Three Essays of Applied Bayesian Modeling: Financial Return Contagion, Benchmarking Small Area Estimates, and Time-Varying Dependence

Three Essays of Applied Bayesian Modeling: Financial Return Contagion, Benchmarking Small Area Estimates, and Time-Varying Dependence

A dissertation presented

by

Andrew Jay Vesper

to

The Department of Statistics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Statistics

Harvard University

Cambridge, Massachusetts

May 2013

Dissertation Advisor: Professor Carl Morris Andrew Jay Vesper

Three Essays of Applied Bayesian Modeling: Financial Return Contagion, Benchmarking Small Area Estimates, and Time-Varying Dependence

Abstract

This dissertation is composed of three chapters, each an application of Bayesian statistical models to particular research questions.

In Chapter 1, we evaluate systemic risk exposure of financial institutions. Building upon traditional regime switching approaches, we propose a network model for volatility contagion to assess linkages between institutions in the financial system. Focusing empirical analysis on the financial sector, we find that network connectivity has dynamic properties, with linkages between institutions increasing immediately before the recent crisis. Out-of-sample forecasts demonstrate the ability of the model to predict losses during distress periods. We find that institutional exposure to crisis events depends upon the structure of linkages, not strictly the number of linkages.

In Chapter 2, we develop procedures for benchmarking small area estimates. In sample surveys, precision can be increased by introducing small area models which "borrow strength" by incorporating auxiliary covariate information. One consequence of using small area models is that small area estimates at lower geographical levels typically will not aggregate to the estimate at the corresponding higher geographical levels. Benchmarking is the statistical procedure for reconciling these differences. Two new approaches to Bayesian benchmarking are introduced, one procedure based on Minimum Discrimination Information, and another for Bayesian self-consistent conditional benchmarking. Notably the proposed procedures construct adjusted posterior distributions whose moments all satisfy benchmarking constraints. In the context of the Fay-Herriot model, simulations are conducted to assess benchmarking performance.

In Chapter 3, we exploit the Pair Copula Construction (PCC) to develop a flexible multivariate model for time-varying dependence. The PCC is an extremely flexible model for capturing complex, but static, multivariate dependency. We use a Bayesian framework to

extend the PCC to account for time dynamic dependence structures. In particular, we model the time series of a transformation of parameters of the PCC as an autoregressive model, conducting inference using a Markov Chain Monte Carlo algorithm. We use financial data to illustrate empirical evidence for the existence of time dynamic dependence structures, show improved out-of-sample forecasts for our time dynamic PCC, and assess performance of dynamic PCC models for forecasting Value-at-Risk.

Contents

Acknowledgments

I am deeply indebted to Dr. Yoonjung Lee for her continued support throughout my graduate school career, even now as her children have claimed most of her time and energy. Dr. Lee played a crucial role in the development of Chapters 1 and 3 of this dissertation. The pair copula conference in Munich was a turning point for my career, and that opportunity would not have been possible without the financial support of Dr. Lee.

Prof. Carl Morris is easily the most insightful and deep-thinking statistician with whom I have had the pleasure to interact. Nearly all of my foundational statistical knowledge has been developed under Prof. Morris's guidance, in courses 210/211/399/392, in consulting work with MLB, and in research. Prof. Morris has also been an important advocate on my behalf with the department, and without his support the completion of this dissertation would have been a much more difficult process.

I am also very thankful for Prof. Stephen Blyth's comments, advice, and willingness to serve on my dissertation committee. The perspective of a financial practitioner has improved the quality and content of this dissertation, particularly Chapter 1. Prof Blyth has been unbelievably supportive of my work, and his early feedback on Chapter 3 has been extremely encouraging. Additionally, I join my colleagues in thanking Prof. Blyth for his endowment of the Dempster Prize.

I want to express my gratitude to the statistical research group at the U.S. Census Bureau, specifically Dr. Tommy Wright, Dr. Jerry Maples, and Dr. Ryan Janicki, for their guidance during an internship in 2011. The research question generating Chapter 2 was developed in my work at the U.S. Census. Any insight contained in Chapter 2 is undoubtedly due to the ongoing collaboration with Dr. Janicki, who is also a constant reminder of the necessity for statistical humility.

I would not be the person I am today without the support of my family (Mom, Dad, Katy, and P.J), my alma mater (UGA), my closest friends (Brooks and Mike), and my roommate (Leah). In terms of my academic career, the course of study for a PhD is a long and daunting process. I can honestly say that all of you have been there to pick me up at moments in the past 6 years when I was feeling especially helpless. Hopefully I have been able to do the same for each of you. Leah, without 3 years of fanfare and Sunday dinners and decompressing long runs (though, occasionally not so decompressing) and 6-second hugs, I never would have made it this far.

Citation

A modified version of Chapter 3 has been previously published:

Vesper, A. (2012). "A time dynamic pair copula construction: with financial applications," *Applied Financial Economics*, 22, 1697-1711.

To my parents, who at least once gave me a kick in the pants.

Chapter 0

Introduction

John Tukey is famously quoted as having said that "the best thing about being a statistician is that you get to play in everyone's backyard." This dissertation is a manifestation of that ideal, presenting results of applied Bayesian statistical methodology in the fields of empirical finance with exploration of systemic risk exposure, small area estimation in sample surveys, and econometric techniques for time-varying dependence structures. While there is no direct link between these topics, the projects all pursue the goal of applying statistics to solve a research question of interest in someone's "backyard." On a personal level, the joy I find in statistics is precisely in the diverse array of potential applications. With each project in a different area of research, a statistician is presented with the opportunity to engage with, learn about, and gain exposure to another field. Accordingly, every new project is an opportunity for personal and professional growth. Because applied projects have unique statistical requirements, important theoretical advances almost always develop from applied work. I would argue that, in many ways, a Ph.D. dissertation in statistics not only could be, but actually *should* be a composition of a variety of applied statistics projects.

This chapter outlines the three projects which comprise the dissertation, briefly describing the research question of interest in a field of application, the statistical tools required to address the question, and the primary conclusions of the study. The brief descriptions given

here are intended to be less technical, with appeal to a wider audience. A substantially more detailed introduction to each of the projects is given in Chapters 1, 2, and 3.

0.1 A Network Model for Volatility Contagion

Systemic risk is broadly defined as the risk of failure of an entire financial system. With the ongoing European sovereign debt crisis, and following the domestic financial crisis, which peaked in 2008 with the failure of Lehman Brothers and Bear Sterns and the government takeover of American International Group (AIG), there has been increased awareness of global systemic risk. Financial institutions, domestically and abroad, have been subjected to intense scrutiny, including required stress testing for evaluating the ability of particular institutions to survive periods of distress.

While an understanding of systemic risk is crucial in the current global financial landscape, relatively few academic studies of systemic risk exposure have been conducted. Because there is no concrete definition of systemic risk, it is difficult to define a structural model of the economy to quantify systemic risk exposure of particular institutions. Generally, researchers have proposed descriptive models for empirical analysis intended to capture specific characteristics of systemic risk. Though such models are neither causal nor theoretically linked to the economy, researchers have been able to demonstrate relationships between the models and predictions for firm characteristics associated with systemic risk. In Chapter 1 a new metric is proposed, with the goal of contributing to the growing catalog of measures of systemic risk exposure.

A key requirement for systemic risk is the existence of linkages between institutions. An institution may fail, but if its risk exposure is isolated, it will present no threat to the health of the greater financial system. As firms become more interconnected, systemic risk increases. The model developed in Chapter 1 is unique in its goal to implement a statistical network model to capture unobserved linkages between institutions on the basis of publicly available

data. Specifically, the model assumes that institutional returns are subject to two distinct volatility regimes: distress and tranquility, and the probability of regime transitions depends on connectivity in the network structure. Thus, the network connections are interpreted as volatility (or distress) contagion. Given this definition, an assessment of the relationship between contagion and systemic risk exposure is presented.

Empirical application of the network model illustrates time dynamic patterns of network connectivity, with linkages between institutions increasing before the recent crisis. Out-of-sample tests show a relationship between characteristics of the network and crisis period losses of institutions. Results indicate that the structure of the network is more influential for institutional losses than merely the number of connections. However, the number of connections are found to be associated with firm characteristics of systemic risk exposure, namely, leverage.

0.2 Benchmarking Bayesian Small Area Estimates

Administrators of sample surveys are often interested in producing estimates at different levels of geographic aggregation. As an illustrative example, suppose a survey is conducted to produce estimates of the counts of impoverished individuals within each county and state in the United States. When reporting these estimates, it is often required that the county estimates be consistent with state estimates; that is, the estimated count of impoverished individuals for each county within a state must sum to the estimated count of impoverished individuals in that state. The direct survey estimates and variances, which are based upon the survey design, will satisfy this constraint.

Precision for estimates can often be improved by introducing small area models, which "borrow strength" by connecting different areas (e.g. counties) and introducing auxiliary covariate information (Rao 2003). A consequence of these models is estimates across different geographic levels will no longer be consistent. The statistical procedure for reconciling these

differences and achieving consistency is *benchmarking*.

Traditional benchmarking procedures adjust lower level point estimates to achieve consistency. Adjustments are typically based on the uncertainty of these estimates. Chapter 2 develops new perspectives for benchmarking estimates from small area models, especially complex models estimated via Markov Chain Monte Carlo (MCMC). The approaches developed in Chapter 2 stand in contrast to existing procedures by adjusting entire posterior distributions, not just point estimates. As a result, all moments of these distributions, including point estimates and variances, will be consistent across geographic aggregations. The two new approaches are based on the principle of Minimum Discrimination and fully conditional Bayesian distributions estimated via MCMC. Other benefits over the traditional procedures include generalization to more than two geographic levels, the ability to compromise between the higher and lower level models, and benchmarked standard errors.

Simulations from the Fay-Herriot model (Fay and Herriot 1979), which is a popular small area model in practice, are used to evaluate the performance of the proposed procedures, and the results suggest that the new approaches outperform existing techniques.

0.3 A Time Dynamic Pair Copula Construction

The copula, a particular choice of statistical model, has taken blame for contributing to the domestic subprime mortgage crisis. Specific limitations of the copula model resulted in a failure to capture dynamic changes in the mortgage landscape, creating important errors in the pricing and rating of specific mortgage backed financial securities.

A copula is a multivariate distribution function with uniform margins. Any multivariate density can be factorized as the product of the marginal densities and a copula function. The copula is a popular modeling framework because it allows the analyst to separate the dependence structure of the multivariate distribution from the marginal distributions. Unfortunately there are several important features of multivariate copula densities which make

them unsuitable for many empirical applications, especially in finance. Notably, the catalog of multivariate copula is relatively sparse. Practitioners often rely on Gaussian copula, which have dependence features incompatible with empirically observed financial returns data. Another potential drawback of the copula is that the dependence structure is constant over time. In reality, financial correlations are constantly changing in response to market shocks, regulatory conditions, product offering changes, and countless other dynamic features of the financial marketplace. Chapter 3 evaluates copula-based methods which allow for time-varying dependence structures in high-dimensional multivariate data.

In econometrics, finance, and statistics, the class of models for multivariate time-varying dependence is primarily limited to Dynamic Conditional Correlation (DCC), proposed by Engle (2002). DCC shares an important limitation of the multivariate copula by typically relying on Gaussian dependence structures. A general goal of Chapter 3 is to develop a more flexible model for time-varying dependence. The application of such a model would not be restricted to finance, but is more widely applicable to any time series of data with dynamic dependence features.

The Pair Copula Construction (PCC), is a recent econometric modeling technique based upon a cascading pairwise conditional factorization of a multivariate copula into bivariate pair copula. Because there is a richer class of bivariate copula densities than multivariate copula densities, this factorization results in an extremely flexible model for multivariate dependence. The methods presented in Chapter 3 extend the PCC to account for time-varying dependence. From a Bayesian perspective the analyst can introduce time dynamics by implementing an autoregressive time series prior distribution for parameters of these bivariate pair copula. By applying time-varying dependence structure to pairs, rather than a multivariate distribution, many of the technical concerns for time-varying covariance matrices are avoided (e.g. positive semi-definiteness) which otherwise limit the modeling capabilities. Sklar (1959) guarantees that the PCC with time-varying bivariate parameters will produce a valid time-varying multivariate distribution.

Chapter 3 applies this Bayesian PCC model with time-varying dependence to financial returns and Value at Risk (VaR) estimates. In both cases there is strong evidence for the existence of time-varying dependence, and the out-of-sample forecast error of the dynamic PCC significantly outperforms the traditional static PCC.

Chapter 1

A Network Model for Volatility Contagion

1.1 Introduction

In light of ongoing global financial uncertainty, systemic risk has become an important topic of empirical research. Broadly, *systemic risk* is the risk of failure of an entire financial system. While an understanding of systemic risk is critical to regulatory and policy decisions for financial markets, a more precise definition of systemic risk has not been established. Despite this lack of clarity, market participants are often able to observe systemic risk. Unfortunately, the observation typically takes place only after a financial crisis has already occurred.

Systemic events of the past few years include the domestic financial crisis and the more recent European sovereign debt crisis. The former peaked in 2008 with the failure of Bear Stearns and Lehman Brothers and the government takeover of American International Group (AIG). The latter resulted in downgrading of government debt to "junk" status and near default in Greece, Spain, Portugal, and Ireland. In both cases, significant bailout funds were established to prevent more widespread catastrophe. Domestically, the federal government

established the Troubled Asset Relief Program (TARP) to purchase assets and equities from financial institutions in an effort to strengthen the financial system. Though the original TARP bill authorized $700 billion in purchases, by the end of 2012, $418 billion was disbursed, of which $405 billion has already been recovered through these investments (United States Department of Treasury 2013). In Europe, central banks allocated hundreds of billions of Euros in bailout funds or other concessions for troubled nations, most notably Greece. These actions, domestically and abroad, are largely perceived as protections against systemic collapse in the event of failure of the protected institutions or sovereignties. Understanding systemic risk is clearly of central importance with respect to the current financial and political landscape.

Paraphrasing Billio et al. (2012a), financial crises can be characterized by four "L"s: leverage, liquidity, losses, and linkages. As a brief case study, consider Long Term Capital Management (LTCM), a hedge fund management firm founded in 1994. (Lowenstein (2001) provides a detailed exposition of the rise and fall of LTCM; the following is a brief summary and the facts presented refer to that source). The trading strategies pursued by LTCM largely sought opportunities for fixed income arbitrage. Because arbitrage is often characterized by very small differences in price, LTCM used leverage to increase profits from these trades. Early in 1998, the firm had a debt to equity ratio over 25 to 1. As a result of the impact of the Russian financial crisis on the bond market and the leveraged positions that LTCM held, between May and August of 1998 LTCM lost nearly $2 billion. These losses were magnified when LTCM was forced to liquidate further trading positions at a loss as investors withdrew from the fund. With billions of dollars in debts still on the balance sheet, and owed to most major Wall Street participants, the Federal Reserve Bank of New York was forced to orchestrate a bailout of LTCM to prevent widespread financial collapse.

The LTCM case study highlights the fact that linkages are a necessary condition for systemic risk. If LTCM had merely suffered massive losses compounded by its leverage and illiquidity, but had no counter-parties on its trades or loans, there would have been no risk

to the wider financial system. Of course, trades require counter-parties, and institutions are necessarily linked, either directly through these trades, shared risk exposure, and correlation among holdings, or indirectly through price effects and liquidity spirals (Brunnermeier and Pederson 2009). It has been established that the extent to which financial institutions are linked to each other provides an indication of the degree of systemic risk present in the financial system (Acharya and Richardson 2009; Allen and Gale 2000; Allen et al. 2012; Brunnermeier and Pederson 2009; Gray 2009; Rajan 2006; Reinhart and Rogoff 2009). As links between institutions become more numerous or more concentrated, susceptibility to systemic events increases.

Conceptually, measures of systemic risk can be constructed through observations, descriptions, and evaluations of the four "L"s. Losses and leverage can be observed directly from financial statements and corporate reports, especially as transparency and regulation has increased globally in recent years. The relationship between liquidity and crises has been an historically active topic of research. Getmansky et al. (2004) and Chan et al. (2006) consider illiquidity in hedge fund returns, showing that serial correlation is a proxy for illiquidity and often increases following negative market shocks. Billio et al. (2012b) find that the effect of liquidity factors on hedge fund returns is dynamic and increases during crisis regimes. Similarly, Boyson et al. (2010) show that liquidity shocks lead to an increase in the probability of hedge fund contagion. For equities, several measures of liquidity have been explored (Amihud and Mendelson 1986; Chordia et al. 2002; Lo and Wang 2000; Pastor and Stambaugh 2003; Sadka 2006).

Relatively little of the academic literature is devoted to the impact of linkages on systemic risk. Three recent studies have proposed measures of systemic risk based on the principle of assessing losses of institutions during periods of time when other institutions are likewise suffering losses. Adrian and Brunnermeier (2011) develop Conditional Value at Risk (CoVaR), which estimates the Value at Risk (VaR) of the financial system conditional on the state of a particular institution. The institution's contribution to systemic risk is the

amount by which CoVaR evaluated conditional on distress of the institution exceeds CoVaR evaluated conditional on the median state of the institution. Similarly, Acharya et al. (2010) propose Systemic Expected Shortfall (SES), which measures the likelihood of an institution to be undercapitalized when the system as a whole is undercapitalized. Huang et al. (2012) compute the Distressed Insurance Premium (DIP), which is the insurance premium required to cover distressed losses in the financial system on the basis of underlying credit default swaps. An institution's individual contribution to systemic risk can be determined based on its marginal impact on DIP.

These approaches are all philosophically consistent with the concept of systemic risk in their intent to evaluate losses conditional on financial distress. However, statistical estimation will only be reliable if other similar periods of distress are represented in the available data. For example, correlations between institutions increase during a period of financial distress (Longin and Solnik 2001). The ability of CoVaR, SES, and DIP to forecast systemic risk exposure may be limited if correlations observed in periods preceding a systemic event are much lower than those during the crisis itself. Billio et al. (2012a) propose an alternative approach to evaluate systemic risk directly and unconditionally by conducting pairwise Granger causality tests (Granger 1969) for all institutions in a financial system. A directed network of connectivity is constructed on the basis of the direction of Granger causality for each pair, and the links between particular institutions are shown to be reflective of their exposure to systemic events. Where CoVaR, SES, and DIP measure connectivity between financial institutions indirectly based on measures of conditional loss, Billio et al. (2012a) explicitly evaluate linkages in the financial system.

This chapter aims to complement existing measures of systemic risk, particularly that of Billio et al. (2012a), by introducing a model for the latent network structure of financial institutions. We assess systemic risk exposure through inference on this network structure and the linkages between institutions within the network. Our approach builds upon the foundation of the regime switching models of Chan et al. (2006) and Billio et al. (2012b),

where institutional returns are assumed to come from either a high or low volatility distributional state. We extend this model by introducing a latent network of the institutions which governs the regime switching process. An institution's probability of entering the high volatility regime in the current time period is greater if it is connected via the latent network to other institutions in the high volatility regime in the previous time period. Section 1.2 provides further details.

We establish a working definition of *volatility contagion* in the context of our model to be the impact, via the network, of a particular institution entering the high volatility regime on the volatility of all institutions. Forecasts of contagion can be constructed from the network and volatility regime parameters of the model. Because the high volatility regime generally corresponds to a crisis state, the network captures two important aspects of systemic risk: the likelihood of a collection of financial institutions being in the high volatility state contemporaneously (losses) and volatility contagion (linkages).

There are several reasons why our modeling approach is advantageous. First, the model provides an illustration of how volatility contagion propagate through the network. We can estimate the effect of an individual institution entering a high volatility regime on overall market volatility. Similarly, we can estimate which institutions are most likely to be impacted by other institutions entering the high volatility regime. Institutions which have a strong impact on others, and institutions which are impacted strongly by others, may be especially important in evaluating systemic risk exposure. Even though such inferences are not causal, they provide a descriptive portrayal of volatility and risk spill-overs in financial markets, and we present results from out-of-sample tests demonstrating these relationships. Additionally, SES, DIP, and CoVaR all assess the magnitudes of losses when other institutions are simultaneously distressed. These measures are only quantifiable to the degree that historical data is representative of similar situations and they may fail to recognize systemic exposures that have not been historically observed, including potential increases in systemic risk exposure during non-crisis periods. Accordingly, the utility of these models in forecast-

ing systemic risk exposure prior to a systemic event occurring may be compromised. By focusing on descriptive measures of connectivity, and not conditioning on observed losses, our approach can recognize linkages even when losses are not occurring. Lastly, by estimating a network structure on a set of institutions we capture more diverse network dynamics than Billio et al. (2012b), who test for Granger causality only on pairs of institutions.

Empirical application of our model reveals that overall connectivity for the hedge fund industry is comparable to the equity market. Focusing on the largest financial institutions in three major sectors - banks, brokers, and insurers - we find that the pattern of connectivity is dynamic over time. The number of linkages increased prior to the domestic financial crisis in 2008. In the post-crisis economy, a changing regulatory landscape has coincided with different patterns of connectivity, with links across industries becoming relatively more common than links within industries. At the institutional level, forecasts show that losses during out-of-sample distress periods are not associated with the number of connections an institution has; however, losses are significantly higher for those most impacted by volatility contagion in the network model and significantly lower for those that are most responsible for such contagion. The structure of connectivity in the network is more important for identifying exposure to systemic risk for individual institutions than the number of connections.

We emphasize that this study is descriptive and provides no causal implication for the relationship between inferred network connectivity and systemic risk exposure, nor does it correspond to any theoretical model of the economy. In particular, there is not an explicit economic interpretation of linkages in our model. The degree of connectivity in the network structure is only a measure of systemic risk to the extent that volatility contagion as defined by our model is indicative of systemic risk. Because of the lack of a structural or causal foundation for the model, any conclusions about the relationship between the model and the economy are inherently speculative. Nonetheless, we will identify several suggestive patterns relating features of the model, volatility contagion, and network connectivity to empirically observed market outcomes.

Given this limitation, the goal of this chapter is not to replace or compete with existing measures, but rather to contribute to a growing catalog of systemic risk metrics. To quote Billio et al. (2012a), "...the risk of [systemic] events is multifactorial and unlikely to be captured by any single metric. After all, how many ways are there of measuring 'stability' and 'public confidence'?" The model we have chosen is somewhat arbitrary and requires several assumptions, but without a concrete definition of systemic risk, all models intended to describe features of systemic risk are necessarily arbitrary. The primary aim of our chosen model is to capture the unobserved network linking firms in the financial industry. Without proprietary trading information, or access to data controlled by central banks, a structural model for a network of institutions is unrealistic. Using only publicly available data, layering a network structure on top of a model for volatility regime switching is just one possible framework to assess linkages between institutions. We have conveniently chosen to interpret these linkages as volatility contagion. Traditionally, volatility contagion is evaluated through multivariate Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, but our model is intentionally less sophisticated, allowing only for two volatility regimes corresponding to periods of distress and tranquility. An alternative interpretation of linkages that is possibly more consistent with the regime switching construction might be *distress contagion*, as opposed to *volatility contagion*. While we emphasize that this model is limited in several ways, it is a unique systemic risk metric in its ability to infer a latent network of linkages between institutions based solely on publicly available data. Out-of-sample tests forecasting institutional losses during the domestic subprime mortgage crisis are generally supportive of the utility of this model for assessing systemic risk exposure of institutions.

The rest of the chapter is organized as follows. In Section 1.2 we fully describe our model of connectivity and its relationship to systemic risk. Section 1.3 and Section 1.4 provide empirical analysis of connectivity in the hedge fund and financial sectors, with discussion of changes in connectivity over time and forecasting crisis period losses. Section 1.5 concludes.

1.2 Model

A natural statistical model for capturing systemic risk exposure due to linkages between institutions is a network model, which is used to describe features of a network of connections. Network models have seen an extremely diverse array of applications: in the social sciences with application to social networking on websites such as Facebook, in the natural sciences with application to protein interactions, in government intelligence with application to terrorist networks, in politics with application to bill co-authorship, in economics with application to labor markets, and many other areas. In finance network models have most frequently been used to assess financial stability. Empirical network modeling has been conducted for assessing direct linkages via contagion (Allen and Gale 2000; Dasgupta 2004; Leitner 2005), indirect linkages via balance sheets (Cifuentes et al. 2005; Lagunoff and Schreft 2001), and how failures of institutions result from mutual claims on each other (Furfine 2003; Upper and Worms 2004; Wells 2004). Babus and Allen (2009) provide a review of network models in finance.

Another common statistical technique in the financial literature is a regime switching model. Regime switching has been used for empirical application in finance for asset pricing (Bekaert and Harvey 1995), asset allocation (Ang and Bekaert 2002; Guidolin and Timmermann 2008), volatility contagion (Billio and Pelizzon 2000, 2003), and dynamic hedge fund risk exposure (Billio et al. 2012b; Chan et al. 2006).

This chapter combines these two statistical tools in the context of a fully Bayesian modeling framework. As described in Section 1.1, our model assumes that underlying financial returns of institutions there exists a latent volatility regime and a latent network structure which links the regime switching transition probabilities of institutions.

We use a fully Bayesian approach to estimate the parameters of the model, making use of a Markov Chain Monte Carlo (MCMC) routine to sample from the posterior distribution. Parameter estimates and posterior intervals can be constructed on the basis of these samples.

Let us establish notation for the following discussion. Consider a multivariate time series

of returns for $i = 1, ..., n$ institutions observed at time periods $t = 1, ..., T$. Let $G_{i,j} = 1$ (for $i < j$) indicate that institutions i and j are connected via the network structure, $I_{i,t} = 1$ indicate that institution i is in the high volatility state at time t, and $Y_{i,t}$ be the return of institution i at time t. $G^{n \times n}$, $I^{n \times T}$, and $Y^{n \times T}$ are the corresponding matrices of parameters and data. The only observed data is contained in Y. G and I are matrices of parameters for the latent network and regime indicators underlying the observed return distribution. The power of the Bayesian modeling framework is the ability to estimate the elements of G and I via MCMC methods, including the ability to quantify the uncertainty of these estimates and generate forecasts based upon the MCMC samples from the posterior distribution.

For the purposes of intuition, we separately outline the three primary components of our model in Section 1.2.2, Section 1.2.3, and Section 1.2.4. Section 1.2.5 provides full details of the model and its estimation, including description of prior distributions on the parameters.

1.2.1 Filtering Returns

Financial returns typically exhibit strong positive correlations, reflective of mutual exposure to risk factors. The goal of our study is to infer linkages and estimate volatility contagion from institution to institution. We desire to avoid mistakenly identifying contagion which merely reflect common market factors. In order to capture the idiosyncratic returns and volatility of an institution, we filter the raw return data by applying a multi-factor model, which explicitly controls for institutional exposure to shared risk factors.

The risk factors and brief definitions are listed in Table 1.1. Market, Large-Small, and Value-Growth are the Fama-French factors (Fama and French 1993) and Momentum is an extension of the Fama-French model which accounts for persistence in return performance (Carhart 1997). The choice of other factors is based upon the empirical applications of Billio et al. (2012b); Boyson et al. (2010); Chan et al. (2006); Fung and Hsieh (2004) to hedge fund data. We include financial sector performance as a sector-specific risk factor for our analysis of connectivity amongst financial institutions.

Table 1.1: Definitions of risk factors used in the factor models for filtering returns data. These variables are obtained from Federal Reserve Economic Data (FRED), Yahoo! Finance, and Ken French's website.

Variable	Definition
Market	Value weighted return on all publicly traded U.S. firms
Large-Small	Difference in returns between large and small portfolios
Value-Growth	Difference in returns between value and growth portfolios
Momentum	Momentum factor
Credit Spread	Difference between Moody's BAA and AAA bond indexes
Term Spread	Difference between 10-year T-bond and 6-month LIBOR
Gold	Return on gold bullion price ($ per troy ounce)
U.S. Dollar	Return on trade weighted U.S. dollar index
Bond	Return on Vanguard bond index
VIX	First difference of VIX index
EM Bond	Return on T. Rowe Price emerging markets bond ETF
EM Equity	Return on T. Rowe Price emerging markets equity ETF
Financial Sector	Return on SPDR financial sector ETF

For institutions $i = 1, \ldots, n$, stepwise regression is conducted on returns to select from these risk factors to be included in the factor model. The residuals from these models, which are the idiosyncratic returns, are used for analysis. The data matrix Y is composed of these filtered returns. Additional details for filtering of empirical data is given in Section 1.3.1.

The filtering primarily aims to prevent the model from attributing contagion to institutions which is merely related to common market factors. The filtering is also a convenient tool to produce idiosyncratic returns, which simplifies the return distribution of the contagion model. A limitation of the filtering is that our list of risk factors is potentially incomplete, and there are many additional factors that may otherwise mitigate contagion in the model. Another drawback of this approach is that we filter returns using a factor model which is not time dynamic, and then we use the filtered returns to estimate a model which is time dynamic. The filtering is based on a linear factor model for the expected return, and the impact of filtering on volatility is minimal. The time dynamics of our model apply primarily to volatility regimes, and the network model links volatility regimes of institutions. Thus, we might expect the network or time dynamic volatility inference not to be substantially

impacted by the filtering.

1.2.2 Return Distribution

Consider the filtered returns. Given $I_{i,t}$, the volatility regime indicator of firm i at time t, we assume that the returns are independent and Normally distributed

$$\begin{aligned} Y_{i,t} &\overset{\text{ind.}}{\sim} \mathcal{N}\left(\mu_{i,t}, \sigma^2_{i,t}\right), \\ \mu_{i,t} &= \mu_{0i}(1 - I_{i,t}) + \mu_{1i} I_{i,t}, \\ \sigma^2_{i,t} &= \sigma^2_i(1 - I_{i,t}) + \sigma^2_i \delta_i I_{i,t}, \end{aligned} \tag{1.1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The return of institution i has mean μ_{0i} and variance σ^2_i in the low volatility state. In the high volatility state the return has mean μ_{1i} and variance $\delta_i \sigma^2_i$, where $\delta_i > 1$. Though we make no restrictions on the mean of the returns in either volatility state, the high volatility regime is associated with financial distress, and in empirical applications the mean return in the high volatility state is typically negative.

While the return distributions are independent conditional on the volatility regime, unconditionally there is dependence between the return distributions. Consistent with the findings of Longin and Solnik (2001), unconditional correlations will be larger when firms are jointly in the high volatility regime. Likewise, the return distributions are Normal and homoskedastic conditional on the volatility regime, but unconditionally the return distributions are mixtures of distributions allowing for non-Normality and heteroskedasticity. A GARCH model would provide a more accurate assessment of the return volatility time dynamics; however, the goal of our model is not to forecast volatility, but simply to identify volatility regime switches between distress and tranquility.

1.2.3 Regime Switching

The Markov regime switching process for the volatility state is governed by the transition probability matrix

$$Pr(I_{i,t}) = \begin{bmatrix} P_{00,t} & P_{01,t} \\ P_{10,t} & P_{11,t} \end{bmatrix}, \tag{1.2}$$

$$\log\left(\frac{P_{01,t}}{1-P_{01,t}}\right) = \alpha_{01} + \beta_{01}\frac{\sum_{j=1}^{n}(G_{i,j}I_{j,t-1})}{n},$$

$$\log\left(\frac{P_{11,t}}{1-P_{11,t}}\right) = \alpha_{11} + \beta_{11}\frac{\sum_{j=1}^{n}(G_{i,j}I_{j,t-1})}{n},$$

$$P_{00,t} = 1 - P_{01,t},\ P_{10,t} = 1 - P_{11,t},$$

for $i = 1, \ldots, n$, $t = 1, \ldots, T$, and $\beta_{01}, \beta_{11} > 0$. Given volatility regime at time $t-1$, $P_{00,t}$ and $P_{11,t}$ are the probabilities of remaining in the low and high volatility states, respectively, at time t. $P_{01,t}$ is the probability of transitioning from the low to high volatility regime, and $P_{10,t}$ is the probability of transitioning from the high to low volatility regime.

We have made the simplifying assumption that the effect of network connections on the log odds of the transition probabilities is the same for all institutions. Alternatively, we could allow each institution to be parameterized by its own $\beta_{01,i}$ and $\beta_{11,i}$. Sensitivity analyses allowing for varying coefficients by institutions in the log odds have shown that it is reasonable to assume $\beta_{01,i} = \beta_{01}$ and $\beta_{11,i} = \beta_{11}$ for all i.

