yahoo! Finance*, which represents the top financial news and research website in the U.S. since January 2008, providing stock quotes, stock exchange rate and report from the main financial markets worldwide. For the analysis we choose the adjusted closing quotations, which account for all corporate actions such as stock splits and dividend distribution, allowing for a more accurate representation of the firm's equity value beyond the simple market price.

Figure 5.1 shows the main features in terms of stationarity (top), mean patterns (middle) and volatility (bottom) of two selected market stock indices (USA NASDAQ and ITALY FTSE MIB, respectively). The non-stationary behavior motivate the analysis of the p -dimensional vector of logarithmic returns y_i with $i = 1, \dots, 415$, defined in (2.1). Beside this, the marginal distribution of log returns shows heavy tails and irregular cyclical trends in the nonparametric estimation of the mean, while EWMA estimates, highlight rapid changes of volatility during the financial crisis that occurred in the recent years. All these results, together with large datasets and high frequency data, typical in financial fields, motivate the use of our approach to obtain a better characterization of the time-varying dependence structure among financial markets.

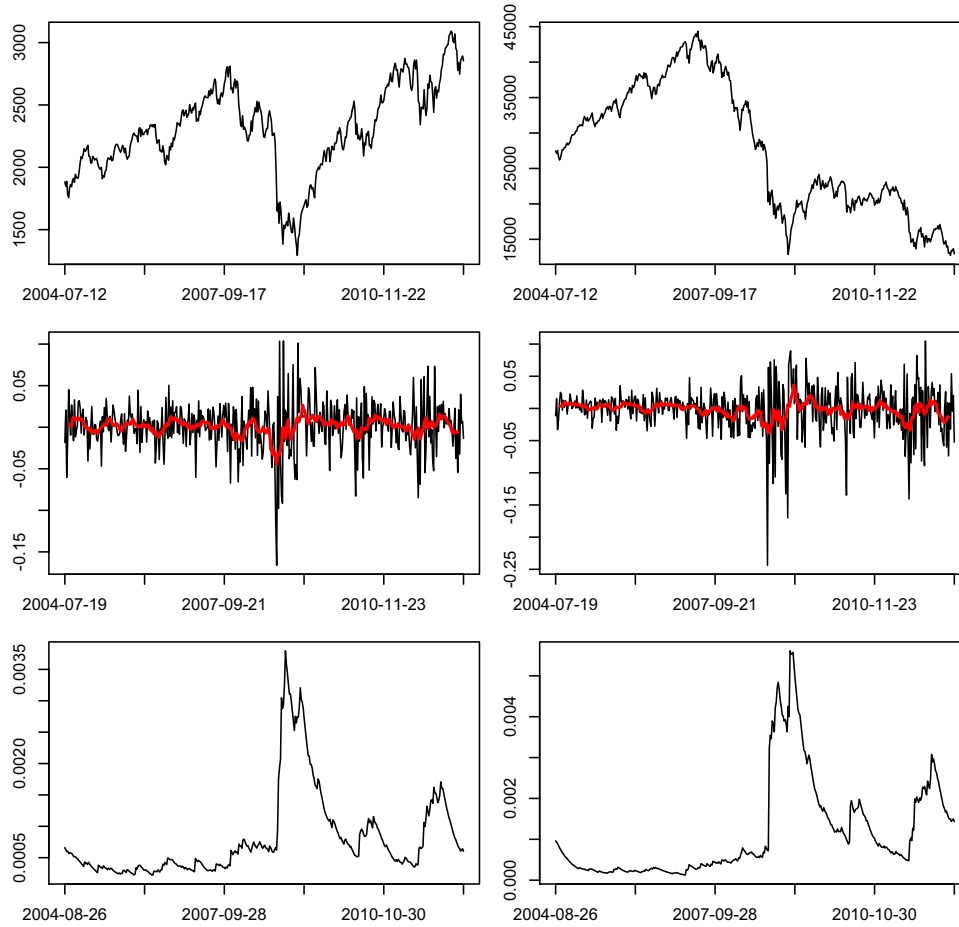


Figure 5.1: Plots of the main features of USA NASDAQ (left) and ITALY FTSE MIB (right). Specifically: observed time series (top), log-returns series (black) with nonparametric mean estimation via 12 week Equally Weighted Moving Average (red) in the middle, EWMA volatility estimates (bottom).

5.1 Heteroscedastic Modeling of National Stock Market Indices

We consider the heteroscedastic model $y_i \sim N_{33}(\mu(t_i), \Sigma(t_i))$, for $i = 1, \dots, 415$, and t_i in the discrete set $\mathcal{T}_o = \{1, 2, \dots, 415\}$, where mean $\mu(t_i)$ and covariance matrix $\Sigma(t_i)$ of the stock market indices at time $t = t_i$ are given in (3.11) and (3.9), respectively.

