Quantitative Finance and Economics

http://www.aimspress.com/journal/QFE

QFE, 2(2): 413–467.
DOI:10.3934/QFE.2018.2.413
Received: 28 November 2017
Accepted: 22 April 2018
Published: 11 June 2018

Research article

Applications of random-matrix theory and nonparametric change-point analysis to three notable systemic crises

David Melkuev[1], **Danqiao Guo**[2] **and Tony S. Wirjanto**[2,3,*]

[1] Canada Pension Plan Investment Board, One Queen Street East, Suite 2500, Toronto, ON M5C 2W5, Canada

[2] Department of Statistics & Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

[3] School of Accounting & Finance, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

* **Correspondence:** Email: twirjant@uwaterloo.ca; Tel: +15198884567 ext. 35210; Fax: +15197461875.

Abstract: This paper studies association between changes in absorption ratio and aggregate market returns in three systemic crises across a broad class of assets. Time series of normalized eigenvalue estimates reveal that crises are characterized by a general breakdown of correlation structure. The structure of return correlations is nonlinear and nonstationary across different asset groups. So we introduce a nonparametric technique to monitor divergence in distributions underlying successive observations of normalized dominant eigenvalue of the returns. Periods of high divergence imply a change in the correlation structure of asset returns. They are found to either precede or coincide with systemic shocks. An additional parametric analysis is provided as an informal check on the results obtained in the paper.

Keywords: global financial crisis; Eurozone sovereign debt crisis; Asian financial crisis; equities; bonds; CDS; contract; principal component analysis; random matrix theory; nonparametric change point analysis
JEL classification numbers: C1, F4, G1

1. Introduction

This paper explores changes in the strength of correlation (or equivalently comovement) between financial asset returns over time. What can perhaps be most reliably said about this phenomenon is that correlations between asset returns tend to increase during epsiodes of downturns in the financial

markets. There is plenty empirical evidence in support of this claim, for example in Ang and Chen (2002); Bouchaud and Potters (2001); Cizeau et al. (2001); Longin and Solnik (2001); Meric et al. (2001); Solnik et al. (1996) and the references therein. In this paper we focus our investigation on return correlations between asset returns in reference to more circumstantial market drawdowns. Specifically we are interested in correlation dynamics of asset returns in relation to systemic shocks which occur in financial markets.

Systemic events have a special characteristic of causing drawdowns that are contagious in that they ultimately propagate to a large number of assets relative to the number of assets immediately affected. In other words, under certain market conditions, devaluation of some assets can cause devaluation of some other assets in the financial system. The notion of financial contagion is not new and is recognized in historic accounts of financial crises dating back several centuries ago. See Reinhart and Rogoff (2011). Viewed from the perspective of a purely systemic crisis it is intuitive to explain correlation asymmetry in asset returns simply as a manifestation of financial contagion itself. However the risk of systemic shocks tends to change over time as the conditions that facilitate contagion are created. Therefore an important research question posed in this paper is whether an increased systemic risk is reflected in rising correlations between asset returns $before$ the onset of a systemic crisis.

To measure the strength of comovement of asset returns we take an approach based on spectral analysis. The idea underlying this approach is to transform coordinates of the original multivariate return data by means of a principle component analysis (PCA), such that the variance along each axis of the new system is maximized. The values along each axis represent linearly uncorrelated factors, and have successively decreasing variance (See Appendix A). Recently this PCA has been used to study changes in correlation between asset returns (Billio et al., 2012; Conlon et al., 2009; Drożdż et al., 2000; Kritzman et al., 2011; Meng et al., 2014; Pan and Sinha, 2007; Zhang et al., 2015). Since each eigenvalue equals the variance of the Principle Component (PC) found through its associated eigenvector, we can define the total risk in the system as the sum of these eigenvalues. Then the proportion of total variance attributed to a small number of PCs indicates the degree of commonality between returns on the assets in question. We estimate the proportion of variance explained by a fixed small number of factors through time and study its fluctuations. A similar statistic is coined an absorption ratio (AR) in Kawahara et al. (2007), which is in reference to the proportion of variance absorbed by a few factors. Studies that have taken this approach commonly find that severe downturns coincide with and are sometimes preceded by increases in the strength of comovement between asset returns. The explanation provided for this phenomenon is that a higher degree of integration among financial institutions allows for a ripple effect that characterizes systemic financial crises. However the literature drawing such a conclusion is mostly based on investigations that are limited to certain asset classes and over certain time periods. Specifically the Global Financial Crisis of 2008 is by far the most common case study for this stream of literature and equities are the most common asset class examined in this regard. The current paper expands the analysis in two important directions : (i) we examine three episodes of crises of systemic nature: the Global Financial Crisis of 2008 (or simply the Financial Crisis), the Eurozone Sovereign Debt Crisis of 2009/2010 (or simply the European Debt Crisis) and the Asian Financial Crisis of 1997 (or simply the Asian Financial Crisis); and (ii) we study a broad class of assets in this paper including equities, which have been much studied, and bonds, credit default swaps and currencies,which have been relatively less studied so far.

Strictly speaking sample eigenvalues are random variables. For this reason we are interested in

distributional aspects of eigenvalue-based statistics (see Appendix B). Results from multivariate statistical analysis include exact expressions about both joint and marginal eigenvalue distributions of Wishart matrices (see Muirhead (1982) for examples.) Although Wishart matrices are useful models of correlation matrices, there are issues in practical applications of exact density representations. Firstly the true covariance matrix is a required parameter that is usually unknown to investigators. Secondly for a non-null covariance structure the expressions are usually difficult to evaluate, especially in a high-dimensional setting. Yet results from random matrix theory (RMT) point to a convergence of the density of sample eigenvalues to a nonrandom density as matrix dimensions go to infinity at a fixed aspect ratio. The main result for Wishart matrices has been known since 1967 as the Marčenko-Pastur law (Marčenko and Pastur, 1967) (See Appendix B.3.). Since then, the limiting eigenvalue density has been shown to exist for many different matrix structures, including covariance matrices of time series with temporal dependence (Jin et al., 2009; Yao et al., 2012) or cross-sectional dependence (Silverstein, 1995). Knowledge that the limiting eigenvalue density exists implies that second-order stationarity in a series of eigenvalue observations can be a reasonable assumption under some conditions and this is a desirable property to have for the purpose of statistical inference. Furthermore the discrepancy between empirical and theoretical eigenvalues points to a deviation of the actual system covariance structure from what is postulated in theory. For example, if the empirical largest eigenvalue exceeds the upper support of the Marčenko-Pastur density, then this can serve as evidence that there is more structure to the data than if the data were generated by a purely multivariate white-noise process.

Recognizing potential drawbacks of assuming linearity for asset returns relationships in general, and given evidence of nonstationarity of the correlation structure of asset returns documented in this paper in particular, we introduce a nonparametric method to gain some insight into change-points in the correlation of asset returns. The problem of detecting changes in the underlying distribution of a stochastic process is commonly known as a *change-point detection*. Its origin can be traced back to as early as the 1930s in the work on the problem of monitoring the quality of manufacturing processes (Wetherhill and Brown, 1991). More recently this problem has been studied in a wide array of fields including econometrics (Andersson et al., 2004, 2006; Andrews, 1996; Berkes et al., 2004; Broemling and Tsurumi, 1987) and finance (Beibel and Lerche, 2000; Chen and Gupta, 1997; Shiryaev, 2002). There are many detection procedures, which differ mainly in being *offline* or *online*. Offline algorithms are algorithms that are applied on a sample of historical observations. The purpose of the offline algorithms is to detect changes in an *a posteriori* fashion within the sample with the lowest probability of error. On the other hand online algorithms are designed for real-time monitoring for changes in a stochastic process. Online detection procedures aim to minimize both the false alarm rate and the detection delay, as each new sample is assumed to incur a cost. In this paper we take an online approach to monitor the strength of asset comovements. However, instead of searching for an algorithm with a decision function on whether change has occurred, we propose to track a measure of divergence between the distributions of successive observations. This addresses the robustness issues that often arise in financial applications due to financial data being inherently noisy.

This paper contributes to the literature on financial systemic crises in a number of ways. Firstly we conduct an empirical investigation into the connection between changes in correlation and systemic risk in three notable financial crises: the Financial Crisis, the European Debt Crisis, and the Asian Financial Crisis. It is important to stress from the outset that, in this paper, we do not explore any

potential linkages that may exist among the three crises, in particular between the Financial Crisis and the European Debt Crisis.

To measure correlation between asset returns, we adapt a variant of the AR, which is computed from the eigenvalues of the sample return correlation matrix. Such a spectral analysis of correlation has been previously used in studies of equities during the Financial Crisis. We expand this analysis by investigating three episodes of systemic crises using a broad class of assets. We find that the relationship between correlation in asset returns and extreme drawdowns is more general than what has been reported so far in the extant literature. In particular, whereas extreme drawdowns have been linked to an increasing comovement in asset returns in the Financial Crisis, consistent with the findings reported in the literature, we find that, in other episodes of systemic crisis studied in this paper, a systemic shock actually results in a decoupling of asset returns, so that systemic crises are associated more with a general breakdown of the correlation structure of asset returns. In addition, given evidence of nonlinearity and nonstationarity in correlation structure of asset returns, we also conduct a change-point analysis on the correlation structure of asset returns by using a nonparametric technique. By estimating the divergence in the distributions of successive groups of samples through time, we identify time periods that are associated with shifts in the correlation structure of asset returns. We find that the divergence score increases either before or coincidentally with systemic financial shocks.

The organization of the remaining parts of this paper is as follows. Section 2 discusses systemic risk and its connection with correlation of asset returns, and provides a selected literature review most relevant to our study. Section 3 describes the data used in our study and performs an exploratory data analysis of the correlation dynamics in reference to three financial crises of systemic nature across a broad range of asset classes. In Sections 4–7 we introduce and use a nonparametric approach based on statistical change-point detection techniques to study the nonlinearity and nonstationarity of correlation structure of asset returns. Section 8 provides an additional parametric analysis as a simple and informal check on the results obtained by using the RMT and the nonparametric approach. Section 9 concludes the paper. Appendix A provides a brief review on spectral analysis of correlation essential to the analysis conducted in this paper. Appendix B presents discussion on the distribution of eigenvalues, focussing on results from the RMT, and, lastly, Appendix C lists stylized properties of eigenvalue-based time series.

2. Selected literature review

There is a number of studies in the literature that are related to our paper in the use of the spectral analysis and the RMT to study the level of correlation between asset returns and its implications on risk management. In particular a recursive PCA scheme has been implemented to gain insight into the temporal evolution of the interconnectedness of asset returns and its relationship with systemic shocks (see Billio et al., 2012; Kritzman et al., 2011; Meng et al., 2014, for examples). However, so far, such an analysis has only been made to illustrate a relationship qualitatively. In this paper we expand this approach by systematically studying the properties of correlation matrix eigenvalues in empirical asset-return data, and determining what, if any, can be statistically inferred from the eigenvalue series of these return data about the systemic risk.

Financial systemic risk is inherently difficult to define and there is no clear consensus in the

literature on a precise definition of it to this date. However proposed definitions in the literature (e.g. International Monetary Fund, 2009; Billio et al., 2012; Bijlsma et al., 2010; De Brandt and Hartmann, 2000) tend to share one common thread that, when realized, systemic risk results in a severe disruption to the financial system. The outcome of the event tends to propagate from local to system-wide shocks through some amplification mechanism, also known as "contagion". Systemic risk is then defined as the probability of such an event occurring. The best recent example of this is the Financial Crisis, in which defaults by a small initial number of financial institutions cascaded throughout the entire U.S. financial industry. In this case the amplification mechanism was due to a high degree of interdependence of institutions' solvency because of contingent-claims like default insurance. Vastly common dependence on catalytic factors like the health of the housing market also played a role (Lewis, 2010) in this crisis. In general it is much easier to recognize systemic risk in an ex-post sense than to define it an ex-ante sense. Unfortunately its recognition often comes only after the risk has been realized, imposing a substantial economic cost to the affected financial system (International Monetary Fund, 2009). Policy makers therefore have a keen interest in detecting an elevation of systemic risk as early as possible in order to take necessary preventive measures.

The connection between systemic risk and correlation is rooted in a wellknown stylized fact of financial returns, which asserts that times of crisis are associated with increased asset return correlations. This relationship has been investigated in a relatively large volume of studies in the literature, such as (Ang and Chen, 2002; Bouchaud and Potters, 2001; Cizeau et al., 2001; Longin and Solnik, 2001; Meric et al., 2001; Solnik et al., 1996). In particular the authors in Ang and Chen (2002) find that, conditional on the negativity of returns in the U.S. equity market, correlations are 11.6% higher than implied by a normal distribution. This is contrasted with correlations that, when conditioned on positive returns, cannot be statistically distinguished from those implied by a normal distribution. While it has been extensively documented that correlations increase during volatile periods, a more interesting question for us is whether a strengthening in correlations between asset returns precedes widespread shocks in the financial system. That is whether contagion is facilitated by a state of high degree of interconnectedness that the system evolves to. In this case the extent to which we can make inference on asset returns and volatility following states of high correlation is of some interest.

To measure the strength of comovement in returns on a class of assets we apply the PCA to the sample correlation matrix and obtain the proportion of total variance explained by every PC. If a relatively small number of PCs explain a relatively large proportion of variance, this is interpreted as a state of high degree of interconnectedness between the assets' returns (See Appendix A). This approach has been used in the past in studies of asset return correlations, including some which analyzed correlation dynamics in the context of systemic risk (Billio et al., 2012; Conlon et al., 2009; Drożdż et al., 2000; Kritzman et al., 2011; Meng et al., 2014; Pan and Sinha, 2007). As stated in Appendix A.4, there are a few important differences between our measure of interconnectedness between the assets' returns and the more commonly used Average Pearson Correlation between these returns. Firstly, the interconnectedness measure accounts for the direction of the strongest variability in the return co-movement, as such it accounts for the component of the asset return correlations, which is not captured in a simple average. As a result it is possible for the average correlation of asset return to decrease over time but for the interconnectedness measure to display an increased value. Another distinction between these two measures is the ability of the interconnectedness measure to

measure the proportion of the variability in asset returns captured in specific directions as opposed to the correlations that capture only bilateral congruence; the interconnectedness measure therefore captures secondary links between asset returns. Thirdly, the interconnectedness measure reflects changes arising from co-movement between the asset returns with high variance, while a simple average correlation between these assets accounts for the levels of volatility implicit in each.