The key feature of the transition probabilities for estimating volatility contagion is their dependence on the network structure. A firm that is more connected to firms in the high volatility regime at time $t-1$ is more likely to transition into the high volatility regime, or remain in the high volatility regime, at time t. Likewise, if a firm is in the high volatility

regime at time $t-1$, the firms to which it is connected will be more likely to enter the high volatility regime at time t. The former description measures the degree to which firms experience volatility contagion; the latter measures the degree to which firms spread volatility contagion. Both will increase as the network connectivity increases. β_{01} estimates the impact of additional high volatility connections on the log-odds of switching into the high volatility regime, and β_{11} estimates this impact for remaining in the high volatility regime.

Conditional on volatility state at time T, forecasts can easily be constructed for the volatility regime at time $T+k$. We will use these forecasts to conduct out-of-sample testing evaluating the role of volatility contagion in crisis period losses of institutions.

1.2.4 Network

Lastly, we must model the network of linkage parameters, G. From the Bayesian perspective, our model on G is essentially a prior distribution on the probability of connections between firms. In the latent space model of Hoff et al. (2002), each institution in the model is assumed to occupy a latent position in the network space. The likelihood of a network connection between institutions in the model depends upon the latent distance between the institutions. We model the network space as a circle, and the angle between institutions on the circle structure is defined as the distance between those institutions. Mathematically,

$$\log\left(\frac{P(G_{i,j}=1)}{1-P(G_{i,j}=1)}\right) = \gamma - ||\theta_i - \theta_j||, \tag{1.3}$$

where $||\theta_i - \theta_j|| = \min(|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|)$, $i = 1, \ldots, n-1$, $j = 1, \ldots, n$, $i < j$.

An attractive feature of the latent space model is that it provides a convenient visualization of the network structure in the circle space. Because this model is simply a prior distribution on G, our contagion inference does not depend on the latent positions on the institutions; however, these latent positions can simplify the presentation of results.

1.2.5 Full Model and Estimation

The previous sections described the three primary components of the model, as shown in Equations (1.1), (1.2), and (1.3). In this section we provide more complete technical details. As a convention, functions will be distinguished by their arguments. Functions denoted by f describe the primary components of the model, and functions denote by π describe prior distributions. Using notation established in prior sections and the convention that bold-faced parameters are vectors, e.g. $\boldsymbol{\sigma^2} = (\sigma_1^2, \ldots, \sigma_n^2)$, the density functions corresponding to these components are given in Equations (1.4), (1.5), and (1.6), respectively.

$$f\left(Y|I, \boldsymbol{\mu_0}, \boldsymbol{\mu_1}, \boldsymbol{\sigma^2}, \boldsymbol{\delta}\right) = \prod_{i=1}^{n}\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_{i,t}^2}} \exp\left[-\frac{1}{2\sigma_{i,t}^2}\left(Y_{i,t} - \mu_{i,t}\right)^2\right], \tag{1.4}$$

$$f(I|G, \alpha_{01}, \alpha_{11}, \beta_{01}, \beta_{11}) = \prod_{i=1}^{n}\prod_{t=1}^{T} P_{00,t}^{(1-I_{i,t-1})(1-I_{i,t})} P_{01,t}^{(1-I_{i,t-1})I_{i,t}} P_{10,t}^{I_{i,t-1}(1-I_{i,t})} P_{11,t}^{I_{i,t-1}I_{i,t}}, \tag{1.5}$$

$$f(G|\gamma, \boldsymbol{\theta}) = \prod_{i<j} P(G_{i,j} = 1)^{G_{i,j}} (1 - P(G_{i,j} = 0))^{1-G_{i,j}}. \tag{1.6}$$

In order to produce a fully Bayesian model, we must additionally define prior distributions for the parameters of the model. While choices of prior distributions are inherently arbitrary, sensitivity analyses have shown that inference is generally not impacted by any particular choice. We have chosen priors to be non-informative to the extent that it is possible.

Priors on return distribution parameters are given by:

$$\pi(\boldsymbol{\mu_0}, \boldsymbol{\mu_1}, \boldsymbol{\sigma^2}) \propto \prod_{i=1}^{n} \frac{1}{\sigma_i^2},$$

$$\delta_i - 1 \sim \text{Gamma}(3, 1), \text{for } i = 1, \ldots, n.$$

$$\pi(\boldsymbol{\delta}) = \prod_{i=1}^{n} \frac{1}{2}(\delta_i - 1)^2 e^{-\delta_i + 1}.$$

Priors on regime switching parameters are given by:

$$\begin{aligned}
\pi(\alpha_{01}, \alpha_{11}) &\propto 1, \\
\beta_{01}, \beta_{11} &\sim \text{Gamma}(10, 10/n_0), \\
\pi(\beta_{01}, \beta_{11}) &= \frac{\left(\frac{10}{n_0}\right)^{10}}{\Gamma(10)} \beta_{01}^{10-1} e^{-\frac{10}{n_0}\beta_{01}} \frac{\left(\frac{10}{n_0}\right)^{10}}{\Gamma(10)} \beta_{11}^{10-1} e^{-\frac{10}{n_0}\beta_{11}}, \\
I_{i,0} &\equiv 0, \text{for } i = 1, \ldots, n.
\end{aligned}$$

n_0 is a fixed value, which we will take to be equal to the dimension of the particular data set.

Priors on network parameters are given by:

$$\begin{aligned}
\gamma &= 0, \\
\theta_1 = 0,\ \theta_2 &\sim \text{Unif}(0, \pi), \\
\theta_i &\sim \text{Unif}(0, 2\pi), i = 3, \ldots, n, \\
\pi(\boldsymbol{\theta}) &\propto 1
\end{aligned}$$

The locations on a circle are only unique up to rotation and reflection. In order to uniquely identify a set of locations, we fix $\theta_1 = 0$ and require $0 < \theta_2 < \pi$. γ is a parameter governing the prior probability of network connections. By fixing $\gamma = 0$, the prior probability of a connection can be no greater than 0.50 and no smaller than 0.04, depending on the latent locations. We find this to be a conservative choice of prior belief in identifying connections.

Combining these prior distributions with Equations (1.4), (1.5), and (1.6) results in the posterior distribution

$$\begin{aligned}
p\left(I, G, \boldsymbol{\mu_0}, \boldsymbol{\mu_1}, \boldsymbol{\sigma^2}, \boldsymbol{\delta}, \alpha_{01}, \alpha_{11}, \beta_{01}, \beta_{11}, \gamma, \boldsymbol{\theta} \middle| Y\right) \propto & \\
f\left(Y | I, \boldsymbol{\mu_0}, \boldsymbol{\mu_1}, \boldsymbol{\sigma^2}, \boldsymbol{\delta}\right) & f\left(I | G, \alpha_{01}, \alpha_{11}, \beta_{01}, \beta_{11}\right) f\left(G | \gamma, \boldsymbol{\theta}\right) \\
\times\ \pi(\boldsymbol{\mu_0}, \boldsymbol{\mu_1}, \boldsymbol{\sigma^2}) & \pi(\boldsymbol{\delta}) \pi(\alpha_{01}, \alpha_{11}) \pi(\beta_{01}, \beta_{11}) \pi(\boldsymbol{\theta}). \qquad (1.7)
\end{aligned}$$

Given this posterior distribution we conduct inference using MCMC methods, specifically Metropolis-Hastings (Metropolis et al. 1953; Hastings 1970) within Gibbs sampling (Gelfand and Smith 1990). At each iteration of the chain, we conditionally update parameters of the model in batches by making a proposal and accepting that proposal on the basis of the Metropolis-Hastings rule. We run the chain for 100,000 iterations using the first 50,000 for burn-in. Based on storage limitations we save every fourth iteration of the chain. Inference is conducted on the basis of posterior means and intervals which can be calculated directly from the MCMC samples from the posterior distribution. We will also use the MCMC output to create out-of-sample forecasts. Diagnostics are conducted to ensure convergence of the MCMC draws to the target posterior distribution.

1.2.6 Simulation Results

According to the structure of the model and the data available, our goal is to estimate a latent space model for a latent network for connections between latent volatility regimes underlying an observed multivariate time series of financial returns. Using MCMC methods with a fully Bayesian approach, we have the technological capability to estimate the parameters of this model. However, because the network, which is of primary interest, is at the third "level" of the model, it is important to assess our ability to actually infer the network structure. In order to evaluate the power of the model to detect connections, and to test the performance and convergence properties of the MCMC routine, we conduct a simulation study.

Consider returns for $n = 12$ hypothetical funds, observed over $T = 250$ time periods. We fix parameters $\boldsymbol{\mu_0}$, $\boldsymbol{\mu_1}$, $\boldsymbol{\sigma^2}$, $\boldsymbol{\delta}$, $\boldsymbol{\theta}$, $(\alpha_{01}, \alpha_{11})$, and (β_{01}, β_{11}) to be consistent with values that are expected to be observed in an empirical application. Alternatively, we could simulate these values from their prior distributions. The simulation results are similar regardless of how the true parameters are created. Given the values for these parameters, we simulate from Equation (1.3) to generate G, then from Equation (1.2) to generate I, and finally from Equation (1.1) to generate Y.

For simulated data Y, we implement the MCMC routine described in Section 1.2.5. We compare our estimates from the MCMC output to the "true" values for the parameters based on the design of the simulation study.

Table 1.2 shows posterior means, 95% posterior intervals, and true values for the parameters $\boldsymbol{\mu_0}$, $\boldsymbol{\mu_1}$, $\boldsymbol{\sigma^2}$, $\boldsymbol{\delta}$, $(\alpha_{01}, \alpha_{11})$, and (β_{01}, β_{11}). Elements of the table highlighted in red indicate the parameters for which the posterior interval does not cover the true value. There is only one such parameter ($\mu_{0,4}$). Posterior means are typically near the true values.

Table 1.2: Simulation results. For each cell in the table: posterior mean (upper left), true value (upper right), posterior 95% interval (bottom). True values not captured by the posterior interval are listed in red.

	μ_{0i}	μ_{1i}	σ_i^2	δ_i
i = 1	0.33 0.15 (0.08, 0.59)	-1.13 -1.14 (-1.65, -0.63)	1.24 1.30 (0.85, 1.79)	8.01 8.81 (5.19, 11.39)
i = 2	-0.07 0.00 (-0.39, 0.25)	-0.42 -0.02 (-2.19, 0.91)	3.80 4.42 (2.93, 5.04)	5.47 4.07 (3.60, 8.12)
i = 3	0.49 0.25 (0.00, 0.98)	-0.34 -0.44 (-1.16, 0.46)	4.14 3.56 (2.96, 5.76)	6.56 6.74 (4.42, 9.33)
i = 4	0.15 0.05 (0.06, 0.24)	-0.58 -0.18 (-1.16, -0.09)	0.42 0.48 (0.33, 0.52)	4.52 3.86 (2.80, 6.82)
i = 5	0.31 0.30 (0.18, 0.43)	-1.01 -1.17 (-1.93, -0.22)	0.71 0.71 (0.54, 0.89)	4.43 5.27 (2.73, 6.78)
i = 6	0.41 0.28 (0.20, 0.62)	-0.97 -1.10 (-1.36, -0.59)	0.90 1.04 (0.60, 1.27)	4.87 3.92 (3.08, 7.20)
i = 7	0.24 0.32 (0.09, 0.38)	-1.58 -1.23 (-2.55, -0.72)	1.10 1.02 (0.90, 1.34)	4.91 4.78 (3.05, 7.69)
i = 8	-0.00 0.03 (-0.12, 0.12)	-0.06 -0.18 (-0.51, 0.37)	0.59 0.58 (0.46, 0.78)	4.78 3.76 (3.08, 7.15)
i = 9	0.04 0.12 (-0.15, 0.23)	-0.72 -0.91 (-1.16, -0.27)	0.71 0.65 (0.49, 1.00)	11.62 12.02 (7.81, 16.28)
i = 10	0.01 0.08 (-0.19, 0.21)	-1.03 -0.14 (-2.16, -0.09)	1.76 1.92 (1.35, 2.28)	6.43 7.32 (4.30, 9.39)
i = 11	0.01 0.01 (-0.13, 0.15)	-0.37 -0.07 (-2.37, 1.58)	1.18 1.25 (0.96, 1.43)	5.01 4.82 (2.40, 8.85)
i = 12	0.20 -0.01 (-0.27, 0.64)	0.00 0.17 (-6.19, 2.75)	9.22 9.56 (7.15, 11.63)	2.83 1.93 (1.52, 5.27)
	α_{01}	α_{11}	β_{01}	β_{11}
	-4.55 -4 (-5.55, -3.74)	1.52 1.5 (0.76, 2.20)	19.96 18 (11.25, 29.85)	5.97 8 (0.45, 13.94)

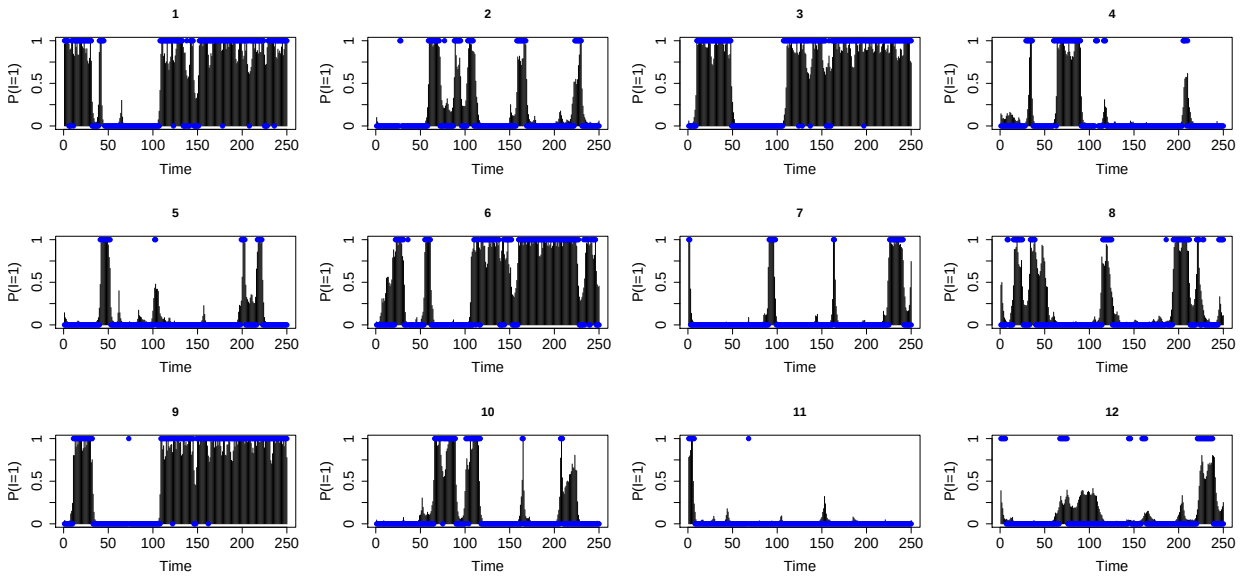

Figure 1.1: Posterior mean probability for each fund to be in the high volatility state at a given point in time. True volatility state at each time point is indicated by a blue circle.

The results for identifying the volatility regime for each fund over time are shown in Figure 1.1, which plots the posterior probability of a fund being in the high volatility state at each time point, with blue dots indicating the true regime state. These posterior probabilities are successful in predicting volatility state, with the potential exception of Fund 12. Note that $\delta_{12} = 1.93$ is relatively small, making it difficult to distinguish between the high and low volatility regimes since the standard deviation of returns is similar in each state.

Given our emphasis on measuring systemic risk, we are most interested in the ability to identify the true network connections in the model. Figure 1.2 and Figure 1.3 compare the inferred network connections to their true values. Keeping in mind that estimation performance is limited by sample size ($T = 250$), results suggest that we have a reasonable degree of power to identify connections. Figure 1.2 shows that posterior probabilities of network connections are relatively high when the funds are truly connected. Because of the pattern of connections, we occasionally infer connections that exist only indirectly in the true network. For example, in the true network Fund 1 is connected to 3, and 3 is connected to 6, but 1 is not connected to 6; however, we estimate according to our MCMC output

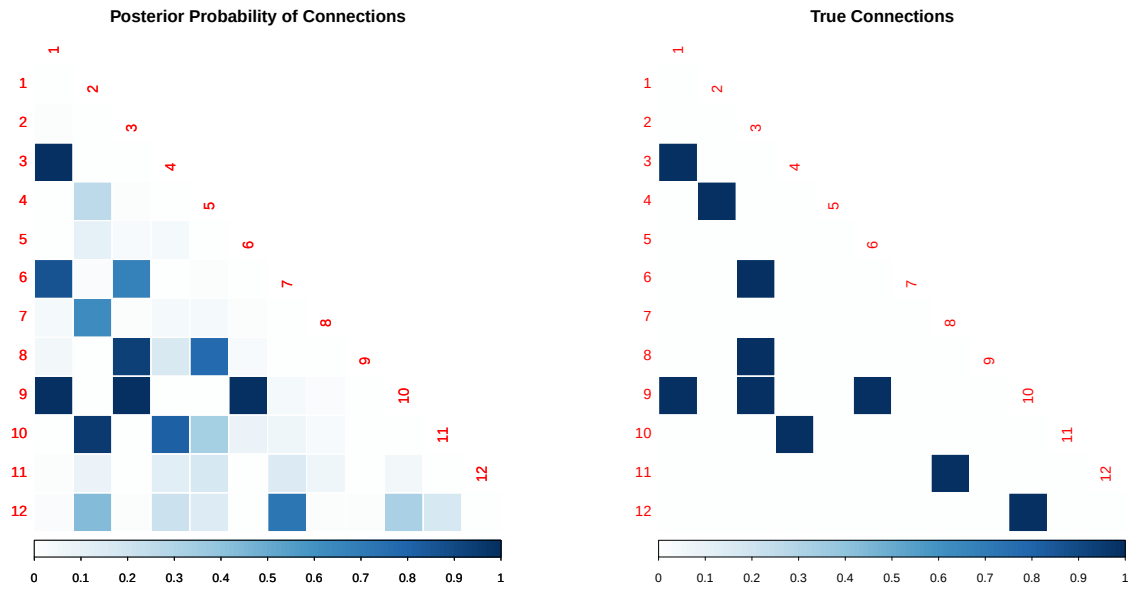

Figure 1.2: Posterior probabilities of connections between funds (left) and true connections (right).

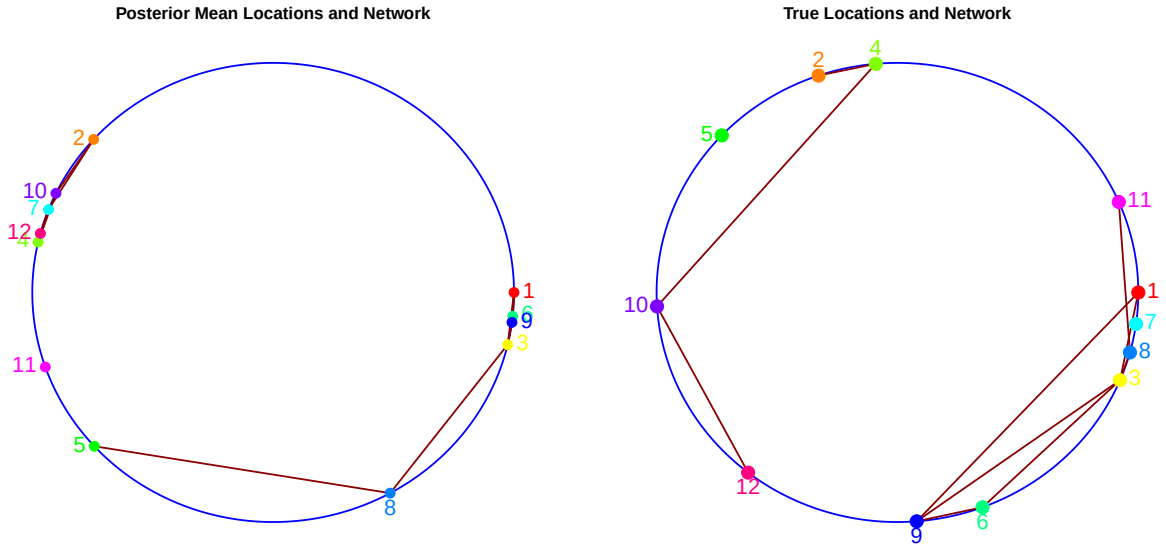

Figure 1.3: Posterior mean location of each fund with connections indicated when the posterior probability exceeds 0.5 (left) and true location and connections between funds (right).

that Fund 1 is connected to 6. This is not a severe limitation of the model, because based on the true network structure, Fund 1 and 6 are indirectly linked. Figure 1.3 compares the posterior mean positions on the circle of each fund with the true positions. Links are drawn

when the posterior probability of a connection exceeds 0.5. Inference on the latent positions is inconsequential to measures of systemic risk exposure, but the model does indicate the ability to cluster funds according to their true locations. Overall, the model shows a capability to infer true network connections.

Satisfied that the model estimation procedure and inference based on MCMC output performs well, we turn to several empirical applications of our model.

1.3 Empirical Application: Hedge Funds

Following the 2007-09 financial crisis in the United States, there is increased awareness of systemic risk in the hedge fund industry. During this time, the industry suffered significant losses, particularly in Q4 of 2008, when a record number of hedge funds liquidated. Recovering from this period of financial distress, the number and size of hedge funds is currently growing. At the end of 2012, hedge funds globally held a record $2.25 trillion assets under management (Hedge Fund Research 2013). As recently as 2007, hedge funds accounted for more than half of trading volume in equity and bond market (Adrian 2007). With the rise of high frequency trading, the relative share of transactions for hedge funds has fallen substantially, but hedge funds still account for a significant proportion of all trades. In addition to the size of the industry, hedge fund portfolios are among the most dynamic investment strategies. Due to limited regulatory and reporting requirements, hedge fund managers have freedom to trade, enter long or short positions, alter strategies, exploit leverage, and pursue investment returns to the extent that they desire at any point in time. The dynamic volatility regimes of our model are intended to capture the dynamic strategic outlook and dynamic risk exposure of hedge funds.

As the industry has grown over the past decade, so has the amount and quality of available data for evaluating the performance of both individual funds and indexes of funds. With higher quality data, academic evaluation of hedge funds has become more common,

especially for risk exposure (Brealey and Kaplanis 2001; Fung and Hsieh 2001; Agarwal and Naik 2004; Gupta and Liang 2005; Bali et al. 2007; Billio et al. 2012b) and systemic risk (Getmansky et al. 2004; Chan et al. 2006; Boyson et al. 2010). This chapter contributes to the growing hedge fund literature by extending existing models to incorporate a network structure.

1.3.1 Data

The Dow Jones Credit Suisse Hedge Fund Index (DJCSHFI) maintains a database of indexed, monthly hedge fund returns. The index tracks about 8,000 individual funds and is restricted to funds with assets under management at least $50 million, a 12-month record, and audited financial statements. The overall index is broken down into several indexes based on trading strategy, and the database accounts for at least 85% of all assets under management in the index universe within each of its subclassifications. We consider monthly index returns from April 1994 to December 2012 for strategy indexes Convertible Arbitrage, Dedicated Short, Emerging Markets, Equity Market Neutral, Event Driven: Distressed, Event Driven: Multi-Strategy, Event Driven: Risk Arbitrage, Fixed Income Arbitrage, Global Macro, Long Short Equity, Managed Futures, and Multi-Strategy.

This data set is limited in a few important ways. Because of the strategic flexibility afforded to hedge fund managers, assignment of individual funds to particular strategy classifications is often subjective. There is also an inherent survivorship bias, as funds that close are removed from the index. According to the website of the DJCSHFI, "in order to minimize [the effect of survivorship bias], the index does not remove funds in the process of liquidation, and therefore captures all of the potential negative performance before a fund ceases to operate." Last, the data consists of indexes of returns, rather than individual fund returns. In the context of the model, it is less clear how to interpret contagion between strategy indexes than it would be to interpret contagion between funds. Despite these limitations, hedge fund index data has seen empirical analysis in several studies (Billio et al.

2012b; Boyson et al. 2010; Chan et al. 2006). We will apply our model to this data set in order to describe the network structure of hedge fund indexes and to demonstrate model performance in a particular financial marketplace.

A brief description of each hedge fund strategy index is given in Appendix A, and summary statistics are shown in Table 1.3. There is considerable heterogeneity in return and risk profiles for the strategy indexes. Average returns range from -3.52% for Dedicated Short, to 12.26% for Global Macro. Dedicated Short has the largest standard deviation at 16.87%, and Equity Market Neutral the lowest at 3.87%. Non-directional strategies tend to have negative skewness and high kurtosis. The indexes all experience their minimal returns during the 2007-2009 crisis period with peak impact in October 2008.

Table 1.3: (April 1994-December 2012) Summary statistics for $N = 225$ monthly hedge fund index returns. Mean and SD are annualized.

Name	Mean (%)	SD (%)	Min (%)	Median (%)	Max (%)	Skew	Kurt.
Convertible Arbitrage	7.62	6.87	-12.59	0.97	5.81	-2.70	19.15
Dedicated Short	-3.52	16.87	-11.28	-0.75	22.71	0.70	4.44
Emerging Markets	8.19	14.54	-23.03	1.17	16.42	-0.84	8.43
Equity Market Neutral	7.67	3.87	-5.61	0.68	3.66	-0.99	8.06
ED: Distressed	10.12	6.52	-12.45	1.13	4.15	-2.23	14.23
ED: Multi-Strategy	8.63	6.78	-11.52	0.90	4.78	-1.78	10.62
ED: Risk Arbitrage	6.41	4.15	-6.15	0.53	3.81	-0.97	7.48
Fixed Income Arbitrage	5.75	5.63	-14.04	0.74	4.33	-4.59	35.34
Global Macro	12.26	9.44	-11.55	1.06	10.60	0.06	7.21
Long Short Equity	9.65	9.81	-11.43	0.79	13.01	-0.02	6.33
Managed Futures	5.92	11.72	-9.35	0.32	9.95	0.04	2.93
Multi-Strategy	7.89	5.33	-7.35	0.80	4.28	-1.71	8.93

As described in Section 1.2.1, the raw hedge fund return data is filtered with a multi-factor model. For the hedge fund data, the factor model includes the factors listed in Table 1.1 with the exception of the Financial Sector and Emerging Markets factors because historical data for these factors does not span the entire sample period. The factor exposures from step-wise regression for each strategy index are given in Table 1.4. The idiosyncratic returns, which are residuals from the multi-factor model, serve as the source of data for analysis.

Table 1.4: Risk exposures of the hedge fund index returns from mutli-factor model with stepwise regression. Standard errors in parentheses.

	Convert. Arb.	Dedicated Short	Emerging Markets	Equity Market Neutral	ED: Dis-tressed	ED: Multi	ED: Risk Arb.	Fixed Income Arb.	Global Macro	Long Short Equity	Managed Futures	Multi
(Intercept)	0.22	1.08	0.14	0.49	0.90	0.35	0.27	0.24	0.22	−0.06	−0.07	0.34
	(0.13)	(0.42)	(0.22)	(0.08)	(0.22)	(0.11)	(0.08)	(0.24)	(0.19)	(0.25)	(0.24)	(0.09)
Market	0.09	−0.91	0.46	0.09	0.23	0.25	0.10	0.10	0.20	0.47		0.11
	(0.04)	(0.05)	(0.05)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)		(0.02)
Large-Small	0.07	−0.28	0.19		0.10	0.11	0.07			0.20		0.06
	(0.04)	(0.05)	(0.06)		(0.03)	(0.03)	(0.02)			(0.03)		(0.03)
Value-Growth	0.08	0.09		0.04	0.06	0.09	0.05	0.06		−0.07		0.06
	(0.04)	(0.05)		(0.02)	(0.03)	(0.03)	(0.02)	(0.03)		(0.03)		(0.03)
Momentum			0.08			0.05			0.09	0.17	0.11	
			(0.04)			(0.02)			(0.03)	(0.02)	(0.04)	
Credit Spread		−1.23			−0.59			−0.44		0.42		
		(0.37)			(0.21)			(0.22)		(0.22)		
Term Spread				−0.11	0.12		−0.11	0.27		−0.21		
				(0.05)	(0.08)		(0.05)	(0.08)		(0.07)		
Gold				−0.03			0.07		0.08			
				(0.02)			(0.02)		(0.05)			
US Dollar	−0.35		−0.63	−0.30	−0.18	−0.22		−0.29	0.27	−0.31	−0.49	−0.24
	(0.10)		(0.19)	(0.06)	(0.08)	(0.08)		(0.08)	(0.16)	(0.08)	(0.18)	(0.08)
Bond	0.21	0.24				−0.15		0.17	0.63	0.17	0.56	
	(0.11)	(0.16)				(0.09)		(0.09)	(0.16)	(0.09)	(0.21)	
VIX	−0.06	−0.15					−0.07			0.04	0.09	
	(0.04)	(0.05)					(0.02)			(0.03)	(0.05)	
R^2	0.24	0.76	0.41	0.35	0.48	0.47	0.44	0.26	0.17	0.78	0.10	0.22

1.3.2 Results

The model described in Section 1.2 is estimated using the filtered hedge fund returns data. Parameter estimates are presented numerically and graphically in this section.

Table 1.5: Posterior means for select parameters from the model. Return distribution parameters (upper panel), where # indicates the posterior mean number of connections in the network model; regime switching parameters (lower panel).

Index	μ_0	μ_1	σ^2	δ	#
Convertible Arbitrage	0.15	-0.76	0.98	9.88	3.36
Dedicated Short	0.08	-0.56	3.87	4.09	3.16
Emerging Markets	0.25	-0.49	3.05	7.70	3.38
Equity Market Neutral	0.01	-0.13	0.48	4.40	2.38
ED: Distressed	0.16	-0.83	0.94	5.51	3.17
ED: Multi-Strategy	0.25	-1.07	0.92	5.49	3.57
ED: Risk Arbitrage	-0.07	0.10	0.46	3.12	3.95
Fixed Income Arbitrage	0.14	-1.25	0.63	11.85	2.74
Global Macro	0.10	-0.14	1.46	8.31	4.36
Long Short Equity	-0.03	0.05	1.05	3.57	3.60
Managed Futures	0.06	-0.51	9.28	2.41	2.29
Multi-Strategy	0.18	-0.63	0.59	8.13	3.50
	α_{01}	α_{11}	β_{01}	β_{11}	
	-4.09	0.38	15.88	8.04	

The upper panel of Table 1.5 shows posterior means for the parameters from the return distribution component of the model. These parameter estimates correspond to monthly returns and have not been annualized. It is clear that the high volatility state corresponds to crisis periods. 10 of the 12 indexes have negative mean returns under the high volatility state. These indexes typically exhibit significantly higher mean returns in the low volatility state relative to the high volatility state. For example, there is a more than one percentage point decrease in mean monthly returns between the low and high volatility states for Fixed Income Arbitrage and Event Driven: Multi-Strategy. Fixed Income Arbitrage has the largest high volatility multiplier, with 11.85 times larger variance in the high volatility state. Because Fixed Income Arbitrage funds are more leveraged than other funds from other strategies, an extreme volatility multiplier is expected. Volatility multipliers are between 2.41 and 11.85,

corresponding to standard deviation multipliers between 1.55 and 3.44.

Posterior means for the regime switching probability parameters are given in the lower panel Table 1.5. Because β_{01} is larger in magnitude than β_{11}, connectivity is more influential for indexes entering the high volatility state than for indexes staying in the high volatility state. The odds of an index entering the high volatility state increase by a factor of 3.76 for each additional connection to another index in the high volatility state.