Posterior computation is performed by first rescaling the predictor space \mathcal{T}_o to $(0, 1]$ and using the same setting of the simulation study, with the exception of the truncation levels fixed at $K^* = 4$ and $L^* = 5$, and the

hyperparameters of the nGP prior for each ξ_{lk} and ψ_k with $l = 1, \dots, L^*$ and $k = 1, \dots, K^*$, set to $a_\xi = a_A = a_\psi = a_B = 2$ and $b_\xi = b_A = b_\psi = b_B = 5 \times 10^7$ to capture also rapid changes in the mean functions according to Figure 5.1. Recalling a key advantage of Fox and Dunson (2011) formulation, the few number of missing values in our dataset does not represent a limitation, since we can update our posterior considering solely the observed data, without introducing approximations. We run 10,000 Gibbs iterations with a burn-in of 2,500.

Similarly to simulation studies, we monitor convergence by combining trace plots examination with the analysis of Gelman-Rubin diagnostic. In this case we consider also Geweke test (Geweke, 1992), which assesses the equality of means of the first and last parts of the Markov chain for each variable, by implementing an equivalent of t -test. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution (i.e. is a standard Z -score). With a significance level $\alpha = 0.05$ the 94.3% of the Z -scores (one for each chain) fall within the interval $[-2, 2]$, showing no evidence against convergence. Similar conclusions are suggested by the examination of the trace plots for $\{\Sigma(t_i)\}_{i=1}^{415}$ and $\{\mu(t_i)\}_{t=1}^{415}$ in Figure 5.2, and from the results of the Gelman-Rubin's diagnostic which shows a potential reduction factor lower than 1.2 in the 95% of the chains, with a median equal to 1.03.

5.2 Posterior results and economic facts

Results from posterior computation provide relevant informations regarding the volatility and co-volatility processes with reference to theory and economic facts.

Posterior distributions for the variances in Figure 5.3 show that we are clearly able to capture the rapid changes in the dynamics of volatility that occur during the world financial crisis of 2008, in early 2010 with the Greek debt crisis and in the summer of 2011 with the financial speculation in government bonds of European countries, together with the rejection of the U.S. budget and the downgrading of the United States rating. Moreover, the resulting marginal distribution of the log returns induced by the posterior mean of $\mu_j(t)$ and $\Sigma_{jj}(t)$, shows that we are also able to accommodate heavy tails as well as cyclical trends for the means.