Since the variance explained by each PC is given by the associated eigenvalue, this analysis can draw on what is known from the RMT regarding the distribution of eigenvalues (See Appendix B.2). For example differences between the empirical density of eigenvalues and the theoretical density of eigenvalues (in purely random data) suggest that there is a common "market" factor to the returns (Laloux et al., 2000). A useful procedure is to compute the proportion of variance explained by a fixed number of PCs through time using the PCA on a rolling window of sample returns. This results in a time series of the correlation measure and reveals the temporal dynamics of the degree of interconnectedness between the assets. Notably (Kritzman et al., 2011) calculate the AR for returns on the MSCI USA equity index. They find that, between January 1998 and October 2010, all 1% worst monthly drawdowns are preceded by a one-standard-deviation spike in the AR. In a related study (Pukthuanthong and Roll, 2009) use an R-square similar to the changes in the AR proposed by (Kritzman et al., 2011) to provide a measure of integration. They point out that integration is only a necessary but not sufficient condition for the identification of periods of heightened systemic risk. As a result (Pukthuanthong and Berger, 2012) extend the integration analysis of (Pukthuanthong and Roll, 2009) to obtain a time-varying measure of systemic risk within international equity markets. They find that an increase in their measure of systemic risk leads periods in which the probabilities of market crashes, and of joint co-exceedances across markets, increase substantially. In an out-of-sample analysis, conditional on the AR exceeding a certain threshold, stocks with higher contribution to the risk of the whole system suffered statistically larger losses during the Financial Crisis (Billio et al., 2012). In Meng et al. (2014) the authors study returns on U.S. real-estate prices in 51 states and use the AR to analyze correlation dynamics. Regarding the question of whether housing bubbles can be identified in advance the authors point to a gradual increase in the largest eigenvalue from 1993. More recently the PCA has been applied to returns on volatilities implied from options in order to assess the systemic importance of various underlying equities (Doris, 2014). Stocks whose correlation matrix of option-implied volatility returns has relatively large eigenvalues are classified as systemic and those with relatively small eigenvalues are classified as idiosyncratic. In addition the authors apply a rolling-window PCA procedure to obtain a time series of the normalized largest eigenvalue and number of eigenvalues exceeding the Marčenko-Pastur law (See Appendix B.3 for discussion of this law). Since the limits of the support of the Marčenko-Pastur law are based on the a white-noise assumption, eigenvalues that exceed the upper bound are considered significant in terms of the content of their information about the system's structure. The authors find that, during high-volatility periods, the largest eigenvalue increases, whereas the number of eigenvalues beyond the theoretical boundary decreases.

The most important conclusion that is common to studies using the AR to analyze the relationship between return comovement and crises is that severe downturns are preceded by or, at least, coincide with increases in the AR. The explanation provided for this phenomenon is that a higher degree of integration among financial institutions facilitates the ripple effect that typically characterizes financial crises. However it should be noted that these conclusions have been reached mostly based on

investigations that are limited to certain asset classes and over certain time periods. Specifically and as mentioned earlier, the Financial Crisis is by far the most common case study and equities are the most commonly studied asset class. In this paper we expand the universe of asset classes to include equities, bonds, currencies and CDS contracts. In addition to the Financial Crisis, we also study two other crises of systemic nature, namely the European Debt Crisis and the Asian Financial Crisis.

3. Data description and exploratory data analysis

3.1. Data description

In this study we focus on three episodes of systemic crises of notable nature, namely the Financial Crisis, the European Debt Crisis, and the Asian Financial Crisis. To investigate each crisis we collect data for groups of assets that potentially have played an important role in its causes and consequences. We also obtain data on at least one broad-based index to be used as a measure of distress for each crisis. A summary of asset groups and indices used is given in Table 7.

The NYSE Financial Index (NYK) covers NYSE-listed common stocks that belong to the Financial Sector according to the Industry Classification Benchmark. Components in the Index represent eighteen countries globally and several industries including banking, insurance, financial services and real estate investment. The market capitalization of NYK components represents a significant portion of the total market capitalization of the Financial Sector in the United States and globally (NYSE Financial Index, 2014). We obtain daily logarithmic returns between January 1, 2003 and December 31, 2013 (for a total of 2769 observations) on 248 stocks that were components of the NYK (at the time the data were being collected for this study).

We also obtain spreads for five-year senior-debt Credit Default Swaps (CDS) contracts on major global financial institutions for the period January 1, 2004–December 31, 2013. The list of entities for this data set is given in Table 1.

Table 1. Institutions included in data set of Financial Sector CDS spreads.

Financial Institutions in CDS Data		
ACE Limited	AIG Group	The Allstate Corporation
American Express	Banco Santander	Banco Bilbao Vizcaya Argentaria
Barclays PLC	BNP Paribas SA	Citigroup Inc
Commerzbank AG	Credit Agricole SA	Credit Suisse Group AG
Deutsche Bank AG	Goldman Sachs Group Inc	HSBC Holding PLC
ING Group	JP Morgan Chase	LCL SA
Lloyds Banking Group	Mitsubishi UFJ Financial Group	Morgan Stanley
Nomura Holding Inc	Royal Bank of Scotland Group	UBS AG

Next daily values of the S&P 500 (denoted as SP500) index are collected for the period matching financial equity and CDS data. We also gather ten-year yields on European Sovereign Debt between January 1, 2006 and December 31, 2013 on bonds issues by the countries in Table 2. All of these countries had a significant exposure to the European Debt Crisis due to their mutual economic ties.

Table 2. Countries included in data set of ten-year bond yields.

Countries Represented in Bond Yield Data		
Austria	Hungary	Portugal
Belgium	Ireland	Spain
Denmark	Italy	Sweden
Finland	Netherlands	Switzerland
France	Norway	United Kingdom
Germany	Poland	

We collect five-year CDS spreads between January 1, 2006 and December 31, 2013 on on European Sovereign Debt issued by the countries in Table 3. Again all of these countries had a significant exposure to the European Debt Crisis due to their close economic ties.

Table 3. Countries included in data set of five-year CDS spreads for government-issued debt.

Countries Represented in CDS Spread Data		
Austria	Hungary	Slovakia
Belgium	Italy	Spain
France	Poland	
Germany	Portugal	

The Bloomberg European Financial Index (BEFINC) is a cap-weighted index of the most highly capitalized European companies that belong to the financial sector and trade on European exchanges. We obtain daily values of the index between January 1, 2006 and December 31, 2013.

We also collect the daily spot exchange rate between the Euro and U.S. dollar (EURUSD) between January 1, 2006 and December 31, 2013. The exchange rate was commonly viewed as efficient in responding to events throughout the crisis due to its implications on the stability, and, hence, the supply-demand balance, of the common currencies (Figure 1).

In addition we gather the spot exchange rate of the domestic currency of various countries/regions in the region of the Asian Financial Crisis versus the U.S. dollar. Included countries/regions are listed in Table 4. The data is daily, spanning from May 31, 1995 to December 31, 1999.

Table 4. Countries/regions (exchange rate tickers) represented in Asian currency data.

Countries Represented in Currency Data	
Australia (AUDUSD)	Philippines (PHPUSD)
Burundi (BNDUSD)	Singapore (SGDUSD)
Indonesia (IDRUSD)	South Korea (KRWUSD)
India (INRUSD)	Taiwan (TWDUSD)
Japan (JPYUSD)	Thailand (THBUSD)

For each country/region in Table 5, we obtain daily closing prices of a major domestic free-float equity index between April 15, 1995 and December 31, 1999. The companies represented collectively in these indices account for the vast majority of economic output in the region directly affected by the Asian Financial Crisis.

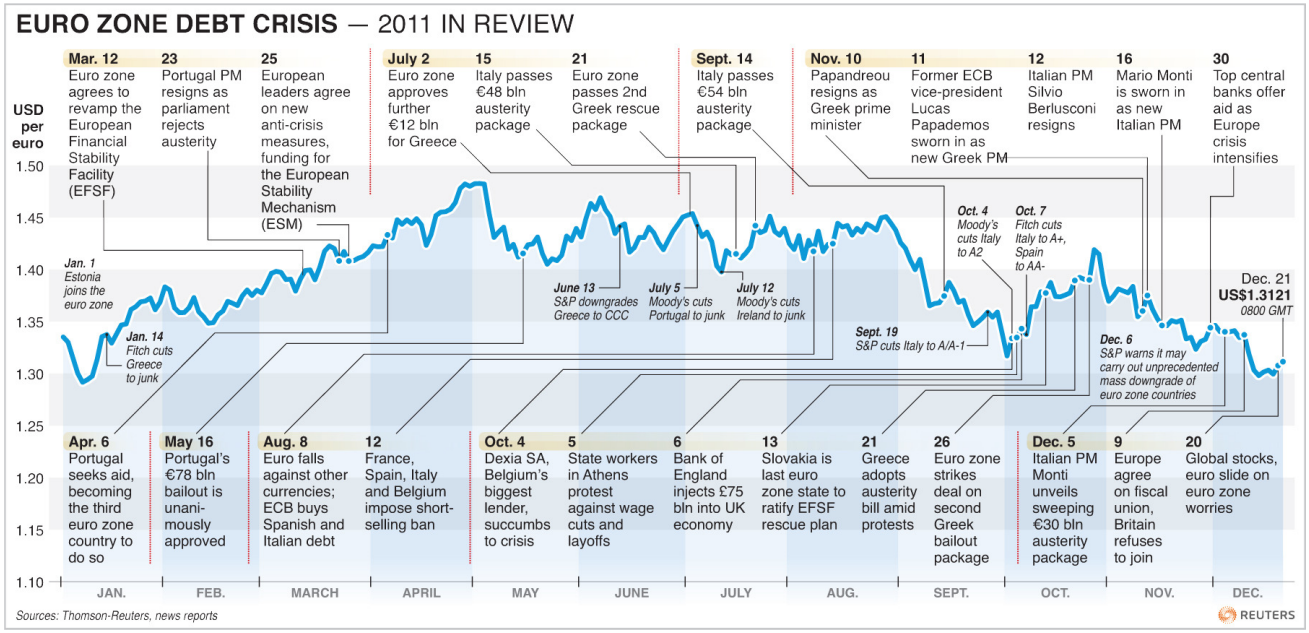

Figure 1. A time line of major events related to the European Debt Crisis and fluctuations in the EURUSD rate. Source: *Reuters*.

Table 5. Countries/regions (equity index tickers) represented in Asian equity data.

Countries Represented in Asian Equity Data	
China (SHCOMP)	Philippines (PCOMP)
Hong Kong (HSI)	Singapore (SGX)
Indonesia (JCI)	South Korea (KOSPI)
Japan (NIKKEI)	Taiwan (TWSE)
Malaysia (KLCI)	Thailand (SET)

Lastly we note that the MSCI AC Asia Pacific Index (MXAS) captures large and mid cap representation across the 13 countries/regions in the Asia Pacific region as listed in Table 6. With 989 constituents, the index covers approximately 85% of the free float-adjusted market capitalization in each country/region. Daily closing prices between April 15, 1995 and December 31, 1999 are collected.

3.2. Exploratory data analysis

3.2.1. Global financial crisis of 2008

The systemic nature of the Financial Crisis is due to links among financial institutions that were established using contingent claims such as the credit default swap (CDS). The web of contractual relationships that resulted from a proliferation of claims contingent on defaults across the financial industry served as a mechanism by which losses would propagate. The interdependence in bank solvency was so profound that over 270 banks collapsed within two years since September 2008, when Washington Mutual Inc. became the biggest bank that fell down on record (Smith and Sidel,

Table 6. Countries/regions represented in MSCI Asia Pacific Index.

Countries/regions Represented in Index for the Asian Financial Crisis		
Australia	China	Philippines
Hong Kong	India	Taiwan
Japan	Indonesia	Thailand
New Zealand	South Korea	
Singapore	Malaysia	

Table 7. Summary of asset groups and indices used in analyzing each crisis episodes. The AR is calculated for each asset group and its relationship with the level of distress is studied using the relevant indices as proxies.

Crisis	**Assets**	**Indices**
Financial Crisis	Equities	SP500
	CDSs	
European Debt Crisis	Bonds	BEFINC
	CDSs	EURUSD
Asian Financial Crisis	Currencies	MXAS
	Equities	

2010). Since these links were formed and strengthened over time, it is plausible that the increased economic dependence among financial institutions resulted in a stronger comovement in the value of the assets of these institutions.

The first data set for the financial sector comprises daily logarithmic returns between January 1, 2003 and December 31, 2012 (for a total of 2769 observations) on 248 stocks that were components of the NYK (as of the time the data were collected). We estimate the correlation matrix by using a rolling window of 252 trading days and perform a PCA to obtain the AR measures. We plot the AR time series using different numbers of eigenvalues in Figure 2. The resulting time series exhibit a highly nonlinear and potentially nonstationary pattern over time. The nonstationarity of these time series is confirmed by evidence obtained from ADF, PP and KPSS unit-root tests (See Appendix C). This is particularly the case for the first eigenvalue, whose relative magnitude varies dramatically, with a range from 25% to 65%. It appears that the comovement in returns has strengthened gradually in the years leading up to the Financial Crisis and remained relatively elevated since that time. Specifically the average before September 1, 2008 was 34% and 53% after. Apart from being richly dynamic the estimated AR time series for the first dominant PC, $\hat{\phi}_1$, as defined in Appendix A.3., explains a relatively large proportion of the variance; it is, at its lowest, about 16 times larger than the upper support of the Marčenko-Pastur distribution with a commensurate aspect ratio parameter. On average, the first PC explains approximately 44% of the variance. The first 25 PCs, a ten-th of the number of variates, account for approximately 70% of the variance on average and as much as 80% at the apex of the crisis.

We also analyze 2051 observations of logarithmic returns on CDS spreads for 24 major global financial institutions. A time series of CDS spreads for some major U.S. financial institutions is plotted in Figure 3. As with our analysis of financial equity returns we plot the AR time series for CDS data using a 252-day rolling window in Figure 4. Again the resulting time series appear to be highly

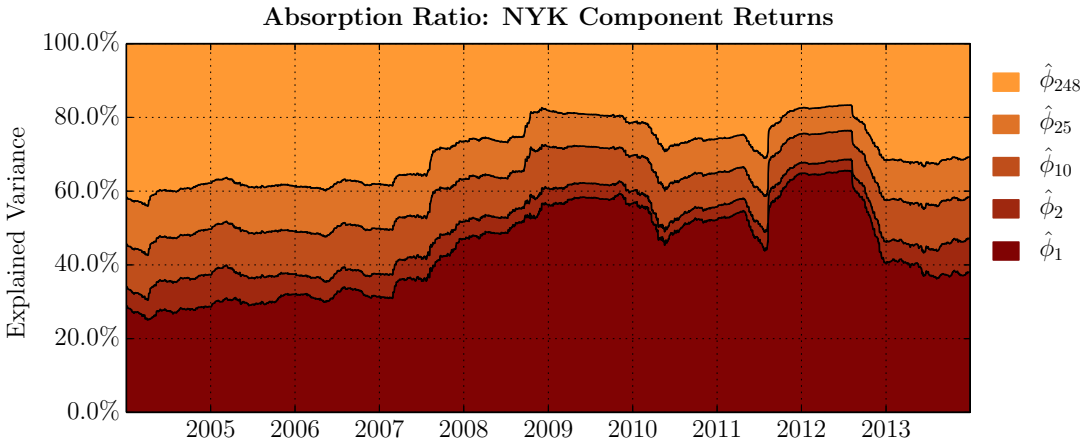

Figure 2. Eigenspectrum analysis for returns on components of the NYSE Financial Sector Index (NYK).