Figure 1.4 illustrates the posterior probability of an index being in the high volatility state at each point in time. During the major crisis periods in the sample, many of the indexes are simultaneously in the high volatility regime: the Russian and LTCM crisis of 1998, the dot-com bubble of 2000-01, and the subprime mortgage collapse of 2007-09. Only Equity Market Neutral and the Event Driven classifications have exhibited any periods of high volatility since 2009. Because the hedge fund industry was much smaller in size in the mid-1990s, much of the observed high volatility in the early part of the sample is an artifact of small sample sizes within indexes during that time period. Indexes with similar time patterns of volatility regimes would be expected to be connected in the network component

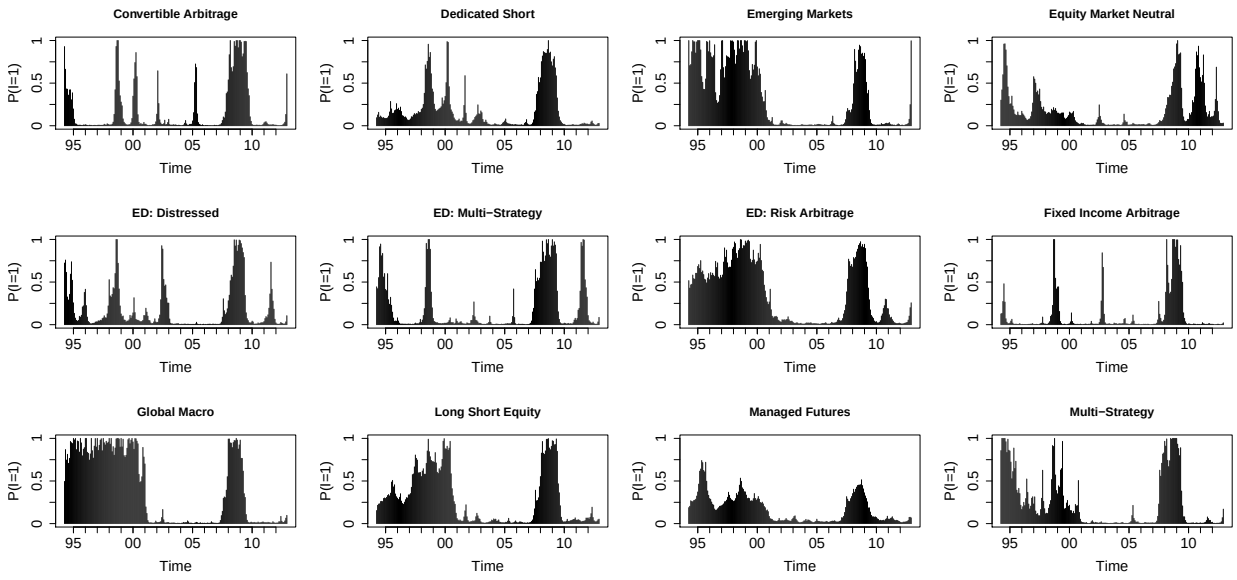

Figure 1.4: Posterior mean probability for each fund to be in the high volatility state at a given point in time.

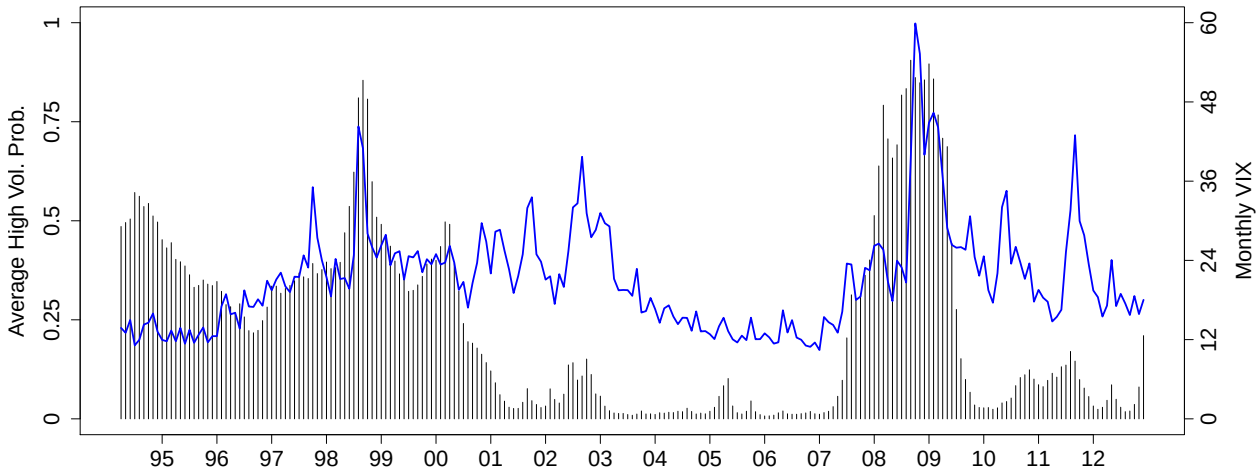

Figure 1.5: Average probability for the 12 funds to be in the high volatility state at a given time point (left axis). VIX volatility index of the S&P 500 (blue, right axis).

of the model, for example, Global Macro and Emerging Markets.

Figure 1.5 aggregates the results in Figure 1.4 by plotting the average probability of indexes being in the high volatility state. The overlay line shows the volatility index of the market (VIX). The volatility regimes of the hedge fund model generally correspond with changes in the VIX, with the notable exception of the VIX spike in 2011. An important feature of the plot is that the average volatility regime relative to the VIX tends to be a leading indicator of financial crises. In early 1998, early 2000, and early 2008 the volatility of the hedge fund model is relatively larger than the VIX, and in all three cases a major market event followed. Since 2009, hedge funds have entered a period of tranquility, especially relative to the market volatility. The final sample point (December 2012) shows a small spike in volatility which could be a precursor of future hedge fund volatility regime switches.

Because high volatility and distress often result in losses, aggregate hedge fund volatility is related to systemic risk indirectly through these losses. The network of volatility contagion provides a more direct, though neither causal nor structural, assessment of systemic risk through linkages. The posterior mean structure of this network is presented in Figure 1.6.

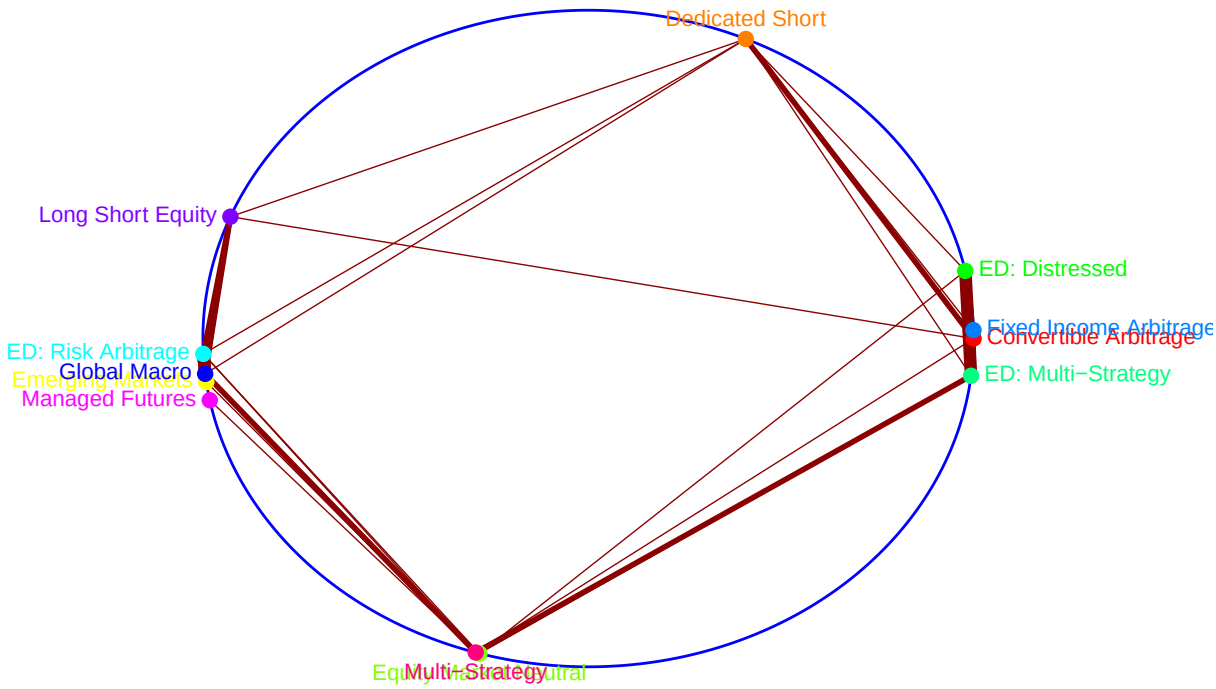

Figure 1.6: Posterior mean location and network connections for each fund. Line thickness indicates greater probability of network connectivity.

The indexes are plotted in their posterior mean latent position on the circle, and the thickness of lines corresponds to the posterior probability of a connection between indexes. The posterior mean number of total connections is 19.7 out of a possible 66, a rate of 30%. Based on the latent positions of the indexes, there are two primary clusters, the first consisting of Long Short Equity, Event Driven: Risk Arbitrage, Global Macro, Emerging Markets, and Managed Futures, and the second consisting of Event Driven: Distressed, Fixed Income Arbitrage, Convertible Arbitrage, and Event Driven: Multi Strategy. The former is a cluster of international equity strategies, and the latter is a cluster of event driven and arbitrage strategies. Event Driven: Risk Arbitrage is the primary deviation from this classification. The only connection between the clusters is a low probability link between Long short Equity and Convertible Arbitrage. The Dedicated Short, Multi-Strategy, and Equity Market Neutral

indexes do not appear to belong to either cluster and have connections to both. Intuitively, it is not surprising that the Multi-Strategy and Dedicated Short strategies are not elements of either cluster. By definition Multi-Strategy funds pursue a variety of trading strategies, and Dedicated Short is unique in its bias towards short positions. The links with the highest posterior probability of a connection are Event Driven Distressed - Event Driven: Multi-Strategy; Emerging Markets - Event Driven: Risk Arbitrage; Emerging Markets - Global Macro; and Event Driven: Risk Arbitrage - Global Macro. Global Macro has the greatest number of connections, while Managed Futures has the fewest.

As a statistical measure of the importance of the network structure in our model we calculate the Deviance Information Criterion (DIC), a Bayesian measure for model comparison (Gelman et al. 2004). Models with smaller DIC are preferred. The DIC for the full model is 9372.1, corresponding to 126.8 effective parameters. For comparison, we consider a regime switching model without the network component. This is equivalent to the full model with $G_{i,j} = 0$ for all i and j. For this reduced model, the DIC is 9416.7 with 109.6 effective parameters. According to the DIC, the network component of the model offers a significant improvement in model fit, even accounting for the 17.2 additional effective parameters in the full model. We conclude that the hedge fund index returns express connectivity and that volatility contagion exists, at least as defined according to our particular modeling framework in the context of hedge fund index returns.

As a basis of comparison of overall connectivity for the hedge fund industry, we collect data for returns on Exchange Traded Funds (ETF) which track sector indexes of the U.S. equities market. These ETFs are managed by iShares and correspond to the Dow Jones sector indexes: Consumer Goods, Consumer Services, Energy, Financial, Financial Services, Healthcare, Industrials, Materials, Real Estate, Technology, Telecommunications, and Utilities. Because the number of ETF under consideration is equal to the number of hedge fund indexes, we can make direct comparisons between the two samples in terms of the number of connections in the network structure. For monthly data from January 2001 to December

2012 we fit the model for both sets of data: ETF and hedge fund indexes. We find nearly identical connectivity. The posterior mean number of connections is 21.10 for the hedge fund data, and 21.11 for the ETF data. In each sample, inferred connectivity accounts for approximately 32% of possible connections. To the extent that our network model for volatility contagion is descriptive of systemic risk, we find that systemic risk exposure is similar in the hedge fund industry to the equity market. In Section 1.4 we provide a more detailed analysis of systemic risk in equities, focusing on the financial sector.

Overall, we find several interesting patterns in the network model for hedge funds. The connectivity is more influential for transitions into the high volatility regime than remaining in the high volatility state; average volatility of the model has historically been a leading indicator for financial crises; network connections exist and are statistically significant; connectivity between strategy indexes are consistent with intuition; and overall volatility contagion and systemic risk exposure, as expressed by the rate of connectivity in the network model, is similar for hedge funds and the equity market. All of these findings are descriptive in the sense that our model does not provide any formal link between the network structure and systemic risk. In Section 1.4 we move towards inferential procedures by conducting out-of-sample forecasts and evaluating the time dynamics of connectivity.

1.4 Empirical Application: Financial Sector

The financial sector of the economy has been subjected to intense scrutiny for its role in the crisis of 2007-09. Acting in a deregulated environment following the repeal of the Glass-Steagall Act, banks have typically been assigned blame for selling Mortgage Backed Securities (MBS) and Collateralized Debt Obligations (CDO) on underlying mortgages that were risky or subprime. The mortgages were packaged together with the intent to reduce risk for certain levels of investors. The securities received high credit ratings, despite the riskiness of the mortgages. When the housing bubble burst, investors and banks suffered significant losses

on MBS and CDO, even on investments that received AAA ratings. The impact on markets was widespread. Consumer and investor confidence plummeted and credit tightened. In 2010, in response to the role of the financial sector in the crisis, the Dodd-Frank regulatory reforms were enacted, including the "Volcker Rule" which prohibits an insured depository institution from engaging in proprietary trading. With this motivation in mind, and parallel to the methods of Acharya et al. (2010) and Billio et al. (2012a), we turn our focus to financial sector equities.

Specifically, we analyze three industries within the financial sector: banks, brokers/dealers, and insurers. These industries have become increasingly connected over the past several years as institutions engage more frequently in non-core activities and investment opportunities become more varied and complex. For example, prior to the crisis AIG began insuring CDO obligations against default for banks, a counterparty relationship that had not been historically observed. Our goal for this empirical analysis is to explore time dynamics of the network of volatility contagion between institutions, as linkages between these institutions and industries change in a pre- and post-crisis economy.

1.4.1 Data

In order to evaluate the time dynamics of network connectivity, we consider five overlapping windows of time. Each consists of four years of weekly returns for 24 financial institutions. The time periods are January 5, 2001 - December 31, 2004; January 3, 2003 - December 29, 2006; January 7, 2005 - December 26, 2008; January 5, 2007 - December 31, 2010; and January 2, 2009 - November 30, 2012. These windows cover time periods before, during, and after the crisis of 2007-2009. For each time period we identify the 8 largest firms in each of the three industries: banks, brokers/dealers, insurers. Industry classification is based on Standard Industry Classification (SIC) codes, and size of firms is based on market capitalization at the mid-point of the time period (e.g. December 31, 2002 for the 2001-04 sample period). Market capitalization is available on University of Chicago's Center for

Table 1.6: Financial sector firms included in the analysis; Y indicates that a firm was included in the analysis for a particular 4-year time period or Out-of-Sample (OOS) analysis. Industry classification: 1 = Banks, 2 = Brokers/Dealers, 3 = Insurance.

Ticker	Name	Industry	01-04	03-06	05-08	07-10	09-12	OOS
AXP	American Express	1	Y	Y	Y			Y
BAC	Bank of America	1	Y	Y	Y	Y	Y	Y
BK	Bank of New York Mellon	1				Y		
BNS	Bank of Nova Scotia	1					Y	
C	Citigroup	1	Y	Y	Y	Y	Y	Y
FNM	Fannie Mae	1	Y	Y				
JPM	JP Morgan	1	Y	Y	Y	Y		Y
RY	Royal Bank Canada	1			Y	Y	Y	Y
TD	Toronto Dominion Bank	1					Y	
UBS	UBS	1	Y	Y	Y	Y	Y	Y
USB	US Bancorp	1	Y	Y	Y	Y	Y	Y
WFC	Wells Fargo	1	Y	Y	Y	Y	Y	Y
AMTD	TD Ameritrade	2	Y	Y		Y	Y	Y
BEN	Franklin Resources	2	Y	Y	Y	Y	Y	Y
BLK	Blackrock	2			Y	Y	Y	Y
CME	Chicago Mercantile Exchange	2		Y	Y	Y	Y	
GS	Goldman Sachs	2	Y	Y				
IVZ	Invesco	2				Y	Y	
LM	Legg Mason	2	Y	Y	Y			Y
MS	Morgan Stanley	2	Y	Y	Y			Y
NYX	NYSE Euronext	2			Y	Y	Y	
SCHW	Charles Schwab	2	Y	Y	Y	Y	Y	Y
SEIC	SEI Investments	2	Y					Y
TROW	T Rowe Price	2	Y	Y	Y	Y	Y	Y
AFL	Aflac	3	Y			Y	Y	
AIG	American International Group	3	Y	Y	Y		Y	Y
ALL	Allstate	3	Y	Y	Y	Y		Y
BRK	Berkshire Hatheway	3	Y	Y	Y	Y	Y	Y
CB	Chubb	3				Y		
HIG	Hartford Financial	3	Y		Y			Y
MET	Metlife	3	Y	Y	Y	Y	Y	Y
MFC	Manulife Financial	3		Y	Y	Y	Y	Y
MMC	Marsh & McLennan	3	Y					
PRU	Prudential Financial	3		Y			Y	
TRV	Travelers	3			Y	Y	Y	Y
UNH	UnitedHealth Group	3	Y	Y	Y	Y	Y	Y
WLP	Wellpoint	3		Y				

Research in Security Prices (CRSP). For inclusion in the sample, firms must have traded until December 31, 2012 and must have a trading history covering the entire period. Weekly returns for the 24 firms are collected on Yahoo! Finance. The firms selected for each time period are shown in Table 1.6. Because the relative size of firms changes over time, different firms may be included in each sample. The out-of-sample analysis in Section 1.4.3

collects data analogously. A consequence of the data collection procedure is that firms that are no longer publicly traded will not be included in the sample. This restriction is potentially important for the analysis because it precludes bankrupt firms which may have been especially exposed to contagion, for example Lehman Brothers in October 2008.

For each sample time period, the returns are filtered according to the multi-factor model described in Section 1.2.1 with the full set of factors listed in Table 1.1.

Analyzing publicly traded equities provides several important benefits. The data is available at a higher frequency, which allows us to estimate models in shorter windows. Hedge fund data is typically reported on a monthly basis. To acquire a reliable sample size for a model of network connectivity, at least 10 years of data would be required. Market conditions evolve much more rapidly, and the ability to estimate the model on narrower intervals of time is essential. We have constructed 4-year time windows, but even shorter windows based on daily data would be possible. We also collect data on individual firms, as opposed to indexes, which allows us to identify specific financial institutions which are most exposed to, or most responsible for, systemic risk.

1.4.2 Network "Dynamics"

We estimate the model separately for each of the five time windows, and Table 1.7 summarizes overall posterior mean network connectivity for each sample period. We find that there are time dynamic patterns in financial sector connectivity.

Table 1.7: Posterior mean number of connections, within industry, between industry, and the percentage of connections within industry.

Time Period	# Connections	# Within	# Between	% Within
2009-12	40.05	12.28	27.76	30.07%
2007-10	42.31	12.52	29.79	29.59%
2005-08	41.83	15.46	26.36	36.97%
2003-06	32.80	10.20	22.60	31.09%
2001-04	34.19	14.08	20.11	41.20%

There are relatively fewer linkages in the pre-crisis periods, with 34.19 and 32.80 posterior mean number of connections in 2001-04 and 2003-06, respectively. The rate of connections spiked in the 2005-08 sample with a 27.5% increase over the 2003-06 sample. Speculatively, this sudden increase in connectivity reflects the elevated exposure to systemic risk in the pre-crisis time period. Market connectivity remained high during and after the crisis. The number of connections in the sample periods 2005-08, 2007-10, and 2009-12 are all similar, corresponding to the time periods immediately before, during, and after the crisis, respectively. With 2009-12 linkages on par with pre-crisis levels in 2005-08, the model suggests that despite regulatory efforts, the connectivity within the financial sector continues to reflect a vulnerability to systemic events.

While the overall rate of connectivity has been relatively stable after the crisis, the structure of the network has changed. In 2005-08, about 37% of all observed connections occured for firms within the same industry. In the 2007-10 and 2009-12 samples, this number falls to about 30%. Firms are increasingly becoming more connected to firms in other industries and less connected to firms in the same industry. There are many possible explanations for this trend, including increased regulatory oversight, more diverse product offerings and non-core investments, and changing business relationships between industries.

Figure 1.7 graphically illustrates the time dynamics of the network structure. For each time window, firms are plotted according to their posterior mean latent position on the circle, color coded by industry. Lines between firms indicate network connections, where the thickness of the line corresponds to the posterior probability of a connection. In the early time periods, the location of firms tends to cluster by industry, but the posterior mean distance between firms within the same industry increases in later time periods. Figure 1.7 also shows an increase in connectivity starting in the 2005-08 sample period, and an increase in cross-industry connections, most visible in 2007-10.

The time dynamics of connectivity in the network provide a measure of the degree of volatility contagion across the financial sector, according to our working definition of conta-

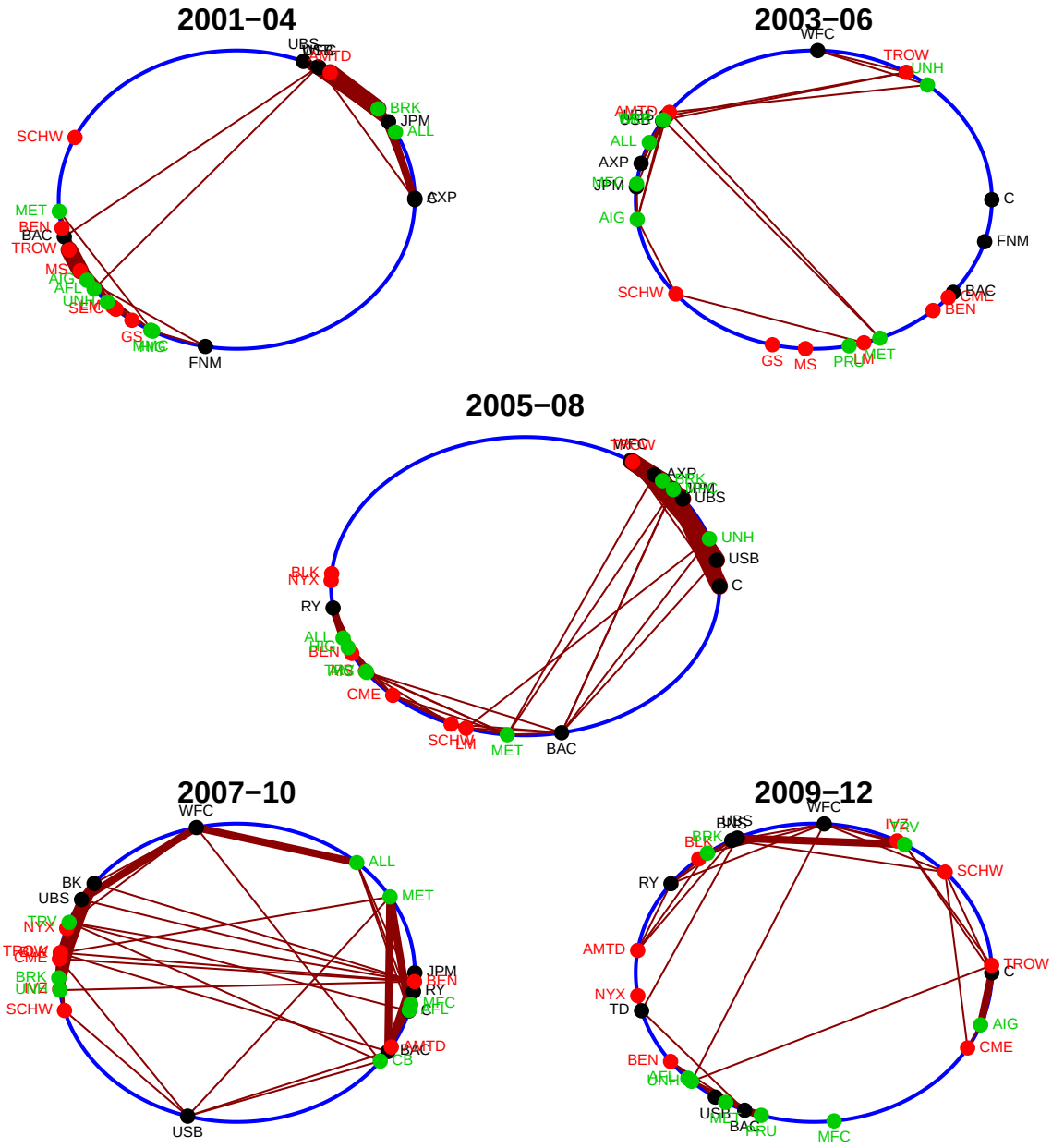

Figure 1.7: Posterior mean locations and network observed in 4-year time periods; line thickness is associated with the posterior probability of a network connection; colors indicate industry groups (black = banks, red = brokers/dealers, green = insurers).

gion. Volatility contagion surged prior to the crisis, and continues to be elevated relative to the pre-crisis period. In Section 1.4.3 we demonstrate the out-of-sample predictive power of our model at the institutional level.

1.4.3 Out-of-sample Prediction

The value of the model for evaluating systemic risk is dependent on its ability to forecast systemic risk exposure at both the institutional and sector levels. Consistent with the approach of Acharya et al. (2010) and Billio et al. (2012a), we evaluate the power of our model for predicting institutional losses on the basis of inferred network properties. The out-of-sample time period July 2007 - December 2008 (monthly) corresponds to the subprime mortgage crisis, allowing us to directly assess model performance in regard to a specific and recent distress period. For the out-of-sample time period, we calculate the maximum percentage loss (Max%Loss) of each firm in the sample. Max%Loss is defined as the largest percent change in market capitalization for the months July 2007 - December 2008, relative to June 2007. The Max%Loss variable definition and notation follows that of Billio et al. (2012a).

Using the same data collection procedure described in Section 1.4.1, we obtain and filter weekly returns for 24 firms for the within-sample period July 6, 2001 - June 29, 2007. We estimate our model on this within-sample data. Using the output from the MCMC estimation routine, we construct forecasts of volatility contagion. At each iteration of the MCMC chain after burn-in:

1. Suppose all firms are in the low volatility regime.

2. Shock the system by forcing firm i into the high volatility regime at time T and beyond (all other firms remain in the low volatility state at time T).

3. According to the parameters of the model at the current iteration of the chain, simulate forward the regime states using Equation 1.2 to forecast the 4-week ahead ($T+4$) volatility regime of each firm following the shock to firm i.

4. Repeat for all firms $i = 1, \ldots, n$.

For firm i, we define Volatility Incurred$_i$ as the average volatility state of firm i at time $T+4$ when another randomly selected firm is shocked into the high volatility state at time T.

Volatility Created$_i$ is defined as the proportion of all firms in the high volatility state at time $T + 4$ when firm i is shocked into the high volatility state at time T. Intuitively, Volatility Incurred is a measure of a firm's exposure to volatility contagion, and Volatility Created is a measure of the volatility contagion for which a firm is responsible. These volatility contagion forecasts are closely related to the number of connections a firm has in the model. More connected firms are expected to have greater exposure to and responsibility for volatility contagion. Our analysis is not sensitive to the choice of the number of weeks ahead to forecast volatility, and observed patterns are similar for 1, 2, 4, and 8 steps.

For our within-sample data, we estimate the model and construct the volatility contagion forecasts. The results are shown in Table 1.8. AIG, Hartford (HIG), and Citigroup (C) experienced the largest crisis period losses. UBS and JP Morgan (JPM) have the largest number of connections, with Legg Mason (LM), United Health (UNH), and Franklin Resources (BEN) having the least. The firms with the largest Volatility Created are generally firms with more connections, notably UBS, JPM, Met Life (MET), AIG, and Charles Schwab (SCHW). Differences in Volatility Created are smaller than differences in Volatility Incurred. While Volatility Incurred for UNH is half that of UBS, the rate of Volatility Created only varies from 0.15 for the lowest firms to 0.20 for the highest firms. The structure of the network has an important impact on the relationship between the number of connections, Volatility Incurred, and Volatility Created. For example, even though Allstate (ALL) only has 2.85 connections, it has the same level of Volatility Created as C which has 13% more connections and 11% more Volatility Incurred.

To conduct inference for the out-of-sample performance of contagion forecasts and connectivity, we consider separate linear regressions predicting Max%Loss for each firm on the basis of the number of connections and on the basis of Volatility Incurred and Volatility Created. Table 1.9 shows coefficient estimates and standard errors for these regressions, which also control for the industry of the firm. In the first regression, the number of connections is not a significant predictor of out-of-sample losses. The coefficient estimate is slightly neg-

Table 1.8: Maximum crisis period loss and posterior means for network statistics for each of the 24 firms used in the out-of-sample analysis. Industry 1 = Banks, 2 = Brokers/Dealers, and 3 = Insurers.

Ticker	Industry	Max%Loss	# Connections	Vol. Incurred	Vol. Created
C	1	0.86	3.23	0.18	0.18
AXP	1	0.70	3.08	0.17	0.18
UBS	1	0.69	5.08	0.24	0.20
BAC	1	0.67	4.08	0.20	0.19
RY	1	0.42	3.64	0.19	0.19
WFC	1	0.33	4.11	0.20	0.19
JPM	1	0.29	5.02	0.23	0.20
USB	1	0.23	3.77	0.19	0.19
MS	2	0.82	3.85	0.19	0.19
LM	2	0.80	0.73	0.11	0.15
BEN	2	0.57	1.70	0.13	0.16
SEIC	2	0.48	4.15	0.21	0.20
TROW	2	0.36	3.21	0.17	0.18
AMTD	2	0.34	3.52	0.18	0.19
SCHW	2	0.27	4.56	0.22	0.20
BLK	2	0.18	2.31	0.14	0.17
AIG	3	0.98	4.70	0.22	0.20
HIG	3	0.92	4.43	0.21	0.20
ALL	3	0.64	2.85	0.16	0.18
UNH	3	0.63	1.40	0.12	0.16
MFC	3	0.56	4.00	0.20	0.19
MET	3	0.52	4.70	0.22	0.20
TRV	3	0.30	3.48	0.19	0.19
BRK	3	0.14	3.39	0.18	0.19

ative, but is statistically indistinguishable from zero. However, in the second regression we find that Volatility Incurred and Volatility Created are both statistically significant predictors of Max%Loss. Given the relatively small sample size of 24 firms, we use a significance level of 0.10, though Volatility Created would also be significant at the 0.05 level. Volatility Incurred is positively associated with Max%Loss, meaning that firms with more exposure to volatility contagion experience significantly greater losses, controlling for the level of Volatility Created. On the other hand, Volatility Created is negatively associated with Max%Loss, meaning that firms with greater responsibility for the spread volatility contagion experience significantly fewer losses, controlling for the level of Volatility Incurred. Combining the re-

Table 1.9: Linear regression models for predicting out-of-sample maximum crisis loss from features of the network estimated on within-sample data. Standard errors in parentheses.

	Max%Loss	Max%Loss
(Intercept)	0.57**	4.88**
	(0.17)	(1.98)
# Connections	−0.01	
	(0.05)	
Industry = Brokers/Dealers	−0.07	−0.07
	(0.14)	(0.13)
Industry = Insurance	0.05	0.11
	(0.13)	(0.12)
Vol. Incurred		16.40*
		(7.93)
Vol. Created		−39.78**
		(18.30)
R^2	0.04	0.23
Num. obs.	24	24

$^{**}p < 0.05$, $^{*}p < 0.1$

sults of these two regression equations, we conclude that it is not the number of connections in the network, but rather the structure of the network, that predicts crisis period losses. Firms exposed to volatility contagion experience greater losses, while firms that spread contagion experience fewer losses. The R^2 from the second regression equation is 0.23, indicative of the strength of this relationship. Note, however, there are many other important economic and financial factors related to institutional crisis period losses for which we do not account.

As a final demonstration of the relationship between inferred network connectivity and firm characteristics, particularly those associated with systemic risk, we regress the number of connections for a firm on balance sheet measures of size, illiquidity, and leverage. Size is defined to be the log of book equity; illiquidity is based upon maturity mismatch, which is the difference in short term debt and cash relative to total liabilities; and leverage is defined to be the ratio of total assets to total equity. These measures are calculated for each firm on the basis of 2006 fiscal year financial reports in CRSP. Although there is not a forecasting element to this analysis, we continue to use the same data and model output. The results

of this regression are shown in Table 1.10.

Table 1.10: Linear regression model for the relationship between firm connectivity and balance sheet characteristics. Standard errors in parentheses.

	Connections
(Intercept)	2.07
	(1.88)
Log(Size)	0.09
	(0.18)
Illiquidity	0.05
	(0.50)
Leverage	0.05**
	(0.02)
R^2	0.30
Num. obs.	24

$^{**}p < 0.05$, $^{*}p < 0.1$

There is a statistically significant, positive correlation between firm connectivity and leverage. More leveraged firms are more exposed to contagion transmission in our model. This relationship provides further evidence for the association of inferred linkages with systemic risk exposure, expressed indirectly through leverage effects. Connectivity is not associated with either size or liquidity of firms.