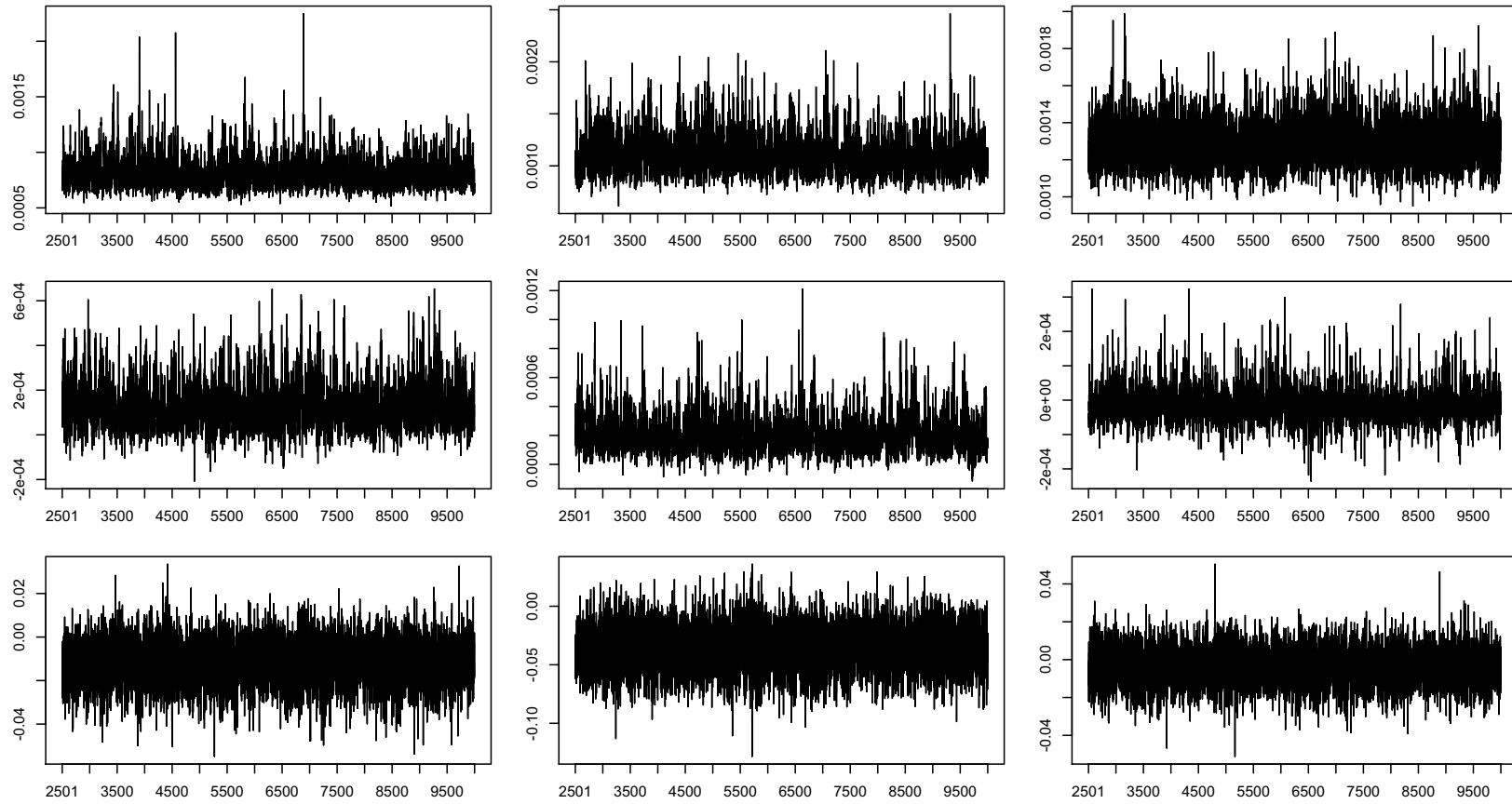


Figure 5.2: Some trace plot for posterior computation in LABNCR model after discarding the first 2,500 iterations as burn-in. Variances (Top), Covariances (Middle), Means (Bottom).

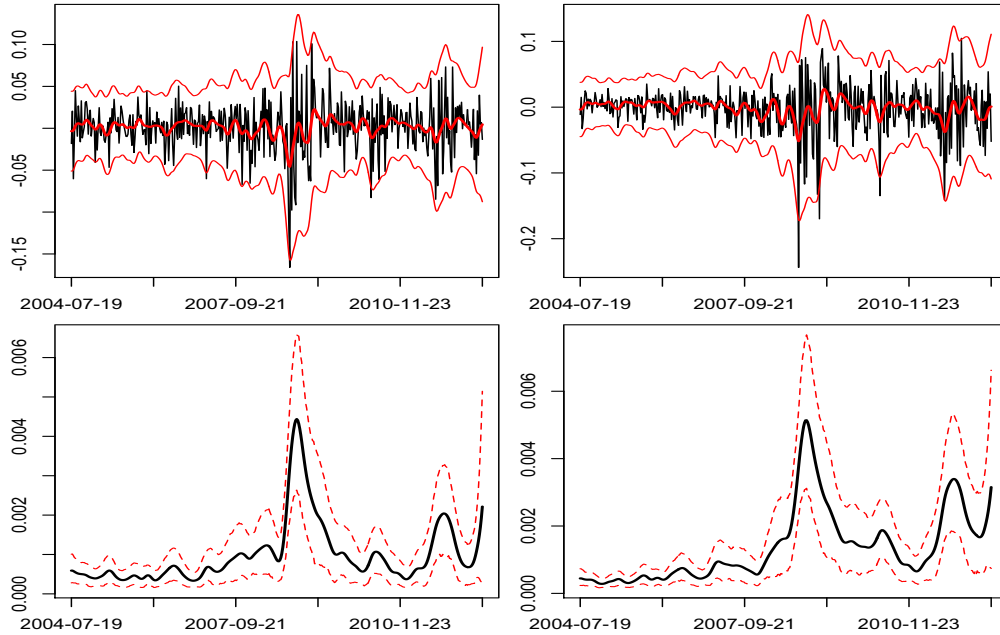


Figure 5.3: Top: Plot for 2 stock market indices, respectively USA NASDAQ (left) and ITALY FTSE MIB (right), of the log returns (black) and the time-varying estimated mean $\{\hat{\mu}_j(t_i)\}_{i=1}^{415}$ together with the time-varying 2.5% and 97.5% quantiles (red) of the marginal distribution for y_{ji} with mean and variance from posterior mean of LABNCR. Bottom: posterior mean (black) and 95% hpd intervals (dotted red) for the volatilities process of the two selected indices.

Important informations about the ability of our model to capture the evolution of world geo-economic structure during different finance scenarios are provided in Figures 5.4, 5.5 and 5.6. From the correlations between NASDAQ and the other stock market indices (based on the posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{415}$ of the covariances function) in Figure 5.4, we can immediately notice the presence of a clear geo-economic structure in world markets, where the dependence between the U.S. and European countries is systematically higher than that of South East Asian Nations (Economic Tigers), showing also different reactions to crises.