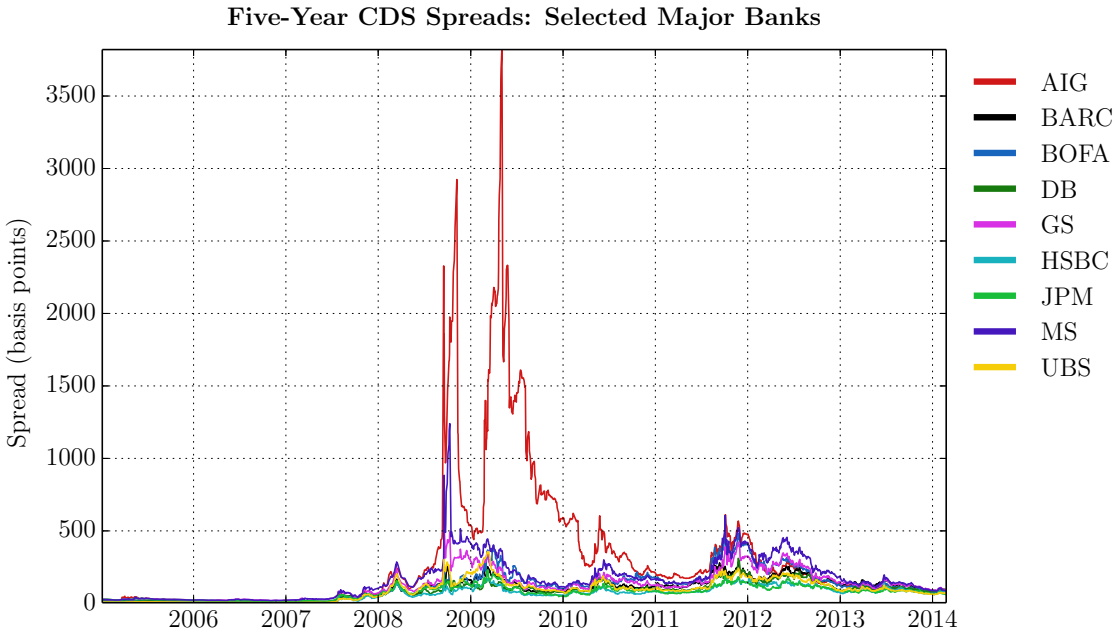

Figure 3. Five-year CDS spreads in basis points on selected major banks.

nonlinear and possibly nonstationary. The latter is supported by statistical evidence obtained from the ADF, PP and KPSS unit-root tests. We also see that correlations have remained strong after the crisis in CDS spread returns. The relative magnitude of the largest eigenvalue varies approximately from 17% to 66% and much higher than the upper bound predicted by the RMT. There is a dramatic spike occurring in June, 2007 that is followed by a steady increase throughout the financial crisis. On average, $\hat{\phi}_1$ explains 47% of the total variance.

To illustrate the relationship between SP500 and the proportion of variance explained by the first PC in our data sets we overlay their values in Figure 5. Sharp increases in $\hat{\phi}_1$ are visually evident before critical events such as major bank failures. This suggests that a systemic risk monitoring scheme based

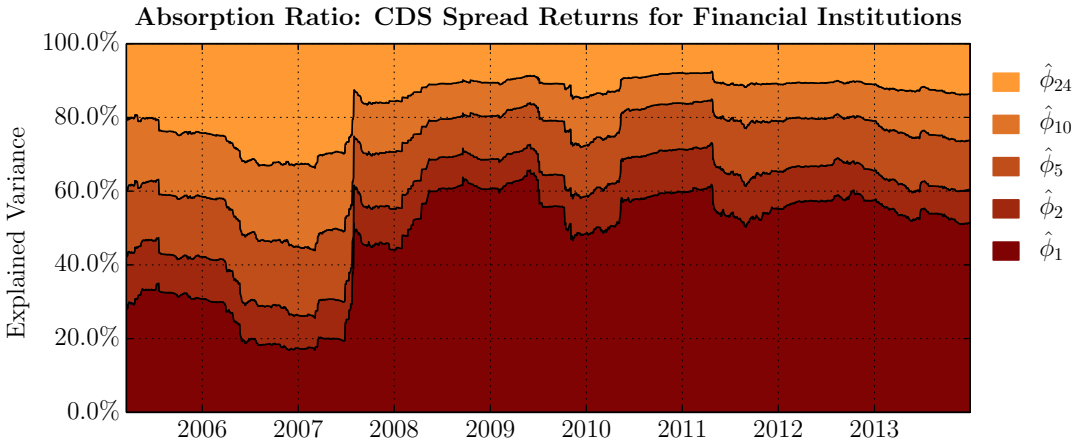

Figure 4. Eigenspectrum analysis for returns on spreads of financial institution CDSs.

on the level of $\hat{\phi}_1$ could potentially raise a red flag before the crisis reached its height. However, if there is any predictive component to the information content of $\hat{\phi}_1$, it is certainly mixed with some coincident response to shocks. For example we observe a dramatic spike in equity correlations in the summer of 2011, which coincides with the credit rating downgrade of U.S. sovereign debt. However there is no apparently significant change in the correlation structure prior to this event.

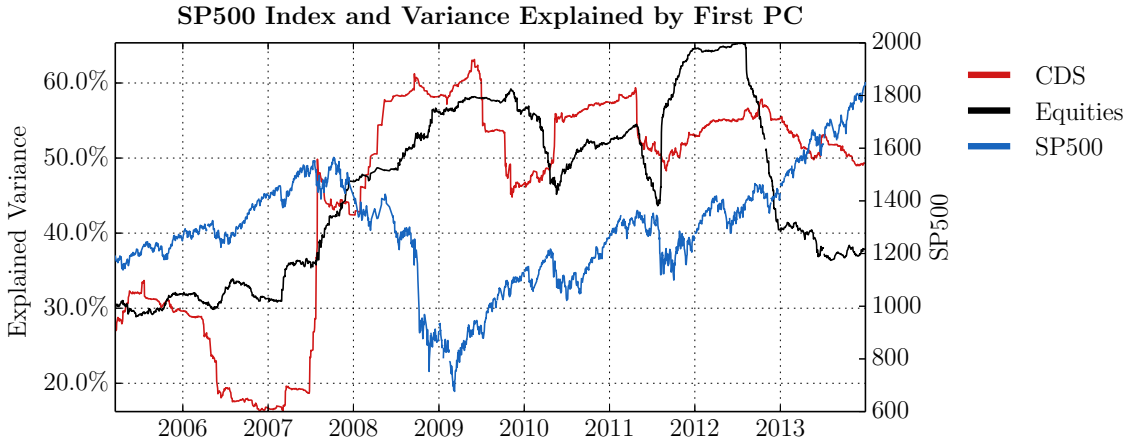

Figure 5. Proportion of variance explained by first PC in CDS and NYK component (equity) data, compared with the level of SP500.

Our observations thus far are consistent with the existing work in the literature for the case of the Financial Crisis; that is periods of turmoil in equity markets appear to have been preceded by or coincide with increases in correlation of asset returns.

3.2.2. Eurozone sovereign debt crisis of 2009/2010

Importantly the European Debt Crisis is an example of a systemic shock in fixed income assets. It is similar to the Financial Crisis in that it was characterized by financial contagion. While the cornerstone of the crisis involved certain members of the currency union, known as peripheral countries, financial

interlinkages meant that countries that are not necessarily in the eurozone would also be affected, by virtue of their membership in the European Union, for example. Thus the analysis in the extant literature and our work so far is consistent with the setting of this crisis. In this paper we extend the analysis by studying the relationship between correlation and contagion in the European sovereign credit market.

In particular we focus on the information contained in returns on sovereign bond yields and CDS spreads of countries in the EU*. The first data set is comprised of daily logarithmic returns on ten-year yields for government debt issued by the countries listed in Table 2. The second data set consists of daily logarithmic returns on five-year CDS spreads on sovereign debt issued by the countries in Table 3.

It should be pointed out that the variates in the system ought to be temporally homogeneous in order to preserve comparability and the power of inferential analysis. A missing value in one variate would require filling or else the time point would need to be discarded altogether. Therefore we have excluded certain countries from the analysis due to limited data availability. For example Greece stands out as the country that had endured the most significant rise in its cost of borrowing, as well as CDS spreads reflecting expectations of near certain default. Greece indeed practically defaulted on its debt as it had undergone restructuring, with private bondholders accepting deep haircuts, and a downgrade of its debt to 'Restricted Default' rating (Fitch cuts Greeces issuer default ratings to RD, 2012). As a result, data on bond yields and CDS spreads was not available after March 9, 2012 and September 16, 2011, respectively.† Comparing results using samples with and without data on Greek debt for available time periods we find that, despite Greece's prominent role in the crisis, the results are not materially different. For this reason we exclude data on Greek debt from our analysis.

The cost of debt and its insurance would reach historically high levels for multiple countries, particularly Greece, Ireland, Italy, Portugal and Spain. A time series plot of yields for selected eurozone countries is presented Figure 6. A similar plot of CDS spreads is presented in Figure 7. It can be seen both in yield and CDS spread data, which are related to sovereign creditworthiness in the eurozone, started to rise in 2009.

In calculating the AR we again keep the size of our rolling estimation window at 252 trading days. Figure 8 plots the temporal evolution of the AR for bonds in this system. The correlation structure of this time series are nonlinear and nonstationary. This visual pattern of nonstationarity is further supported by statistical evidence obtained from the ADF, PP and KPSS unit-root tests. However we observe that the correlation dynamics in this case are very different from our results for the Financial Crisis. There is a notable decline in the AR of a small number of eigenvectors through the crisis. Returns on sovereign bond yields in Europe were highly correlated before the Financial Crisis, with $\hat{\phi}_1$ explaining approximately 80% of the variance. Interestingly, near the end of 2008, when the Financial Crisis was at its peak, the strength of comovement began to decrease steadily. There was also a sharp shift in the structure at the start of May 2010, when the first PC decreased in dominance compared with the rest. The aspect ratio is 0.0675 and the Marčenko-Pastur density with this parameter has an upper support of 1.58, so that the limiting theoretical (normalized) maximum of the largest eigenvalue, assuming white-noise data, is approximately 9.3% or about 7 times smaller than what we observe on

*The vast majority of the countries were *de jure* EU members (at the time when the data are being collected for a preliminary study in August of 2014), although we also study countries who have adopted provisions in order to participate in the EU single market without membership.

†Greek bond yield data is missing for approximately one year.

average.

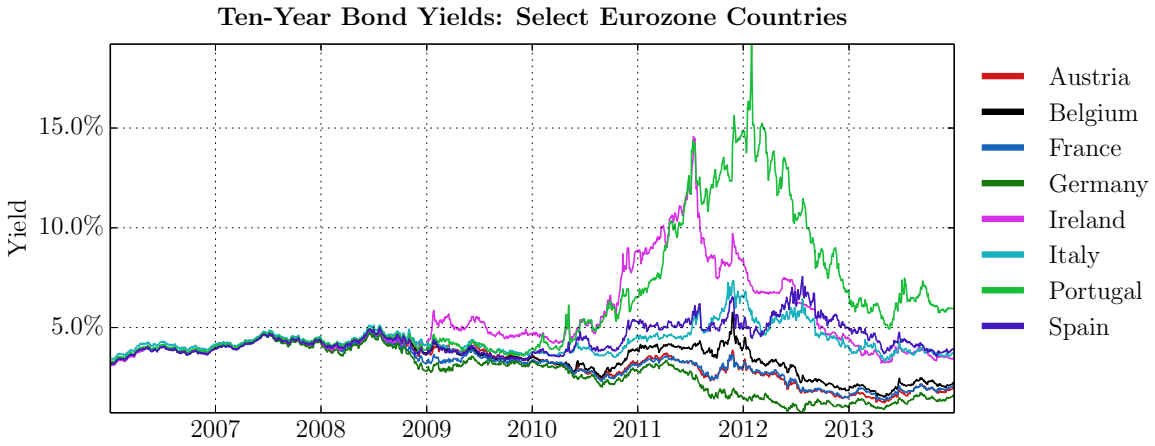

Figure 6. Ten-year yields on sovereign debt issued by selected eurozone member states.

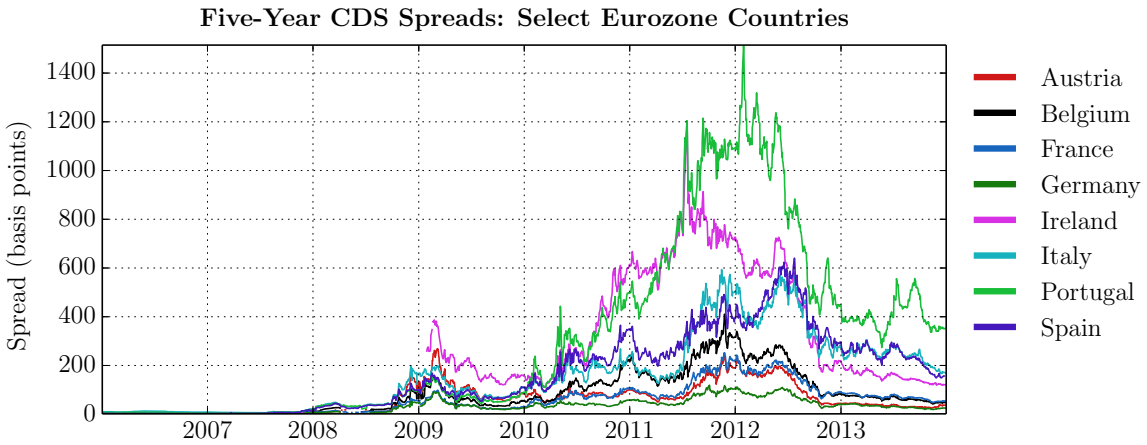

Figure 7. Five-year CDS spreads in basis points on selected eurozone member states.

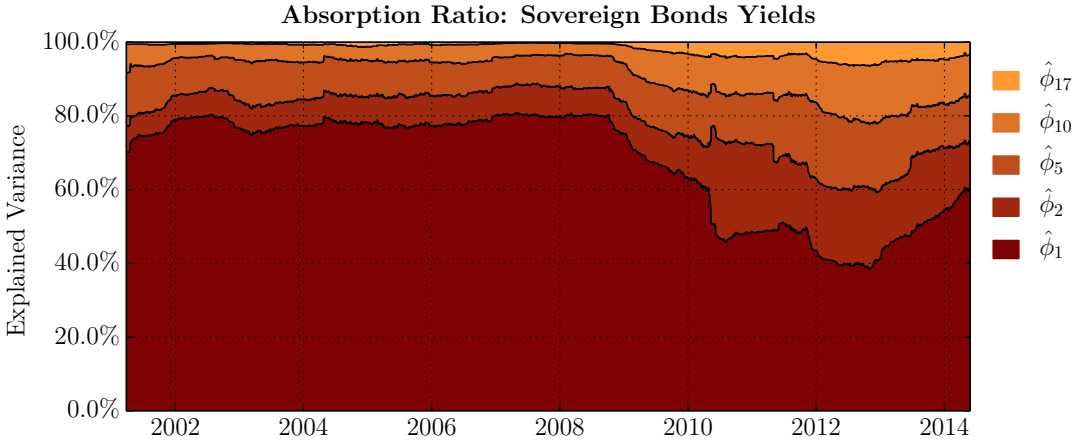

Figure 8. Eigenspectrum analysis for returns on eurozone sovereign bond yields.

It can also be seen that $\hat{\phi}_1$ began to increase through the year of 2013, as yields and CDS spreads have come down in response to improving economic conditions. Our results suggest that the comovement in bond yields is found to be weakened during the period of the crisis due to flight to quality. Said otherwise bond yield fluctuations have diverged as bonds of certain countries became an alternate, safer, source of yield vis-à-vis those experiencing fiscal distress. For example, German, French and Austrian bond yields were decreasing, while Portuguese, Italian, Irish and Spanish yields were rising for certain periods (see Figure 6). This is interesting because decreasing yields should reflect an improvement in the issuing entity's creditworthiness. Analysis of the creditworthiness of countries is beyond the scope of this work. However, when we construct ARs for the subset of sovereign bonds of the core and peripheral countries separately,[‡] we find a discernibly increasing risk of default among core countries, due not least to uncertainties regarding the stability of the euro currency. The mere exposure of core country governments and banks to distressed debt put, *ceteris paribus*, an upward pressure on core country yields as well. Thus the divergence in yield fluctuations is in contrast to the fundamental risk-return relationship.

Next we consider the AR for the CDS spreads, plotted in Figure 9. The correlation dynamics for sovereign debt CDS spread fluctuations resemble their analogue for debt issued by our sample of financial institution in Section 3.2.1. We see a gradual increase in the AR starting in the latter half of 2008 and continuing through the year of 2009. In this shift $\hat{\phi}_1$ rises from about 23% to near 60%. This implies a significant increase in the strength of comovement between CDS spreads. We also find that comovement has begun a gradual yet significant weakening starting in late 2012. Based on a comparison with results from the RMT, as before there is evidence of non-random structure since the largest eigenvalue is, on average, about 4 times greater than the theoretical upper bound.