The overall conclusion from the out-of-sample test is that specific elements of our network model may be predictive of institutional losses during the recent financial crisis, which is consistent with the findings of Billio et al. (2012a). We find that the number of linkages in the network is less important than how firms are linked in the network. The results reveal an important distinction between exposure to volatility contagion and responsibility for volatility contagion. Further analysis shows that leverage is positively associated with the degree of contagion of individual firms.

Again, we emphasize that these conclusions are neither causal nor linked structurally to the economy. We have considered only a single crisis period, and we have chosen a particular construction for measures of Volatility Incurred and Volatility Created. All conclusions are descriptive in nature. However, we believe that we have collected enough suggestive evidence

to convincingly demonstrate a relationship between linkages and volatility contagion in our model and characteristics of systemic risk exposure.

1.5 Conclusion

We have proposed a network model linking the volatility regime transitions of institutions, capturing both exposure to and responsibility for the spread of volatility contagion. Empirical analysis shows that the network model is a statistically significant improvement over the regime switching model, which is evidence for the existence of volatility contagion in the network. Focusing on the financial sector, we find that patterns of linkages between institutions are dynamic, with increases prior to the subprime mortgage crisis and connections across industries becoming more common after the crisis. Out-of-sample forecasts show that firms that are more susceptible to contagion undergo significantly more losses during systemic events, and firms that spread contagion experience significantly lower losses. Overall, we find that the network model captures certain features of systemic risk exposure for the market and for individual institutions for a specific crisis period. The degree and structure of connectivity have dynamic properties that reflect historically observed patterns of systemic risk.

There are several limitations to this study. Most importantly, the model does not reflect any real economic mechanisms explaining linkages between institutions in the marketplace. Because the study is descriptive, we cannot make any causal conclusions with respect to inferred linkages or their relationship to systemic risk exposure. We have addressed this concern by conducting out-of-sample tests illustrating the predictive power of linkages for forecasting crisis period losses of institutions. Nonetheless, more informative network models can be constructed by incorporating counterparty exposure, marketplace transactions between institutions, and other real-time financial information. This information is not publicly available, but regulatory organizations have access to this type of data and the mandate

to evaluate and restrict systemic risk in the marketplace.

We motivated the model in reference to the four "L"s of systemic risk. Our model accounts for losses, through the high volatility crisis regime, and linkages, through the network. The model is limited by its failure to account for leverage and liquidity of institutions, although we have shown that leverage effects are captured indirectly by the degree of connectivity. An extension of our model to account for additional elements of systemic risk exposure would be to introduce covariates in the network model. For example, firms that have less liquid portfolios may be more likely to make connections in the marketplace. Incorporating covariates into the model is straightforward: the log-odds of a connection would simply be linear in these covariates, in addition to the latent distance of institutions. Covariates in the network model may also include financial statement information or data regarding counterparties.

The regime switching component of our model accounts for dynamics of the returns. However, the network structure itself is static over time. As we have discussed throughout the chapter, the nature, type, and number of relationships between institutions, is constantly evolving, especially in the financial sector. Our empirical analysis has avoided this issue to the extent possible by estimating the network on relatively short time windows (four years in the case of the financial sector study). Daily data would allow for even shorter windows. A more comprehensive solution would be to allow for certain elements of the network to also be dynamic over time. This is potentially an area of future research.

In the post-crisis economy, banks continue to be subjected to scrutiny for their contributions and exposure to systemic risk. Regulatory and reporting requirements have increased following the passage of Dodd-Frank, and the Federal Reserve administers stress tests to determine whether banks have enough capital to survive a severe economic crisis. Despite the limitations of our study, the results confirm the need for ongoing evaluation of systemic risk, especially in the financial sector. Connections between institutions across industries have increased in the post-crisis economy, relative to the number of connections within industries.

Given the degree of connectivity observed in our model, we would speculate that the risk of volatility contagion and systemic risk remains as great today as in the pre-crisis economy.

Chapter 2

Benchmarking Bayesian Small Area Estimates

2.1 Introduction

In many surveys, it is of interest to estimate a parameter at different geographic levels, such as at the county, state, and national level. For example, the Small Area Health Insurance Estimates (SAHIE) program at the United States Census Bureau produces estimates of the count of uninsured individuals within states and counties, cross-classified by age, gender, income, and additionally by race for state estimates. The American Community Survey (ACS) is the primary source of data used for the project. Because the ACS is a large-scale survey with over three million households sampled annually, the direct survey estimates of the number of uninsured individuals at the national or state levels are typically very reliable. However, sample sizes in some domains may not be large enough to produce reliable direct estimates for specific small areas. Here, *small area* refers to any subpopulation of interest, such as a geographic area or socio-demographic group, for which the domain-specific sample is not large enough for reliable direct estimation. In the case of SAHIE, consider the difficulty of estimating the number of uninsured men between the ages of 40 and 64 with income at

or below 138% of the poverty level in Lawrence County, South Dakota, which has a county population size just over 20,000 and an ACS sample size within the county of only a few hundred households. (In 2009 SAHIE estimated that 40.0% of such residents in Lawrence County were uninsured, with standard error of 8.6%).

For sample surveys, precision for estimates within small areas can be increased by introducing a small area model (Rao, 2003), which borrows strength by connecting different areas and incorporating auxiliary covariate information. When multiple geographic levels are of interest, conceptually, it makes sense to implement a joint model for the higher and lower levels. From a practical standpoint, this is nearly impossible due to the complexity of the models, the number of parameters to be estimated, and the fact the available data and parameters of interest are not necessarily nested from one level to the next. For this reason, the small area modeling is done separately at distinct geographic levels. These model-based estimates can differ substantially from direct estimates, especially for areas with small sample sizes. An important consequence is that the model-based estimates for the smaller geographic areas typically do not aggregate to the corresponding direct survey or model-based estimate of a larger geographic area. There are several reasons why this inconsistency is undesirable for survey administrators: government funds may be appropriated on the basis of model estimates (e.g. to counties within states), reporting agencies' credibility with the general public or legislators may be impacted, and the degree of inconsistency may be a reflection of model misspecification. *Benchmarking* is the statistical procedure which adjusts model-based estimates to satisfy constraints.

Existing benchmarking procedures have been developed from a frequentist perspective primarily for application to linear random effects models. Random effects models are closely related to the classical small area model introduced by Fay and Herriot (1979). Wang et al. (2008) review benchmarking procedures for random effects models, providing simulation results and a mathematical formulation which unifies several different procedures as minimizers of weighted Mean Square Error (MSE). For more general Bayesian small area models, Datta

et al. (2011) extend the results of Wang et al. (2008) to produce a Bayes rule benchmarked estimator, and Toto and Nandram (2010) illustrate Bayesian benchmarking with sampling weights. Nandram et al. (2011) develop a Bayesian benchmarking procedure applicable in the case of the one-way random effects model.

The approaches described in this chapter stand in contrast to existing methods. Rather than obtaining point estimates by minimizing a loss function subject to a benchmark constraint for linear random effects models, general Bayesian small area models are benchmarked by modifying the posterior distributions themselves at the different geographic levels. This approach has the advantage of guaranteeing that further characteristics, such as the estimated variances at each level of the model also agree.

Additionally, this chapter aims to develop a new perspective for benchmarking estimates which considers a *flexible* benchmarking constraint. Previously existing benchmarking procedures adjust lower geographic level estimates to be consistent with a higher geographic level estimate, which is held fixed. Because there is often uncertainty associated with the higher level estimate, it is possible that inference can be improved by allowing the higher and lower level estimates to be adjusted simultaneously in the benchmarking procedure. This flexible constraint allows the lower level estimates to inform the adjustment of the estimate of the higher level total, and vice versa. With a fixed constraint, there is no mechanism for the lower level estimates to influence the higher level estimate. In this way, the flexible constraint introduces new information into the benchmarking procedure, creating potential for improved performance.

This chapter illustrates two new benchmarking procedures for modifying the posterior distribution for general Bayesian small area models, either by explicitly describing the form of the modified posterior distributions or through approximations based on MCMC output. It is shown that in both procedures, either a fixed or a flexible benchmarking constraint can be used. Section 2.2 provides background on existing benchmarking results and discusses the links between these procedures. Section 2.3 introduces the new benchmarking methods

for Bayesian small area models with either a flexible or a fixed constraint. In Section 2.4, a simulation study is presented comparing the two methods to each other and to existing benchmarking procedures. Concluding remarks are given in Section 2.5.

2.2 Background

Throughout this chapter, let $\hat{\theta}$ denote a direct survey estimate of the higher level (for example, state) parameter of interest, θ, and let $\hat{\theta}_i$ denote direct survey estimates of the lower level (for example, county) parameters, θ_i, for $i = 1, \ldots, m$. θ and $(\theta_1, \ldots, \theta_m)$ are assumed to be related through some benchmarking constraint. For example, traditional benchmarking literature addresses the linear constraint

$$\sum_{j=1}^{m} c_j \theta_j = \theta,$$

for known constants $c_1, \ldots, c_m$. In the case of SAHIE, θ represents the number with health insurance in a state, and θ_i is the number with health insurance in county i in the state. The linear constraint then holds with $c_i \equiv 1$ for all i.

Small area models are often necessary when the sample size associated with the direct estimates is insufficient for reliable direct estimation. One popular small area model is the Fay-Herriot model (Fay and Herriot, 1979)

$$\begin{aligned} \hat{\theta}_i | \theta_i, D_i &\overset{ind.}{\sim} N(\theta_i, D_i), \\ \theta_i | \mathbf{x}_i, \boldsymbol{\beta}, A &\overset{ind.}{\sim} N(\mathbf{x}_i^T \boldsymbol{\beta}, A), \end{aligned} \tag{2.1}$$

where D_i is the variance of $\hat{\theta}_i$ and is assumed known according to the survey design, $\mathbf{x}_i$ is an observed vector of covariates, and $\boldsymbol{\beta}$ and A are unknown parameters. This model is a linear random effects model and is chosen for illustrative purposes due to its connections with the existing benchmarking literature and because it forms the foundation of many small area models used in practice. We emphasize that the new benchmarking approaches we develop

are suitably general to be applied to any Bayesian small area model. For concreteness, we will often refer to the Fay-Herriot model (2.1), and we will consider this model in our simulation studies.

From a Bayesian perspective, inference can be made through the posterior distribution by using a prior, such as $\pi(\boldsymbol{\beta}, A) \propto A^{-a-1}e^{-b/A}$. The choice $a = -1$ and $b = 0$ results in a non-informative, improper prior on $(\boldsymbol{\beta}, A)$, but a proper posterior. This particular prior is known as Stein's harmonic prior, which is scale-invariant, admissible, and conjugate (Morris and Lysy 2012). When Bayesian models are used, the posterior distribution often is intractable, and instead is only known up to a normalizing constant. In such cases, a Gibbs sampler (Gelfand and Smith, 1990) or MCMC algorithm must be used, as is true of the SAHIE model. The posterior means and the quantiles of the posterior distribution can be used to formulate point and interval estimates for the parameters $(\theta_1, \ldots, \theta_m)$. We denote these point estimates by $(\tilde{\theta}_1, \ldots, \tilde{\theta}_m)$.

From a frequency perspective, estimates of the parameters are constructed by computing the Best Linear Unbiased Predictor (BLUP):

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} (D_i + A)^{-1} \mathbf{x_i} \mathbf{x_i'}\right]^{-1} \left[\sum_{i=1}^{n} (D_i + A)^{-1} \mathbf{x_i} \hat{\theta}_i\right]$$

$$\gamma_i = \frac{A}{D_i + A}$$

$$\tilde{\theta}_i = \mathbf{x_i^T}\hat{\beta} + \gamma_i(\hat{\theta}_i - \mathbf{x_i^T}\hat{\beta})$$

If A is unknown, an estimator of A is substituted, resulting in the Empirical BLUP (EBLUP). For relatively simple linear random effects models, Bayesian and frequentist inference will typically be similar.

Whether the parameters are estimated from a frequentist or Bayesian standpoint, the Fay-Herriot model (2.1) produces small area estimates $\tilde{\theta}_i$ of θ_i for $i = 1, \ldots, m$. The esti-

mate of the higher level parameter, $\tilde{\theta}$, may be the direct survey estimate or model-based. Typically the estimates $\tilde{\theta}$ and $(\tilde{\theta}_1, \ldots, \tilde{\theta}_m)$ themselves will not satisfy the benchmarking constraint, that is, $\sum_{j=1}^m c_j \tilde{\theta}_j \neq \tilde{\theta}$. Benchmarking procedures systematically adjust the point estimates to produce new estimates θ^* and $(\theta_1^*, \ldots, \theta_m^*)$, such that $\sum_{j=1}^m c_j \theta_j^* = \theta^*$. Traditional approaches, including those discussed below, assume $\theta^* = \tilde{\theta}$ and adjust only the lower level estimates. The assumption of a fixed higher level estimate will be relaxed in Section 2.3, where we consider flexible benchmarking constraints.

A straightforward benchmarking procedure is the "raking" or "ratio" adjustment,

$$\theta_i^* = \frac{\tilde{\theta}_i}{\sum_{j=1}^m c_j \tilde{\theta}_j} \theta^*,$$

for $i = 1, \ldots, m$. This procedure is simple and adjustments are made proportional to the size of an estimate. However, raking does not account for the uncertainty associated with estimates or the model from which $(\tilde{\theta}_1, \ldots, \tilde{\theta}_m)$ are generated. For example, certain types of survey designs over-sample from under-represented domains. In such cases, the raking procedure would not be appropriate. The over-sampled areas would have proportionally less variability and should optimally receive smaller benchmarking adjustments relative to their size. In practice, the raking procedure may be a reasonable approach because survey design-based variance is often roughly proportional to the size of the area. You et al. (2004) provide an example of the raking procedure applied to estimating the Canadian census undercoverage.

Wang et al. (2008) establish a criterion for choosing the "best" predictors for $(\theta_1, \ldots, \theta_m)$ which satisfy the benchmarking constraint. In particular, they use weighted mean square error

$$\sum_{j=1}^m \phi_j E(\theta_j^* - \theta_j)^2,$$

with weights $\phi_i, i = 1, ..., m$. In the Fay-Herriot model (2.1), Wang et al. (2008) show that the estimator which minimizes the weighted mean square error subject to the benchmarking

constraint $\sum_{j=1}^{m} c_j \theta_j^* = \sum_{j=1}^{m} c_j \hat{\theta}_j$ is

$$\theta_i^* = \tilde{\theta}_i + \left(\sum_{j=1}^{m} \phi_j^{-1} c_j^2 \right)^{-1} \phi_i^{-1} c_i \left(\theta^* - \sum_{j=1}^{m} c_j \tilde{\theta}_j \right),$$

for $i = 1, \ldots, m$. Note that this estimator allocates the excess difference between the higher and lower level estimates, $\theta^* - \sum_{j=1}^{m} c_j \tilde{\theta}_j$, to each component θ_i^* on the basis of $\left(\sum_{j=1}^{m} \phi_j^{-1} c_j^2 \right)^{-1} \phi_i^{-1} c_i$.

This benchmarking procedure unifies, through the choice of weights ϕ_i, several procedures developed previously in the literature. Notably, the choice $\phi_i = D_i^{-1}$ gives the Isaki et al. (2000) benchmarked estimates; Pfeffermann and Barnard (1991) construct a benchmarked predictor equivalent to the choice of $\phi_i = c_i[\text{cov}(\tilde{\theta}_i, \sum_{j=1}^{m} c_j \tilde{\theta}_j)]^{-1}$; and when $\phi_i = [\widehat{\text{var}}(\tilde{\theta}_i)]^{-1}$ the resulting estimator is the one used by Battese et al. (1988). We will refer to these estimators as ITF, PB, and BHF, respectively. Roughly speaking, ITF assigns the excess (or shortage) to each small area proportional to the survey design variance within the area. BHF, on the other hand, assigns excess on the basis of the variance of the estimator $\tilde{\theta}_i$. PB assigns the excess proportional to the covariance between the area estimate and the estimate of the total. All three procedures are similar in that the excess is allocated on the basis of uncertainty. The procedures differ in how the measure of uncertainty is determined. BHF and PB account for modeling uncertainty, whereas ITF is strictly based on the uncertainty of the survey design. PB accounts for uncertainty beyond BHF by allowing ϕ_i to include the covariance terms.

Additionally, the choice of weight $\phi_i = \frac{c_i}{\tilde{\theta}_i}$ results in the ratio estimator. As previously discussed, the RATIO procedure allocates excess proportional to the size of the estimate.

Datta et al. (2011) approach the benchmarking problem from the Bayesian perspective. With this approach, the goal is to minimize the weighted posterior expected squared error under the benchmarking constraint

$$\sum_{j=1}^{m} \phi_j E\left[(\theta_j^* - \theta_k)^2 \mid \hat{\theta}, \hat{\theta}_1, \ldots, \hat{\theta}_m \right],$$

where $\sum_{j=1}^{m} c_j \theta_j^* = \theta^*$. The solution to this problem is identical to that of Wang et al. (2008) for linear predictors, though from the Bayesian point of view, the same solution is more generally applicable to a wider class of models and non-linear predictors. As above, this benchmarking approach unifies commonly used procedures through the choice of the weights ϕ_i.

Another approach to benchmarking, not based on a minimization criterion, is to choose $\hat{\boldsymbol{\beta}}^*$ such that the resulting predictors automatically satisfy the benchmarking constraint. We will refer to this method, developed primarily for unit-level models by You and Rao (2002) but also applicable to area level models, as YR.

These existing procedures have limitations which make them unsuitable for many practical applications. The SAHIE model is an example for which the current benchmarking technology is insufficient. Variously, the currently available benchmarking procedures may fail to estimate the variance of the benchmarked estimators, may not benchmark both first and second moments, may not be directly generalizable beyond linear random effects models, may be applicable for only two geographic levels (e.g. state and nation, but not county, state, and nation), or may not allow for possibly flexible benchmarking constraints where the higher level estimate can be simultaneously adjusted with the lower level estimates. All of these points allow for potential improvement in benchmarking.

2.3 Methods

In this section we will develop two new benchmarking procedures which address the limitations of existing methods. In particular, both of the proposed procedures modify the posterior distributions of the small area models to satisfy constraints, rather than strictly modifying the point estimates. By adjusting entire distributions, higher moments can be benchmarked, and the benchmarked standard errors of the benchmarked point estimates can be directly calculated using the adjusted posterior distributions. Moreover, each of these model-based

approaches to benchmarking is suitably general to be applied to any Bayesian small area model with either flexible or fixed constraint.

2.3.1 Minimum Discrimination Information

Define

$$\boldsymbol{\theta} = \begin{bmatrix} \theta & \theta_1 & \theta_2 & \dots & \theta_m \end{bmatrix}^T.$$

For any Bayesian small area model, conditional on the survey and auxiliary data, inference can be made through the joint posterior distribution $\pi(\boldsymbol{\theta})$. In practical applications, the different geographic levels are modeled separately. Consequently, $\pi(\boldsymbol{\theta}) = \pi_h(\theta)\pi_l(\theta_1, ..., \theta_m)$, where π_h and π_l are the respective posterior distributions for small area models at the higher and lower geographic levels. Where appropriate, the higher level model may simply be based upon the survey design. To estimate the joint posterior π, MCMC samples are taken independently from π_h and π_l. Accordingly, θ is uncorrelated with θ_i for all i, under π.

The first of the proposed procedures is based on the principle of Minimum Discrimination Information (MDI) as described by Kullback (1959). Kullback-Leibler (K-L) divergence is a measure of the "distance" between two probability distributions (Kullback and Liebler, 1951). The principle of MDI implies that given benchmarking constraints on $\boldsymbol{\theta}$, with posterior distribution π, a new distribution π^* should be chosen to be as close to π as possible in terms of the K-L divergence while satisfying the constraints in expectation. In other words, π^* should be as hard to discriminate from π as possible, and the information gain from incorporating the benchmarks as small as possible. MDI is closely related to the principle of maximum entropy (Jaynes, 1957).

The K-L divergence between π and π^* is defined as

$$KL(\pi^*, \pi) = E\left(\log \frac{\pi^*}{\pi}\right) = \int \pi^*(\boldsymbol{\theta}) \log \frac{\pi^*(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta}.$$

The goal is to choose an adjusted "posterior" distribution, π^*, which minimizes this quantity subject to the benchmarking constraints in expectation. In the case of SAHIE, analysts wish

to benchmark first and second moments. These constraints are given by

$$T_1(\boldsymbol{\theta}) = \theta - \sum_{j=1}^{m} c_j \theta_j, \quad T_2(\boldsymbol{\theta}) = \theta^2 - \left(\sum_{j=1}^{m} c_j \theta_j \right)^2.$$

Thus, applying the MDI principle and incorporating the benchmarking constraints results in π^*, given by the optimization problem:

$$\min_{\pi^*(\boldsymbol{\theta})} KL(\pi^*, \pi) \text{ such that } \int T_1(\boldsymbol{\theta})\pi^*(\boldsymbol{\theta}) = \int T_2(\boldsymbol{\theta})\pi^*(\boldsymbol{\theta}) = 0. \tag{2.2}$$

MDI provides a principled approach for benchmarking Bayesian small area models from an information theoretic perspective. Theoretically, for any small area model we can numerically compute the appropriate adjusted distribution π^* and construct further MCMC draws from π^*. However, numerically solving for the MDI distribution is challenging and time consuming, and repeating an MCMC procedure may be prohibitively expensive in terms of computational time if the original model itself is complex and computationally intensive.

Alternatively, we can derive simple, closed form solutions by assuming the posterior distribution for $\boldsymbol{\theta}$ is Normal. In many small area models the direct survey estimates are assumed to follow a Normal distribution. If the data itself is assumed to be Normal, and the influence of the prior is minimal, it is reasonable to expect the posterior to be approximately Normal. Also, if the overall sample size is large, the Bernstein-von Mises theorem states that the posterior distribution is approximately Normal. The shrinkage properties of small area models may introduce some skewness, but in simulations we have found the impact of the skewness to be minimal. Even if the posterior distribution is not precisely Normal, this assumption makes the MDI approach analytically tractable, and it can then be used as a tool to adjust the small area estimates in a systematic and principled way to achieve the benchmarking constraint. From the practitioner's standpoint, we will show that the MDI procedure with Normality approximation is straightforward to implement, requiring only the knowledge of posterior means and covariances of the parameters.

Given MCMC draws from any small area model posterior distribution for θ_i, for $i =$

$1, \ldots, m$ and θ, we compute the Normal approximation to the joint posterior by calculating the mean and covariance of these draws. We define the following notation

$$\boldsymbol{R} = \begin{bmatrix} 1 & -c_1 & -c_2 & \ldots & -c_m \end{bmatrix}^T$$

$$\boldsymbol{C} = \begin{bmatrix} c_1 & c_2 & \ldots & c_m \end{bmatrix}^T$$

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \tilde{\theta} & \tilde{\theta}_1 & \tilde{\theta}_2 & \ldots & \tilde{\theta}_m \end{bmatrix}^T$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_s \end{bmatrix},$$

where $\tilde{\boldsymbol{\theta}}$ is the vector of posterior means of $\boldsymbol{\theta}$, σ^2 is the posterior variance of θ, and $\boldsymbol{\Sigma}_s$ is the $m \times m$ dimensional posterior covariance matrix of $(\theta_1, \ldots, \theta_m)$. We further define $\sigma_s^2 = \boldsymbol{C}^T \boldsymbol{\Sigma}_s \boldsymbol{C}$, which is the posterior variance of $\sum_{j=1}^m c_j \theta_j$.

If the (approximate) posterior distribution for $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} \sim N_{m+1}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\Sigma}).$$

The benchmarked "posterior" distribution π^* satisfying the MDI principle is the solution to the optimization problem given in Equation (2.2):

$$\boldsymbol{\theta}^* \sim N_{m+1}\left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\Sigma}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\boldsymbol{R}^T\tilde{\boldsymbol{\theta}}, \begin{bmatrix} \frac{2\sigma_s^2\sigma^2}{\sigma_s^2+\sigma^2} & 0 \\ 0 & \left(\left(\frac{\sigma_s^2-\sigma^2}{2\sigma_s^2\sigma^2}\right)\boldsymbol{C}\boldsymbol{C}^T + \boldsymbol{\Sigma}_s^{-1}\right)^{-1} \end{bmatrix} \right).$$

The benchmarked point estimates and their benchmarked variances are given by the mean and variances of this MDI distribution. The details of the calculations are given in Appendix B.

The benchmarked MDI point estimate, $\tilde{\boldsymbol{\theta}} - \boldsymbol{\Sigma}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\boldsymbol{R}^T\tilde{\boldsymbol{\theta}}$, or equivalently, $\tilde{\boldsymbol{\theta}} - \boldsymbol{\Sigma}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\left(\tilde{\theta} - \sum_{j=1}^m c_j\tilde{\theta}_j\right)$, can be linked to several existing benchmarking procedures. It is identical to the so-called GR-estimator of Knottnerus (2003), which was further explored for domain estimation by Sostra and Traat (2009). And similar to the procedures described

in Section 2.2, the MDI approach allocates excess, $\tilde{\theta} - \sum_{j=1}^{m} c_j \tilde{\theta}_j$, to each of the areas on the basis of statistical uncertainty, which is $\boldsymbol{\Sigma R}(\boldsymbol{R}^T \boldsymbol{\Sigma R})^{-1}$. This measure of uncertainty for the MDI procedure differs from those in Section 2.2 by incorporating the variability of the national level estimate through the flexible benchmarking constraint. Accordingly, estimates for both higher and lower geographic levels are adjusted, and imposing the flexible constraint can be viewed as a compromise between the higher and lower level models.

Before any adjustment, there are two ways to estimate the mean and variance of θ – by directly using the higher level model through the posterior moments $E_h(\theta) = \tilde{\theta}$ and $Var_h(\theta) = \sigma^2$, or by using the lower level model in conjunction with the benchmarking constraint through the posterior moments $E_l\left(\boldsymbol{C}^T \boldsymbol{\theta}_s\right) = \boldsymbol{C}^T \tilde{\boldsymbol{\theta}}_s$ and $Var_l\left(\boldsymbol{C}^T \boldsymbol{\theta}_s\right) = \boldsymbol{C}^T \boldsymbol{\Sigma_s C} = \sigma_s^2$. Notice that for the MDI distribution π^*, the posterior mean is a weighted combination of $\tilde{\theta}$ and $\boldsymbol{C}^T \tilde{\boldsymbol{\theta}}_s$ while the posterior variance is the harmonic mean of σ^2 and σ_s^2. Furthermore, if $\sigma_s^2 > \sigma^2$, which would be expected in practice, the matrix $\frac{\sigma_s^2 - \sigma^2}{2\sigma_s^2 \sigma^2} \boldsymbol{C C}^T$ is non-negative definite, hence $\boldsymbol{\Sigma}_s \geq \left(\frac{\sigma_s^2 - \sigma^2}{2\sigma_s^2 \sigma^2} \boldsymbol{C C}^T + \boldsymbol{\Sigma}_s^{-1}\right)^{-1}$, and precision for benchmarked lower level estimates will be improved by using the flexible benchmarking constraint.

In practice, the higher level estimate may not be modeled if the direct estimate itself is assumed to be reliable. In such cases, a fixed benchmarking constraint may be appropriate, with $\sigma^2 = D$ (the design-based variance), and $\tilde{\theta} = \hat{\theta}$ (the direct survey estimate). Under these fixed benchmarking constraints, the first and second moment restrictions for the lower-level model, respectively, are

$$T_1^{\text{fix}}(\boldsymbol{\theta}_s) = \hat{\theta} - \sum_{j=1}^{m} c_j \theta_j, \quad T_2^{\text{fix}}(\boldsymbol{\theta}_s) = D - \left(\sum_{j=1}^{m} c_j \theta_j\right)^2,$$

where

$$\boldsymbol{\theta_s} = \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_m \end{bmatrix}^T.$$

Further define

$$\tilde{\boldsymbol{\theta}}_s = \begin{bmatrix} \tilde{\theta}_1 & \tilde{\theta}_2 & \dots & \tilde{\theta}_m \end{bmatrix}^T.$$

Then, the MDI distribution, π_l^*, for $\boldsymbol{\theta_s}$ nearest to the original posterior, π_l, satisfying $\int T_1^{\text{fix}}(\boldsymbol{\theta}_s)\pi_l^*(\boldsymbol{\theta}_s) = \int T_2^{\text{fix}}(\boldsymbol{\theta}_s)\pi_l^*(\boldsymbol{\theta}_s) = 0$ is

$$\boldsymbol{\theta_s}^* \sim N_m\left(\tilde{\boldsymbol{\theta}}_s + \boldsymbol{\Sigma}_s\boldsymbol{C}\left(\boldsymbol{C}^T\boldsymbol{\Sigma}_s\boldsymbol{C}\right)^{-1}\left(\hat{\theta} - \sum_{j=1}^{m} c_j\tilde{\theta}_j\right), \left(\left(\frac{\sigma_s^2 - D}{D\sigma_s^2}\right)\boldsymbol{C}\boldsymbol{C}^T + \boldsymbol{\Sigma}_s^{-1}\right)^{-1}\right).$$

The benchmarked point estimate, $\tilde{\boldsymbol{\theta}}_s + \boldsymbol{\Sigma}_s\boldsymbol{C}\left(\boldsymbol{C}^T\boldsymbol{\Sigma}_s\boldsymbol{C}\right)^{-1}\left(\hat{\theta} - \sum_{j=1}^{m} c_j\tilde{\theta}_j\right)$ is equivalent to the PB estimator, and consequently is also a Bayes rule under weighted squared error loss.

MDI provides a model-based and information theoretic approach to satisfying the benchmarking constraint in expectation based on the Normal approximation to the posterior distribution. It is closely related to some existing benchmarking techniques, but extends the current literature by allowing for a flexible benchmarking restriction and satisfying second moment benchmarking rules. Importantly for practitioners, the procedure is straightforward to implement, requiring only the knowledge of posterior means and covariances of the parameters.

2.3.2 Full Conditional

The MDI approach produces modified "posterior" distributions which satisfy certain benchmarking properties, such as consistency of the means and the variances at different geographic levels, according to the restrictions of the optimization problem in Equation (2.2). In this section, an alternative algorithm is presented for creating a fully consistent posterior distribution by embedding the benchmarking constraint into the model itself without requiring any additional distributional assumptions.

As in Section 2.3.1, consider the joint posterior distribution, conditional on auxiliary data and other parameters, $\pi(\boldsymbol{\theta}) = \pi_h(\theta)\pi_l(\theta_1, \ldots, \theta_m)$. Consider also a general benchmarking constraint, which relates the parameters through a function g: $\theta = g(\theta_1, \ldots, \theta_m)$. As we have throughout the chapter, for ease of presentation we assume a linear benchmarking constraint, $g(\theta_1, \ldots, \theta_m) = \sum_{j=1}^{m} c_j\theta_j$.

In order to incorporate the benchmarking constraint into the model, we make use of the

non-singular transformation

$$\begin{aligned} Y_1 &= \theta_1 \\ &\vdots \\ Y_{m-1} &= \theta_{m-1} \\ Y_m &= g(\theta_1, \ldots, \theta_m) = \sum_{j=1}^{m} c_j \theta_j = \theta. \end{aligned}$$

The Jacobian of this transformation is constant, so the posterior distribution of the transformed variables is

$$\tilde{\pi}(Y_1, \ldots, Y_m) \propto \pi_h\left(Y_m\right) \pi_l\left(Y_1, \ldots, Y_{m-1}, \frac{1}{c_m}\left(Y_m - \sum_{j=1}^{m-1} c_j Y_j\right)\right).$$

Using MCMC methods, we can sample from the posterior distribution $\tilde{\pi}$ for $Y_1, \ldots, Y_m$. In turn, we construct samples for benchmarked parameters $\boldsymbol{\theta}^* = (\theta^*, \theta_1^*, \ldots, \theta_m^*)$ via the inverse transformation:

$$\theta^* = Y_m,\ \theta_1^* = Y_1,\ \ldots, \theta_{m-1}^* = Y_{m-1},\ \theta_m^* = \frac{1}{c_m}\left(Y_m - \sum_{j=1}^{m-1} c_j Y_j\right).$$

Each draw from the joint posterior distribution of $\boldsymbol{\theta}^*$ will satisfy the benchmarking constraint. Accordingly, θ^* will be fully consistent with $(\theta_1^*, \ldots, \theta_m^*)$, in the sense that all characteristics (including all moments) of their respective posterior distributions will be compatible with the benchmarking constraint. Point estimates, standard errors, and posterior intervals for $\boldsymbol{\theta}$ can be constructed based on the draws from the posterior distribution of $\boldsymbol{\theta}^*$. We will refer to this approach as Full Conditional (FULL).