Figure 5.5 confirms the above considerations showing how Western countries shows more connection with countries closer in terms of geographical, political and economic structure; the same holds for Eastern countries where we observe a reversal of the colored curves. As expected, Russia is placed in a middle path between the two blocks. A further element that our model captures about the structure of the markets is shown in Figure 5.6. The time-varying regression coefficients obtained from the standard formulas of

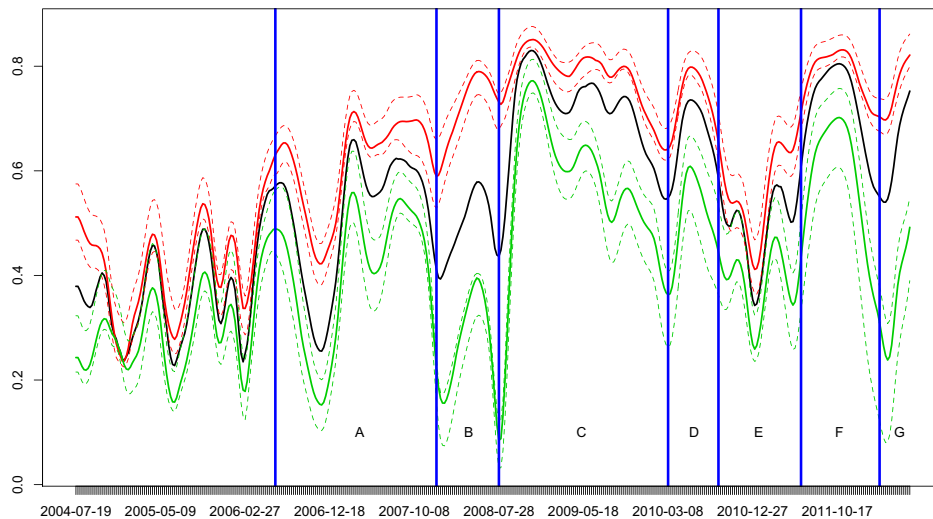


Figure 5.4: Black line: For USA NASDAQ median of correlations with the other 32 world stock indices based on posterior mean of $\{\Sigma(t_i)\}_{i=1}^{415}$. Red lines: 25th, 75th (dotted lines) and 50th (solid line) quantiles of correlations between USA NASDAQ and European countries (without considering Greece and Russia which present a specific pattern). Green lines: 25th, 75th (dotted lines) and 50th (solid line) quantiles of correlations between USA NASDAQ and the countries of Southeast Asia (Asian Tigers and India). The timeline is divided in windows that relate to the main financial events of the recent years. Specifically: event A corresponds to the burst of U.S. housing bubble, event B to the concrete risk of failure of the first U.S. credit agencies (Bear Stearns, Fannie Mae and Freddie Mac), event C to the world financial crisis after the Lehman Brothers' bankruptcy, event D to the Greek debt crisis, event E to financial reform launched by Barack Obama and EU efforts to save Greece (the two peaks represent Irish debt crisis and Portugal debt crisis, respectively), event F to the worsening of European sovereign-debt crisis and the rejection of the U.S. budget, finally G to the crisis of credit institutions in Spain and the growing financial instability Eurozone.

the conditional normal distribution based on the posterior mean of $\{\mu(t_i)\}_{i=1}^{415}$ and $\{\Sigma(t_i)\}_{i=1}^{415}$, highlight clearly the increasing dependence between European countries with higher crisis in sovereign debt and Germany, which plays a central role in Euro zone as expected.

The flexibility of the proposed approach and the possibility of accommodating varying smoothness in the trajectories over time, allows to obtain a good characterization of the dynamic dependence structure according with the major theories on financial crisis. Figure 5.4 shows that the change of regime in correlations occurs exactly in correspondence of burst of U.S. housing bubble (A), in the first half of 2006. Moreover we can immediately notice that the correlations among financial markets increase significantly during