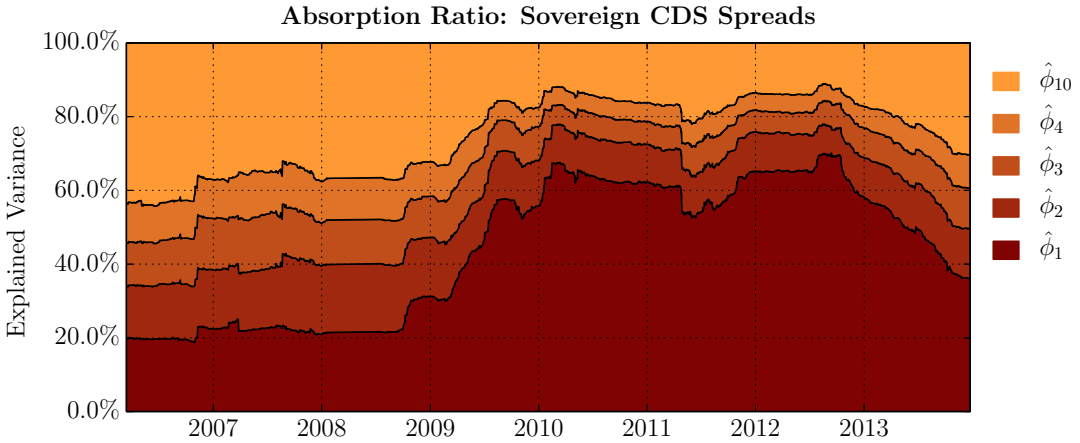

Figure 9. Eigenspectrum analysis for returns on eurozone sovereign CDS spreads.

There are notably different comovement dynamics on the sovereign CDS and the underlying bond market during the European Debt Crisis.[§] Briefly our results suggest an inverse relationship where bond yield fluctuations diverged during market distress and converged in times of relative tranquility. On the other hand CDS spreads returns converged in times of distress and diverged otherwise, in broad agreement with what we have observed when analyzing correlations during the Financial Crisis. These

[‡]We thank one of the referees for this suggestion.
[§]We thank one of the anonymous referees for making this point.

relationships are visualized in Figure 10, where the largest eigenvalue of each asset class is compared with BEFINC, and likewise in Figure 11 for the EURUSD rate.

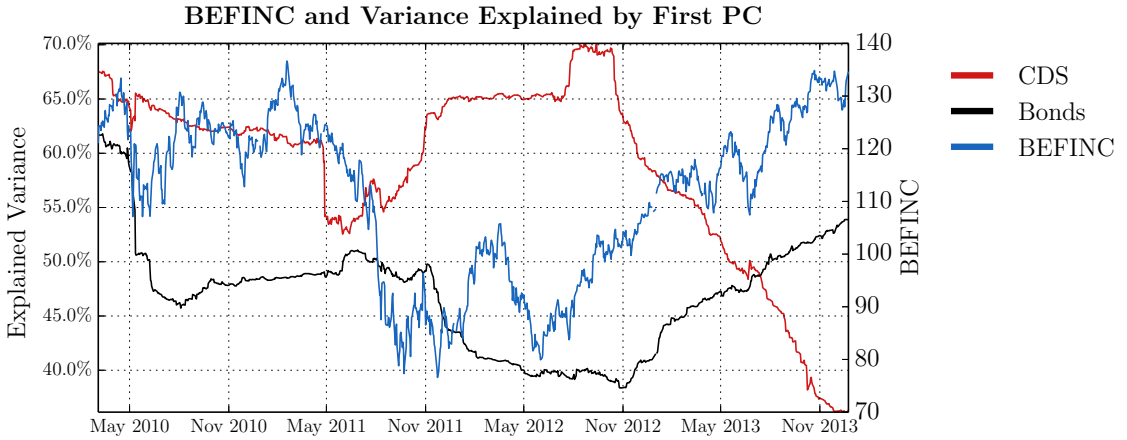

Figure 10. Proportion of variance explained by first PC in CDS and bond yield data, compared with the level of BEFINC.

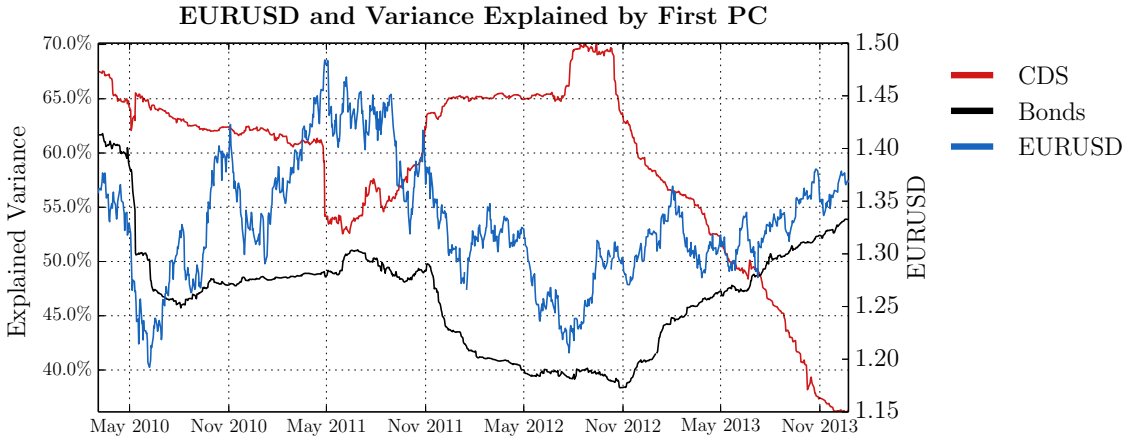

Figure 11. Proportion of variance explained by first PC in CDS and bond yield data, compared with the level of EURUSD.

It is important to reiterate that, in this paper, we study the European Debt Crisis as a separate crisis despite its clear connection with the events associated with the Financial Crisis. This is because our interest in this paper is in the correlation dynamics of a basket of assets that were central to the sovereign debt crisis and their relationship with a broader measure of the system's economic stability. Before a recognition of fiscal trouble in the eurozone, such measures reflect distress from the Financial Crisis that is independent from the effects of what is now deemed as a loss of confidence in the ability of multiple countries to make good on loans. Therefore, in studying the connection between correlations and market turmoil as per a broad, observable indicator, our period of study for this crisis begins on March 9, 2009.[¶]

[¶]Major equity indices in the U.S. have reached their low on this date and it has been used as a starting point for a post crisis analysis in reference to the Financial Crisis (Daniel and Moskowitz, 2016).

3.2.3. Asian Financial Crisis of 1997

In 1997 the global financial system experienced the ripples of a systemic shock to the economies of a few countries in East Asia. In particular in late 1996 and early 1997 Thailand's currency, the baht, was subject to speculative attacks from traders. Short sellers had begun to recognize the weakness of the baht in terms of its relatively low demand compared to the U.S. dollar, which was required to service debt, and a diminishing ability of the government to defend it with (still depleting) foreign currency reserves. In July 1997, when the Thai government was no longer able to defend the baht, it was allowed to float freely. The result was an immediate sharp decrease in its value. The baht continued to depreciate through the year, making the debt burden of Thai companies increasingly difficult to manage to the point that many of them were forced into bankruptcy. At the same time these events were coupled with a significant decline in the Stock Exchange of Thailand. Panic spread to neighbouring countries as various Asian currencies experienced similar speculative attacks. Within two months depleting foreign exchange reserves forced Malaysia, Singapore and Indonesia to drop the peg of their currencies to the U.S. dollar. In each case the move was met with a sharp devaluation of the local currency and equities. The shock would significantly affect South Korea, Japan, Taiwan, Philippines, Laos. This can be seen in Figure 12 and Figure 13, where we plot the returns currencies and equity indices of several Asian countries/regions. Other countries/regions were affected to a lesser extent, such as China, which saw reduced growth rates following the crisis.

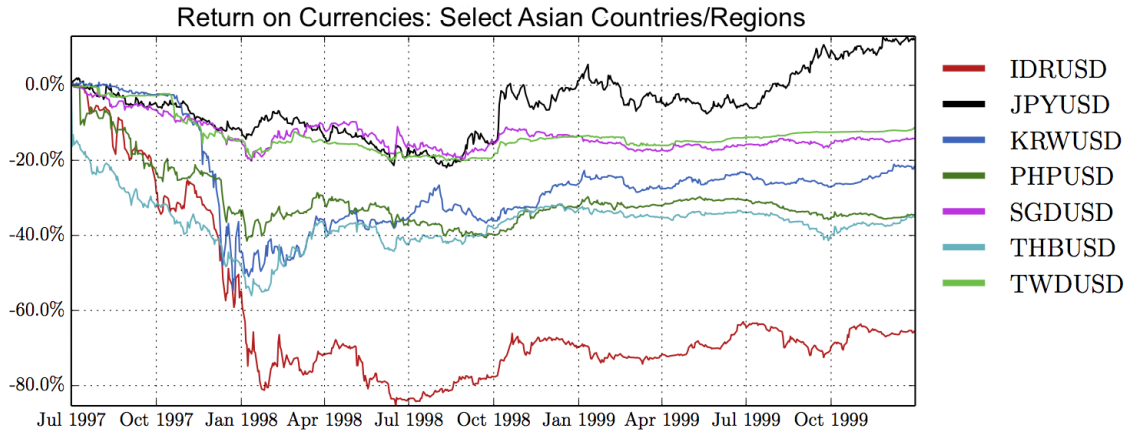

Figure 12. Returns on currencies of Asian countries/regions affected by the Asian Financial Crisis.

The developments in this particular crisis make it a compelling subject of investigation in our study due to the spectacular contagion effect. We again seek to gain insight into the correlation dynamics of tradable assets and potential relationships with returns in times of distress. The assets that we investigate are a basket of currencies of Asian countries/regions and a basket of major stock indices. It is important to note that for the period that a given currency is pegged, the price discovery process is critically altered by government intervention. This affects any conclusions drawn about the dynamics of the AR prior to the crisis. However the returns of pegged currencies are not identically zero as can be seen in Figure 14. The currencies fluctuate within a narrow band under normal conditions. During an unusual trading activity, such as speculative attacks, fixed exchange rate bands can be breached. Therefore there is still useful information content in currency returns even under a fixed exchange-rate

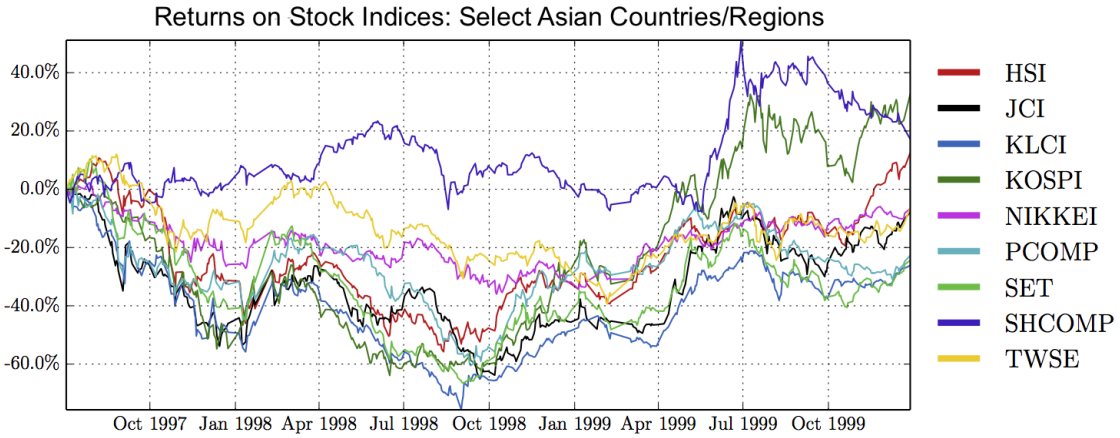

Figure 13. Returns on equities of Asian countries/regions affected by the Asian Financial Crisis.

regime.

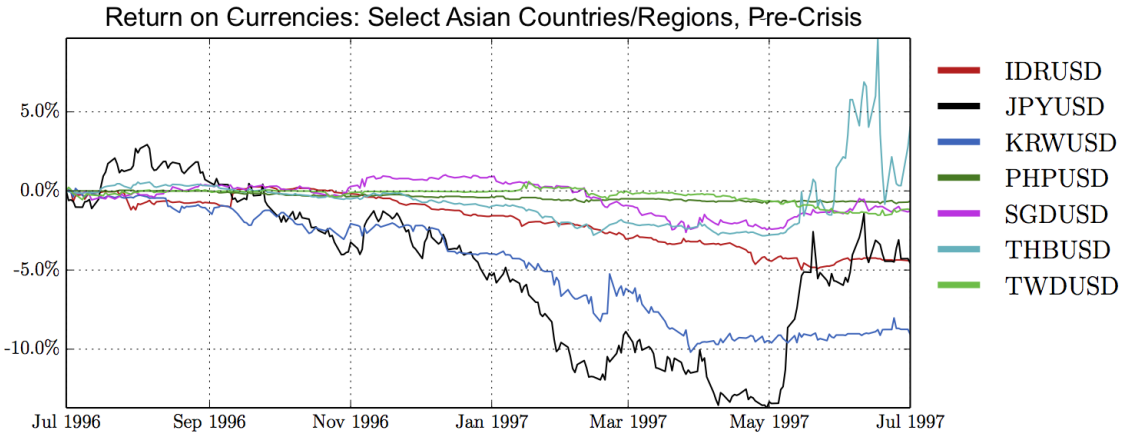

Figure 14. Returns on currencies of select Asian countries/regions before local currency pegs to the US dollar were dropped by affected countries.

As in the previous analysis we maintain the size of our rolling estimation window for the AR at 252 trading days. Figure 15 plots its the temporal evolution for returns on Asian currencies versus the U.S. dollar. Importantly the correlation structure of the returns appears to be highly nonlinear and possibly also nonstationary. The latter is further reinforced by statistical evidence obtained from ADF, PP and KPSS unit-root tests. We observe a strengthening in correlation beginning in mid to late 1997. Similar features are present in equity data which we use to generate the AR series In Figure 16. The aspect ratio of the correlation matrices for both asset baskets is $y = \frac{10}{252}$. Nevertheless the distribution of eigenvalues differs with equity returns being more tightly coupled as $\hat{\phi}_k$ is higher in equities for each k. In particular $\hat{\phi}_1$ is 31% for equities and 21% higher on average over the same period. With the aspect ratio at hand the upper support of the Marčenko-Pastur density is 14%.

Both equities and currencies appear to have increasingly correlated returns in times of crisis. An

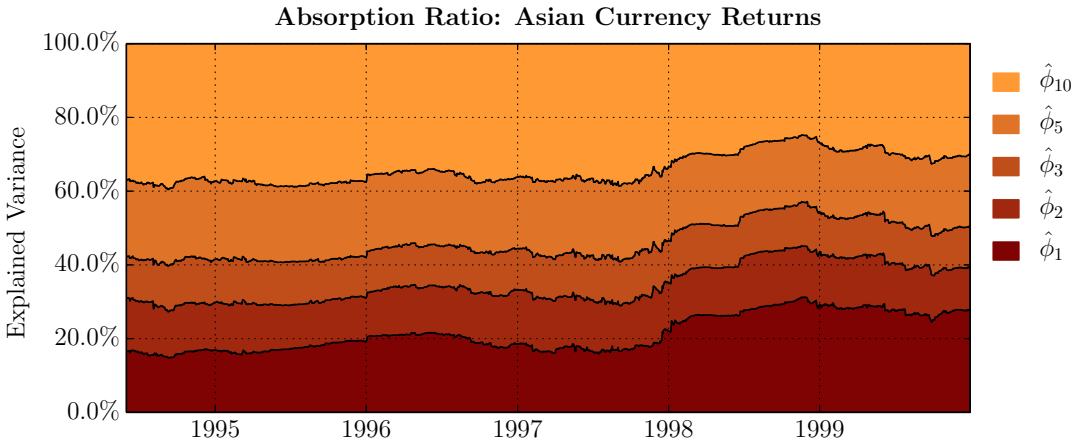

Figure 15. Eigenspectrum analysis for returns on currencies of countries/regions in Asia that were significantly affected by the Asian Financial Crisis.