As an illustration of the Full Conditional procedure, suppose that the higher level model is $\hat{\theta} \sim N(\theta, D)$, and the lower level model is the Fay-Herriot model (2.1) with a prior distribution $\pi\left(\boldsymbol{\beta}, A\right) \propto A^{-a-1} e^{-b/A}$, where a, b are constants. As before, we are assume the survey design-based variances are known.

Enforcing the linear constraint $\theta = \sum_{j=1}^{m} c_j \theta_j$ with transformation as described above

results in the modified joint posterior distribution

$$\tilde{\pi}\left(Y_1,\ldots,Y_m,\boldsymbol{\beta},A\right) \propto A^{-m/2-a-1}\exp\Bigg\{ -\frac{b}{A}-\sum_{j=1}^{m-1}\frac{1}{2D_j}\left(\hat{\theta}_j-Y_j\right)^2 \\ -\frac{1}{2D_m}\left(\hat{\theta}_m-\frac{Y_m-\sum_{j=1}^{m-1}c_jY_j}{c_m}\right)^2-\sum_{j=1}^{m-1}\frac{1}{2A}\left(\mathbf{x}_j\boldsymbol{\beta}-Y_j\right)^2 \\ -\frac{1}{2A}\left(\mathbf{x}_m\boldsymbol{\beta}-\frac{Y_m-\sum_{j=1}^{m-1}c_jY_j}{c_m}\right)^2-\frac{1}{2D}\left(\hat{\theta}-Y_m\right)^2\Bigg\}.$$

This modified posterior distribution does not belong to a standard family. It does, however, admit simple full conditional distributions, which are given in terms of the benchmarked elements of $\boldsymbol{\theta}^*$, inverse transformed from $(Y_1,\ldots,Y_m)$:

$$\theta^* \mid \bullet \sim N(\mu_m, V_m)$$
$$\mu_m = V_m\left(\frac{c_m\hat{\theta}_m+\sum_{j=1}^{m-1}c_j\theta_j^*}{c_m^2D_m}+\frac{c_m\mathbf{x_m^T}\beta+\sum_{j=1}^{m-1}c_j\theta_j^*}{c_m^2A}+\frac{\hat{\theta}}{D}\right)$$
$$V_m = \left(\frac{1}{c_m^2D_m}+\frac{1}{c_m^2A}+\frac{1}{D}\right)^{-1}$$

$$\theta_i^* \mid \bullet \sim N(\mu_i, V_i), i=1,...,m-1$$
$$\mu_i = V_i\left(\frac{\hat{\theta}_i}{D_i}+\frac{\mathbf{x_i^T}\beta}{A}+\frac{\theta^*-c_m\hat{\theta}_m-\sum_{j=1,j\neq i}^{m-1}c_j\theta_j^*}{\frac{c_m^2D_m}{c_i}}+\frac{\theta^*-c_m\mathbf{x_m^T}\beta-\sum_{j=1,j\neq i}^{m-1}c_j\theta_j^*}{\frac{c_m^2A}{c_i}}\right)$$
$$V_i = \left(\frac{1}{D_i}+\frac{1}{A}+\frac{c_i^2}{c_m^2D_m}+\frac{c_i^2}{c_m^2A}\right)^{-1}$$

$$\theta_m^* = \frac{\theta^*-\sum_{j=1}^{m-1}c_j\theta_j^*}{c_m}$$

$$\beta \mid \bullet \sim N(\mu_\beta, V_\beta)$$
$$\mu_\beta = \left(\sum_{j=1}^{m}\boldsymbol{x_j^T}\boldsymbol{x_j}\right)^{-1}\left(\sum_{j=1}^{m-1}\boldsymbol{x_j}^T\theta_j^*+\boldsymbol{x_m^T}\frac{\theta^*-\sum_{j=1}^{m-1}c_j\theta_j^*}{c_m}\right)$$
$$V_\beta = A\left(\sum_{j=1}^{m}\boldsymbol{x_j^T}\boldsymbol{x_j}\right)^{-1}$$

$$A \mid \bullet \sim IG\left(\frac{m}{2}+a, b+\frac{1}{2}\sum_{j=1}^{m-1}\left(\theta_j^*-\mathbf{x_j^T}\beta\right)^2+\frac{1}{2}\left(\mathbf{x_m^T}\beta-\frac{\theta^*-\sum_{j=1}^{m-1}c_j\theta_j^*}{c_m}\right)^2\right)$$

Using these conditional distributions, a Gibbs sampler can easily be implemented to draw samples from the benchmarked posterior distribution. Here, the bullet notation indicates conditioning on all data, parameters, and hyperparameters.

There is some intuition to be gained from examining these conditional distributions. Notably, the mean of the conditional distribution of each θ_j^* is the weighted average of the direct estimate, the regression estimate, and the direct estimate and regression estimates according to the benchmarking constraint. The conditional mean incorporates information from both the higher and lower level models through the benchmarking constraint. Likewise the estimate for θ^* is a linear combination of estimates for the parameter, and is also a compromise between the higher and lower level models. As a result the Full Conditional approach can be considered a flexible benchmarking procedure.

There are several additional features of this approach that make it attractive for practitioners. It is generally applicable to any Bayesian small area model for each geographic level and requires no distributional assumptions, there is no additional difficulty of implementing the MCMC routine, the procedure generalizes to multiple geographic levels, more general benchmarking constraints can be enforced, and every draw from the posterior satisfies the benchmarking constraint so that the procedure is fully consistent and the standard errors of the point estimates are benchmarked.

The Full Conditional method can also be used in conjunction with a fixed benchmarking constraint. In this case, the Full Conditional becomes a two-step procedure. Under the fixed benchmarking constraint the higher level posterior distribution remains unchanged. Thus, in the first step, we sample θ^* from π_h.

The second step is to construct a posterior sample from the lower level distribution, given the constraint on the lower level parameters according to the value of θ^* sampled in the first step. Using the same transformation $Y_1, \ldots, Y_m$ as defined above, we must compute the

distribution of $Y_1, \ldots, Y_{m-1}$ conditional on the value $Y_m = \theta^*$,

$$\tilde{\pi}_{cond}(Y_1, \ldots, Y_{m-1} | Y_m = \theta^*) \propto \pi_l \left(Y_1, \ldots, Y_{m-1}, \frac{1}{c_m} \left(\theta^* - \sum_{j=1}^{m-1} c_j Y_j \right) \right).$$

MCMC samples for $\theta_1^* = Y_1, \ldots, \theta_{m-1}^* = Y_{m-1}, \theta_m^* = \frac{\theta^* - \sum_{j=1}^{m-1} c_j \theta_j^*}{c_m}$ can be constructed on the basis of MCMC samples for $Y_1, \ldots, Y_{m-1}$ from $\tilde{\pi}_{cond}$.

These two steps are iterated to produce posterior samples from the benchmarked distributions for the elements of $\boldsymbol{\theta}$, which can be used to produce point estimates, standard errors, and posterior intervals. Of note, it can be shown that if both π_h and π_l are Normal distributions and a fixed benchmarking constraint is used, the MDI and FULL approaches are equivalent.

As we have emphasized throughout the chapter, the flexible benchmarking constraint allows benchmarked higher and lower level parameter estimates to be a compromise between the higher and lower level models. Intuitively we would expect the flexible constraint to outperform the fixed constraint due to the compromise and information exchange between the two levels of the model. However, there are several features of the FULL procedure that suggest that this may not be the case. Implementing FULL with flexible constraint results in different conditional posterior distributions depending on the choice of which small area is labeled θ_m, suggesting the role of the flexible constraint may also depend on this choice. On the other hand, FULL with fixed constraint does not depend on the labeling of the areas. Moreover, while a traditional fixed benchmarking procedure fails to account for the uncertainty of the higher level model, the FULL procedure samples sequentially, first from the higher level model, and then the lower level model conditional on the higher level draw. The uncertainty of the higher level model is incorporated into this procedure, which may place the fixed constraint on par with the flexible constraint in the context of the FULL procedure. The simulation results in Section 2.4 will shed more light on the performance of the procedures and the role of the flexible constraint for both MDI and FULL.

2.4 Simulation Study

2.4.1 Setup

In order to assess the performance of the benchmarking procedures we have described in Sections 2.2 and 2.3, we will conduct a simulation study based on the Fay-Herriot model (2.1). Population parameters are chosen to reflect a national survey with state estimates of interest. Table 2.1 shows the approximate populations for the $m = 50$ states, as of 2000, in addition to the assumed survey sample sizes within each state. The sample sizes (n_i) are proportional to the square roots of the state populations. The general setup of this simulation borrows heavily from Wang et al. (2008). We have chosen to use a similar setup for the purposes of replication, with the goal of extending their results by including more benchmarking procedures in the simulation. Moreover, because the simulation setup reflects the actual demographics of the United States, it is a realistic simulation scenario for empirical

Table 2.1: Population and sample sizes for simulation study

State	Pop. (10,000s)	Sample size	State	Pop. (10,000s)	Sample size	State	Pop. (10,000s)	Sample size
50	3364	58	33	529	23	16	196	14
49	2116	46	32	529	23	15	169	13
48	1936	44	31	529	23	14	169	13
47	1600	40	30	484	22	13	169	13
46	1225	35	29	441	21	12	121	11
45	1225	35	28	441	21	11	121	11
44	1156	34	27	441	21	10	121	11
43	1024	32	26	400	20	9	121	11
42	841	29	25	400	20	8	100	10
41	841	29	24	361	19	7	81	9
40	784	28	23	324	18	6	81	9
39	729	27	22	324	18	5	81	9
38	625	25	21	289	17	4	64	8
37	625	25	20	289	17	3	64	8
36	576	24	19	256	16	2	64	8
35	576	24	18	256	16	1	49	7
34	576	24	17	225	15			

application. Results are consistent for other simulations from the same model with different hypothetical sampling designs.

Recall the Fay-Herriot model (2.1), which we assume is the true model for the state level estimates:

$$\hat{\theta}_i|\theta_i, D_i \overset{ind.}{\sim} N(\theta_i, D_i),$$
$$\theta_i|\mathbf{x}_i, \boldsymbol{\beta}, A \overset{ind.}{\sim} N(\mathbf{x}_i^T\boldsymbol{\beta}, A),$$

with true values $\mathbf{x_i^T} = (1, z_i)$, $\boldsymbol{\beta^T} = (6, 3)$, $z_i = \text{Pop}_i^{0.2} - 1/50\sum_{j=1}^{50}\text{Pop}_j^{0.2}$, $A = 1$, and $D_i = \frac{16}{n_i}$, for $i = 1, \ldots, 50$. A total of 50,000 simulated samples for $(\theta_1, \ldots, \theta_{50})$ and $(\hat{\theta}_1, \ldots, \hat{\theta}_{50})$ are generated from this model. For each sample, the national level total is $\hat{\theta} = \sum_{j=1}^{50} c_j\hat{\theta}_j$, where $c_i = \frac{\text{Pop}_i}{\sum_j \text{Pop}_j}$. Likewise, $\theta = \sum_{j=1}^{50} c_j\theta_j$. This formulation implies that the survey design-based variance for the national total is $D = \sum_{j=1}^{50}\frac{16c_j^2}{n_j}$.

We assume the survey variances D_i are known, and assign a prior $f(\boldsymbol{\beta}, A) \propto A^{-a-1}e^{-b/A}$. By choosing $a = -1$ and $b = 0$, we have a non-informative prior which we expect to provide very similar results to the frequentist approach. Based on the simulated data, $(\hat{\theta}_1, \ldots, \hat{\theta}_{50})$, we estimate the parameters of the model: $(\theta_1, \ldots, \theta_{50})$, $\boldsymbol{\beta^T}$, and A, using MCMC methods. We do not model the national total, but assume that $\hat{\theta} \sim N(\theta, D)$, according to the survey design. After fitting the model, we benchmark the state estimates to the national total with benchmarking constraint $\theta^* = \sum_{j=1}^{50} c_j\theta_j^*$. We then compare these benchmarked estimates $\theta^*, (\theta_1^*, \ldots, \theta_{50}^*)$ to their true values $\theta, (\theta_1, \ldots, \theta_{50})$.

We evaluate the performance of a number of different benchmarking procedures: unbenchmarked and unadjusted small area estimates estimates, $(\tilde{\theta}_1, \ldots, \tilde{\theta}_{50})$, which are included as a point of comparison (UN), raking/ratio (RATIO), Pfefferman-Bernard (PB), Isaki-Tsay-Fuller (ITF), Battese-Harter-Fuller (BHF), You-Rao (YR), Full Conditional with flexible constraint (FULL), Full Conditional with fixed constraint (FULL*), and minimum discrimination with flexible constraint (MDI). Point estimates for MDI with fixed constraint (MDI*) are equivalent to PB.

We are interested in several different performance metrics for these procedures. Primarily, we want to assess the accuracy of the estimators at the state level. We also examine the accuracy of the national estimate and the associated impact of the flexible benchmarking constraint for MDI and FULL. Additionally, we assess the coverage rates of intervals and standard errors based on benchmarked second moments for the estimators MDI, MDI*, FULL, and FULL*.

Due to computational time, the FULL and FULL* procedures were only conducted for the first 20,000 simulations. All other methods were conducted for the full 50,000 simulations.

2.4.2 Results

For the state level of the model we consider the weighted squared error criterion,

$$\frac{1}{50}\sum_{j=1}^{50}\phi_j(\theta_j^* - \theta_j)^2,$$

where $\phi_i = \frac{A+D_i}{AD_i}$. ϕ_i^{-1} is the variance of $\hat{\theta}_i$ with known A. Scaling by the variance of the predictor is necessary to compare across states which are of different sizes. For the national level of the model, we evaluate the un-weighted squared error criterion, $(\theta^* - \theta)^2$. To aggregate over the simulations, we compute the MSE, which is simply the average of these criterion for the 50,000 simulations.

Table 2.2 shows the results for the national level of the model. For the benchmarking procedures with fixed constraint, the estimator of θ is the "Survey" design-based estimator, $\hat{\theta}$, with known variance D. All of the traditional methods, MDI*, and FULL* use a fixed constraint; whereas, MDI and FULL use a flexible benchmarking constraint.

The top panel of Table 2.2 shows the MSE of the three different predictors used to estimate the national total. The FULL estimator has minimal MSE, but MDI also outperforms the Survey estimator. While the differences in MSE are quite small, from a practical standpoint these are not unimportant. Benchmarking adjustments themselves are typically quite small, and are typically even smaller when the model fits the data well. Consequently,

Table 2.2: MSE and coverage rates of nominal 95% intervals based on benchmarked second moments (top panel), and p-values from pairwise tests of differences in MSE for national predictors (bottom panel).

	Survey	MDI	FULL
MSE	0.018560	0.018540	0.018448
Coverage	0.9489	0.9486	0.8344
MDI	**0.00**	-	0.14
FULL	0.33	0.14	-

potential gains from implementing different benchmarking procedures are also quite small, especially when all procedures under consideration are expected to perform reasonably well. In these simulations, the unadjusted estimates of the national and state estimates differ by 1% on average. The MDI procedure outperforms the unadjusted estimate of the national total by 0.1% of MSE, and FULL outperfoms the unadjusted estimate by 0.6%. Relative to the size of the adjustment required, the gains in MSE are reasonably large.

The bottom panel of Table 2.2 shows p-values for pairwise tests for differences in MSE for the two proposed procedures. In particular, we note that the MDI procedure statistically outperforms the Survey estimate. There are no significant differences between FULL and either MDI or Survey. Based on the MSE and these pairwise tests, we would conclude that the flexible benchmarking constraint implemented via the MDI procedure actually allows us to improve the performance of our benchmarked estimates of the national total, relative to the Survey design-based estimate. Moreover, the anecdotal evidence according to the MSE of the FULL procedure suggests it may be the strongest estimator, though this result is not statistically significant.

Throughout the chapter, we have emphasized the ability of the proposed procedures to benchmark both first and second moments. The top panel of Table 2.2 also gives the coverage rates of nominal 95% intervals based on benchmarked second moments for FULL and MDI. Coverage based on the Survey design are shown for comparison. Both the Survey and MDI procedures have coverage rates consistent with the 95% nominal rate. However, the FULL method covers the true national total only 83.44% of the time. The second moment

Table 2.3: Weighted MSE and coverage rates of nominal 95% intervals based on benchmarked second moments for applicable estimators (top panel), and p-values from pairwise tests of differences in weighted MSE for proposed state predictors and other predictors (bottom panel).

	UN	MDI	FULL	FULL*	RATIO	MDI*/PB	BHF	ITF	YR
MSE	1.06835	1.06834	1.06628	1.06478	1.06837	1.06836	1.06837	1.06839	1.07778
Coverage	0.9478	0.9470	0.9463	0.9536	-	0.9470	-	-	-
MDI	**0.04**	-	0.31	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
MDI* / PB	0.37	**0.00**	0.40	**0.00**	**0.04**	-	0.37	**0.00**	**0.00**
FULL	0.31	0.24	-	**0.00**	0.46	0.40	0.43	0.55	**0.00**
FULL*	**0.00**	**0.00**	**0.00**	-	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

benchmarking of the FULL procedure does not produce a reliable standard error for the national level estimate. Even though the FULL procedure has poor coverage rate at the national level of the model, we will show that coverage rates are acceptable at the state level of the model.

Overall, the simulation results for the national total indicate the superior performance of the MDI estimator, with statistically significantly lower MSE than the Survey design and much more accurate confidence interval coverage rates than FULL.

The results for the state level of the model are shown in Table 2.3. We emphasize again that even seemingly very small differences in the MSE can be important in practice. As seen in the top panel of Table 2.3, the FULL* and FULL procedures have the lowest weighted MSE, followed in order by MDI, UN, MDI*/PB, BHF, RATIO, ITF, and YR. The You-Rao procedure performs much worse than any of the other predictors, likely because it has been heavily adapted to apply to the area level models in this setting.

The bottom panel of Table 2.3 shows pairwise p-values for tests of significant differences in MSE for each of the four proposed procedures compared to each of the other procedures under consideration. FULL* significantly outperforms all other estimators at the state level. While the FULL method has the second lowest value for MSE, there is not statistical evidence to suggest that this procedure outperforms any other, with the exception of YR. The MDI approach, is statistically indistinguishable from FULL, and statistically worse than FULL*, but otherwise has significantly lower MSE than the unadjusted estimates and all of the

traditional procedures. These results are striking because all of the traditional procedures have higher MSE than the unadjusted small area estimates, where MDI and FULL* are significant improvements over the unadjusted small area estimates.

The choice of weight for the MSE criterion could impact the relative rankings of some of the predictors. The results discussed in Section 2.2 suggest that for the choice of ϕ_i presented here, we would expect the BHF method to have minimal MSE among the predictors with fixed constraint. In the simulation, amongst the traditional estimators, BHF is actually second best, behind the MDI*/PB procedure, though the difference was not statistically significant. As a sensitivity check, the performance of the benchmarking procedures was assessed for different choices of ϕ_i. Although the rankings for traditional procedures sometimes changes in accordance with the optimization rules described in Section 2.2, the comparisons with MDI and FULL are consistent with the results as presented above.

In addition to the benchmarked point estimates, we also evaluate the benchmarked second moments of the estimators FULL, FULL*, MDI, and MDI*. As the top panel of Table 2.3 indicates, the nominal 95% coverage rates of the benchmarked second moment estimators are all near 95%. We have included the coverage rates for the unadjusted estimates as a point of comparison, though neither the first or second moments are benchmarked. To further explore the benchmarking of second moments, Figure 2.1 plots the percentage change in the standard error estimate of each state against the average state estimate. Changes in standard error are computed relative to the unadjusted standard error.

For the MDI, MDI*, and FULL* procedures, the benchmarked second moment of the national level estimate exceeds that of the unadjusted state level estimates. Accordingly, the benchmarked standard errors of the state level estimates generally must increase relative to the unadjusted standard errors. For the FULL procedure, the opposite is true. The pattern of adjustments for MDI, MDI*, and FULL* are monotonic, with adjustments to the standard errors being larger for larger states. For the FULL* procedure, the pattern of standard error adjustments in non-monotonic. Even though the procedure increases the uncertainty of the

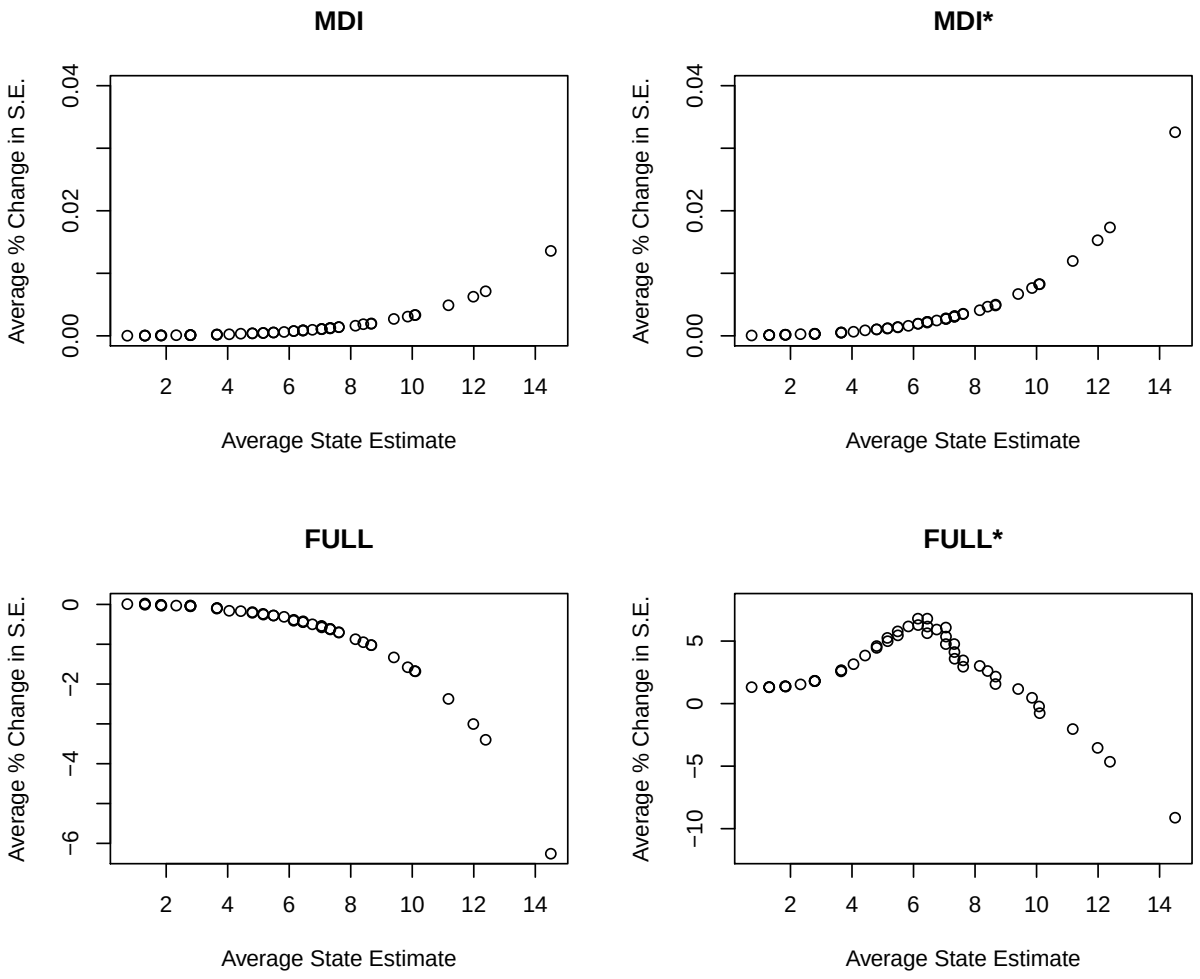

Figure 2.1: Illustrating adjustments for benchmarked standard errors. Plots of the average percent change in standard error from the unadjusted standard error.

state level of the model in total, individual larger states actually show declines in standard errors. MDI and MDI* show very small adjustments to standard errors, with a maximum change of 0.04% relative to unadjusted standard errors; whereas, FULL and FULL* have some adjustments exceeding 5%.

Overall, the simulation results suggest that the proposed procedures perform well. MDI is a significant improvement over the traditional estimators at both the national and state levels according to the MSE criterion. The FULL model is statistically indistinguishable from any other procedure, with the exception of YR. However, FULL* statistically outperforms all

other procedures at the state level, including MDI. On the other hand, because FULL* uses a fixed benchmarking constraint, MDI produces a significantly better estimate at the national level. The simulations indicate that MDI and FULL* have the best performances, and there is a trade off between the two in terms of optimizing performance at either the national level or the state level. Importantly, both MDI and FULL* significantly outperform the unadjusted estimates at the state level, while traditional procedures perform slightly worse than the unadjusted estimates. In other words, benchmarking estimates using the principled methods described in Section 2.3 significantly improves the state level estimates, and in the case of MDI with flexible constraint, the national estimate as well. Throughout, we have emphasized the flexible constraint and the importance of benchmarking both first and second moments. Table 2.2 provides evidence that the flexible benchmarking constraint improves performance at the national level by sharing information between the higher and lower level models. Though FULL* has a fixed constraint, it also incorporates higher level uncertainty into the benchmarking procedure, producing superior results at the state level. Coverage rates for the MDI and FULL* procedures are consistent with the nominal rate, suggesting the second moment estimates are adequate. A drawback of the FULL procedure is that the coverage rate at the national level is much lower than the nominal rate, though coverage rates at the state level satisfy the 95% nominal rate.

2.4.3 Misspecification

All of the simulation results presented thus far have ignored the issue of model misspecification. True values were drawn from the same model used for estimation. In practice, the true model is never known. Practitioners may be hesitant to employ a flexible benchmark constraint out of concern for model misspecification. Benchmarking model-based state estimators to a reliable (and fixed) survey-based national estimate is often more appealing than allowing the national estimate to be influenced by the model-based estimates. In this section we address the issue of misspecification.

First, we make the argument that model misspecification is not the primary consideration for the specific goal of evaluating the performance of benchmarking procedures. While omitted variables, distributional assumptions, and other related issues will always result in some minor degree of model misspecification, post-estimation diagnostics (e.g. posterior predictive checks, residual plots, etc.) protect against misspecification. A responsible analyst will not conduct inference based on a severely misspecified model. If misspecification is a major concern, then the choice of benchmarking procedure will not be a crucial element of the estimation procedure because benchmarking adjustments are small, but the impact of a misspecified model could be quite substantial. While this section presents results from a misspecified model for purposes of performance evaluation, we believe the simulation results in the preceding section are more reflective of benchmarking decisions in practice.

With that caveat, we consider 10,000 additional simulations from the true model described in Section 2.4.1. Now we introduce model misspecification by withholding the covariate z_i, and we estimate the model with only an intercept. Intuitively, this model will shrink all estimators towards the overall area average, even though, in truth, larger areas tend to have larger estimates. Obviously in a real application we would be able to identify this type of model misspecification using a simple residual plot. This model is chosen strictly for illustration.

Table 2.4 replicates the top panel of Table 2.3 for the misspecified model. The pairwise test of p-values is omitted because essentially all differences in this table are statistically significant.

There are a few major points to consider. First, the degree of misspecification is huge. The MSE from the correctly specified and misspecified model is 1.06835 and 1.84753, respec-

Table 2.4: For the misspecified model: Weighted MSE and coverage rates of nominal 95% intervals based on benchmarked second moments for applicable estimators.

	UN	MDI	FULL	FULL*	RATIO	MDI*	BHF	ITF	YR
MSE	1.84753	1.84427	1.84427	1.84463	1.87027	1.84290	1.84278	1.83420	1.95587
Coverage	0.9487	0.9497	0.9480	0.9538		0.9500			

tively. Comparatively, the MSE for the direct survey estimates is 1.98093. If we did not have to estimate the parameter A, the MSE would be 1. This suggests that the correct model is about twice as efficient as the survey estimates. There is a 7% loss of efficiency for estimating A and an additional 72% loss of efficiency due to misspecification. While the MDI, FULL, and FULL* procedures are significantly under performing relative to the traditional benchmarking procedures at the state level, the differences in MSE are not large in magnitude given the degree of misspecification. For more reasonable amounts of misspecification, we would expect the proposed procedures to perform equally as well.

Intuitively, when the model is correctly specified, we can add information through benchmarking by simultaneously adjusting the national and state level estimates because both estimates are reliable to the extent that we know the uncertainty associated with each estimate. When the model is misspecified, the national estimate, which is not model-based, continues to be reliable. The model-based state estimates become biased and their standard errors increase. In such cases, it is preferable to adjust the state estimates and hold the national estimate fixed. The proposed procedures are principled, in the sense that as the state estimates' uncertainty increases with misspecification, the benchmarked national estimate will be nearer to the survey design-estimate. Unfortunately, in this scenario the degree of misspecification biases state estimates to the extent that the procedures cannot overcome these biases. For more moderate misspecification we expect the new methods will still perform well.

Coverage rates of our proposed procedures continue to be consistent with the nominal rate, despite misspecification.

2.5 Conclusion

We began this chapter by motivating the benchmarking problem from the perspective of Bayesian models applied to ACS data by the SAHIE group at the United States Census

Bureau. The small area models used in practice to produce estimates have become sufficiently complex that MCMC procedures are typically necessary to estimate the parameters of the model. We have proposed two benchmarking procedures for such situations. MDI is a principled, information theoretic approach to solving the benchmarking problem. A major advantage of MDI is that it is simple to implement, relying only on posterior means and covariances. The Full Conditional approach is a self-consistent benchmarking procedure whereby draws from the posterior are conducted via MCMC conditional on the benchmarking constraint being satisfied. While this procedure adds a thin layer of technical complexity when used in practice, it results in no additional computation time and is shown to outperform other procedures at the state level. Both of these procedures result in modified posterior distributions, not simply adjustments of point estimates, which satisfy benchmarking constraints, including both first and second moments. The procedures are generally applicable to any Bayesian small area model, have the ability to implement the flexible benchmarking constraint, and can easily be generalized to more than two geographic levels.

Simulations indicate the superior performance of the MDI and FULL* procedures. These are the only benchmarking procedures that statistically outperform unadjusted estimates, and all other traditional benchmarking procedures, at the state level. MDI also is superior to the Survey design-based estimate at the national level. Some of this improvement for MDI is due to allowing for a flexible benchmarking constraint. For these two procedures, simulations show that the principled imposition of the benchmarking constraint actually improves the performance of estimators, even over the unadjusted estimates. While additional simulations show that the performance of the procedures is sensitive to the degree of misspecification, we argue that if the analyst believes that the model is misspecified (due to posterior predictive checks, for example), the analysts' efforts should focus on modifying the model itself, not the choice of benchmarking procedure.

In the end, no clear statistical recommendation between MDI and FULL* can be given. MDI has the better performance at the national level; FULL* has the better performance

at the state level. MDI is simpler to implement in practice, though the FULL* procedure is not much more difficult than the original MCMC algorithm. Where MDI relies on the approximate Normality assumption to derive closed form solutions, FULL* requires no additional distributional assumptions. The Normality assumption is likely partially responsible for the loss of efficiency of the MDI procedure at the state level, relative to FULL*. Though, both of these procedures are model-based, in the sense that entire posterior distributions are adjusted to achieve constraints. The choice between the two depends primarily on the taste of the practitioner.

An important limitation of the MDI procedure is the requirement of Normality to achieve closed-form solutions. If we relax this assumption, we can still derive the MDI benchmarked distribution using numerical methods. However, in order to determine the moments of the benchmarked distribution, we would typically be required to sample from this distribution using an MCMC algorithm. Only with an infinite number of MCMC iterations will these draws exactly satisfy the benchmarking constraint. Thus, further benchmarking adjustments may be required for an MCMC chain with a finite number iterations. Additionally, misspecification is an important limitation in the context of the proposed procedures, especially FULL which is completely model-based. If the model is misspecified, a model-based benchmarking procedure will be susceptible to the misspecification. An advantage of the FULL* procedure is that it is partially protected against misspecification by using a fixed benchmarking constraint. Lastly, we note that we have presented simulation results only the context of a particular small area model. Our model-based procedures may be more or less effective under different model conditions. Caution in the selection of benchmarking procedures is especially important as models become more complex.

As a theoretical result, we have shown that the MDI principle is closely linked to other benchmarking procedures. Specifically, MDI* is equivalent to PB, and is a Bayes rule with the correct choice of weight in the loss function.