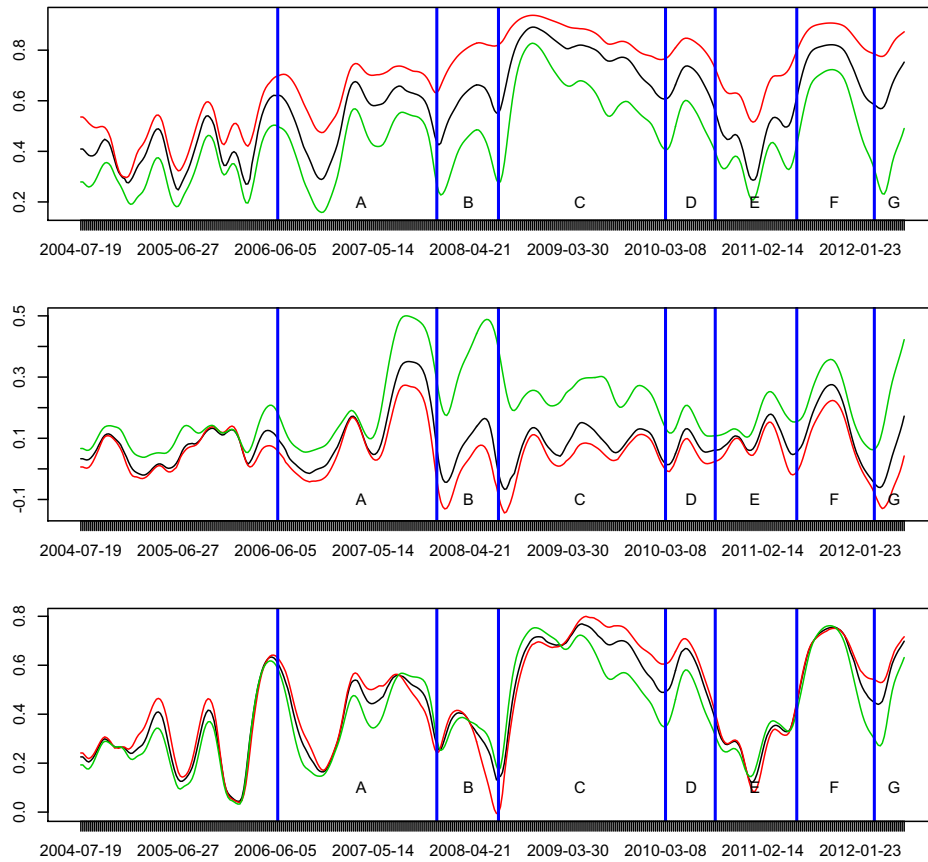


Figure 5.5: For 3 selected Stock Market Indices, respectively GERMANY DAX30 (top), CHINA SSE Composite (middle) and RUSSIA RTSI Index (bottom), plot of the median of the correlations based on posterior mean of $\{\Sigma(t_i)\}_{i=1}^{415}$ with the other 32 world stock indices (black), the European countries without considering Greece and Russia (red) and the Asian Tigers including India (green).

the crisis, showing a clear international financial contagion effect in agreement with other theories on financial crisis (see, e.g., Baig and Goldfajn, 1999 and Stjin and Forbes, 2009). As expected the persistence of high levels of correlation is evident during the global financial crisis between late-2008 and end-2009 (C), at the beginning of which, our approach also capture a dramatic change in the correlations between the U.S. and Economic Tigers, which interestingly lead to levels close to those of Europe. Further rapid changes are identified in correspondence of Greek crisis (D), the worsening of European sovereign-debt crisis and the rejection of the U.S. budget (F), and the recent crisis of credit institutions in Spain, together with the growing

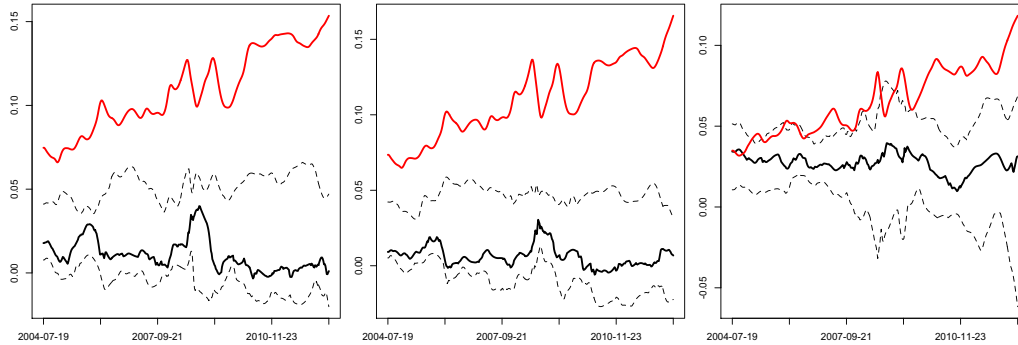


Figure 5.6: For 3 of the European countries more subject to sovereign debt crisis, respectively ITALY (left), SPAIN (middle) and GREECE (right), plot of 25th, 50th and 75th quantiles of the time-varying regression parameters based on posterior mean $\{\hat{\Sigma}(t_i)\}_{i=1}^{415}$ with the other countries (black) and Germany (red).

financial instability in Eurozone (G). Finally, even in the period of U.S. financial reform launched by Barack Obama and EU policy responses through rescue packages to ensure financial stability in Europe (E), we can notice two peaks representing Irish debt crisis and Portugal debt crisis, respectively.

5.3 Updating and Predicting

The possibility to quickly update the estimates and the predictions as soon as new data arrive, represents a crucial aspect to obtain quantitative informations about the future scenarios of the crisis in financial markets. To answer this goal, we apply the online updating algorithm presented in Section 3.4, to the new set of weekly observations $\{y_i\}_{i=416}^{422}$ from 02/07/2012 to 13/08/2012, conditioning on posterior estimates from the Gibbs sampler based on observations $\{y_i\}_{i=1}^{415}$ available up to 25/06/2012. We initialized the simulation smoother algorithm with the last 8 observations of the previous sample.