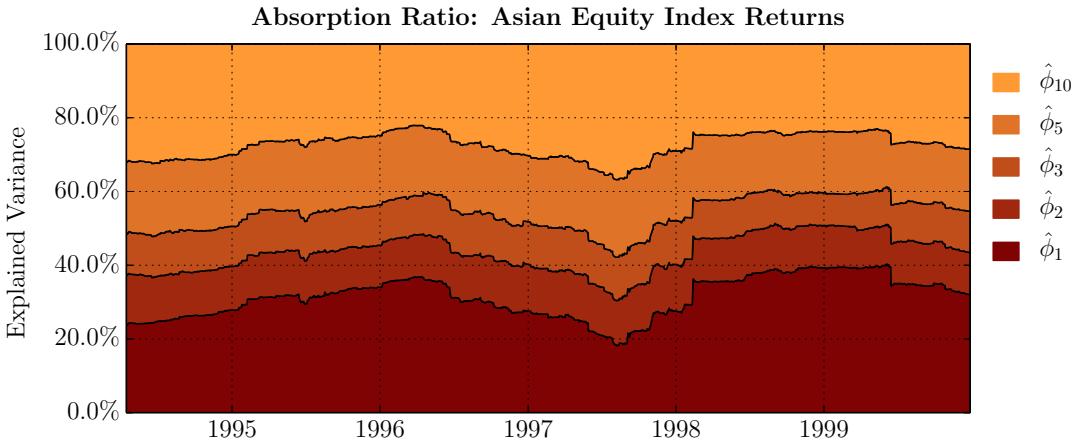

Figure 16. Eigenspectrum analysis for returns on major stock indices of countries/regions in Asia that were significantly affected by the Asian Financial Crisis.

inverse relationship between the AR and the MXAS index can be seen in Figure 17. For equities, this reinforces our earlier findings from Section 3.2.1. It should be pointed out that currencies have been relatively unexplored in this type of study. Apart from the directional relationship there is the important question of whether correlations tend to strengthen before crisis episodes and, thereby, foster the type of contagion that characterizes systemic events. Figure 17 provides evidence to the contrary. Returns on equities and currencies in the region had been decoupling from mid 1996 until the onset of the crisis.

Plotting a time series of the AR provides a visual summary of the strength of comovement in the data variates. In this section we found some evidence for a relationship between the AR and the level of distress in the system as measured by a broad index during various crises. We found that returns on stocks and CDS spreads of financial institutions appeared to have an inverse relationship with the NYSE Financial Sector Index during the Financial Crisis. Furthermore correlations in these assets had been increasing before the crisis attained its peak. Returns on CDS spreads of European sovereign debt had a similar behaviour around the Eurozone Sovereign Debt Crisis. However returns on yields

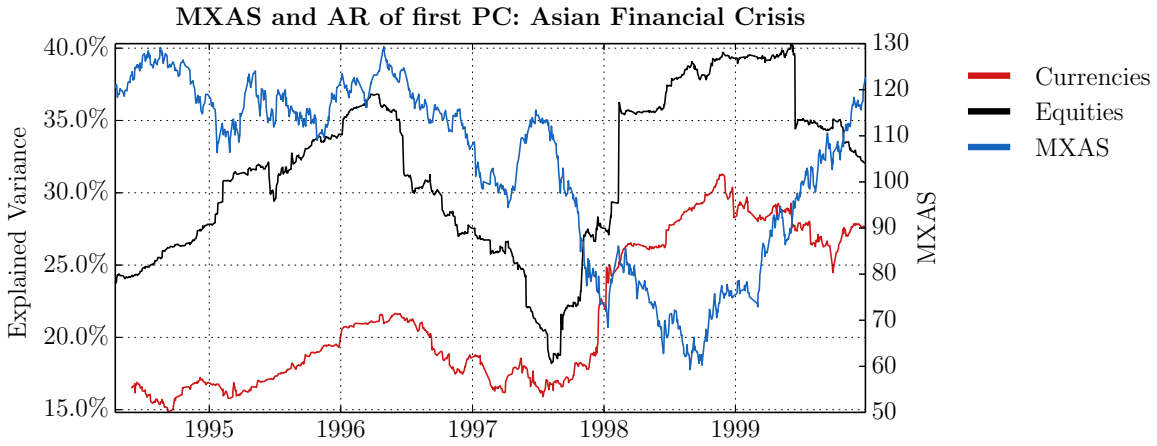

Figure 17. Proportion of variance explained by first PC in currency and equity yield data, compared with the level of MXAS.

of bonds issued by these countries had become less correlated. Finally higher correlation in stock and currency returns around the Asian Financial Crisis appears to be associated with higher levels of distress.

The AR time series are obtained through the PCA on return observations using the covariance method in a rolling-window fashion. That the AR series is dynamic suggests that individual asset data is not second-order stationary. This supports a known stylized fact that financial data exhibits a pattern of nonstationarity, which greatly impedes the robustness of models fitted to historical samples using standard methods.

4. Change-point analysis

Change-point detection is a process of identifying abrupt temporal changes in a stochastic process (Basserville and Nikiforov, 1993). Change is determined in terms of distributional properties of an underlying process. The problem of detecting abrupt changes in the statistical behavior of an observed signal or time series is a classical one, whose provenance dates back at least to the work in the 1930s on the problem of monitoring the quality of manufacturing processes (Wetherhill and Brown, 1991). More recently this problem has been studied in a wide array of fields including econometrics (Andersson et al., 2004, 2006; Andrews et al., 1996; Berkes et al., 2004; Broemling and Tsurumi, 1987), environmental science (Petterson, 1998), finance (Beibel and Lerche, 2000; Chen and Gupta, 1997; Shiryaev, 2002), image analysis (Trivedi and Chandramouli, 2005), medical diagnosis (Petzold et al., 2004) and network security (Tartakovsky et al., 2006; Thottan and Ji, 2003), among others. Change-detection serves a broad range of purposes and detection procedures are commensurately diverse. Perhaps, at the most basic level, they differ in being *offline* or *online*. Offline algorithms are algorithms which are applied to a sample of historical observations. The purpose of this tool is to detect changes in an *a posteriori* fashion within the sample with the smallest probability of error. In particular online algorithms are designed for real-time monitoring for changes in a stochastic process. Online detection procedures aim to minimize both the false alarm rate and the detection delay, as each new sample is assumed to incur a cost. We will briefly discuss the online change-detection here.

Consider a sequence $\{X_t\}$ of i.i.d. observations and an associated minimal filtration $\{\mathcal{F}_t\}$. Suppose that X_t obeys one of the following hypotheses:

$$H_0: \quad X_t \sim P, \quad t = 1, 2 \ldots \tag{1}$$
$$H_1: \quad X_t \sim Q, \quad t = 1, 2, \ldots \tag{2}$$

where P and Q are two distinct distributions with probability density functions p and q, respectively. Let $L(X) = \dfrac{q(X)}{p(X)}$ be the likelihood ratio and define

$$S_t = \sum_{i=1}^{t} \log L(X_i). \tag{3}$$

Since $S_t = S_{t-1} + \log L(X_t)$, for $t > 1$ we have

$$\mathbb{E}\left[S_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[S_{t-1} + \log L(X_t)|\mathcal{F}_{t-1}\right] \tag{4}$$
$$= S_{t-1} + \mathbb{E}\left[\log L(X_t)|\mathcal{F}_{t-1}\right] \tag{5}$$
$$= S_{t-1} + \mathbb{E}\left[\log L(X_t)\right], \tag{6}$$

where the second equation holds because S_{t-1} is $\mathcal{F}_{t-1}$-measurable and third equation holds because $\log L(X_t)$ is independent of $\mathcal{F}_{t-1}$. Now, under some probability measure F with density f, we have

$$\mathbb{E}_F\left[\log L(X_t)\right] = \int_{-\infty}^{\infty} \log\left(\frac{q(x)}{p(x)}\right) f(x)\,\mathrm{d}x \tag{7}$$
$$= \int_{-\infty}^{\infty} \log\left(\frac{q(x)f(x)}{p(x)f(x)}\right) f(x)\,\mathrm{d}x \tag{8}$$
$$= \int_{-\infty}^{\infty} \log\left(\frac{f(x)}{p(x)}\right) f(x)\,\mathrm{d}x - \int_{-\infty}^{\infty} \log\left(\frac{f(x)}{q(x)}\right) f(x)\,\mathrm{d}x. \tag{9}$$

Under H_0 we have $F = P$, so that the first term becomes zero and we are left with

$$\mathbb{E}_P\left[\log L(X_t)\right] = -\int_{-\infty}^{\infty} \log\left(\frac{p(x)}{q(x)}\right) p(x)\,\mathrm{d}x \tag{10}$$
$$= -D_{KL}(P \parallel Q), \tag{11}$$

where $D_{KL}(P \parallel Q)$ denotes the Kullback-Liebler divergence between P and Q. On the other hand, under H_1, we have $F = Q$ and

$$\mathbb{E}_Q\left[\log L(X_t)\right] = \int_{-\infty}^{\infty} \log\left(\frac{q(x)}{p(x)}\right) q(x)\,\mathrm{d}x \tag{12}$$
$$= D_{KL}(Q \parallel P). \tag{13}$$

Recall that $D_{KL}(A \parallel B) \geq 0$ for any distributions A and B, with strict equality for $A = B$. Thus in equation (6) we have that $\mathbb{E}\left[\log L(X_t)\right] < 0$ under H_0 and $\mathbb{E}\left[\log L(X_t)\right] > 0$ under H_1. It then follows that, under H_1, $\mathbb{E}\left[S_k|\mathcal{F}_l\right] \geq S_l \;\forall\; l \leq k$, so $\{S_t, \mathcal{F}_t\}$ is a submartingale. Similarly under H_0,

$\mathbb{E}[-S_k|\mathcal{F}_l] \geq -S_l \; \forall \; l \leq k$, so $\{S_t, \mathcal{F}_t\}$ is a supermartingale. In fact, in both cases, we have almost sure divergence [70] with

$$S_t \xrightarrow{a.s.} \begin{cases} -\infty & \text{under } H_0 \\ \infty & \text{under } H_1 \end{cases}. \tag{14}$$

If we consider the situation that

$$X_t \sim \begin{cases} P & t < k \\ Q & t \geq k \end{cases}, \tag{15}$$

i.e., there is a change-point in the distribution of X_t at time $k < \infty$, then this change will be reflected as a change in the sign of the mean value of the log-likelihood ratio. In the setting described by expression (15), a typical objective of an online monitoring scheme is to detect the change while minimizing both the time delay in raising an alarm after the change and the rate of false alarms. Of course there is a tradeoff between the two performance criteria and studies in the area of stochastic control have sought to optimize it. For instance Shiryaev (1978) adopted a Bayesian approach whereby the change point is a random variable with a prior distribution. Lorden (1971) formulated the problem with a deterministic change point and sought to minimize the worst-case detection delay subject to a lower-bound on the mean time between false alarms. In both cases the authors arrive at optimal solutions. The cumulative sum (CUSUM) procedure (Page, 1954), which was originally proposed for continuous inspection schemes, was also proved to be optimal for Lorden's formulation by Moustakides (1986). The CUSUM statistic is defined as

$$g_t = \max_{0 \leq k \leq t}(S_t - S_k) \tag{16}$$

$$= S_t - \min_{0 \leq k \leq t} S_k \tag{17}$$

$$= \max\left(g_{t-1} + \log L(X_t), 0\right), \tag{18}$$

where $S_0 \equiv 0$ and S_k is the log-likelihood ratio defined in equation (3). The stopping rule is

$$\tau = \inf\left\{t \geq 1 : S_t - \min_{0 \leq j \leq n} S_j \geq b\right\}, \tag{19}$$

with $b > 0$ representing an alarm threshold. From equation (17) it becomes clear that fact that the expectation of the log-likelihood ratio changes signs under different distributions is a key statistical property. It guarantees the almost sure divergence of g_t so that the threshold b will be crossed after the change-point. The effectiveness of the algorithm depends on the size of that change, i.e.,

$$D(P \parallel Q) = \mathbb{E}_Q[\log L(X_t)] - \mathbb{E}_P[\log L(X_t)] \tag{20}$$

$$= D_{KL}(P \parallel Q) + D_{KL}(Q \parallel P) \tag{21}$$

which is noted to also be the *symmetrized Kullback-Liebler divergence* between P and Q. A larger value of D indicates greater dissimilarity between the distributions. That it is symmetric implies $D(P \parallel Q) = D(Q \parallel P)$, so the statistic can be thought of as a distance more intuitively than the asymmetric Kullback-Liebler divergence.

There are a couple of challenges that are common in change-point detection procedures. Firstly information about the distribution of the series – either before the change, after the change or both – is assumed to be known in parametric methods. In many practical cases this is not the case and, so, assumptions about the underlying distribution need to be made. Secondly the use of thresholds in the decision function introduces subjectivity and threshold parameters likely need to be tuned periodically and between different applications. When one or both of the distributions P and Q are unknown, we can resort to nonparametric methods of change-point detections. This approach is often undertaken due to its practicality, even when optimality cannot be established formally (see Kawahara et al., 2007; Liu et al., 2013; Yamada et al., 2013, for examples). To alleviate robustness issues in dealing with procedures that involve a binary decision on whether change has occurred, we suggest to monitor the symmetrized divergence statistic $D(P \parallel Q)$. As stated before this statistic represents the degree of dissimilarity between two distributions. Many detection procedures rely on the change of sign of the expected log-likelihood ratio after a change-point. This statistic, which is equal to the difference between $\mathbb{E}[\log L(X)]$ before and after the change, has therefore been used as a detectability index. Here detectability can be defined in terms of the performance of change-point detection procedures. For example the average number of samples taken before a decision can be made in some online detection procedures is proportional to the Kullback-Liebler divergence (Basserville and Nikiforov, 1993). In the following section we introduce a nonparametric approach for estimating $D(P \parallel Q)$.

5. Nonparametric monitoring procedure

We adapt a nonparametric estimation method for multivariate stochastic processes in Liu et al. (2013). Let $\boldsymbol{x}_t \in \mathbb{R}^d$ be a d-dimensional time-series and let $\boldsymbol{X}_t = [\boldsymbol{x}_t', \boldsymbol{x}_{t+1}', ..., \boldsymbol{x}_{t+k-1}']' \in \mathbb{R}^{dk}$ be a batch sequence of k observations of $\boldsymbol{x}_t$. Note that $(\cdot)'$ indicates transposition, so $\boldsymbol{x}_t$ is a time-varying vector of length dk. In what follows, instead of using a single observation of $\boldsymbol{x}_t$ as a sample instance, $\boldsymbol{X}_t$ is treated as a sample instance in order to capture the temporal correlation that empirical data tends to exhibit. To detect a change in the underlying distribution a measure of dissimilarity is computed between two groups of temporally-spaced samples as follows. Let $\mathcal{X}_t = \{\boldsymbol{X}_t, \boldsymbol{X}_{t+1}, ..., \boldsymbol{X}_{t+n-1}\}$ and $\mathcal{X}_{t+n} = \{\boldsymbol{X}_{t+n}, \boldsymbol{X}_{t+n}, ..., \boldsymbol{X}_{t+2n-1}\}$ and suppose that $\mathcal{X}_t \sim P$ and $\mathcal{X}_{t+n} \sim Q$. Denote with $D_{\mathrm{KL}}(P \parallel Q)$ the Kullback-Liebler divergence between two distributions P and Q with densities $p(\boldsymbol{X})$ and $q(\boldsymbol{X})$ respectively. Then we are interested in the quantity

$$D(P \parallel Q) = D_{\mathrm{KL}}(P \parallel Q) + D_{\mathrm{KL}}(Q \parallel P) \tag{22}$$

with

$$D_{\mathrm{KL}}(P \parallel Q) = \int p(\boldsymbol{X}) \log\left(\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right) \mathrm{d}\boldsymbol{X}. \tag{23}$$

Figure 18 illustrates the structure of the samples, where the two groups used to estimate divergence are comprised of n samples, each of which is a batch of length k.