Future work in this area would extend the MDI and Full Conditional methods to three

levels of aggregation. In this chapter we have considered benchmarking state estimates to a national total. In practice, as is the case for SAHIE, we often must benchmark county estimates to states and state estimates to a national total. For three geographic levels, gains due to flexible benchmarking may be even larger. Extensions of the MDI procedure to the multivariate case would also be of importance for practical applications.

Chapter 3

A Time Dynamic Pair Copula Construction

3.1 Introduction

In financial models, understanding the dependence structure of multivariate data is crucially important. The task of modeling such data is made more difficult by the fact that the simplifying assumption of Multivariate Normality is typically inappropriate. A popular method for capturing flexible multivariate dependency is the copula. A copula is a multivariate distribution function with Uniform margins. Any joint distribution can be decomposed into its marginal distributions, which describe the univariate behavior of each variable, and a copula, which describes the dependence between those variables (Sklar 1959).

Copulas have seen application in several areas of financial research. Most famously, Li (2000) used copulas to price Collateralized Debt Obligations (CDOs). This application has since been partially blamed for the market collapse of 2008 (Salmon 2009). Other copula-based descriptions of default correlation include Laurent and Gregory (2005) and Frey and McNeil (2003). There have also been applications of copulas to asset allocation with asymmetric dependency (Patton 2004), the calculation of the Value-at-Risk of a portfolio

(Embrechts et al. 2003), the spread of financial contagion (Rodriguez 2007), and option pricing (van den Goorbergh et al. 2005).

Unfortunately there are only a handful of readily available and tractable multivariate copula functions. Most common are the multivariate Gaussian and t-copulas, which often fail to fully capture the complexities of certain types of data. For example, the Gaussian and t-copulas would be insufficient for modeling data characterized by asymmetric correlation, extreme tail dependence, or heavy tails. As a result, traditional applications of the multivariate copula to financial data are limited. A more recent econometric technology for dealing with complex multivariate dependency is the pair-copula construction (PCC), which takes advantage of the fact that there is a much richer class of bivariate copula functions (e.g. Gaussian, t, Frank, Gumbel, Clayton, Joe, and rotations of each of these). The PCC creates flexible higher dimensional copulas using bivariate copula as building blocks. The multivariate joint distribution is systematically factored into conditional pair-wise components, between which the dependence structure is modeled by an appropriate bivariate copula. Section 3.2 provides technical details.

The PCC was proposed by Joe (1996), formalized and applied to Gaussian copulas by Bedford and Cooke (2001), and extended to a wider class of bivariate pairs by Aas et al. (2009). While the typical multivariate copula has the same dependence structure for any pair of margins, the PCC allows the analyst of multivariate data to model dependence between individual pairs distinctly, even with possibly different types of bivariate copula components for each pair. Accordingly, the PCC offers substantial flexibility in modeling, and typically a better fit to financial data, than traditional multivariate copulas.

There has been a great deal of recent work developing the PCC for application in finance. However, the technique has not gained much traction in empirical analysis. This fact is likely due to two important drawbacks of the current PCC model: limited dimensionality and inability to account for time dynamic dependencies.

In this chapter we propose a sophisticated Bayesian model for the PCC that allows

for time dynamics and addresses dimensionality. We model a transformation of parameters of certain bivariate pair components as a first order autoregressive (AR(1)) time series model. Additionally, we allow for Bayesian identification of conditional pairwise independence. Using the PCC and time dynamic bivariate copula building blocks, our model is a flexible multivariate copula model for capturing time-varying dependence in high dimension. Though there are already many time-varying bivariate copula in the literature, there are few, if any, time-varying multivariate copula in more than two dimensions.

The rest of the chapter is organized as follows. Section 3.2 briefly reviews the basics of the PCC. In Section 3.3 we extend the PCC and introduce our Bayesian model of conditional independence and time varying association, including the results of a simulation illustrating the effectiveness of our estimation procedure. Section 3.4 and Section 3.5 show applications to financial returns data for out-of-sample prediction of equity correlation and Value at Risk (VaR) estimation. Section 3.6 concludes.

3.2 Background

Consider d-dimensional multivariate data $\mathbf{X} = (x_1, ..., x_d)$ with joint distribution function $F(x_1, ..., x_d)$, joint density $f(x_1, ..., x_d)$, marginal distributions $F_1(x_1), ..., F_d(x_d)$, and marginal densities $f_1(x_1), ..., f_d(x_d)$. Let $u_1, ..., u_d$ be independently and identically distributed (i.i.d.) Uniform(0,1). A copula is a multivariate distribution with Uniformly distributed marginals. Sklar (1959) shows that for any F there exists a copula, C, such that

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)),$$

and it is easy to show that this copula is

$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d)),$$

with copula density

$$c(u_1, ..., u_d) = \frac{f(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}.$$

The joint density is

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) f_1(x_1) \cdots f_d(x_d),$$

illustrating the fact that the copula separates the multivariate dependence from the marginal distributions.

For $d = 2$ the joint density can be written simply as

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1) f_2(x_2)$$

with bivariate (pair) copula c_{12}. The conditional density is

$$f_{1|2}(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1),$$

for the same pair copula.

More generally, for a vector $\mathbf{v}$ of dimension k, where $1 < k < d - 1$, with v_j an arbitrary component of $\mathbf{v}$ and $\mathbf{v_{-j}}$ the vector $\mathbf{v}$ with v_j removed, we have

$$f(x|\mathbf{v}) = c_{xv_j|\mathbf{v_{-j}}}(F(x|\mathbf{v_{-j}}), F(v_j|\mathbf{v_{-j}})) f(x|\mathbf{v_{-j}}),$$

where $c_{xv_j|\mathbf{v_{-j}}}$ is a pair copula. Joe (1996) shows, for any x and $\mathbf{v}$,

$$F(x|\mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v_{-j}}}(F(x|\mathbf{v_{-j}}), F(v_j|\mathbf{v_{-j}}))}{\partial F(v_j|\mathbf{v_{-j}})}.$$

Using this general form for an arbitrary conditioning, combined with the fact that the density f can be factorized as the product of conditional distributions:

$$f(x_1, ..., x_d) = f_d(x_d) f_{d-1|d}(x_{d-1}|x_d) f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d) \cdots f_{1|2\cdots d}(x_1|x_2, ..., x_d),$$

it becomes obvious that any multivariate density can be re-constructed as the product of pair copula.

For example with $d = 3$, one possible factorization of the joint density is given by

$$\begin{aligned} f(x_1, x_2, x_3) = & \, c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) \\ & \times c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_2|x_3)) \\ & \times f_1(x_1) f_2(x_2) f_3(x_3), \end{aligned}$$

which is the product of marginals and conditional bivariate copula densities. In general, such a formulation is known as the PCC (or vine). The bivariate copula in the PCC need not necessarily be in the same class. For example, c_{12} could be Gaussian, while c_{23} and $c_{13|2}$ are Frank. For further details on the factorization producing the PCC see Aas et al. (2009).

There are many possible PCC for d-dimensional data. There are different ways to factor the density and different permutations of the labeling of the variables for each factorization. To organize the possible choices of PCC, Bedford and Cooke (2002) introduced a tree representation of the vine structure, known as regular vines. Within the class of regular vines, there are two sub-classes, C- and D-vines which are most commonly used in conjunction with the PCC (Kurowicka and Cooke 2004). Throughout this chapter we will concentrate on D-vines, but our model applies equally well to any regular vine structure. We choose D-vines simply to have a consistent representation that will offer an intuitive interpretation in our models of financial returns. Figure 3.1 illustrates the D-vine construction assuming the margins are Uniform; Berg and Aas (2009) and Aas et al. (2009) include similar figures.

The five dimensional D-vine in Figure 3.1 shows the factorization of the joint density into its conditional copula pairs and, implicitly, the margins. We will often refer to the $d-1$ branches of the vine (or equivalently of the PCC). The branches correspond to the degree of conditioning. Labeled in Figure 3.1 as T_1, T_2, T_3, T_4, the k-th branch of the vine indicates the pair copula is conditioned on $k-1$ other variables. Each branch consists of $d-k$ nodes, which correspond to the conditional bivariate copula of the PCC. For example, with 5 dimensions, the fourth branch has one node, which is conditional on three other variables. Typically in the D-vine structure, the lowest branches, those with the least conditioning variables, have

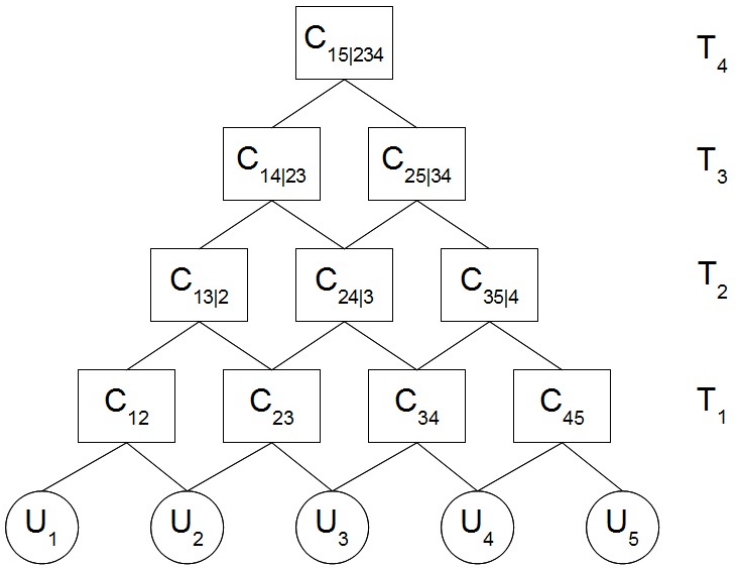

Figure 3.1: D-vine for PCC with Uniform margins in 5 dimensions

the most influence on the dependence structure in the PCC.

The D-vine PCC results in $d(d-1)/2$ pair copula, in addition to d marginal distributions. As the dimensionality of the problem increases, the parameterization due to the pair copula increases quadratically. Even in moderate dimension, it quickly becomes computationally difficult to estimate the parameters of the PCC. Many early examples of the PCC were estimated only for dimension three or four. For financial applications, we are typically interested in a much larger number of variables. In the next section we will develop a model that addresses the curse of dimensionality in the PCC and introduces time dynamics.

3.3 Model

We take a Bayesian approach to estimating and fitting our new PCC model. With the advent of computing power and Markov Chain Monte Carlo (MCMC) methods due to Metropolis et al. (1953) and Hastings (1970), Bayesian models have become extremely common in statistics and econometrics. The popularity of Bayesian methods is largely due to the flexibility in approaching complex problems which are highly parameterized and difficult to estimate using traditional statistical tools. Bayesian inference is often preferable to classical meth-

ods because the model estimation procedure via MCMC results in independent draws from the joint posterior distribution of the parameters. Point estimates and credible intervals are easily constructed from these MCMC draws. In the case of classical maximum likelihood estimates, often we produce estimates of parameters, but quantifying uncertainty of the parameters can be a difficult problem. A fully Bayesian model often requires additional computational effort, but given the flexibility to fit more complex models, the benefit in model performance often far outweighs the cost of computer time. Though Bayesian models are popular in the wider statistical literature, they have had somewhat limited application to copulas. Pitt et al. (2006) and Dalla Valle (2009) use Bayesian inference for traditional multivariate copulas. More recently, Min and Czado (2010) have advocated using Bayesian methods for the PCC, and Min and Czado (2011) and Smith et al. (2010) use a Bayesian approach to introduce model selection to the PCC. We extend these Bayesian models for the PCC to further develop parsimony and to introduce time dynamics.

Before outlining our Bayesian model, it is important to note a few simplifying assumptions that we have made without loss of generality. First, we assume Uniformly and independently distributed margins. The typical strategy for applying the PCC to empirical data is to first transform the marginal distributions to be i.i.d. Uniform. For certain types of data this may be as simple as using the empirical cumulative distribution function (CDF). For other types of data the transformation is more complicated. For multivariate financial time series, practitioners often use an autoregressive moving average-generalized autoregressive conditional heteroskedasticity (ARMA-GARCH) model to account for marginal serial dependence, and then transform the residuals of the marginal models to be Uniform. This approach is implemented by Min and Czado (2010) with data on Euro swap rates. At the model building stage we are not concerned with the marginal distributions, precisely because we can always transform the margins to be Uniform. When we apply our model to financial data we will be careful to specify the correct marginal distributions as a first step in our analysis.

We also make the assumption that all bivariate pair copula are t. Though our model

applies equally well to different choices of pair copula building blocks, we do not wish to address the issue of model selection and goodness-of-fit. It is generally accepted among practitioners of the PCC that the t-copula performs well when dealing with financial data. Dißmann et al. (2013) provide discussion of model selection for the PCC.

Lastly, we restrict our attention to D-vines. The D-vine representation is unique only up to the labeling/ordering of the variables. While the choice of ordering may have important impact, for the purposes of this chapter, we ignore this potential issue of goodness-of-fit. In our applications, we make some effort to link important pairs in the vine, but our results are not greatly impacted by the choices we make.

By ignoring these issues, we are not intending to downplay their importance. Quite the opposite, we emphasize that they can often be crucial for accurately capturing the dependence of multivariate data. Our proposed model improvements do not address any of these points, and for the purposes of this chapter we will largely ignore them.

In Sections 3.3.1, 3.3.2, and 3.3.3 we motivate our final model by separately outlining the intuition and development of the components of the model: basic PCC, PCC with conditional independence, and PCC with time dynamics. In Section 3.3.4 we combine these ideas and methods to produce our proposed time dynamic, parsimonious PCC.

3.3.1 PCC Basics

As a starting point, we describe the basic components of a Bayesian PCC. In Section 3.2 we reviewed some of the mechanics of the factorization of the joint density which produced the PCC. Building from that foundation, Aas et al. (2009) showed for a d-dimensional D-vine PCC with uniform margins

$$f(x_1, ...x_d) = \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j|i+1,...,i+j-1}(F(x_i|x_{i+1}, ..., x_{i+j-1}), F(x_{i+j}|x_{i+1}, ..., x_{i+j-1})),$$

where j identifies the branch of the vine, and i specifies the nodes of the branch. Figure 3.1 is a graphical representation of this formula.

In our formulation, each of the copula in the likelihood is assumed to be a bivariate t-copula. The t-copula has two parameters, typically denoted ν and ρ, the degrees of freedom and association, respectively. Intuitively, the association parameter is similar to a correlation and the degrees of freedom parameter controls the strength of tail dependence. For smaller degrees of freedom, the tail dependence is greater, and the likelihood of jointly extreme outcomes is greater. As $\nu \to \infty$, the t-copula converges to the Gaussian copula, and as $\nu \to \infty$ and $\rho \to 0$ the t-copula converges to the conditional independence copula. For Uniform u_1, u_2, the density of the bivariate t-copula is

$$c(u_1, u_2) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)}{\nu\pi dt(x_1,\nu)dt(x_2,\nu)\sqrt{1-\rho^2}}\left(1+\frac{x_1^2+x_2^2-2\rho x_1 x_2}{\nu(1-\rho^2)}\right)^{-\frac{\nu+1}{2}},$$

where $x_1 = t_\nu^{-1}(u_1), x_2 = t_\nu^{-1}(u_2)$, and dt and t_ν^{-1} are the probability density function (pdf) and quantile function for the univariate Student-t distribution.

Consider a multivariate time series over T time periods in d dimensions, $X_{d\times T}$. The likelihood for our multivariate time series data with Uniform margins is

$$\prod_{t=1}^{T}\prod_{j=1}^{d-1}\prod_{i=1}^{d-j} c_{i,i+j|i+1,\ldots,i+j-1}(F(x_{i;t}|x_{i+1;t},\ldots,x_{i+j-1;t}), F(x_{i+j;t}|x_{i+1;t},\ldots,x_{i+j-1;t}); \nu_{i,i+j}, \rho_{i,i+j}).$$

The likelihood is relatively straight forward, but some complexity is introduced by the necessity of computing the conditional distribution functions, which are functions of the bivariate t-copula parameters. Aas et al. (2009) provide an algorithm for evaluating the D-vine likelihood.

In order to conduct Bayesian inference, we need to place priors on the parameters of the likelihood. For the basic formulation, the typical prior distributions on the parameters of the PCC are non-informative, Uniform priors:

$$\rho_{i,i+j} \sim U(-1,1)$$
$$\nu_{i,i+j} \sim U(1,30),$$

for $j = 1,\ldots,d-1$ and $i = 1,\ldots,d-j$.

We restrict ν to be between 1 and 30 for computational and technical reasons. If the degrees of freedom is less than 1, there are potential numerical instabilities in evaluating quantiles of the bivariate copula distribution. If the degrees of freedom parameter is above 30 and we do not restrict its value a priori, the posterior mode is typically very large and convergence is slow. We can safely restrict ν to be less than 30 because there is very little difference between the t-copula with 30 degrees of freedom and the t-copula with, say, 1000 degrees of freedom. Both can be closely approximated by the Gaussian copula, which is a t with infinite degrees of freedom. The choice of 30 as an upper bound is somewhat arbitrary, though inference is generally not affected by different choices. It is important to accurately estimate ν if it is small. For low degrees of freedom, there is high tail dependency between the variables, which is often of interest in financial models, especially for estimating default correlation.

There are many other possible choices for the priors on the copula parameters. We typically have at least some prior information: associations tend to be lower for the branches of the vine with more conditioning, associations in financial models are typically positive, and the degrees of freedom is usually larger with more conditioning. While a weakly informative prior could be formulated to reflect some of this information, the model is typically not sensitive to the choice of prior, and we defer to the non-informative, Uniform prior that is the standard for Bayesian PCC models (Min and Czado 2010).

3.3.2 Parsimony

There are $d(d-1)/2$ pair copula in the D-vine, which corresponds to $d(d-1)$ parameters if the building blocks of the vine are all bivariate t-copula. As a result, the PCC quickly becomes intractable in even moderate dimensions. Some recent research has focused on reducing the parameterization of the PCC. Brechmann et al. (2012) use likelihood ratio methods to truncate regular vines. In their method, if the copula is k-truncated, all bivariate copula conditional on at least k variables are assumed to be the independence copula. Min and

Czado (2011) implement a Reversible Jump MCMC model selection algorithm to choose between different specifications associated with more parsimonious PCC formulations. A simpler strategy is applied by Smith et al. (2010), introducing an auxiliary indicator variable for conditional copula independence. Our approach will be an extension of the Smith et al. (2010) method with a moderately informative prior that encourages conditional independence and overall parsimony.

Let $I_{i,i+j} = 1$ if copula pair $(i, i+j)$ is conditionally independent and $I_{i,i+j} = 0$ otherwise, for $j = 1, ..., d-1$ and $i = 1, ..., d-j$. In the event that $I_{i,j+i} = 1$, the bivariate t-copula parameters are $\rho_{i,i+j} = 0$ and $\nu_{i,i+j} \to \infty$. For the purpose of computations, and given our upper bound on ν specified by the prior, we let $\nu_{i,i+j} = 30$ given the independence copula. If $I_{i,i+j} = 0$, there are no restrictions on the pair copula parameters, other than those imposed by the prior distribution.

This specification is essentially the same as in Smith et al. (2010). However, we advocate the choice of an informative prior:

$$P(I_{i,i+j} = 1) = \alpha \exp(-\beta(d-j-1)),$$

where α and β are appropriately chosen hyperparameters. The prior is intended to promote parsimony within the PCC, especially at the higher branches of the vine. For example, with $d = 6$, $\alpha = .9$, and $\beta = .1$, the vine has five branches with associated prior probability of pair copula independence at each branch of $0.40, 0.49, 0.60, 0.74, 0.90$, respectively. The prior is conveniently chosen to ensure, with $\beta > 0$, that higher branches of the vine have greater a priori likelihood of conditional independence. α and β are fixed tuning hyperparameters which control the baseline probability of independence, and the rate of change of the probability of independence from branch to branch, respectively.

This choice of prior is intuitive, powerful, and incorporates more a priori information. It is established amongst PCC practitioners that pair copula on higher branches of the vine, with more conditioning variables, are far more likely to demonstrate conditional indepen-

dence. Even if not independent, they are less influential and typically near independent. If one of the major drawbacks to using PCC in empirical financial applications is the curse of dimensionality, we think it best to take aggressive action to reduce the parameterization of the PCC. Our proposed prior does exactly that, by placing a lot of prior weight on independence. In other words, the data must strongly indicate dependence to move away from the prior probability for independence. Accordingly, we achieve a much more parsimonious model with far fewer parameters to estimate. At lower levels of the vine where the conditional independence assumption is more crucial, we are less informative with our choice of prior. We prefer this approach to that of Smith et al. (2010) because it accounts for additional prior information. Our approach is also more flexible than truncation, allowing individual pairs to be conditionally independent, rather than entire branches of the vine.

All together, the component of our model addressing the issue of dimensionality of the parameter space is as follows, for $j = 1, ..., d-1; i = 1, ..., d-j$:

$$\begin{aligned}
&\text{If } I_{i,i+j} = 1,\ c^*_{i,i+j}(u_i, u_{i+j}; \nu_{i,i+j}, \rho_{i,i+j}) = 1 \\
&\text{If } I_{i,i+j} = 0,\ c^*_{i,i+j}(u_i, u_{i+j}; \nu_{i,i+j}, \rho_{i,i+j}) = c_{i,i+j}(u_i, u_{i+j}; \nu_{i,i+j}, \rho_{i,i+j}) \\
&P(I_{i,i+j} = 1) = \alpha \exp(-\beta(d-j-1)),
\end{aligned}$$

where c^* is the bivariate t-copula to be used in the model and α and β are fixed hyperparameters.

3.3.3 Time Dynamics

The biggest contribution of our model is extending the PCC to account for time dynamics. The fact that PCC methods have been static with respect to time is a severe limitation to their empirical application, especially in finance. It is well documented that financial returns exhibit correlations which are not constant (Longin and Solnik 1995). Any reasonable model of financial returns must be able to account for changes over time. A failure to do so is one of the reasons the copula model of Li (2000) helped lead to the financial crisis of

2008 when it was (mis-)applied to pricing CDOs of mortgage securities. In order to be a serious consideration for financial models, the PCC must account for a dynamic dependence structure.

There are a handful of non-copula based multivariate models for volatility that address time-varying correlation. Most notably, Engle (2002) introduces the Dynamic Conditional Correlation (DCC) GARCH model, and Yu and Meyer (2006) develop a stochastic volatility model with time-varying correlation coefficients. Each of these models has drawbacks. First, they fail to account for non-linear dependence structures by typically relying on Gaussian distributions. In financial applications, correlations tend to be asymmetric, with lower tail dependence much greater than upper tail dependence, a feature of the data these models would fail to capture. More importantly, empirical applications for these methods typically are given only in two or three dimensions. Copula methods provide an improvement by allowing for a more flexible dependence structure, and the PCC provides a framework to capture time dynamics in much higher dimensions.

Our approach builds on the stochastic bivariate copula literature, and applies it to the PCC to create a flexible, dynamic, and high dimensional modeling framework for financial time series data. Early efforts for introducing time-varying copula parameters allowed for structural breaks or regime switching. Dias and Embrechts (2004) suggest a test statistic based on a Likelihood Ratio test for the null hypothesis of no structural break at each point in time. If the hypothesis is rejected, we would conclude that a structural break exists and account for this feature of the data in our modeling procedure. Similarly, regime switching copulas allow for a finite number of parameter states. A variable for the indicator of the current state is assumed to follow a Markov chain, and the model estimates the current state at each point in time, in addition to the parameter values associated with each state. Garcia and Tsafack (2008) and Chollete et al. (2008) each use the regime switching framework in empirical applications. Both regime switching and structural breaks are somewhat unsatisfactory because they allow for only a finite number of parameter states. More attractive

methods include the stochastic autoregressive copulas (SCAR) due to Hafner and Manner (2012), observation driven models proposed by Patton (2006), and semiparametric dynamic copulas introduced by Hafner and Reznikova (2010). These methods have been applied only to bivariate copula and are insufficient for data in more than two dimensions. However they do have a significant advantage of allowing the copula parameter to vary freely over time, rather than restricting it to a finite number of states. The SCAR model transforms the copula parameter and then assumes an underlying AR(1) process for the transformed parameter over time, using an importance sampler to estimate. Our proposed procedure is motivated by SCAR, as we will discuss in further detail below. For more details on the observation driven models and the semi-parametric approach, as well as a thorough review of dynamic copula models, see Manner and Reznikova (2012).

All of the methods described above concern a single copula parameter. The t-copula, which we use for our PCC construction, is parameterized by both an association and a degrees of freedom. We assume the degrees of freedom is constant over time, and that the association parameter is time dynamic. In part, this assumption is made to ensure that our model is suitably applicable to other bivariate copula, which have only a single parameter. This assumption is also necessary from a modeling standpoint. Intuitively, the pair copula captures two features of the dependence structure of the data: association and tail dependence. If we were to allow the two t-copula parameters to be dynamic, at each point in time the association and degrees of freedom parameters could be chosen arbitrarily to fit the two elements of dependence perfectly. In essence, we would be overfitting the data. For a t-copula with fixed degrees of freedom, or for a single parameter pair copula family, the single time-varying parameter can no longer exactly capture both features of the dependence structure. Even though the degree of freedom parameter is not time varying, for a t-copula, the tail dependence depends on both ρ and ν, and the tail dependence itself will be time varying. Other bivariate copula have a single parameter, so this issue only arises in the special case of t-copula pairs.

One of the drawbacks of the current dynamic copula literature, beyond the restriction to the bivariate case, is the difficulty of estimation. For the truly dynamic methods (SCAR, SDC, observation driven), computation is time consuming and challenging; take as an example the Efficient Importance Sampler required of the SCAR procedure (Hafner and Manner 2012). The Bayesian approach we take will simplify estimation. Another huge advantage of the Bayesian methodology is that we can produce credible intervals for the time dynamics of our model. Quantifying the uncertainty from any of the current methods would be nearly impossible.

Our model is most similar to the SCAR model. A very natural, common sense approach to time dynamics is to allow the parameter to follow a traditional time series model, and we will likewise proceed in this fashion. Conveniently, introducing time dynamics in our Bayesian model requires simply assigning the appropriate prior to the pair copula parameter of interest. Recall the likelihood for the D-vine PCC with pair t-copula, but allow the association parameter to vary with time:

$$\sum_{t=1}^{T}\sum_{j=1}^{d-1}\sum_{i=1}^{d-j}\log\left(c_{i,i+j|i+1,\ldots,i+j-1}(F(x_{i;t}|x_{i+1;t},\ldots,x_{i+j-1;t}),F(x_{i+j;t}|x_{i+1;t},\ldots,x_{i+j-1;t}))\right),$$

where $c_{i,i+j|i+1,\ldots,i+j-1}$ is parameterized by $\nu_{i,i+j}$ and $\rho_{i,i+j;t}$.

For convenience, here we suppress the $(i, i+j)$ subscript for parameters of the pair copula. The time-varying bivariate t-copula parameter $\rho_t,\ t = 1,\ldots,T$ is restricted to be in $(-1,1)$. The function

$$\Lambda(x) = \log\left(\frac{1+x}{1-x}\right)$$

transforms $x \in (-1,1)$ to $\mathbb{R}$. To introduce time dynamics for the association parameter, we allow the prior on the ρ_t be such that $\Lambda(\rho_t) - \Lambda(\rho_{t-1})$ follows an AR(1) Gaussian process. In particular, for $t = 2,\ldots,T$, let $\Lambda(\rho_t) - \Lambda(\rho_{t-1}) = \lambda_t$, and

$$\lambda_t = \phi\lambda_{t-1} + \epsilon_t;\ \epsilon_t \sim N(0,\sigma^2),$$

where ϕ and σ^2 are hyperparameters. As a technical point, we have the additional prior

specification $\rho_1 \sim U(-1, 1)$. Because we are assuming the degrees of freedom of the t-copula pairs are constant, we use the standard prior, $\nu \sim U(1, 30)$.

The basic intuition of our time dynamic approach is that we cannot hope to capture large changes in the association parameter from one time point to the next. There simply is not enough information in the data to be able to accurately estimate a correlation-type value from a single observation in time. We have to introduce relatively strong prior information. That prior information comes in the form of the time series model, and it insists that the association at the next time point is similar to the association at the current time point. Additionally, if ϕ is positive, increases (decreases) in association tend to be followed by further increases (decreases). Both pieces of information are consistent with our intuition of financial returns. When combined with our credible intervals, our model provides a very accurate portrayal of how the dependence structure of a multivariate data set evolves over time.

In the next section we will put the pieces of our model together and discuss our Bayesian estimation procedure.

3.3.4 Full Model

To this point, we have described three different models. The basic PCC, the PCC with conditional independence indicators, and the PCC with time dynamic parameters. Now we combine these components to produce an estimable and tractable model for empirical, financial applications. After finalizing the model we discuss our Bayesian estimation procedure. As we have been to this point, we continue to assume Uniformly distributed margins without loss of generality.

The log likelihood for our parsimonious, time dynamic association parameter, D-vine PCC is

$$\sum_{t=1}^{T}\sum_{j=1}^{d-1}\sum_{i=1}^{d-j}\log\left(c^*_{i,i+j|i+1,\ldots,i+j-1}(F(x_{i;t}|x_{i+1;t},\ldots,x_{i+j-1;t}),F(x_{i+j;t}|x_{i+1;t},\ldots,x_{i+j-1;t}))\right),$$

where the conditional cumulative distributions are functions of parameters, and for c a bivariate t-copula,

$$c^*_{i,i+j;t}(u_{i;t}, u_{i+j;t}) = 1, \qquad \text{if } I_{i,i+j} = 1$$
$$c^*_{i,i+j;t}(u_{i;t}, u_{i+j;t}) = c_{i,i+j}(u_{i;t}, u_{i+j;t}; \nu_{i,i+j}, \rho_{i,i+j;t}), \qquad \text{if } I_{i,i+j} = 0$$

The prior distributions on the parameters of our model are as follows.

For $j = 1, ..., d-1; i = 1, ..., d-j$:

$$\nu_{i,i+j} \sim U(1, 30).$$

For $j = 1; i = 1, ..., d-1$:

$$I_{i,i+1} = 0.$$

For $j = 2, ..., d-1; i = 1, ..., d-j$:

$$P(I_{i,i+j} = 1) = \alpha \exp(-\beta(d-j-1)).$$

For $j = 1; i = 1, ..., d-1; t > 1$, $\Lambda(x) = \log\left(\frac{1+x}{1-x}\right)$, $\Lambda(\rho_{i,i+1;t}) - \Lambda(\rho_{i,i+1;t-1}) = \lambda_{i,i+1;t}$, and $\epsilon_t \sim N(0, \sigma^2)$ independently:

$$\lambda_{i,i+1;t} = \phi\lambda_{i,i+1;t-1} + \epsilon_t.$$

For $j = 1; i = 1, ..., d-1$:

$$\rho_{i,i+1;1} \sim U(-1, 1).$$

For $j = 2, ..., d-1; i = 1, ..., d-j$:

$$\rho_{i,i+j;t} = \rho_{i,i+j} \sim U(-1, 1).$$

The Bayesian PCC posterior distribution of interest is the product of the likelihood and the densities of the prior distributions. We will use MCMC methods to compute random samples from the posterior distribution, which in turn will allow us to conduct inference and make predictions from the model.

We take a relatively simple approach to combining time dynamics with parsimony. Our model allows time dynamics only at the first branch of the vine. For the pair copula on the first branch, we fix the indicator of conditional independence to be zero. For all other copula, i.e. those conditional on at least one variable, we do not allow time dynamics, but we do allow for conditional independence. There are several reasons we construct the model in this way. By separating the time-varying estimation from conditional independence, we do not require the indicator for independence to have a time component. Also, we reduce computational time by restricting dynamics to the first branch. The model is suitably flexible that we could estimate time dynamics at additional branches of the D-vine, at the cost of computational time and effort. We argue that because the dependence structure is determined primarily at the first level of the vine, understanding changes over time in the parameters of the copula at the first branch is of primary interest. Applications have shown that the copula parameters typically reflect near-independence at other branches of the vine, suggesting the influence of time dynamics beyond the first branch is minimal.