Figure 5.7 shows, for 3 selected Stock Market Indices, the new observed log returns $\{y_{ji}\}_{i=416}^{422}$ (black) together with the mean and the 2.5th and 97.5th quantiles of the marginal distribution (red) and conditional distribution of $y_{ji}|y_i^{-j}$ with $y_i^{-j} = \{y_{qi}, q \neq j\}$, from the standard formulas of the multivariate normal distribution (green), based on the posterior mean of the updated $\{\Sigma(t_i)\}_{i=416}^{422}$ and $\{\mu(t_i)\}_{i=416}^{422}$ after 5,000 Gibbs iterations with a burn-in of 500. Examination of the trace plots for the time-varying means and covariance matrices showed no evidence against convergence. From these results,

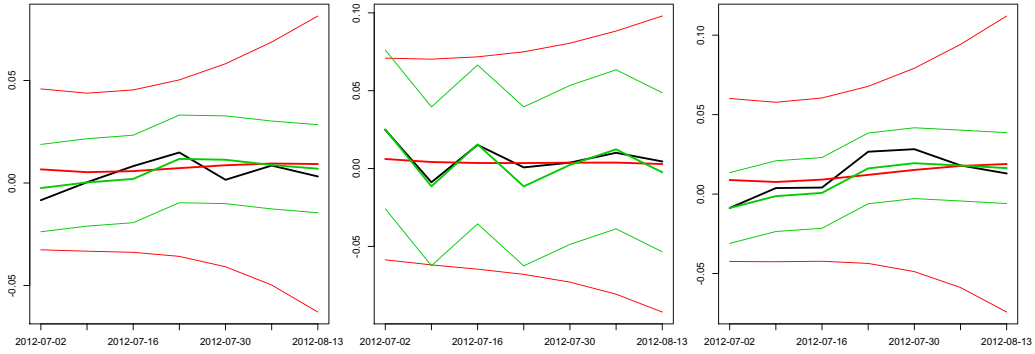


Figure 5.7: For 3 selected Stock Market Indices: USA NASDAQ (left), INDIA BSE30 (middle) and FRANCE CAC40 (right), plot of the observed log returns (black) together with the mean and the 2.5th and 97.5th quantiles of the marginal distribution (red) and conditional distribution given the other 32 world stock indices (green) based on the posterior mean of $\{\Sigma(t_i)\}_{i=416}^{422}$ and $\{\mu(t_i)\}_{i=416}^{422}$, from the online updating procedure for the new observation from 02/07/2012 to 13/08/2012.

we can clearly notice the good performance of our proposed online updating algorithm in obtaining a good characterization for the distribution of new observations. Also note that the multivariate approach, together with flexible model for the mean and covariance, allows for significant improvements when the conditional distribution of an Index given the others are analyzed.

To obtain further informations about the predictive performance of our LABNCR, we can easily use our online updating algorithm to obtain h step-ahead predictions for $\Sigma(t_{T+h|T})$ and $\mu(t_{T+h|T})$ with $h = 1, \dots, H$. In particular, referring to Durbin and Koopman (2001), we can generate the forecasts $\hat{\Sigma}(t_{T+h|T})$ and $\hat{\mu}(t_{T+h|T})$ for $h = 1, \dots, H$ merely by treating $\{y_i\}_{i=T+1}^{T+H}$ as missing values in the proposed online updating algorithm. Here, we consider the one step ahead prediction (i.e. $H = 1$) problem for the new observations. More specifically, for each i from 415 to 421, we update the mean and covariance functions conditioning on informations up to t_i through the online algorithm, and then obtain the predicted posterior distribution for $\Sigma(t_{i+1|i})$ and $\mu(t_{i+1|i})$ by adding to the sample considered for the online updating a last column y_{i+1} of missing values.

Figure 5.8, shows the boxplots of the one step ahead prediction errors for the 33 National Stock Market indices obtained as the difference between the predicted values $\tilde{y}_{j,i+1|i}$ and, once available, the observed log returns $y_{j,i+1}$ with $i + 1 = 416, \dots, 422$ corresponding to weeks from 02/07/2012 to 13/08/2012. In (a) we forecast the future log returns with the unconditional