The densities $p(\boldsymbol{X})$ and $q(\boldsymbol{X})$ are unknown and must be estimated in a nonparametric fashion. A naive approach would be to estimate each of them separately and then compute the ratio. However since knowing $p(\boldsymbol{X})$ and $q(\boldsymbol{X})$ implies knowing their ratio but not vice-versa, estimating the ratio directly is an easier task. In particular a method proposed in Sugiyama et al. (2008) to estimate the density ratio is employed in this paper.

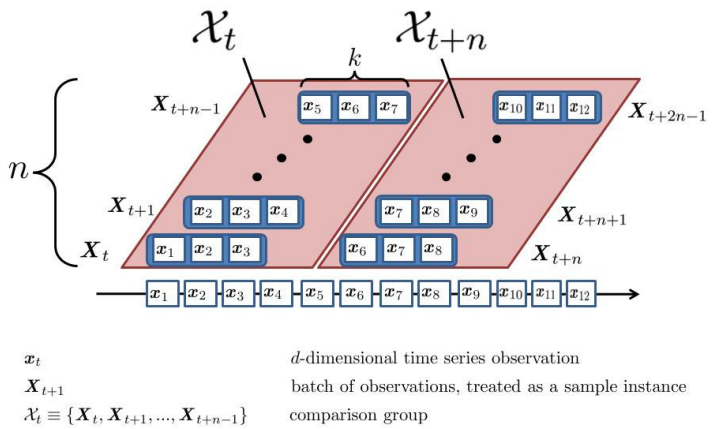

Figure 18. Structure of samples used to calculate $D_{\mathrm{KL}}(P \parallel Q)$.

6. Direct density-ratio estimation

Let us model the density ratio $\dfrac{p(\boldsymbol{X})}{q(\boldsymbol{X})}$ by

$$g(\boldsymbol{X};\boldsymbol{\theta}) = \sum_{i=1}^{n} \theta_i K(\boldsymbol{X}, \boldsymbol{X}_i), \tag{24}$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)'$ is a scaling vector parameter and K is the Gaussian kernel

$$K(\boldsymbol{X}, \boldsymbol{X}') = e^{\frac{\|\boldsymbol{X}-\boldsymbol{X}'\|^2}{2\sigma^2}}, \tag{25}$$

where the kernel width, $\sigma > 0$, is to be determined by cross-validation. The parameters $\boldsymbol{\theta}$ are to be learned from the data as those that minimize the Kullback-Liebler divergence between the distributions with densities $p(\boldsymbol{X})$ and $g(\boldsymbol{X};\boldsymbol{\theta})q(\boldsymbol{X})$. That is,

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int p(\boldsymbol{X}) \log\left(\frac{p(\boldsymbol{X})}{g(\boldsymbol{X};\boldsymbol{\theta})q(\boldsymbol{X})}\right) d\boldsymbol{X} \tag{26}$$

$$= \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ \int p(\boldsymbol{X}) \log\left(\frac{p(\boldsymbol{X})}{q(\boldsymbol{X})}\right) d\boldsymbol{X} - \int p(\boldsymbol{X}) \log\left(g(\boldsymbol{X};\boldsymbol{\theta})\right) d\boldsymbol{X} \right\} \tag{27}$$

$$= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \int p(\boldsymbol{X}) \log\left(g(\boldsymbol{X};\boldsymbol{\theta})\right) d\boldsymbol{X} \tag{28}$$

$$= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ \frac{1}{n} \sum_{j=1}^{n} \log\left(\sum_{i=1}^{n} \theta_i K(\boldsymbol{X}_j, \boldsymbol{X}_i) \right); \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \theta_i K(\boldsymbol{X}_j, \boldsymbol{X}_i) = 1, \boldsymbol{\theta} \geq 0 \right\}, \tag{29}$$

where in (29) the integral is approximated by the empirical estimate and the constraints are added to respect that $g(\boldsymbol{X};\boldsymbol{\theta})q(\boldsymbol{X})$ is a probability density function. Since the problem in (29) is convex, there

exists a global optimum that can be obtained using a gradient-projection method, among others. Then the density ratio can be estimated as

$$\widehat{g}(\boldsymbol{X}) = \sum_{i=1}^{n} \widehat{\theta}_i K(\boldsymbol{X}, \boldsymbol{X}_i) \tag{30}$$

and an estimate of $D_{\mathrm{KL}}(P \parallel Q)$ is given by

$$\widehat{D}_{\mathrm{KL}}(P \parallel Q) = \frac{1}{n} \sum_{i=1}^{n} \log \widehat{g}(\boldsymbol{X}_i). \tag{31}$$

As a simple test of convergence we estimate D_{KL} for two batches $\boldsymbol{y}_1 = [y_{11}, \ldots, y_{1n}]$ and $\boldsymbol{y}_2 = [y_{21}, \ldots, y_{2n}]$ of n observations simulated independently from the processes $Y_{it} = \theta_i Y_{it-1} + \epsilon_{it}$ with Gaussian white noise. Table 8 shows that when $\theta_1 = \theta_2$ the estimate converges to zero, its true value, as n grows large. If $\theta_1 \neq \theta_2$ then $D_{KL} > 0$ and $\hat{D}_{KL}$ converges to a positive constant as n grows albeit at a slow rate (see Table 9.)

Table 8. Estimate of the divergence score between two samples of observations $\boldsymbol{y}_1 = [y_{11}, \ldots, y_{1n}]$ and $\boldsymbol{y}_2 = [y_{21}, \ldots, y_{2n}]$ for different values of n, where $Y_{it} = 0.7Y_{it-1} + \epsilon_{it}$ and ϵ_{it} are independent Gaussian white noise series for $i = 1, 2$.

Convergence test for equal distributions

n	$\hat{D}_{KL}(p(Y_1) \parallel q(Y_2))$
500	0.0286
1000	0.0189
2000	0.0076
4000	0.0051
8000	0.0008

Table 9. Estimate of the divergence score between two samples of observations $\boldsymbol{y}_1 = [y_{11}, \ldots, y_{1n}]$ and $\boldsymbol{y}_2 = [y_{21}, \ldots, y_{2n}]$ for different values of n. The processes are $Y_{it} = \theta_i Y_{it-1} + \epsilon_{it}$, where $\theta_1 = 0.7, \theta_2 = 0.3$ and ϵ_{it} are independent Gaussian white noise series for $i = 1, 2$.

Convergence test for unequal distributions

n	$\hat{D}_{KL}(p(Y_1) \parallel q(Y_2))$
500	0.0652
1000	0.0620
2000	0.0597
4000	0.0628
8000	0.0605

This method is completely nonparametric and has the advantage of estimating the density ratio directly, and not the pre- and post-change densities independently. However its slow rate of convergence may introduce some bias in the results given the finite sample size in practice. The

kernel width and scaling vector are chosen systematically, whereas the parameters which determine the sliding window shape, n and k, are not. These latter two parameters control the size of the rolling windows $\mathcal{X}_t$ and $\mathcal{X}_{t+n}$ and, thus, act as smoothing parameters as well. The estimated statistic $D(P \parallel Q)$ is to be interpreted as a measure of the distance between P and Q.

7. Changes in correlation structure

In this section we relate the concept of statistical change-point detection to the structure of correlations in asset returns. In previous sections we have described the structure of correlations in terms of the strength of comovement as captured by the relative magnitude of the eigenvalues of the sample correlation matrix. As discussed in Appendix A the eigenvalue density of certain models of random matrices tends to a non-random density in the limit that the matrix dimensions grow to infinity. If these densities are known exactly, then parametric change-detection techniques can be applied in an online fashion, such that a trade-off between the detection delay and false alarm rate is optimized. In practice this is most often not the case, but, assuming that a limiting density exists, we can attempt to estimate points of high divergence in an online fashion using the nonparametric procedure discussed in the previous section.

Let us briefly revisit the absorption ratio (AR) time series $\{\phi_1(t)\}$ used in Section 3.2. For purely random, independently generated Wishart matrices, we expect that observed instances of the first (normalized) eigenvalue would form a stationary series. In fact we would expect this to be true in the limit as the matrix dimensions grow to infinity (with a constant aspect ratio) in all cases where a limiting spectral density exists. In testing the level of the AR for stationarity, we find repeatedly for different assets and markets that the series is nonstationary, even over relatively short periods of time. This is entirely consistent with the notion that financial crises are associated with a breaking correlation structure. However nonstationarity in the level of the AR may lead change-detection procedures to produce noisy results. On the other hand the series $\Delta\phi_1$ have more desirable properties in the sense that nonstationarity due to noise becomes less influential. Increasing divergence between successive observations of the first-differenced series also maintains the intuitive interpretation of emerging trends in the AR. We shall use the procedure in Sections 4-6 to investigate abrupt changes in correlation structure associated with financial crises.

We compute $\Delta\phi_1$ for each of the asset groups studied to this point. We use values of $n = k = 30$ in computing divergence for all cases[‖] and plot $\Delta\phi_1$ and the divergence score $D(P \parallel Q)$. We also overlay a plot of the relevant index to compare the divergence score with market conditions. The divergence score increased sharply for financial sector equities and CDS spreads in 2007 (see Figure 19 and Figure 20). These spikes follow larger, positive values of $\Delta\phi_1$, representing strengthening of correlations which occurred well before the S&P 500 index reached its crisis lows.

During the European Debt crisis the divergence score increased sharply in response to downward shifts in the AR. We find that these shifts are not reflected in the divergence score in bond yield data until the crisis has reached a very developed state (see Figure 21). With CDS spreads, based on our data, there is greater potential in detecting a changing correlation structure in an online manner before large drawdowns. This can be observed in Figure 22 where a large drawdown in EURUSD starting in

[‖]Following the suggestion made by one of the referees, we also experiment with a limited range of values for n and k, and find that the results are qualitatively unchanged as long as we stay within the range the $(15, 75)$ interval.

the summer of 2011 is preceded by a spike in the divergence score.

Finally estimating divergence between successive batches of observation of $\Delta\phi_1$ in Asian currencies, we find that divergence increased notably before the start of the Asian Financial Crisis (see Figure 23). This can also be seen in Figure 24, where the results for equity data in this crisis episode are similar, except that the divergence score is noisier. Overall there is an expected lag between observable persistent directional and magnitude changes in $\Delta\phi_1$ and the divergence score because it is estimated from historical observations. We find evidence that despite this lag a likelihood-based test for a change in the eigenvalue distribution of the correlation matrix of a fixed asset basket between different time intervals may provide early warning about impending crises. However financial data is inherently noisy and this results in some robustness issues. For example results may vary depending on the sampling parameters n and k.

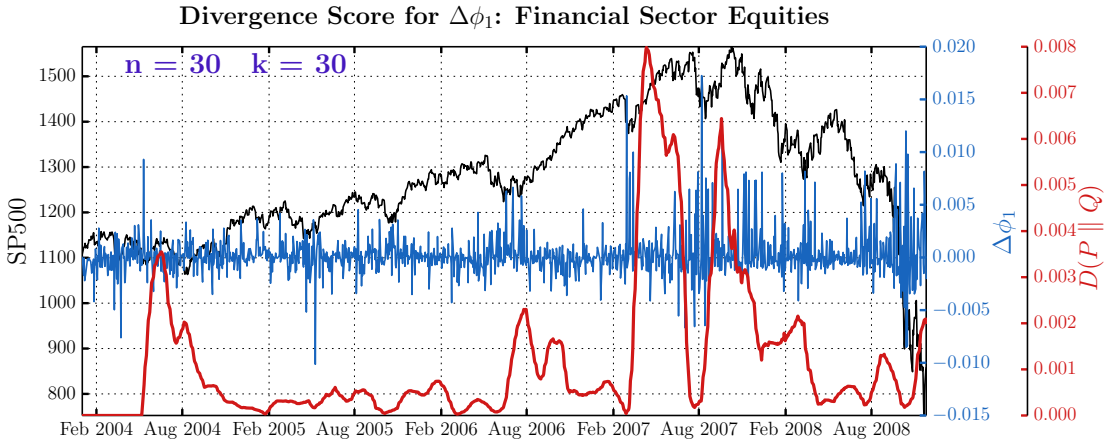

Figure 19. Divergence score for the absorption ratio series of the first principal component in NYK constituent stocks.

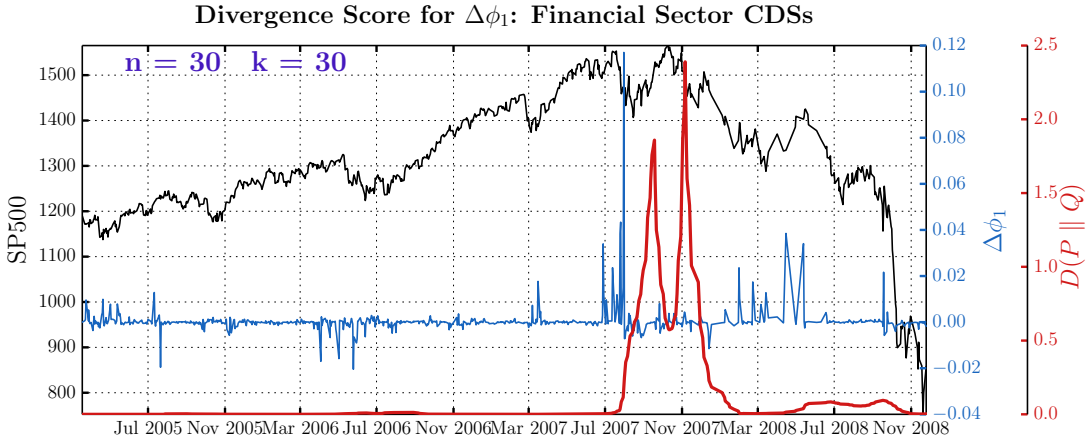

Figure 20. Divergence score for the absorption ratio series of the first principal component in bank debt CDS spreads.