A potential drawback for the model is the difficulty presented by the choice of the hyperparameters α, β, ϕ, and σ^2. Ideally, we would like to assign each a hyperprior distribution and infer the distribution of the hyperparameters as part of the estimation procedure. Unfortunately the data do not provide enough information to produce reliable estimates for the hyperparameters. While the choice of α and β is not critical, the choice of ϕ and σ^2 is potentially important. Roughly speaking, ϕ determines the momentum of the evolution of the association parameter over time (i.e. to what degree increases are followed by further increases), and σ^2 controls how quickly the parameter can change from time point to time point. Estimates of the association parameter are mildly sensitive to our choices for these hyperparameters. We emphasize that we are primarily interested in the overall pattern of the time variation of the parameters, rather than exact inference for any given point in time, and that the impact on out-of-sample forecasts is not large. From this standpoint, we do not view the choice of ϕ and σ^2 as a major drawback for the implementation of our model;

though, they should be chosen carefully and sensitivity analyses may be appropriate.

Overall, the model provides a convenient way to incorporate time dynamics and parsimony into a Bayesian PCC. Allowing time dynamics for the copula pairs which primarily determine the dependence structure is a major improvement over the current technology, and introducing conditional independence allows us to estimate the parameters of the PCC in relatively high dimension.

3.3.5 Estimation Procedure

To estimate the parameters of our model and conduct inference we use an MCMC algorithm: Metropolis-Hastings (Hastings 1970) within a Gibbs sampler (Geman and Geman 1984). At each iteration of our Markov sampling chain, we conditionally update batches of the parameters using the Metropolis-Hastings accept/reject rule. The following steps outline the procedure for each batch of parameters in our model. For simplicity, we suppress the subscript notation for the parameters.

Constant ρ: Among the constant ρ parameters for which the pair-copula is not conditionally independent ($I = 0$), we randomly select a subset, use a truncated (between -1 and 1) Normal jumping rule for each to produce the proposal, and apply the Metropolis-Hastings acceptance rule. For large models there may be as many as several hundred constant ρ parameters. In that case it may be best to update in sub-batches. Min and Czado (2010) illustrate the effectiveness of the simple Bayesian updating procedure for the parameters of a static PCC. Our approach to update only a subset of the parameters at each iteration performs well, considering the large number of iterations in our algorithm.

ν: The updating rule for the ν parameters is identical to that of the constant ρ parameters, except that the Normal jumping rule is truncated between 1 and 30.

Dynamic ρ_t: The dynamic ρ_t parameters are the most difficult to estimate in terms of computational time. At each iteration of the MCMC algorithm, we conditionally update each of the dynamic ρ_t. Because these parameters follow a time series model, the proposal

must be carefully constructed so that changes in the ρ_t over time are consistent with ϕ and σ^2. In order to accomplish this, we propose jumps at the difference of the differences of the transformed associations, $\Lambda(\rho_t)$. This jumping rule allows for small perturbations of the pattern of the evolution of the ρ_t over time which are consistent with the time series model. Additionally we propose a truncated (between -1 and 1) Normal jump for ρ_1 to shift the ρ_t vertically, a move that cannot be accounted for by the time series portion of the model alone. We apply the Metropolis-Hastings acceptance rule to this proposal.

I: We randomly select a subset of the I parameters which are not fixed to be 0 a priori. If the chosen $I = 0$ at the current iteration of the chain, we create the proposal for the bivariate copula pair whereby $I = 1$ and ν and ρ are fixed to be 30 and 0, respectively. If the chosen $I = 1$ at the current iteration of the chain, we create the proposal for that bivariate copula pair whereby $I = 0$ and ν and ρ jump from their values of 30 and 0, respectively, using a truncated Normal rule as described above. After constructing the proposal, we use the Metropolis-Hastings acceptance rule. A minor drawback of this procedure is that there is very little difference between a conditionally independent pair and a conditionally dependent pair with degrees of freedom near 30 and association near 0. This is not a major issue because it will have essentially no impact on inference. However, in an effort to improve computationally efficiency, it is important to be able to infer conditionally independent pairs. This is one of the reasons we may choose to aggressively assign a prior probability which encourages conditional independence, especially for branches of the vine with more conditioning.

Using a sufficient number of MCMC iterations, after discarding an appropriate number of "burn-in" iterations, we can compute posterior means and modes as point estimates for the parameters of interest. We can also create credible intervals. As previously mentioned, a point of emphasis for the Bayesian procedure is the ability to quantify the uncertainty of our estimates, especially for the dynamic parameters.

3.3.6 Simulation Study

We conclude the Modeling section with a simulation study illustrating the performance of our MCMC scheme. For Uniform margins with dimension $d = 9$ and time series of length $T = 250$, we randomly generate the true parameters of our model. We then simulate data from the model and run our MCMC estimation algorithm. Below, we present the results of our simulation, comparing our inferred parameters to the true values.

First, we point out that $d = 9$ is moderate dimensionality, and the model performs equally well in much higher dimensions. We have emphasized that the conditional independence component of our model encourages parsimony and allows us to apply the PCC to high dimensional data. Our empirical applications will use higher dimensional data, but for the purposes of our simulation we have chosen $d = 9$ simply because we can present the results of our simulation more clearly. Of course, as the dimensionality increases, the computational time increases as well.

Inference for the parameters which are constant over time is good. The Bayesian approach has been shown to reliably estimate the values of PCC parameters by Min and Czado (2010) and Smith et al. (2010). Because our sample size for the multivariate data is only $T = 250$, our estimation somewhat lacks in precision for tail dependence. Credible intervals for the degrees of freedom parameters are wide. On the other hand, inference for the association parameter is quite good, even for the relatively small sample size.

More critically, we focus on the simulation results for the conditional independence indicator and the time dynamic association parameters. Using the posterior mean of $P(I = 1) > 0.5$ as a cutoff, the estimation procedure for I correctly identifies 24 out of the 28 indicators. For 2 of these errors, we predict $I = 0$ when in fact $I = 1$. This mistake is not costly in terms of inference. For each of these pair copula, we estimate a low association and a high degrees of freedom, which is essentially identical to the independence copula. A potentially bigger problem is the 2 errors in which we predict $I = 1$ when in truth $I = 0$. Because the sample size is only 250, a few mis-identifications of conditional independence is not surprising. In

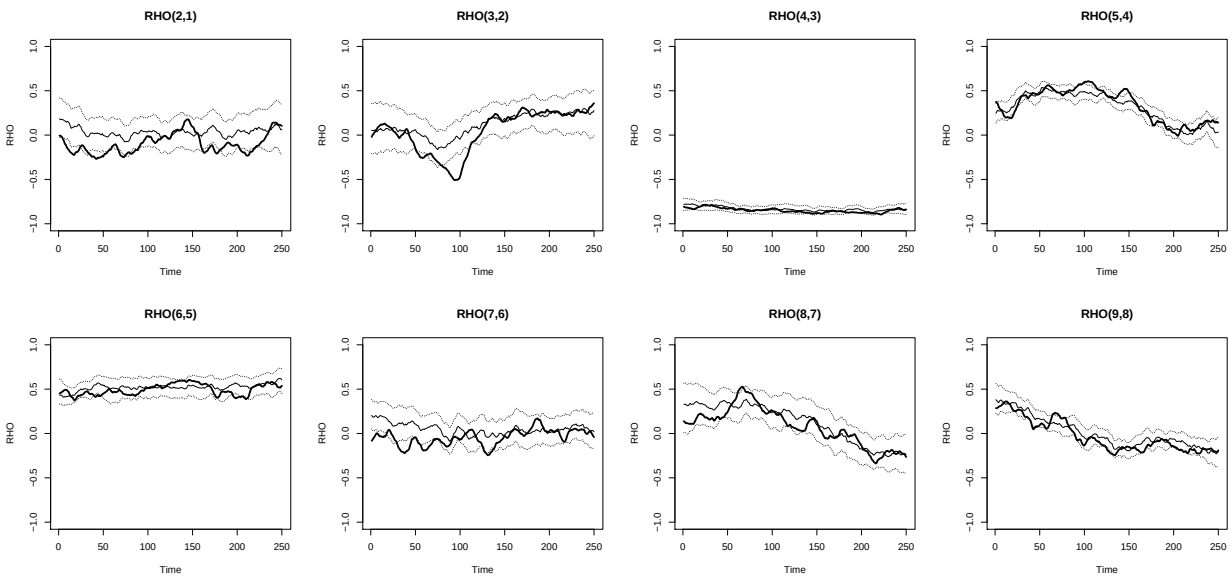

Figure 3.2: Posterior means for dynamic association parameters (solid), true values (dark solid), credible intervals (dotted).

fact, the true parameter values for these two pair copula have relatively low associations and relatively high degrees of freedom. So again, the mistakes are not critical. For the conditional independence there is a trade off between model accuracy and computational time. The more conditional independence the model infers, the faster the MCMC algorithm converges, at the cost of becoming slightly less accurate. Overall, we are satisfied with the performance of the conditional independence portion of the model, and we believe it strikes the appropriate balance between precision and computational effort. In higher dimensions, our simulations have shown that the error rate is similar, but the usefulness of this modeling technique is much more apparent. As dimensionality increases, the conditional independence becomes much more valuable in terms of reducing computational time.

The results of the simulation for the dynamic association are shown in Figure 3.2. The most important feature of the posterior inference is that the posterior means of the association at each time point follow the overall pattern of the true values. This is most apparent in the downward trend of $\rho_{9,8;t}$ and the U-shaped relationship for $\rho_{5,4;t}$. While it is generally true that we cannot estimate large changes in association in short amounts of time (consider

$\rho_{3,2;t}$ between time points 50 and 100), the fact that we can capture the over-arching changes in association over time is a big improvement over the current static time technology. Additionally, the 95% credible intervals provide us a picture of the uncertainty inherent in our model, and they do indeed cover the true association parameter about 95% of the time. Another notable feature of these results is that the accuracy of our estimation increases as time passes. Our model performs better near $T = 250$ than near $T = 1$, with narrower credible intervals. This is likely due to the fact that it takes some time for the time series model to calibrate. The association looks backward in time. As we move forward, our estimates are based on more and more data. This relationship is important because our out-of-sample inference will be based on estimates near the end of the sample. The fact that our model performs especially well at later time points suggests that we should expect improvements in step-ahead predictions, which is the ultimate goal for financial models. Satisfied that our MCMC algorithm for conducting inference for our time dynamic, parsimonious PCC is performing well, we turn to a few empirical applications.

3.4 Application: Out-of-sample Equity Correlation

The basic PCC is valuable because it captures extremely flexible dependence structures in multivariate data. For financial time series, the PCC is a powerful tool for modeling dependence. Naturally, our expectation is that the time dynamic PCC we have outlined thus far will outperform the current PCC technology in capturing the dependence structure of time series data. The improvement is only valuable from a practitioners standpoint, however, if it performs better in out-of-sample prediction. To test the merit of the time dynamic PCC relative to the basic PCC, we compare the out-of-sample prediction for one year ahead, monthly equity correlation for a random sample of stocks from the S&P 100. We choose correlation as the benchmark statistic because it is a natural measure of the dependence structure of a multivariate data set.

To collect data for this empirical application we took a random sample of 16 firms from the S&P 100 (as of January 1, 2011) with at least a 21 year history starting in 1990. The number of firms was arbitrarily chosen to reflect the ability of the model to perform in relatively high dimensions. Table 3.1 lists the companies selected by this procedure. Monthly equity returns from 1990 to 2010 for these 16 names were collected from Yahoo! Finance. In order to assess the out-of-sample predictive properties of the model, we withhold the 2010 observations. The final data set we will use for estimation is a multivariate time series of equity returns for 16 firms and 240 monthly time points (Jan. 1990- Dec. 2009). We will use the remaining 12 observations for out-of-sample comparisons (Jan. 2010 - Dec. 2010).

Table 3.1: The 16 firms randomly selected for the empirical analysis

Symbol	Name	PCC Label
AEP	American Electric Power	7
AXP	American Express	10
BAC	Bank of America	9
CVX	Chevron	8
DELL	Dell	4
INTC	Intel	3
JNJ	Johnson & Johnson	16
K	Coca-Cola	15
ORCL	Oracle	5
S	Sprint	1
SLE	Sara Lee	14
SO	Southern Company	6
T	AT&T	2
TGT	Target	12
USB	US Bancorp	11
WAG	Walgreen's	13

Up to this point, we have assumed the marginal distributions of our data have been Uniformly distributed. For empirical analysis this is clearly not a viable assumption. The typical strategy is to first transform the data so that the marginals are Uniform. In order to forecast on the original scale, the model predictions can then be back-transformed to the desired scale. With financial time series data we must confront the additional issue of time dependence inherent in the data. It has become commonplace in the PCC literature for

financial applications to use an ARMA-GARCH model applied to each of the margins, then transform the standardized residuals to the Uniform distribution using the empirical CDF, and estimate the PCC according to these independent Uniform random variables. The order of the ARMA and GARCH components are chosen to ensure that marginal independence is achieved. While this strategy is probably sufficient for most applications, we take a slightly more advanced approach. Because the choice of the marginal distributions is not of primary concern for our model or our results, we only briefly outline the motivation behind our more sophisticated procedure for transforming the margins to be Uniform. We remark that for the subsequent application of our model to Value at Risk (Section 3.5), the choice of volatility model is potentially more important.

The ARCH model was introduced by Engle (1982) and generalized (GARCH) by Bollerslev (1986) to account for heteroskedasticity in the volatility of time series data. Since then, several stylized facts of financial volatility have been developed in the literature: persistence, mean reversion, clustering, and asymmetry (Engle and Patton 2001). We adopt the Threshold-GARCH model (TGARCH) which is a class of volatility model that accounts for each of these properties. Angelidis et al. (2004) show that the more flexible the GARCH model, the better it performs in terms of volatility forecasting. This model is used frequently in the literature for Value at Risk; see Fantazzini (2008) for example. The TGARCH model contains, as a special case, the GARCH specification.

Keep in mind that the dependence between the margins, which is the focus of this chapter, is captured by the PCC after the appropriate time series model is chosen to transform the margins to Uniform. Our final model for the margins is an AR(1)-TGARCH(1,1). The specification of this model is given by:

$$
\begin{aligned}
y_t &= \mu + \phi_1 y_{t-1} + \tilde{\epsilon}_t \\
\tilde{\epsilon}_t &= \eta_t \sqrt{h_t},\ \eta_t \sim f(0,1) \\
h_t &= \omega + \alpha_0 \tilde{\epsilon}_{t-1}^2 + \gamma \tilde{\epsilon}_{t-1}^2 D_{t-1} + \beta h_{t-1},
\end{aligned}
$$

where y_t is the returns data, the tilde notation for ϵ distinguishes it from the ϵ in the PCC, $D_{t-1} = 1$ if $\tilde{\epsilon}_{t-1} < 0$, and 0 otherwise. The asymmetry enters through D, positive shocks ($\tilde{\epsilon}_t > 0$) and negative shocks ($\tilde{\epsilon}_t < 0$) have different impacts on volatility (γ). The choice of f is left to the analyst, here we use the Student's-t distribution. The standardized residuals from this model are transformed to the Uniform distribution. Ljung-Box tests show that the margins are independent after transformation, and goodness-of-fit tests confirm approximate Uniformity.

After transformation we run the MCMC algorithm for our PCC model with Uniform margins. We point out that the ordering of the names in the PCC construction for the D-vine will have some minimal impact on model inference. We ignore this minor issue of optimal model selection by intuitively linking related firms (e.g. by industry) in the ordering, as indicated in Table 3.1. The results of our model fitting are summarized in Table 3.2 and Figure 3.3.

Before discussing the out-of-sample predictions, we highlight some interesting features of the fitted model. Table 3.2 shows posterior means for the parameters of our PCC which are constant over time, including an indication for the posterior probability of conditional independence of each pair. As the table shows, conditional independence significantly reduces the burden of parameter estimation, especially at the higher levels of the vine. We used a fairly conservative prior for conditional independence, but a more liberal prior would encourage even greater parsimony. We highlight the fact that several of the degrees of freedom parameters are estimated to be less than 10. Low values for degrees of freedom are suggestive of a relatively high degree of tail dependence for these pair copula. This tail dependence would not be captured by using Normal pair copula in the PCC, validating our choice of pair t-copula.

Of greater interest, Figure 3.3 plots the posterior mean at each time point for the dynamic association parameters of the pair copula, and offers convincing evidence that time dynamics exist in empirical applications. While it is true that for a handful of the pairs the association

Table 3.2: Posterior means for each bivariate copula pair in the PCC for the S&P 100 returns data. ν is listed above ρ, and ** indicates conditional independence, i.e. $P(I = 1) > .5$; x indicates dynamic parameter.

(i,i+j)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	19.7 x														
3	24.2 0.11	15.6 x													
4	25.6 0.06	** **	24.6 x												
5	** **	27.4 0.09	15.0 0.33	21.4 x											
6	** **	21.9 0.27	** **	23.8 -0.06	8.7 x										
7	** **	24.7 0.10	** **	** **	** **	16.1 x									
8	** **	** **	** **	** **	25.4 0.04	27.6 0.05	12.9 x								
9	** **	** **	24.0 0.19	** **	15.7 0.20	** **	** **	23.5 x							
10	** **	** **	25.9 0.05	17.9 0.12	23.2 0.08	** **	** **	23.1 0.22	13.9 x						
11	** **	24.9 0.11	** **	** **	** **	28.1 0.05	** **	25.9 0.04	16.0 0.53	8.2 x					
12	** **	** **	** **	25.0 0.08	24.1 0.07	23.0 -0.10	** **	** **	22.8 0.14	8.2 0.20	15.9 x				
13	** **	9.4 0.23	** **	** **	** **	24.2 0.04	** **	** **	26.3 -0.03	** **	26.9 0.08	20.0 x			
14	** **	** **	** **	** **	** **	** **	25.1 0.17	** **	** **	25.7 0.06	23.4 0.15	27.1 0.08	24.0 x		
15	** **	** **	** **	27.1 0.09	** **	** **	24.1 0.11	** **	** **	25.8 0.09	18.8 0.22	27.0 0.08	25.0 0.09	13.1 x	
16	** **	** **	** **	** **	** **	** **	** **	** **	** **	24.92 0.07	26.46 0.03	** **	19.92 0.31	18.17 0.30	11.83 x

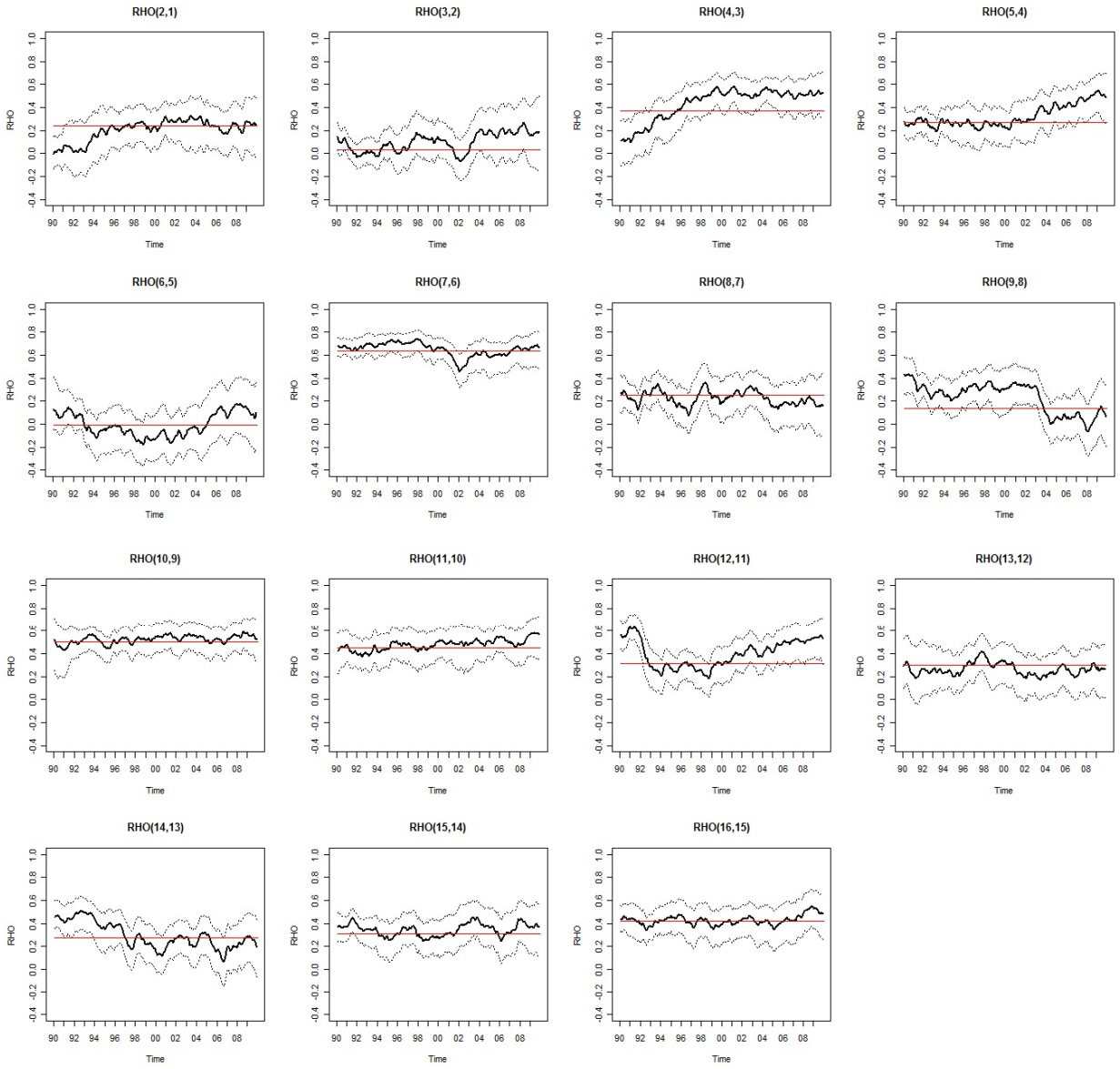

Figure 3.3: Dynamic correlation component of the PCC for 16 randomly chosen S&P 100 stocks. Posterior mean (dark), credible interval (dotted), posterior mean for equivalent model with constant association parameter (solid).

parameter is roughly constant over time, in general associations in equity returns appear to be increasing. This trend is especially visible during the period from 2000 to 2009. The results are consistent with our knowledge that equity correlations have grown over the past two decades. The time dynamics also capture shocks to association at a specific time points for a few of the pairs: 2001 for $\rho_{7,6}$, 2002 for $\rho_{9,8}$, and 1991 for $\rho_{12,11}$.

For comparison, Figure 3.3 also includes the posterior mean of the association from the static time model. Inference for time dynamics, as indicated by the 95% credible intervals, is not very precise. For most time points, the credible interval for dynamic association includes the static association value. This suggests that the dynamic model may provide little improvement over the static model. There are several reasons we would disagree with that assessment. Associations between equity returns have historically not changed drastically over time. Even moderate differences between the dynamic and static results are potentially informative. Additionally, the firms for this application were chosen randomly. In a more focused analysis associations may be larger and potentially more sensitive to major economic shocks. In that case dynamics will show larger differences from the static model. Lastly, if we are most interested in out-of-sample performance, we should focus on the model results at the end of the time period under consideration. Our forecasts will be based only on the last few time points. At the terminal time, predictions from the dynamic model become more precise, and they often differ substantially from the static model. Several of the credible intervals do not contain the static paramater value at the end point ($\rho_{4,3}$, $\rho_{5,4}$, and $\rho_{12,11}$). For out-of-sample prediction, time dynamics will provide a substantial improvement, especially in relation to the firms most directly impacted by the pair copula whose dynamic associations differ from the static estimates at the last time point.

As an additional point of comparison between the static and dynamic models, we consider a few Bayesian model selection criteria. In order to compare two competing models, the most common procedure is the computation of a Bayes factor. The Bayes factor is simply the posterior odds. For data, D, and dynamic and static models, M_{dyn} and M_{stat}, respectively,

the Bayes factor can be written $P(D|M_{dyn})/P(D|M_{stat})$. Raftery (1996) provides a granular scale of interpretation on the two times logarithm scale. For our data and fitted models, we compute a Bayes factor on the twice log scale of 33.2. The Raftery interpretation of the Bayes factor suggests the posterior odds are 'very strongly' in favor of the dynamic model. We also consider a related value, the Deviance Information Criterion (Gelman et al. 2004). We prefer a model with minimal DIC. In this case the dynamic model has a DIC -1164.0 compared to -1137.6 for the static model. Again, convincing evidence for the choice of the dynamic model. We note that the more complex model comes at a cost of an additional 31 effective parameters. According to the DIC, the gain in model fit outweighs the cost of the parameters. Both of these traditional Bayesian model selection methods conclusively favor the dynamic model to the static model, further evidence that time dynamics for the dependence structure are crucial to PCC modeling of financial time series.

Now, we turn to a comparison of the out-of-sample predictions of the dynamic model and the static model. These two PCC are identical except that the association parameters are all constant over time in the static model. In particular, the AR(1)-TGARCH(1,1) margins are the same for each model. We wish to use the models to forecast out-of-sample equity correlation for the 12-month period Jan. 2010 to Dec. 2010. The following steps outline the formulation of the out-of-sample equity correlation forecasts from the dynamic model based on our MCMC output. For each iteration of the chain after burn-in:

1. Based on the current estimates parameter values of the time series model and current estimates of the dynamic association, simulate ahead 12 steps to predict $\hat{\rho}_{i,i+1;T+1}, ..., \hat{\rho}_{i,i+1;T+12}$ for $i = 1, ..., d-1$.

2. Based on the current values of the constant-time model parameters, and the 12-step ahead predictions for dynamic associations from Step 1, simulate from the PCC for each of the 12 months forward. Aas et al. (2009) provides an algorithm for simulating from a D-vine PCC.

3. Transform the simulations from the PCC back to the original scale of the equity returns using the AR(1)-TGARCH(1,1) model to create forecasted returns monthly for 2010.

4. Compute the correlation of the monthly returns for the forecasted values.

Given all of these correlation forecasts, we compute the total Mean Square Error (MSE) by summing the MSE for each pair-wise correlation forecast relative to the observed 2010 monthly pair-wise correlation.

The only difference between the forecasting procedure for the static time model is that we do not have to simulate step-ahead predictions for the time dynamic parameters. Instead, we assume the current estimates from the static MCMC model are the same in the 12-months forward. Note that we have once again taken advantage of our Bayesian specification and the MCMC algorithm to produce estimates that would be much more difficult to obtain with a traditional approach. Though they are not necessary for the comparisons we hope to make, we could easily construct credible intervals for our correlation forecasts, which would provide a measure of predictive uncertainty. For our analysis, the total MSE for the two models is summarized in Table 3.3. Using the time dynamic model results in an approximately 10% reduction in MSE relative to the static model.

Table 3.3: MSE for 12-step ahead equity correlations

Model	MSE	% Diff.
Dynamic	0.151	10.13%
Static	0.168	-

We argue that a 10% reduction in MSE is a substantial improvement over the current PCC technology. First, the static PCC is already an excellent model for financial time series due to its flexible dependence structure. Any improvement at all is noteworthy. Secondly, in financial models better predictions generally correspond with better pricing models, better forecasts, and better strategies, which all lead to bigger profits. This example is not purely illustrative. There are a handful of financial derivatives based on underlying equity correlations (e.g. correlation swaps). From this standpoint, 10% less error could lead to meaningful

monetary gains. Additionally, we might expect bigger improvements for a different set of firms. We may see a modest 10% improvement for randomly chosen stocks across sectors in this particular example, but potentially much bigger gains within sectors.

Though we do not show the output of the model here, the analysis was repeated similarly with daily data for 16 randomly selected financial firms in 2008. The same model resulted in a 13.7% decline in MSE for equity correlation predictions.

Based on the application of the dynamic PCC to this empirical data set, we conclude that our model provides an important improvement over the current PCC technology. We are better able to forecast out-of-sample equity correlation, and the time-dynamic parameters indeed show evidence of changes over time. Both the Bayes factor and DIC validate this conclusion. The dynamic PCC offers a statistically significant improvement to the current static PCC technology.

3.5 Application: Value at Risk

Another important potential application for the dynamic PCC is in VaR. VaR is a widely used assessment of the risk of loss for a financial portfolio. Historically, VaR has been used by banks and regulators to quantify the market risk of the holdings of financial institutions. VaR is defined as the maximal portfolio loss at a given time period, for a pre-specified level of confidence. For detailed background on the history and use of VaR in financial risk management, we refer the reader to Jorion (2000). There are numerous articles considering dynamic modeling for VaR, notably Fantazzini (2008) and Angelidis et al. (2004). We focus on a comparison of VaR between static and dynamic PCC, we defer technical details for the interested reader to those references cited here.

This empirical analysis considers an equally weighted portfolio consisting of one share of three different financial sector stocks: Bank of America, Citigroup, and Wells Fargo. We collect daily data from January 1, 2005 to December 31, 2010, a total of 1510 observations.

These three firms were arbitrarily, but strategically, chosen to be included in the portfolio. We intentionally focus on the financial sector and this particular time period because it will be most affected by the financial crisis of 2008, with the expectation that time dynamics may play an especially important role. We simply chose three large and recognizable companies for the purposes of intuition without conducting any data snooping.

We limit this application to a dimensionality of three because we must fit the dynamic PCC model more than 1000 times, which would be excessively time consuming in high dimension. For a single estimate of VaR, rather than conducting a study over thousands of days, the model is quite capable of creating VaR estimates in higher dimensions.

Our aim is to calculate one-day-ahead 95% VaR forecasts. The data set consists of 1510 time points. For day 501 to day 1510, we use the 500 preceding observations to estimate our PCC and calculate the one-day-ahead VaR. We then compare the VaR to the observed gain/loss of the portfolio on the same day. In the end, we have more than 1000 VaR calculations, for which we know whether the portfolio loss truly exceeded the stated VaR. Ultimately, we will assess the accuracy of the VaR for the static PCC and the dynamic PCC. As in the equity correlation application, the margins are modeled by an AR(1)-TGARCH(1,1), and the standardized residuals are transformed to be Uniform before fitting the PCC.

In order to produce an estimate of the one-day-ahead VaR at each time point we must simulate ahead by one time point from the PCC. The procedure for simulating forward and calculating the VaR is very similar to the algorithm described in the equity correlation application. The steps are as follows for each iteration of the MCMC chain:

1. Based on the current estimates parameter values of the time series model and current estimates of the dynamic association, simulate ahead 1 step to predict $\hat{\rho}_{i,i+1;T+1}$ for $i = 1, ..., d-1$.

2. Based on the current values of the constant-time model parameters and the simulated dynamic parameter values from the previous step, simulate once from the PCC.

3. Transform each of the simulated Uniform margins back to the original scale of the stock prices using the AR(1)-TGARCH(1,1) model to create forecasted stock prices at time $T+1$ for each of the three firms.

4. Sum these forecasts for each of the three stocks and compute the forecasted portfolio loss from the previous time period. A portfolio gain is represented by a negative value for portfolio loss.

Based on these simulated portfolio losses, we compute the 95th percentile of the portfolio loss distribution to create the estimated VaR, which we then compare to the realized portfolio loss using the observed data for the subsequent day. Note that at each time point, we only use available information to produce a one-step ahead VaR forecast, which we then compare to the actual outcome. This entire procedure is repeated for each day 501 to day 1510 based on the previous 500 daily observations.

To assess the differences between our time-constant and time-dynamic models, we take the approach of Angelidis et al. (2004). First we check that each model adequately covers the true portfolio loss 95% of the time, and then we compare the accuracy of the VaR forecasts among adequate models.

For both the dynamic PCC and the static PCC models, the one-step-ahead 95% VaR threshold captures the true portfolio loss 93.9% of the time. This coverage rate is reasonably close to the nominal rate of 95%. We would fail to reject a statistical test comparing this coverage rate to 95% using the procedure outlined in Kupiec (1995), and we conclude that both the static and dynamic PCC are adequate for volatility forecasting.

To distinguish between the two adequate procedures on the basis of the accuracy of the VaR estimates, we consider the loss function proposed by Lopez (1998). For observed portfolio loss at time $t+1$, y_{t+1}, and estimated $\text{VaR}_{t+1|t}$, the loss, Ψ_{t+1}, is defined as:

$$\Psi_{t+1} = \begin{cases} 1 + (y_{t+1} - \text{VaR}_{t+1|t})^2 & \text{if } y_{t+1} < \text{VaR}_{t+1|t} \\ 0 & \text{if } y_{t+1} \geq \text{VaR}_{t+1|t}, \end{cases}$$

for $t = 500, ..., 1509$. The total loss for each model is the sum of the loss function evaluated at each time point.