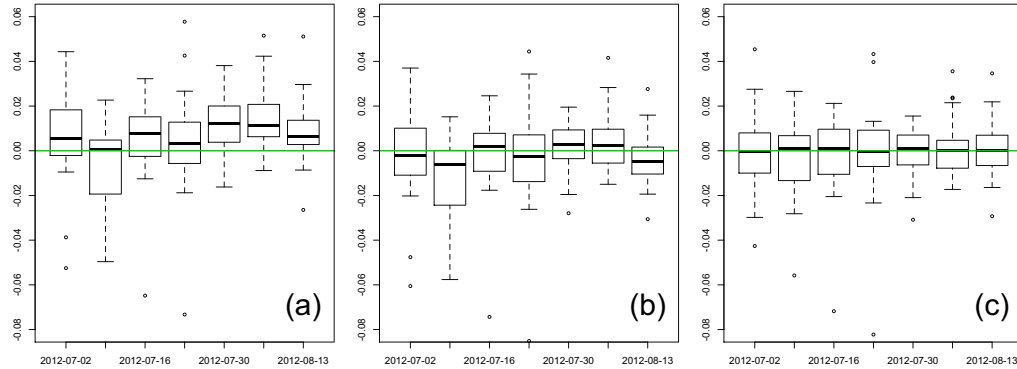


Figure 5.8: Boxplot of the one step ahead prediction errors for the 33 National Stock Market Index, where the predicted values are respectively: (a) unconditional mean $\{\tilde{y}_{i+1}\}_{i=415}^{421} = 0$, (b) marginal mean of the one step ahead predictive distribution using the online updating procedure for $\{\tilde{y}_{i+1|i}\}_{i=415}^{421}$, (c) conditional mean given the log returns of the other 32 Stock Market Indices at $i + 1$ of the one step ahead predictive distribution using the online updating procedure for $\{\tilde{y}_{i+1|i}\}_{i=415}^{421}$. Predictions for (b) and (c) are based on the posterior mean of $\{\Sigma(t_{i+1})\}_{i=415}^{421}$ and $\{\mu(t_{i+1})\}_{i=415}^{421}$ of our LABNCR.

mean $\{\tilde{y}_{i+1}\}_{i=415}^{421} = 0$, which is what is often done in practice under the general assumption of zero mean, stationary log returns. In (b) we consider $\tilde{y}_{i+1|i} = \hat{\mu}(t_{i+1|i})$, the posterior mean of the one step ahead predicted non-parametric mean, obtained from the previous proposed approach after 5,000 Gibbs iteration with a burn-in of 500. Finally in (c) we suppose that the log returns of all Stock Market Indices except that of country j (i.e., $y_{j,i+1}$) become available at t_{i+1} and, considering $y_{i+1} \sim N_p(\hat{\mu}(t_{i+1|i}), \hat{\Sigma}(t_{i+1|i}))$, with $\hat{\mu}(t_{i+1|i})$ and $\hat{\Sigma}(t_{i+1|i})$ posterior mean of the one step ahead predictive distribution for $\mu(t_{i+1|i})$ and $\Sigma(t_{i+1|i})$ respectively, we forecast $\tilde{y}_{j,i+1}$ with the conditional mean of $y_{j,i+1}$ given the other log returns at time t_{i+1} .

Comparing boxplot in (a) with those in (b) we can see that our model allows to obtain improvements also in terms of prediction. Furthermore, by analyzing the boxplot in (c) we can notice how our ability to obtain a good characterization of the time-varying covariance structure, can play a crucial role also in improving forecasting, since it enters into the standard formula for calculating the conditional mean in the normal distribution.

Discussion

In this work, we have presented a generalization of Bayesian Nonparametric Covariance Regression in order to obtain a better characterization for mean and covariance temporal dynamics. The founding element of our approach is the assumption of nGP prior for the random functions for the dictionary elements ξ_{lk} and for the function controlling the mean structure ψ_k , in order to allow locally adaptive smoothing both for the time-varying covariance and mean functions.

Maintaining simple conjugate posterior updates and tractable computations in large- p settings from Fox and Dunson (2011) latent factor model formulation for y_i , our model increases significantly the flexibility of previous approaches as it allows to capture even dramatic changes both in mean and covariance dynamics, improving predictive performance, and leading to conditional distribution able to accommodate even heavy tails. Beside these key advantages, the state space formulation for nGP prior enables us to develop a fast online updating algorithm particularly worthy in application with high frequency data, that can be easily used also to make inference on h steps ahead predictive distributions for y_i .

We compared our approach with the main competing alternative, through a simulation study, showing the better performance of LABNCR, which also highlights good results in the online updating. The application to the problem of capturing temporal and geo-economic structure between the main financial markets, demonstrates the utility of our approach and the improvements that can be obtained in the analysis of multivariate financial time series with reference to (i) heavy tails, (ii) cyclical trends in the mean structure, (iii) dramatic changes in mean and covariance functions, (iii) high dimensional dataset, (iv) online updating with high frequency data and (v) predictions.

Although we focused our attention on multivariate financial time series, the proposed approach can be easily considered in other fields of research. In

medical applications, for example, the evaluation of patients condition over time often leads to the availability of large quantities of online time-varying indicators, whose joint analysis can provide important information regarding the progress of a disease. Other important examples of high dimensional multivariate time series, in which the dynamic analysis of the dependence structure plays a crucial role, can be found in computer science, meteorology as well as bioinformatics. In all cases mentioned, our model could be particularly worthy in obtaining a flexible characterization of the dynamic evolution of the structure of dependence among the time-varying variables analyzed. Moreover the proposed approach can be considered also when t is an arbitrary predictor value. From this point of view, a direct extension of LABNCR relates to the generalization of the model to accommodate a multivariate predictor space.