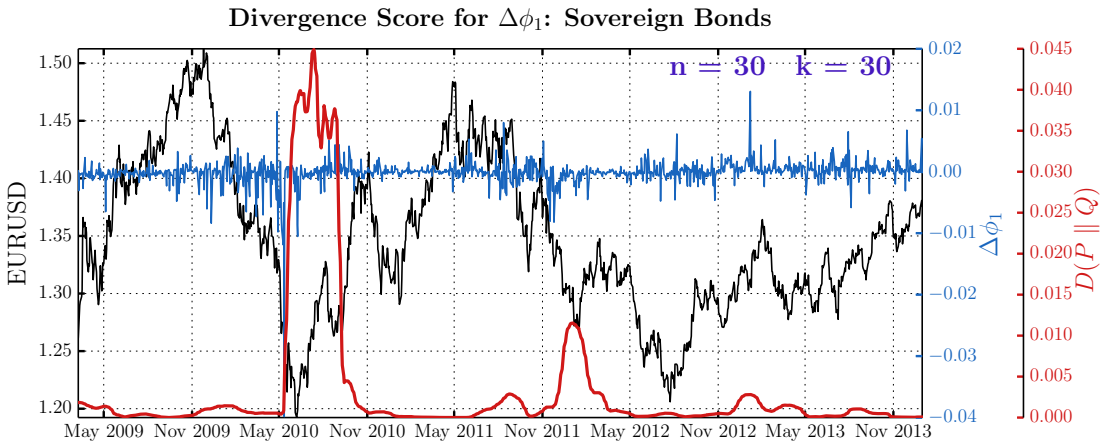

Figure 21. Divergence score for the absorption ratio series of the first principal component in European sovereign bond yields.

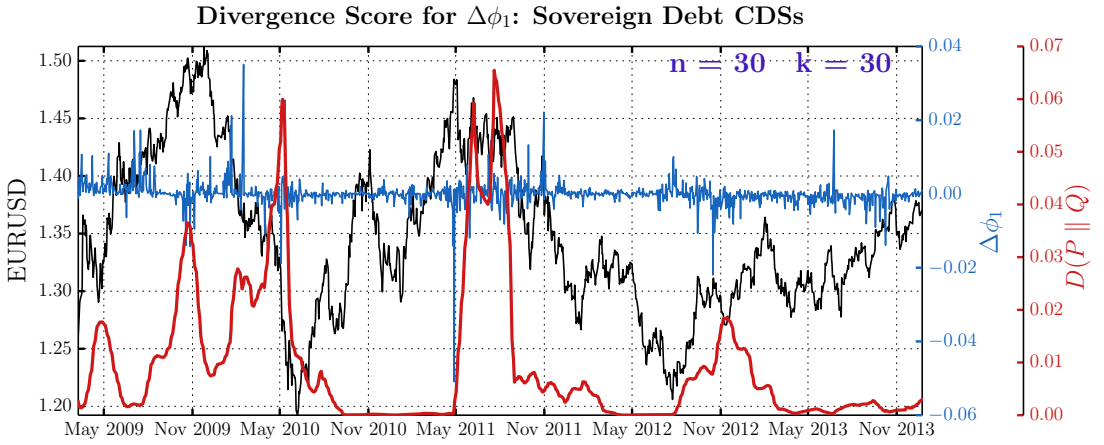

Figure 22. Divergence score for the absorption ratio series of the first principal component in European sovereign debt CDS spreads.

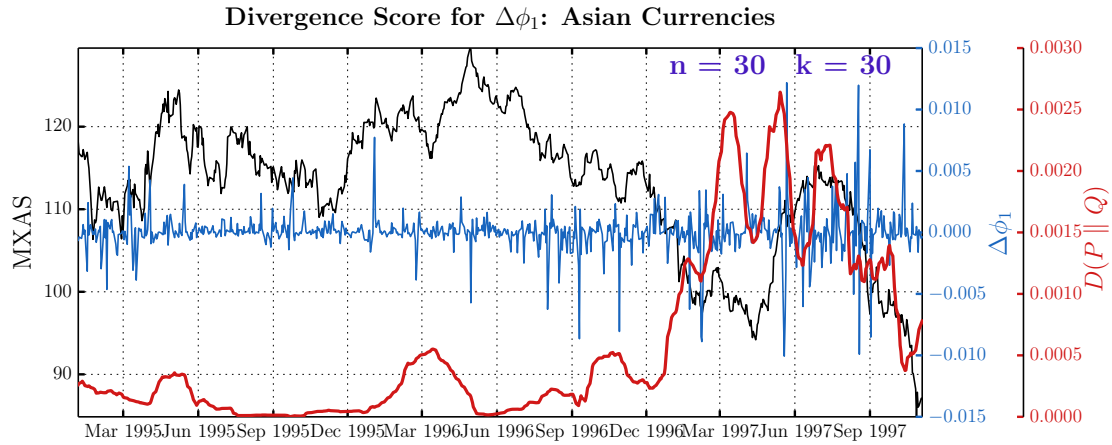

Figure 23. Divergence score for the absorption ratio series of the first principal component in Asian currencies.

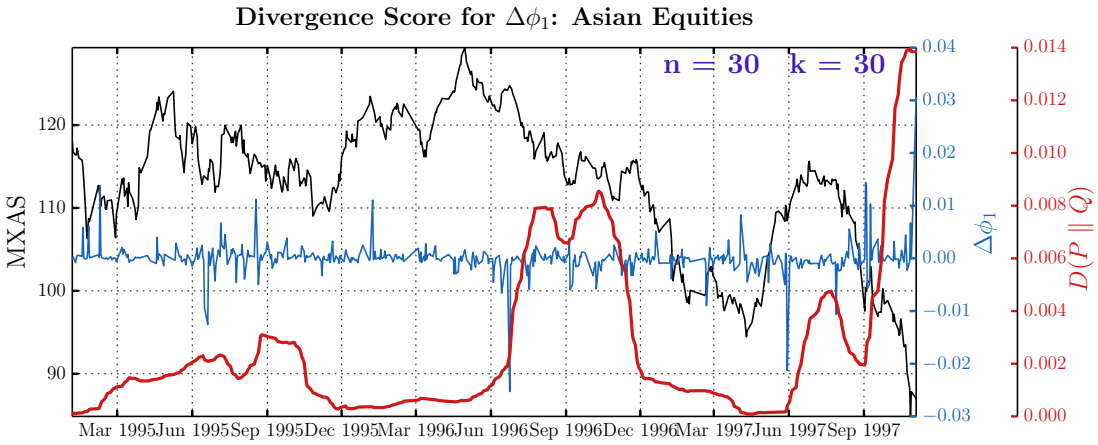

Figure 24. Divergence score for the absorption ratio series of the first principal component in Asian equities.

8. Supplementary analysis through a parametric lense

The noparametric results from Section 7 suggest that the divergence scores of different asset classes may precede or, at least, coincide with a particular crisis period. In this section we provide an additional parametric analysis on the relationship between the AR and broad market conditions as a step towards a formal validation check. This analysis is especially relevant since the major empirical contribution of this paper is to investigate shifts in the comovement of a large set of asset classes for the three notable crises.[**]

Our specific objective in this section is to determine whether there are significant associations between changes in the AR and the returns or volatility of a broad index. The former corresponds to a standardized shift in the AR proposed by (Kritzman et al., 2011) and the latter serves as a proxy for the financial condition of a system in distress. In particular we are interested in investigating whether (i) changes in the AR Granger-cause returns or volatility of a broad index, and (ii) for a particular asset class, changes in the AR are leading, coincidental or lagging indicators for conditions of market fragility for a particular crisis.

Consider a bi-variate Vector Autoregression (VAR) model (Granger, 1969)

$$r_Y(t) = \alpha_0 + \sum_{i=1}^{k} \alpha_i r_Y(t-i) + \sum_{i=1}^{k} \beta_i \Delta\hat{\phi}_1(t-i) + \epsilon(t) \tag{32}$$

$$\Delta\hat{\phi}_1(t) = \gamma_0 + \sum_{i=1}^{k} \gamma_i \Delta\hat{\phi}_1(t-i) + \sum_{i=1}^{k} \delta_i r_Y(t-i) + \eta(t) \tag{33}$$

where $\hat{\phi}_1(t)$ is the estimated change in the AR of the dominant PC, as defined in Appendix A.3,[††] $r_Y(t)$

[**]We thank one of the referees for this suggestion.

[††]Following the suggestion made by one of the referees, we also experiment with a similar measure proposed by (Kritzman et al., 2011), which is $\Delta\hat{\phi}(t) = (\hat{\phi}(15day) - \hat{\phi}(1year)/\sigma_{\hat{\phi}(1year)}$, where $\hat{\phi}(15day)$ and $\hat{\phi}(1year)$ are the 15-day and 1-year moving averages of $\hat{\phi}(t)$ respectively, and $\sigma_{\hat{\phi}(1year)}$ is the standard deviation in $\hat{\phi}(t)$ over the previous year. Overall we find that the results are qualititatively very similar to ours.

is the return on index Y, k is the maximum number of lagged observations included in the model, $\epsilon(t)$ and $\eta(t)$ are rendered approximately serially uncorrelated respectively by choice of lag lengths. A null hypothesis that $\Delta\hat{\phi}_1$ does not Granger-cause r_Y is rejected at a chosen significance level if any of the coefficients of lagged r_Y are jointly significantly different from zero with those of lagged $\Delta\hat{\phi}_1$ in (32), according to an F-test. Similarly a null hypothesis that r_Y does not Granger-cause $\Delta\hat{\phi}_1$ is rejected at a chosen significance level if any of the coefficients of lagged $\Delta\hat{\phi}_1$ are jointly significantly different from zero with those of lagged r_Y in (33) according to an F-test. A variable that Granger-causes another one is considered to be a leading source of information for predicting it.

We fit the model with ordinary least squares and perform the tests separately for returns, r_Y, and realized volatility, σ_Y. We compute changes in the AR, returns and realized volatilities with data at daily frequency. The maximum lag in the fitting of each VAR model is determined by the Akaike Information Criterion (AIC) to ensure that the residuals $\hat{\epsilon}(t)$ and $\hat{\eta}(t)$ are approximately serially uncorrelated. In addition we also adjust the reported standard errors for the estimates of the parameters in (32) and (33) for (i) potential heteroskedasticity due to the estimated dependent variable in (32) by using White's Heteroskedasticity Consistent Covariance Matrix Estimates proposed in White (1980), and (ii) potential generated-regressor bias due to the estimated right-hand side variables in (32) and (33) by using Murphy and Topel's asymptotic covariance matrix estimates proposed by Murphy and Topel (1985).

Components underlying the AR and the corresponding index depend on the crisis episode and are commensurate with Section 3.2.1 - Section 3.2.3. These are also summarized in Table 7. Results of the Granger causality tests are presented in Table 10. The table corresponds to the results for each market index, Y. For the Financial Crisis, changes in the AR of stocks that were components of the NYK are found to Granger-cause realized volatility of the broader market. There is also evidence to reject the no Granger-causality null hypothesis based on CDS data. Lastly in the case of index returns we find similar evidence in support of the Granger causation from changes in the AR of both equities and CDSs to indexes proxing for conditions of market distress. For the European Debt Crisis, we similarly find evidence of Granger causation from changes in the AR of both BEFINC and EURUSD to realized volatility, consistently across bond yields and CDS spreads. There is also sufficient evidence to reach a similar conclusion with respect to index returns for this particular crisis. Lastly for the Asian Financial Crisis, we emphatically reject the null hypothesis that changes in the AR of currency or equity returns do not Granger-cause the returns or volatility of the MXAS. Overall the test results support the idea that changes in the AR Granger-cause returns or realized volatilities for each crisis episode.[‡‡]

Given that changes in the AR Granger-cause the conditions for market fragility, as a next step, we investigate whether for a particular asset class, changes in the AR are leading, coincidental or lagging the indicators for adverse conditions of the markets. That is, we examine the statistical significance of the lead-lag effects between changes in the AR and indicator for conditions of market distress based on the approach proposed by Dimson (1979). The Dimson regression for return of index Y is specified as

$$r_Y(t) = \alpha_0 + \sum_{i=-k}^{k} \beta_i \Delta\hat{\phi}_1(t-i) + \epsilon(t), \tag{34}$$

[‡‡]Although not reported here (to save space), we also conduct Granger-causality tests for each of the three crisis episodes using data from weekly, monthly and quarterly frequencies. In general we find that the strength of the one-way Granger-causation reported for data from daily frequency diminishes monotonically when data from lower frequencies are used in the analysis.

Table 10. Granger-causality test results for the Financial Crisis, the Eurozone Sovereign Debt Crisis: BEFINC and EURUSD, and the Asian Financial Crisis.

	$\Delta\hat{\phi}_1(t) \rightarrow Y$		$\Delta\hat{\phi}_1(t) \leftarrow Y$	
	F-stat	p-value	F-stat	p-value
Index: SP500				
Equities				
Return	3.96	0.00***	0.34	0.79
Volatility	4.01	0.00***	0.30	0.82
CDSs				
Return	2.42	0.08*	1.13	0.32
Volatility	4.08	0.00***	1.51	0.20
Index: BEFINC				
Bonds				
Return	4.02	0.00***	1.67	0.17
Volatility	2.89	0.03**	1.31	0.26
CDSs				
Return	3.59	0.01**	0.11	0.89
Volatility	2.92	0.03**	0.29	0.82
Index: EURUSD				
Bonds				
Return	3.82	0.00***	1.11	0.34
Volatility	2.19	0.08*	1.36	0.25
CDSs				
Return	2.70	0.04**	0.74	0.52
Volatility	2.71	0.04**	1.37	0.25
Index: MXAS				
Currencies				
Return	2.10	0.09*	1.26	0.28
Volatility	3.05	0.00***	0.57	0.72
Equities				
Return	3.97	0.00***	0.97	0.40
Volatility	4.07	0.00***	1.37	0.25

* statistically significant at 10%

** statistically significant at 5%

*** statistically significant at 1%

Note: For the Financial Crisis, each row represents a test of Granger causality between either returns or realized volatility of SP500 and changes in the AR as derived from equity or CDS data. For the European Debt Crisis, each row represents a test of Granger causality between either returns or realized volatility of BEFINC and EURUSD and changes in the AR as derived from bond or CDS data. For the Asian Financial Crisis, each row represents a test of Granger causality between either returns or realized volatility of MXAS and changes in the AR as derived from currency or equity data. The F-stat and p-value found in each test are given for both directions of Granger causality, with $\Delta\hat{\phi}_1(t) \rightarrow Y$ denoting that changes in $\hat{\phi}_1$ Granger-cause the index variable. Conversely $\Delta\hat{\phi}_1(t) \leftarrow Y$ denotes that the index variable Granger-causes changes in $\hat{\phi}_1$.

where ϵ is an independently and identically distributed error term for an appropriate choice of lead/lag length k. For volatility we simply replace $r_Y(t)$ by $\hat{\sigma}_Y(t)$ as the dependent variable in (3). The coefficients β_i with negative subscripts are lag coefficients and those with positive subscripts are lead coefficients. If any lag coefficient estimates are found to be significantly different from zero, then it can be inferred that returns or volatility lag changes in the AR, or, equivalently, changes in the AR lead returns or volatility. On the other hand, rejecting the null hypothesis for any of the lead coefficient estimates would support the notion that changes in the AR lag returns or volatility. There is also a contemporaneous coefficient, β_0, the significance of which would imply that changes in the asset class's ARs occur simultaneously with returns or volatility in the system as captured by the index. Finally, if, together with a rejection of the null hypothesis with the contemporaneous coefficient estimate, we find that both lead and lag coefficient estimates are significantly different from zero, then the relationship is considered to be informationally efficient.

We fit the model with ordinary least squares with $k = 1$. In addition, as in the case of the Granger causality tests, we compute changes in the AR, returns and realized volatilities at the daily frequency. Thus we study the effects of a one-period lead or lag over the observation window. Components underlying the AR and the corresponding index depend on the crisis episode and are again maintained as before. Regression results are presented in Tables 11 and 12. We correct the standard errors of the parameter estimates in the Dimson regression for (i) potential serial correlation in the residual terms in (34) due to insufficient leads and lags and potential heteroskedasticity in the residual terms in (34) due to the estimated volatility as the dependent variable in the Dimson regression by using Newey-West's heteroskedasticity and autocorrelation consistent covariance matrix estimates (Newey and West, 1987), and (ii) for potential generated-regressor bias due to the presence of estimated independent variables in (34) by using Murphy and Topel's asymptotic covariance matrix estimates proposed in Murphy and Topel (1985).