An exception is said to occur if the realized loss exceeds the model-based 95% VaR threshold. This choice of loss does not penalize the VaR estimate if an exception does not occur. If an exception does occur, the penalty incorporates the magnitude of the exception. Total loss is equal to the total number of exceptions plus the sum of the squared magnitudes of the exceptions.

We use the Diebold-Mariano procedure to create a hypothesis test for a comparison of the VaR predictions from the static and dynamic models (Diebold and Mariano 1995). Though both models have the same number of exceptions for the time period under consideration, the Diebold-Mariano test reveals a statistically significant improvement in the accuracy of the VaR forecasts from the dynamic PCC as opposed to the static PCC, under the Lopez (1998) loss function. Total loss for the dynamic model is 2339.50, compared to 2416.76 for the static model. This difference represents a 3.2% decline in loss. The p-value from the Diebold-Mariano test is 0.005. Hence, we would reject the null hypothesis of equal predictive accuracy, and we conclude that the dynamic PCC is superior to the static PCC for the accuracy of its VaR forecasts in this empirical application.

This result provides further statistical evidence that time dynamics for the PCC is a crucial and necessary extension to the current PCC technology. While this is only a simple example, and a 3.2% improvement is relatively small; for VaR calculations in higher dimensions, we expect the time dynamics will be even more influential.

3.6 Conclusion

Despite its flexibility, the PCC lacking in time dynamics has been slow to gain wide appreciation in empirical application. We believe the time dynamic PCC is an important extension to the existing PCC technology, making it more viable in real-world situations. While there

is a substantial literature on bivariate copula dynamics, to our knowledge there is none for multivariate copulas in more than two dimensions. Our idea is simply to extend the dynamic bivariate copula literature to multivariate copulas through the PCC, which uses bivariate copula as building blocks for its multivariate structure. Because financial models require a flexible dependence relationship and the ability to capture the effects of economic shocks over time, a dynamic PCC can be a powerful tool for multivariate financial time series. We have illustrated the impact of our model on out-of-sample equity correlation forecasts and one-step-ahead VaR calculations. In both cases, the dynamic model provided a statistically significant improvement over the current static PCC.

A key to the success of our technique is the Bayesian perspective from which we have undertaken our model development. The Bayesian framework has allowed us to not only estimate a highly parameterized and complex model with relative simplicity, but also to quantify the uncertainty associated with the model. From a classical statistics point of view, fitting a model such as the one we have proposed would be extremely difficult, if not impossible.

Though it is certainly a step forward, there are several drawbacks and areas of future research related to our dynamic PCC. The MCMC estimation algorithm is computationally intensive and time consuming. Though the computer time is not prohibitively expensive, it would be worthwhile to explore a more efficient sampling routine. Because of the computational issues, we have limited the time dynamics to the first branch of the vine. If further research can improve the MCMC procedure, we could include time dynamics for more of the parameters of the model, perhaps resulting in even better performance. Nonetheless, most of the dependence structure is captured by the first branch, so gains from allowing additional time-varying parameters would likely be minimal. Perhaps a bigger concern is the fact that we must fix the hyperparameters associated with the time series model for the dynamic parameters. Our simulations have shown that the choice of these hyperparameters does not effect overall inference and has very little impact on our out-of-sample prediction, but the

within-sample estimates are somewhat sensitive to these values. Ideally we could assign a hyperprior to these hyperparameters and estimate them within the model. Unfortunately this has proven extremely difficult. In our empirical examples, we have made every effort to choose these values carefully based on sensitivity analyses. Lastly, we are somewhat limited by the dimensionality of the data. The conditional independence component of our model encourages parsimony and allows us to estimate the model in relatively high dimension. A model for dynamic PCC that can be estimated in 100 dimensions would be extremely valuable. Our model is satisfactory for dimensionality on the order of 25 to 30.

The time dynamic PCC we have introduced is a powerful tool with wide-ranging applications. While we have illustrated two applications in finance: equity correlation and VaR; there are several other financial products whose pricing depends on time dynamics and flexible dependence. Some examples include default correlation, Collateralized Debt Obligations, basket default swaps, and portfolio and risk management. Of course there are also many potential areas of application outside of the field of finance.

Taking into account the array of potential applications, considering the intuitive value inherent in a time dynamic dependence structure, and having illustrated two empirical applications where the dynamic PCC statistically outperforms the static PCC, we view time dynamics as an essential component of any successful PCC application.

Appendix A

Appendix to Chapter 1

The Dow Jones Credit Suisse Hedge Fund Index is classified into broad index categories. We offer a brief description of each of these classifications.

Convertible Arbitrage: A market neutral investment strategy in which fund managers simultaneously enter long positions on convertible securities (e.g. bonds that can be converted into equity under certain conditions) and short positions on the stock of the same issuer. By carefully managing the portfolio, the combined position is intended to be insensitive to price changes of the underlying security, while still taking advantage of inefficiencies in the pricing of the convertible asset.

Dedicated Short Bias: Such funds typically take on more short positions than long positions, maintaining a net short portfolio over time, and typically focusing on firms with weak cash-flows. The key strategic component of risk management for these types of funds is appropriately holding off-setting long positions, and incorporating stop-loss strategies.

Emerging Markets: These funds focus their investments on financial instruments in developing economies. Examples of emerging market economies include India, China, Latin America, much of Southeast Asia, and parts of Africa and Eastern Europe. Within the Emerging Markets classification there are many different specific investment strategies.

Equity Market Neutral: Portfolio managers pursuing this strategy seek to maintain a portfolio that exploits investment opportunities unique to a subset of stocks, while maintaining an overall portfolio that is insensitive to price movements of the broader range of stocks through a careful balancing of long and short positions.

Event Driven: These funds rely heavily on upcoming or foreseen market events, which investors believe have caused inefficiencies in the pricing of securities affected by these events. Examples of market events include mergers, litigation, asset sales, corporate restructuring, etc. Within this classification, specific strategies include **Distressed**, focusing on companies facing difficult situations (such as bankruptcy), **Risk Arbitrage**, attempting to capture spreads in merger prices of the companies involved after the terms of the merger have been announced, and **Multi-Strategy**, a combination therein.

Fixed Income Arbitrage: This investment strategy seeks steady growth and low volatility by trying to take advantage of small pricing spreads in fixed income securities, such as corporate bonds, U.S. Treasuries, and interest rate swaps.

Global Macro: Managers adherent to the global macro investment strategies will typically rely on a top-down approach in which global political and macroeconomic events drive investments. Typically, the focus is on price and currency movements in the global market. Unlike many of the other categories, this particular trading strategy relies more on knowledge and experience of the managers, as opposed to mathematical models.

Long Short Equity: Investors in this category will take both long and short positions in equity, often hedging across specific sectors of the market about which the fund managers are particularly knowledgeable. For example, the portfolio may be long the stock of one automobile manufacturer and short another.

Managed Futures: These funds tied to futures markets (commodities, equities, and currencies) often employ global trading initiatives reliant upon historical prices and market trends. Positions are often highly leveraged.

Multi-Strategy: Such funds do not specialize in a particular trading strategy, but instead

attempt to capitalize on current market conditions by incorporating a number of the strategies discussed here. This category may also include individual funds which are difficult to otherwise classify.

Appendix B

Appendix to Chapter 2

Using the same notation as in the main text, we assume the posterior distribution, π, conditional on data and auxiliary information, is $\boldsymbol{\theta} \sim N_{m+1}(\tilde{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$. Given the discussion in the main text, we are interested in computing the distribution π^* which has minimal K-L divergence from the posterior distribution π subject to the benchmarking constraints, as shown in Equation (2.2). Kullback (1959) shows that if π is an exponential family, the solution π^* is a member of the same family. Hence, $\boldsymbol{\theta^*} \sim N_{m+1}(\boldsymbol{\mu^*}, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\mu^*}$ and $\boldsymbol{\Sigma}^*$ are chosen to satisfy Equation (2.2). Using the properties of the Multivariate Normal distribution, it is easy to show that

$$KL(\pi^*, \pi) = \frac{1}{2}\left(\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^*) + (\tilde{\boldsymbol{\theta}} - \boldsymbol{\mu^*})^T\boldsymbol{\Sigma}^{-1}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\mu^*}) + \log\frac{\det\boldsymbol{\Sigma}}{\det\boldsymbol{\Sigma}^*} - (m+1)\right).$$

Note that the choice of $\boldsymbol{\mu^*}$ which minimizes the equation does not depend on the choice of $\boldsymbol{\Sigma}^*$. Moreover, the choice of $\boldsymbol{\Sigma}^*$ which minimizes the equation does not depend on either $\boldsymbol{\mu^*}$ or $\tilde{\boldsymbol{\theta}}$. To quote Kullback (1959), the K-L divergence is "expressible as the sum of two components, one due to the difference in means, the other due to the difference in variances and covariances." In other words, rather than solving Equation (2.2) subject to both benchmarking constraints simultaneously, we can solve for $\boldsymbol{\mu^*}$ according to T_1 and independently solve for $\boldsymbol{\Sigma}^*$ according to T_2. Further, because the choice $\boldsymbol{\Sigma}^*$ does not depend on $\tilde{\boldsymbol{\theta}}$, we can

make the simplifying assumption that $\tilde{\boldsymbol{\theta}} = \mathbf{0}$ when imposing T_2 without loss of generality.

We will exploit the following result from Kullback (1959) to perform the necessary calculations. For general restriction T, the solution to the minimization problem

$$\min_{\pi^*(\boldsymbol{\theta})} \int \pi^*(\boldsymbol{\theta}) \log \frac{\pi^*(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta} \text{ such that } \int T(\boldsymbol{\theta}) \pi^*(\boldsymbol{\theta}) = 0$$

is given by

$$\pi^*(\boldsymbol{\theta}) \propto e^{\tau^* T(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}),$$

where τ^* is the solution to $\frac{d}{d\tau} \log M_2(\tau^*) = 0$ and $M_2(\tau) = \int e^{\tau T(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$.

B.1 Flexible Benchmarking Constraint

Recall the benchmarking constraints for the first and second moments are given by

$$T_1(\boldsymbol{\theta}) = \theta - \sum_{j=1}^{m} c_j \theta_j = \boldsymbol{R}^T \boldsymbol{\theta}, \quad T_2(\boldsymbol{\theta}) = \theta^2 - \left(\sum_{j=1}^{m} c_j \theta_j \right)^2.$$

The first moment condition, T_1, can be used to calculate $\boldsymbol{\mu^*}$, which is the mean of π^*, the solution to

$$\min_{\pi^*(\boldsymbol{\theta})} \int \pi^*(\boldsymbol{\theta}) \log \frac{\pi^*(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta} \text{ such that } \int T_1(\boldsymbol{\theta}) \pi^*(\boldsymbol{\theta}) = 0.$$

Based on the moment generating function of the multivariate Normal distribution,

$$M_2(\tau) = E\left[e^{\tau T_1(\boldsymbol{\theta})}\right] = E\left[e^{\tau \boldsymbol{R}^T \boldsymbol{\theta}}\right] = \exp\left[\tau \tilde{\boldsymbol{\theta}}^T \boldsymbol{R} + \frac{1}{2} \tau^2 \boldsymbol{R}^T \boldsymbol{\Sigma} \boldsymbol{R}\right].$$

The following equation can be solved for τ^*:

$$\frac{\partial}{\partial \tau} \log M_2(\tau^*) = 0 \Rightarrow \tilde{\boldsymbol{\theta}}^T \boldsymbol{R} + \tau^* \boldsymbol{R}^T \boldsymbol{\Sigma} \boldsymbol{R} = 0 \Rightarrow \tau^* = -\tilde{\boldsymbol{\theta}}^T \boldsymbol{R} (\boldsymbol{R}^T \boldsymbol{\Sigma} \boldsymbol{R})^{-1}.$$

The MDI distribution, π^*, is

$$\pi^*(\boldsymbol{\theta}) \propto e^{\tau^* T_1(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})$$

$$\propto \exp\left\{-\tilde{\boldsymbol{\theta}}^T \boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{\theta} - \frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\theta}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta} - 2\tilde{\boldsymbol{\theta}}^T\left(\boldsymbol{\Sigma}^{-1} - \boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\boldsymbol{R}^T\right)\boldsymbol{\theta}\right]\right\}.$$

The Multivariate Normal MDI distribution satisfying the first moment benchmarking constraint has mean

$$\boldsymbol{\mu}^* = \tilde{\boldsymbol{\theta}} - \boldsymbol{\Sigma}\boldsymbol{R}(\boldsymbol{R}^T\boldsymbol{\Sigma}\boldsymbol{R})^{-1}\boldsymbol{R}^T\tilde{\boldsymbol{\theta}}. \tag{B.1}$$

The second moment condition, T_2, can be used to calculate $\boldsymbol{\Sigma}^*$, which is the covariance of π^*, the solution to

$$\min_{\pi^*(\boldsymbol{\theta})} \int \pi^*(\boldsymbol{\theta}) \log \frac{\pi^*(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta} \text{ such that } \int T_2(\boldsymbol{\theta})\pi^*(\boldsymbol{\theta}) = 0.$$

Without loss of generality, assume $\tilde{\boldsymbol{\theta}} = \mathbf{0}$. Under this assumption,

$$\boldsymbol{C}^T\boldsymbol{\theta}_s = \sum_{j=1}^{m} c_j\theta_j \sim N(0, \sigma_s^2),$$

where $\sigma_s^2 = \boldsymbol{C}^T\boldsymbol{\Sigma}_s\boldsymbol{C}$. Thus, we have the following independent distributions:

$$\frac{\left(\boldsymbol{C}^T\boldsymbol{\theta}_s\right)^2}{\sigma_s^2} \sim \chi_1^2, \quad \frac{\theta^2}{\sigma^2} \sim \chi_1^2.$$

Using the moment generating function of the Chi-Square distribution,

$$M_2(\tau) = E\left[e^{\tau T_2(\boldsymbol{\theta})}\right] = E\left[e^{\tau\left(\theta^2 - \left(\boldsymbol{C}^T\boldsymbol{\theta}_s\right)^2\right)}\right] = E\left[e^{\tau\theta^2}\right] E\left[e^{-\tau\left(\boldsymbol{C}^T\boldsymbol{\theta}_s\right)^2}\right]$$
$$= (1-2\tau\sigma^2)^{-\frac{1}{2}}(1+2\tau\sigma_s^2)^{-\frac{1}{2}}$$

Solving $\partial \log M_2(\tau)/\partial\tau = 0$ for τ gives $\tau^* = (\sigma_s^2 - \sigma^2)/(4\sigma_s^2\sigma^2)$.

The MDI distribution, π^* is

$$\pi^*(\boldsymbol{\theta}) \propto e^{\tau^* T_2(\boldsymbol{\theta})}\pi(\boldsymbol{\theta})$$

$$\propto \exp\left[\tau^*\left(\theta^2 - \left(\boldsymbol{C}^T\boldsymbol{\theta}_s\right)^2\right) - \frac{1}{2\sigma^2}\theta^2 - \frac{1}{2}\boldsymbol{\theta_s}^T\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\theta_s}\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2} - 2\tau^*\right)\theta^2\right]\exp\left[-\frac{1}{2}\boldsymbol{\theta_s}^T\left(2\tau^*\boldsymbol{C}\boldsymbol{C}^T + \boldsymbol{\Sigma}_s^{-1}\right)\boldsymbol{\theta_s}\right].$$

This factorization reveals that the Multivariate Normal MDI distribution satisfying the second moment benchmarking constraint has covariance matrix

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \frac{2\sigma_s^2\sigma^2}{\sigma_s^2+\sigma^2} & \mathbf{0} \\ \mathbf{0} & \left(\frac{\sigma_s^2-\sigma^2}{2\sigma_s^2\sigma^2}\boldsymbol{C}\boldsymbol{C}^T + \boldsymbol{\Sigma}_s^{-1}\right)^{-1} \end{bmatrix}. \tag{B.2}$$

Combining Equations (B.1) and (B.2), the MDI distribution satisfying the first and second moment benchmarking restrictions is $\boldsymbol{\theta}^* \sim N_{m+1}\left(\boldsymbol{\mu^*}, \boldsymbol{\Sigma}^*\right)$.

B.2 Fixed Benchmarking Constraint

If the higher level parameters are assumed fixed and known, the moment restrictions are

$$T_1^{\text{fix}}(\boldsymbol{\theta}_s) = \hat{\theta} - \sum_{j=1}^{m} c_j\theta_j, \quad T_2^{\text{fix}}(\boldsymbol{\theta}_s) = \left(D + \hat{\theta}^2\right) - \left(\sum_{j=1}^{m} c_j\theta_j\right)^2.$$

Using the same procedures as above, the first and second moments of the MDI distribution can be computed separately, and it can be shown that the MDI distribution is

$$\boldsymbol{\theta_s}^* \sim N_m\left(\tilde{\boldsymbol{\theta}}_s + \boldsymbol{\Sigma}_s\boldsymbol{C}\left(\boldsymbol{C}^T\boldsymbol{\Sigma}_s\boldsymbol{C}\right)^{-1}\left(\hat{\theta} - \sum_{j=1}^{m} c_j\tilde{\theta}_j\right), \left(\left(\frac{\sigma_s^2 - D}{D\sigma_s^2}\right)\boldsymbol{C}\boldsymbol{C}^T + \boldsymbol{\Sigma}_s^{-1}\right)^{-1}\right).$$

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44, 182–198.

Acharya, V., Pedersen, L., Philippon, T., and Richardson, M. (2010), "Measuring systemic risk," Working Paper, NYU.

Acharya, V. and Richardson, M. (eds.) (2009), *Restoring Financial Stability: How to Repair a Failed System*, John Wiley & Sons, New York, NY.

Adrian, T. (2007), "Measuring risk in the hedge fund sector," *Federal Reserve Bank of New York Current Issues in Economics and Finance*, 13, 1–7.

Adrian, T. and Brunnermeier, M. (2011), "CoVaR," NBER Working Paper Series: 17454.

Agarwal, V. and Naik, N. (2004), "Risk and portfolio decisions involving hedge funds," *Review of Financial Studies*, 17, 63–98.

Allen, F., Babus, A., and Carletti, E. (2012), "Asset commonality, debt maturity and systemic risk," *Journal of Financial Economics*, 104, 519–534.

Allen, F. and Gale, D. (2000), "Financial contagion," *Journal of Political Economy*, 108, 1–33.

Amihud, Y. and Mendelson, H. (1986), "Asset pricing and the bid-ask spread," *Journal of Financial Economics*, 17, 223–249.

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

Angelidis, T., Benos, A., and Degiannakis, S. (2004), "The use of GARCH models in VaR estimation," *Statistical Methodology*, 1, 105–128.

Babus, A. and Allen, F. (2009), "Networks in finance," in *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*, eds. Kleindorfer, P. and Wind, J., Wharton Publishing Company, Upper Saddle Creek, NJ.

Bali, T., Gokcan, S., and Ling, B. (2007), "Value at risk and the cross-section of hedge fund returns," *Journal of Banking and Finance*, 30, 1135–1166.

Battese, G. E., Harter, R. H., and Fuller, W. A. (1988), "An error-components model for prediction of county crop areas using survey and satellite data," *Journal of the American Statistical Association*, 83, 28 – 36.

Bedford, T. and Cooke, R. (2001), "Probability density decomposition for conditionally dependent random variables modeled by vines," *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.

— (2002), "Vines - a new graphical model for dependent random variables," *Annals of Statistics*, 30, 1031–1068.

Bekaert, G. and Harvey, C. (1995), "Time-varying world market integration," *Journal of Finance*, 50, 403–444.

Berg, D. and Aas, K. (2009), "Models for construction of multivariate dependence: a comparison study," *European Journal of Finance*, 15, 639–659.

Billio, M., Getmansky, M., Lo, A., and Pelizzon, L. (2012a), "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, 104, 535–559.

Billio, M., Getmansky, M., and Pelizzon, L. (2012b), "Dynamic risk exposure in hedge funds," *Computational Statistics and Data Analysis*, 56, 3517–3532.

Billio, M. and Pelizzon, L. (2000), "Value-at-risk: a multivariate switching regime approach," *Journal of Empirical Finance*, 7, 531–554.

— (2003), "Volatility shocks and spillover before and after EMU in European stock markets," *Journal of Multinational Financial Management*, 13, 323–340.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Boyson, N., Stahel, C., and Stulz, R. (2010), "Hedge fund contagion and liquidity shocks," *Journal of Finance*, 55, 1789–1816.

Brealey, R. and Kaplanis, E. (2001), "Hedge funds and financial stability: an analysis of their factor exposures," *Journal of International Finance*, 4, 161–187.

Brechmann, E., Czado, C., and Aas, K. (2012), "Truncated regular vines in high dimensions with application to financial data," *Canadian Journal of Statistics*, 40, 68–85.

Brunnermeier, M. and Pederson, L. (2009), "Market liquidity and funding liquidity," *Review of Financial Studies*, 22, 2201–2238.

Carhart, M. (1997), "On persistence in mutual fund performance," *Journal of Finance*, 52, 57–82.

Chan, N., Getmansky, M., Haas, S., and Lo, A. (2006), "Do hedge funds increase systemic risk?" *Federal Reserve Bank of Atlanta: Economic Review*, Fourth Quarter, 49–80.

Chollete, L., Heinen, A., and Valdesogo, A. (2008), "Modeling international financial returns with a multivariate regime switching copula," CORE Discussion Paper.

Chordia, T., Roll, R., and Subrahmanayam, A. (2002), "Order imbalance, liquidity and market returns," *Journal of Financial Economics*, 65, 111–130.

Cifuentes, R., Ferrucci, G., and Shin, H. (2005), "Liquidity risk and contagion," *Journal of the European Economic Association*, 3, 556–566.

Dalla Valle, L. (2009), "Bayesian copulae distributions with application to operational risk management," *Methodology and Computing in Applied Probability*, 11, 95–115.

Dasgupta, A. (2004), "Financial contagion through capital connections: a model of the origin and spread of bank panics," *Journal of European Economic Association*, 2, 1049–1084.

Datta, G. S., Ghosh, M., Steorts, R., and Maples, J. (2011), "Bayesian benchmarking with applications to small area estimation," *Test*, 20, 574 – 588.

Dias, A. and Embrechts, P. (2004), "Change-point analysis for dependence structures in finance and insurance," in *Risk Measures for the 21st Century*, ed. Szego, G., Wiley, Chichester, England.

Diebold, F. and Mariano, R. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Dißmann, J., Brechmann, E., C.Czado, and Kurowicka, D. (2013), "Selecting and estimating regular vine copulae and application to financial returns," *Computational Statistics and Data Analysis*, 59, 52–69.

Embrechts, P., Hoing, A., and Juri, A. (2003), "Using copulae to bound value-at-risk fro functions of dependent risks," *Finance and Stochastics*, 7, 145–167.

Engle, R. (1982), "Autoregressive conditional heteroskedasticity with estimates of United Kingdom inflation," *Econometrica*, 50, 987–1008.

— (2002), "Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. and Patton, A. (2001), "What good is a volatility model?" *Quantitative Finance*, 1, 237–245.

Fama, E. and French, K. (1993), "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

Fantazzini, D. (2008), "Dynamic copula modeling for Value at Risk," *Frontiers in Finance and Economics*, 5, 72–108.

Fay, R. E. and Herriot, R. A. (1979), "Estimates of income from small places: an application of James-Stein procedures to census data," *Journal of the American Statistical Association*, 74, 269 – 277.

Frey, R. and McNeil, A. (2003), "Dependent defaults in models of portfolio credit risk," *Journal of Risk*, 6, 59–92.

Fung, W. and Hsieh, D. (2001), "The risk in hedge fund strategies: theory and evidence from trend followers," *Review of Financial Studies*, 14, 313–341.

— (2004), "Hedge fund benchmarks: a risk based approach," *Financial Analysts Journal*, 60, 65–80.

Furfine, C. (2003), "Interbank exposures: quantifying the risk of contagion," *Journal of Money, Credit, and Banking*, 35, 111–128.

Garcia, R. and Tsafack, G. (2008), "Dependence structure and extreme comovements in international equity and bond markets with portfolio diversification effects," Working paper, EDHEC Risk Asset Management Research Centre.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398 – 409.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004), *Bayesian Data Analysis*, Chapman and Hall/CRC, Boca Raton, FL.

Geman, S. and Geman, D. (1984), "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Getmansky, M., Lo, A., and Makarov, I. (2004), "An econometric analysis of serial correlation and illiquidity in hedge fund returns," *Journal of Financial Economics*, 74, 529–610.

Granger, C. (1969), "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 37, 424–438.

Gray, D. (2009), "Modeling financial crises and sovereign risks," *Annual Review of Financial Economics*, 1, 117–144.

Guidolin, M. and Timmermann, A. (2008), "International asset allocation under regime switching, skew and kurtosis preferences," *Review of Financial Studies*, 20, 889–935.

Gupta, A. and Liang, B. (2005), "Do hedge funds have enough capital?" *Journal of Financial Economics*, 77, 219–253.

Hafner, C. and Manner, H. (2012), "Dynamic stochastic copula models: estimation, inference and applications," *Journal of Applied Econometrics*, 27, 269–295.

Hafner, C. and Reznikova, O. (2010), "Efficient estimation of a semiparametric dynamic copula model," *Computational Statistics and Data Analysis*, 54, 2609–2627.

Hastings, W. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

Hedge Fund Research (2013), "HFR Global Hedge Fund Industry Report," Published January 18, 2013.

Hoff, P., Raftery, A., and Handcock, M. (2002), "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, 97, 1090–1098.

Huang, X., Zhou, H., and Zhu, H. (2012), "Systemic risk contributions," *Journal of Financial Services Research*, 42, 55–83.

Isaki, C. T., Tsay, J. H., and Fuller, W. A. (2000), "Estimation of census adjustment factors," *Survey Methodology*, 26, 31 – 42.

Jaynes, E. T. (1957), "Information theory and statistical mechanics," *Physical Review*, 106, 620 – 630.

Joe, H. (1996), "Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters," in *Distributions with Fixed Marginals and Related Topics*, eds. Ruschendorf, L., Schweizer, B., and Taylor, M., Institute of Mathematical Statistics, Hayward, CA.

Jorion, P. (2000), *Value at Risk: The New Benchmark for Managing Financial Risk, 2nd Ed.*, McGraw Hill, New York, NY.

Knottnerus, P. (2003), *Sample Survey Theory: Some Pythagorean Perspectives*, New York: Springer.

Kullback, S. (1959), *Information Theory and Statistics*, New York: Wiley.

Kullback, S. and Liebler, R. A. (1951), "On information and sufficiency," *Annals of Mathematical Statistics*, 22, 79 – 86.

Kupiec, P. (1995), "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives*, 3, 73–84.

Kurowicka, D. and Cooke, R. (2004), "Distribution-free continuous Bayesian belief nets," Proceedings Mathematical Methods in Reliability Conference.

Lagunoff, R. and Schreft, L. (2001), "A model of financial fragility," *Journal of Economic Theory*, 99, 220–264.

Laurent, J.-P. and Gregory, J. (2005), "Basket default swaps, CDO's, and factor copulas," *Journal of Risk*, 7, 103–122.

Leitner, Y. (2005), "Financial networks: contagion, commitment and private sector bailouts," *Journal of Finance*, 60, 2925–2953.

Li, D. (2000), "On default correlation: a copula function approach," *Journal of Fixed Income*, 9, 43–54.

Lo, A. and Wang, J. (2000), "Trading volume: definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 13, 257–300.

Longin, F. and Solnik, B. (1995), "Is the correlation of international equity returns constant: 1960-1990," *Journal of International Money and Finance*, 14, 3–26.

— (2001), "Extreme correlation of international equity markets," *Journal of Finance*, 56, 649–676.

Lopez, J. (1998), "Methods for evaluating Value-at-Risk estimates," Federal Reserve Bank of New York, Economic Policy Review.

Lowenstein, R. (2001), *When Genius Failed*, Random House, New York, NY.

Manner, H. and Reznikova, O. (2012), "A survey on time-varying copulas: specification, simulations, and application," *Econometric Reviews*, 31, 654–687.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, 21, 1087–1092.

Min, A. and Czado, C. (2010), "Bayesian inference for multivariate copulas using pair-copula constructions," *Journal of Financial Econometrics*, 8, 511–546.

— (2011), "Bayesian model selection for multivariate copulas using pair-copula constructions," *Canadian Journal of Statistics*, 39, 239–258.

Morris, C. and Lysy, M. (2012), "Shrinkage estimation in multilevel Normal models," *Statistical Science*, 27, 115–134.

Nandram, B., Toto, M. C. S., and Choi, J. W. (2011), "A Bayesian benchmarking of the Scott-Smith model for small areas," *Journal of Statistical Computation and Simulation*, 81, 1593 – 1608.

Pastor, L. and Stambaugh, R. (2003), "Liquidity risk and expected stock returns," *Journal of Political Economy*, 111, 642–685.

Patton, A. (2004), "On the out-of-sample importance of skewness and asymmetric dependence for asset allocation," *Journal of Financial Econometrics*, 2, 130–168.

— (2006), "Modelling asymmetric exchange rate dependence," *International Economic Review*, 47, 527–556.

Pfeffermann, D. and Barnard, C. H. (1991), "New estimators for small-area means with applications to the assessment of farmland values," *Journal of Business and Economic Statistics*, 9, 73 – 84.

Pitt, M., Chan, D., and Kohn, R. (2006), "Efficient Bayesian inference for Gaussian copula regression models," *Biometrika*, 93, 537–554.

Raftery, A. (1996), "Hypothesis testing and model selection via posterior simulation," in *Markov Chain Monte Carlo in Practice*, eds. Gilks, W., Richardson, S., and Spiegelhalter, D., Chapman and Hall, New York, NY.

Rajan, R. (2006), "Has finance made the world riskier?" *European Financial Management*, 12, 499–533.

Rao, J. N. K. (2003), *Small Area Estimation*, New York: Wiley.

Reinhart, C. and Rogoff, K. (2009), *This Time is Different: Eight Centuries of Financial Folly*, Princeton University Press, Princeton, NJ.

Rodriguez, J. (2007), "Measuring financial contagion: a copula approach," *Journal of Empirical Finance*, 14, 401–423.

Sadka, R. (2006), "Momentum and post-earnings-announcement drift anomalies: the role of liquidity risk," *Journal of Financial Economics*, 80, 309–349.

Salmon, F. (2009), "Recipe for disaster: the formula that killed Wall Street," *Wired Magazine*, 23 February 2009.

Sklar, M. (1959), "Fonctions de répartition à n dimensions et leurs marges." *Publ. Inst. Statist. Univ. Paris*, 8, 229–231.

Smith, M., Min, A., Almeida, C., and Czado, C. (2010), "Modeling longitudinal data using a pair-copula decomposition of serial dependence," *Journal of the American Statistical Association*, 105, 1467–1479.

Sostra, K. and Traat, I. (2009), "Optimal domain estimation under summation restriction," *Journal of Statistical Planning and Inference*, 139, 3928 – 3941.

Toto, M. C. S. and Nandram, B. (2010), "A Bayesian predictive inference for small area means incorporating covariates and sampling weights," *Journal of Statistical Planning and Inference*, 140, 2963 – 2979.

United States Department of Treasury (2013), "Troubled asset relief program: monthly report to Congress (December 2012)," Published January 10, 2013.

Upper, C. and Worms, A. (2004), "Estimating bilateral exposures in the German interbank market: is there danger of contagion," *European Economic Review*, 48, 827–849.

van den Goorbergh, R., Genest, C., and Weker, B. (2005), "Bivariate option pricing using dynamic copula models," *Insurance: Mathematics and Economics*, 37, 101–114.

Wang, J., Fuller, W. A., and Qu, Y. (2008), "Small area estimation under a restriction," *Survey Methodology*, 34, 29 – 36.

Wells, S. (2004), "U.K. interbank exposures: systemic risk implications," *The Journal of Monetary Economics*, 2, 66–77.

You, Y. and Rao, J. N. K. (2002), "A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights," *Canadian Journal of Statistics*, 30, 431 – 439.

You, Y., Rao, J. N. K., and Dick, P. (2004), "Benchmarking hierarchical bayes small area estimators in the Canadian census undercoverage estimation," *Statistics in Transition*, 6, 631 – 640.

Yu, J. and Meyer, R. (2006), "Multivariate stochastic volatility models: Bayesian estimation and model comparison," *Econometric Reviews*, 25, 361–384.