Bibliography

- Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics* 18, 338–357.
- Alexander, C. (2001). Orthogonal GARCH. *Mastering Risk* 2, 21–38.
- Baig, T. and Goldfajn, I. (1999). Financial Market Contagion in the Asian Crisis. *Staff Papers, International Monetary Fund* 46, 167–195.
- Bernanke, B.S. (2007). Global Imbalances: Recent Developments and Prospects. <http://www.federalreserve.gov/newsevents/speech/bernanke>. Bundesbank Lecture, Berlin, Germany.
- Bhattacharya, A. and Dunson, D.B. (2011). Sparse Bayesian infinite factor models. *Biometrika* 98, 291–306.
- Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988). A capital-asset pricing model with time-varying covariances. *Journal of Political Economy* 96, 116–131.
- Baily, M.N. and Elliott D.J. (2009). The US Financial and Economic Crisis: Where Does It Stand and Where Do We Go From Here? Business and Public Policy at Brookings.
- Brunnermeier, M.K. (2009). Deciphering the Liquidity and Credit Crunch 2007–2008. *Journal of Economic Perspectives* 23, 77–100.
- Burns, P. (2005). Multivariate GARCH with Only Univariate Estimation. <http://www.burns-stat.com>.
- Carvalho, C.M., Lucas, J.E., Wang, Q., Chang, J., Nevins, J.R. and West, M. (2008). High-dimensional sparse factor modeling - Applications in gene expression genomics. *Journal of the American Statistical Association* 103, 1438–1456.

- Chib, S., Omori, Y. and Asai, M. (2009). Multivariate stochastic volatility. In *Handbook of Financial Time Series*, pages 365–400.
- Ding, Z. (1994). Time series analysis of speculative returns. PhD thesis, University of California, San Diego.
- Donoho, D.L. and Johnstone, J.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, 425–455.
- Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press Inc., New York.
- Durbin, J. and Koopman, S. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89, 603–616.
- Engle, R.F. and Kronos, K.F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Fama, E.F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business* 38, 34–105.
- Financial Crisis Inquiry Commission (2011). This Crisis was Avoidable—a Result of Human Actions, Inactions and Misjudgments; Warning Signs Were Ignored. *Releases Report on the Causes of the Financial Crisis*.
- Fox, E. and Dunson, D.B. (2011). Bayesian Nonparametric Covariance Regression. *arXiv:1101.2017*.
- Gallo, G.M. and Pacini, B. (2002). *Metodi quantitativi per i mercati finanziari* Carocci editore, Roma.
- Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–511.
- Geweke, J. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies* 9, 557–587.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments. In J. M. Bernardo, J. O. Berger, A. P. Dawiv, and A. F. M. Smith, eds., *Bayesian Statistics*, volume 4, Oxford, UK: Clarendon Press.

- Guermat, C. and Harris, R. (2002). Robust Conditional Variance Estimation and Value-atRisk. *Journal of Risk* 4, 25–41.
- Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies* 61, 247–264.
- Krugman, P. (2002). Dubya's Double Dip?. *The New York Times*. Retrieved 30 Jul 2012.
- Lopes, H.F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* 14, 41–67.
- Nakamoto, M. (2007). Citigroup chief stays bullish on buy-outs *Financial Times*.
- Philipov, A. and Glickman, M.E. (2006a). Multivariate stochastic volatility via Wishart processes. *Journal of Business & Economic Statistics* 24, 313–328.
- Philipov, A. and Glickman, M.E. (2006b). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews* 25, 311–334.
- Prado, R. and West, M. (2010). *Time Series: Modeling, Computation, and Inference*. Chapman & Hall / CRC, Boca Raton, FL.
- Predetti, A. (2006). *I numeri indici. Teoria e pratica*. XI ed., Giuffrè, Milano.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, 289–306.
- Stijn, C. and Forbes, K. (2009). *International Financial Contagion: An overview of the Issues*. Springer.
- Taylor, J.B. (2009). The Financial Crisis and the Policy Responses: An Empirical Analysis of What Went Wrong. *NBER, Working Paper 14631*.
- Tsay, R.S. (2005). *Analysis of Financial Time Series*. II ed., Wiley.
- van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17, 549–564.
- West, M. (2003). Bayesian factor regression models in the large p , small n paradigm. *Bayesian Statistics* 7, 723–732.

Witter, L. (2006). The No-Money-Down Disaster. *Barron's*.

Zhu, B. and Dunson, D.B. (2012). Locally Adaptive Bayes Nonparametric Regression via Nested Gaussian Processes. *arXiv:1201.4403*.