The table corresponds to results with respect to a specific market index. It has been noted here as well as the existing literature that return correlations among stocks seemed to have been increasing before the Financial Crisis, leading to the conjecture that a measure like shifts in the AR could have given an early warning signal. We are able to find a statistically significant lead relationship between changes in the AR of components of the NYK and one-period-ahead return or volatility of the SP500 index (see Table 11). This is also true for changes in the AR of CDS spreads of major financial institutions. In addition coefficients in CDS data are found to be statistically not significant, when changes in the AR are either lagging or contemporaneous with indicators for conditions of market distress. In Section 3.2.2 a qualitative inspection of comovement in yields suggested a divergence pattern associated with the crisis. This is confirmed by the regression of realized volatility of both BEFINC and EURUSD, as the corresponding coefficients are estimated to be negative. However we observe overall less consistency in leading AR coefficient estimates for this crisis. Specifically, realized volatility of BEFINC was negatively but not significantly related to coincidental changes in the bond yield's AR, while realized volatility of EURUSD was positively and significantly related to coincidental changes in bond yield correlation. We also note that small p-values for lagging coefficients and large p-values for leading or coincident coefficients in regressing realized volatility are evident from the results for the Asian Financial Crisis. However we find that this lag relationships with equities, although statistically significant, are much weaker than those with currencies.

Overall we find that changes in the AR lead returns or volatility for financial sector equities and

Table 11. Regression results for the Financial Crisis and the Asian Financial Crisis.

Independent Variable	**Equities**			**CDSs**		
	Coefficient	*t*-stat	*p*-value	Coefficient	*t*-stat	*p*-value
SP500 - Leading						
Return	0.21	0.99	0.32	-0.11	-0.69	0.48
Volatility	0.06	0.49	0.61	-0.03	-0.42	0.66
SP500 - Coincident						
Return	0.02	0.13	0.88	0.46	0.69	0.48
Volatility	0.44	0.99	0.32	-0.00	-0.01	0.99
SP500 - Lagging						
Return	0.04	2.99	0.00***	-0.31	-2.30	0.02**
Volatility	-0.01	-2.18	0.03**	0.17	1.98	0.04**
Independent Variable	**Currencies**			**Equities**		
	Coefficient	*t*-stat	*p*-value	Coefficient	*t*-stat	*p*-value
MXAS - Leading						
Return	-0.20	-1.02	0.30	-0.62	-0.79	0.42
Volatility	0.02	0.16	0.86	-0.23	-1.60	0.10
MXAS - Coincident						
Return	0.02	0.13	0.88	-0.51	-1.02	0.30
Volatility	0.20	1.07	0.29	-0.03	-0.21	0.82
MXAS - Lagging						
Return	-0.01	-2.99	0.00***	0.31	1.76	0.08*
Volatility	0.20	2.96	0.00***	0.16	1.69	0.09*

Note: * statistically significant at 10%

** statistically significant at 5%

*** statistically significant at 1%

Each row represents a *t*-test that a regression coefficient is significantly different from zero. For the Financial Crisis, either returns or realized volatility of SP500 is regressed on lead and lag observations of $\Delta\hat{\phi}_1$ as derived from equity or CDS data. For the Asian Financial Crisis, either returns or realized volatility of MXAS over daily frequencies is regressed on lead and lag observations of $\Delta\hat{\phi}_1$ as derived from currency or CDS data. In the top section the independent variable leads $\Delta\hat{\phi}_1$ by one period, in the middle section they are contemporaneous, and in the bottom section the independent variable lags $\Delta\hat{\phi}_1$ by one period.

CDS. For the European Debt Crisis, the relationship between changes in the AR and returns or realized volatility is found to be coincidental for bond yield data and leading for CDS spreads. Lastly for the Asian Financial Crisis, changes in the AR of both Asian currencies and its equity returns lead returns or realized volatility, although the former relationship is much more significant than the latter.* Thus these test results using a parametric approach are in broad conformity with the nonparametric results in Section 7. The latter suggests that the divergence scores of different asset classes may precede or, at least, coincide with a particular crisis episode.

*Although not reported here (to save space), we also conduct the lead, coincident and lag tests for each of the three crisis episodes using data from weekly, monthly and quarterly frequencies. In general we find that the result are qualitatively very similar to the results obtained with data from the daily frequency.

Table 12. Regression results for the Eurozone Sovereign Debt Crisis: BEFINC.

Independent Variable	Bonds			CDSs		
	Coefficient	*t*-stat	*p*-value	Coefficient	*t*-stat	*p*-value
BEFINC - Leading						
Return	-0.03	-0.16	0.86	-0.03	-0.13	0.89
Volatility	-0.25	-0.97	0.32	0.01	-0.19	0.84
BEFINC - Coincident						
Return	-0.03	-0.11	0.90	0.32	1.30	0.19
Volatility	-0.20	-1.09	0.27	0.02	0.16	0.87
BEFINC - Lagging						
Return	0.38	2.26	0.02*	0.06	2.12	0.03**
Volatility	-0.21	-2.55	0.01**	0.31	2.17	0.03**
Independent Variable	**Bonds**			**CDSs**		
	Coefficient	*t*-stat	*p*-value	Coefficient	*t*-stat	*p*-value
EURUSD - Leading						
Return	0.12	1.11	0.26	0.01	0.25	0.79
Volatility	-0.16	-0.94	0.35	0.02	0.54	0.58
EURUSD - Coincident						
Return	0.02	2.60	0.00***	0.01	0.14	0.88
Volatility	0.09	1.89	0.05*	-0.08	-1.11	0.26
EURUSD - Lagging						
Return	0.23	0.87	0.38	-0.09	-2.89	0.00***
Volatility	-0.10	-1.73	0.08*	0.11	1.87	0.06*

Note: * statistically significant at 10%

** statistically significant at 5%

*** statistically significant at 1%

Each row represents a *t*-test that a regression coefficient is significantly different from zero. For the European Debt Crisis, either returns or realized volatility of BEFINC and EURUSD is regressed on lead and lag observations of $\Delta\hat{\phi}_1$ as derived from bonds or CDS data. In the top section the independent variable leads $\Delta\hat{\phi}_1$ by one period, in the middle section they are contemporaneous, and in the bottom section the independent variable lags $\Delta\hat{\phi}_1$ by one period.

9. Conclusions

Prior studies have pointed to an increasing strength of comovement in returns on assets such as stocks and property before the Financial Crisis. Using the proportion of variance explained by the first principal component in a basket of assets – the Absorption Ratio (AR) – as a measure of correlation, they have asserted that large drawdowns are associated with increasing correlations, sometimes in advance of a financial turmoil, and that the state of higher correlations facilitates contagion.

In this paper we are able to confirm the findings in the literature for the Financial Crisis. In applying a similar framework to two other notable crises that involved asset classes that were previously not considered, we find that, generally, financial distress is not necessarily characterized by

increasing correlations, but by a breakdown of the existing correlation structure that may also result in diverging returns. For example we find evidence that bond yields of countries affected by the European Debt Crisis had diverged. The estimated AR time series are found to exhibit a visibly nonlinear and potentially nonstationary pattern over time than would be expected according to the RMT if asset returns were to be purely randomly distributed. This is consistent with a long-standing held view that financial asset returns often have nonlinear relationships with risk factors, and their distributions are often nonstationary. Given the above finding we conduct a change-point analysis using a nonparametric technique. By estimating the divergence in the distributions of successive groups of samples through time we are able to identify time periods that are associated with shifts in the correlation structure of asset returns. In particular we find that the divergence score increases either before or coincidentally with the systemic shocks. We also provide an additional parametric analysis to serve as an informal device to validate the results obtained by using the RMT and the nonparametric change-point analysis.

As in most empirical studies there are a number of caveats to the results presented in this paper. Firstly the potential for financial contagion should depend on the strength of comovement of asset values in the system. However the use of the PCA to measure this as the relative magnitude of the largest eigenvalue assumes that comovements are linear in nature. In practice financial asset returns and volatility may be nonlinear, especially during periods of extreme shocks (Billio et al., 2012). While our assumptions reflect a simplified market, they allow for a tractable analysis of a high-dimensional problem in finance. Secondly in performing the PCA we estimate the correlation matrix as the sample correlation matrix of the latest 252 trading days. Estimating covariances (and correlations) is a challenging task in itself. There is a variety of approaches that impose different degrees of structure on covariance matrices. Whereas ours is completely free of structure, risk-factor based techniques may produce substantially different results. There are also issues with estimating the true eigenvalues of the correlation matrix and comparing them with results from the RMT, when their underlying assumptions are not satisfied. For example, if the matrix dimensions are too small, then the ideal properties of estimated eigenvalues will not be compatible with theoretical limits. Thirdly the quantitative analysis of the relationship between correlation and crises in this paper relies on measures of financial distress. To this end we select broad indices that were believed to reflect this information accurately and efficiently. If the chosen proxy for financial distress does not possess these properties, then it would serve as a poor input into the analysis.

Fourthly the AR time series are obtained through the PCA on return observations using the covariance method in a rolling-window fashion. That the AR series is richly dynamic suggests that individual asset data may not be second-order stationary. This supports an already known stylized fact that financial data exhibits nonstationarity, which can greatly impede the robustness of the models fitted to historical samples. One response to this last caveat is to conduct a change-point analysis using a nonparametric technique. By estimating divergence in the distributions of successive groups of samples through time, we are able to identify time periods that are associated with shifts in the correlation structure. This allows us to determine whether a certain divergence score increases either before or coincidentally with systemic shocks in the financial markets. In addition we also have shown in the paper that the ARs typically increase in crisis periods, indicating that spikes in the ARs may serve as an indicator of market fragility. However, we have also found that a sharp increase in AR would not necessarily lead to a sharp decline in asset returns. Unless we have data with a fairly

long time period around each of the crisis episodes at our disposal, it would be difficult to detect extreme market downturns in the markets by using spikes in the ARs. In addition we have also witnessed that the divergence score for Asian equities increases well before the onset of the Asian crisis. But over a longer stretch of time-series, it is possible that an increase in the divergence score may not be associated with a substantial drop in the equity market, or sharp increases and decreases in the ARs may occur asymmetrically. We plan to inestigate these issues[†] once a longer stretch of time-series data are at our disposal. Lastly financial return distributions are known to exhibit fat tails due to the occurence of rare but extreme shocks in financial markets. The assumption that return and realized volatility are normally distributed underestimates the probability of such events. Therefore hypothesis tests in the VAR and Dimsons regressions, where this assumption is implicitly made, may lead to conclusions that may not be entirely robust.

Acknowledgments

This research is funded by SSHRC Insight Grant (# 435-201-0682). We thank Adam Kolkiewicz and Chengguo Weng for providing valuable feedbacks on the early version of this paper. Conversation with Phelim Boyle at various stages of this research is also gratefully acknowledged. All of the remaining errors and shortcomings are ours.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

Allez R, Bouchaud JP (2011) Individual and collective stock dynamics: intraday seasonalities. *New J Phys* 13: 345–349.

Anderson TW (1963) Asymptotic theory for principal component analysis. *Ann Math Stat* 34: 122–148.

Andersson E, Bock D, Frisén M (2004) Detection of turning points in business cycles. *J Bus Cycle Manage Anal* 1: 93–108.

Andersson E, Bock D, Frisén M (2006) Some statistical aspects of methods for detection of turning points in business cycles. *J Appl Stat* 33: 257–278.

Andrews DWK, Lee I, Ploberger W (1996) Optimal changepoint tests for normal linear regression. *J Econometrics* 70: 9–38.

Ang A, Chen J (2002) Asymmetric correlations of equity portfolios. *J Financ Econ* 63: 443–494.

Bai Z, Zhou W (2008) Large sample covariance matrices without independence structures in columns. *Stat Sinica* 18: 425–442.

[†]We thank one of the referees for pointing this out to us.

Bai ZD, Silverstein JW (2010) *Spectral Analysis of Large Dimensional Random Matrices*, Second Edition, Springer, New York.

Basserville M, Nikiforov I (1993) *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Beibel M, Lerche HR (2000) A new look at optimal stopping problems related to mathematical finance. *Stat Sinica* 7: 93–108.

Bejan A (2005) Largest eigenvalues and sample covariance matrices. M.Sc. dissertation, Department of Statistics, The University of Warwick.

Berkes I, Gombay E, Horváth L, et al. (2004) Sequential change-point detection in GARCH(p, q) models. *Economet Theor* 20: 1140–1167.

Biely C, Thurner S (2008) Random matrix ensembles of time-lagged correlation matrices: derivation of eigenvalue spectra and analysis of financial time-series. *Quant Financ* 8: 705–722.

Bijlsma M, Klomp J, Duineveld S (2010) Systemic risk in the financial sector: A review and synthesis. *CPB Netherland Bureau of Economic Policy Analysis Paper 210*.

Billio M, Getmansky M, Lo AW, et al. (2012) Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J Financ Economet* 104: 535–559.

Bouchaud JP, Potters M (2001) More stylized facts of financial markets: leverage effect and downside correlations. *Physica A* 299: 60–70.

Broemling LD, Tsurumi H (1987) *Econometrics and Structural Change*, Marcel Dekker, New York.

Capuano C (2008) The option-iPoD. The probability of default implied by option prices based on entropy. *IMF*.

Chen J, Gupta AK (1997) Testing and locating variance change-points with application to stock prices. *J Am Stat Assoc* 92: 739–747.

Chordia T, Swaminathan B (2000) Trading volume and cross-autocorrelations in stock returns. *J Financ* 55: 913–935.

Cizeau P, Potters M, Bouchaud JP (2001) Correlation structure of extreme stock returns. *Quant Financ* 1: 217–222.

Conlon T, Ruskin HJ, Crane M (2009) Cross-correlations dynamics in financial time series. *Physica A* 388: 705–714.

Constantine AG (1963) Some non-central distribution problems in multivariate analysis. *Ann Math Stat* 34: 1270–1285.

Daniel K, Moskowitz T (2016) Momentum crashes. *J Financ Econ* 122: 221–247.

Davis RA, Pfaffel O, Stelzer R (2014) Limit theory for the largest eigenvalue of sample covariance matrices with heavy-tails. *Stoch Proc Appl* 124: 18–50.

De Brandt O, Hartmann P (2000) Systemic risk: A survey. *European Central Bank*.

Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 74: 427–431.

Dimson E (1979) Risk measurement when shares are subject to infrequent trading. *J Financ Econ* 7: 197–226.

Doris D (2014) *Modeling Systemic Risk in the Options Market*. Ph.D. Thesis, Department of Mathematics, New York University, New York, NY.

Drożdż S, Grumer F, Ruf F, et al. (2000) Dynamics of competition between collectivity and noise in the stock market. *Physica A* 287: 440–449.

Edelman A, Persson PO (2005) Numerical methods for eigenvalue distributions of random matrices. *Math*.

Edelman A, Rao NR (2005) Random matrix theory. *Acta Numer* 14: 233–297.

Franses PH, van Dijk D (2000) *Non-Linear Time Series Models in Empirical Finance*. Cambridge University Press, New York, NY.

Geman S (1980) A limit theorem for the norm of random matrices. *Ann Probab* 8: 252–261.

Gopikrishnan P, Rosenov B, Plerou V, et al. (2001) Quantifying and interpreting collective behavior in fnancial markets. *Physi Rev E* 64: 035106.

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

International Monetary Fund (2009). *Global Financial Stability Report; Responding to the Financial Crisis and Measuring Systemic Risks*. Washington, D.C.

James AT (1960) The distribution of the latent roots of the covariance matrix. *Ann Math Stati* 32: 874–882.

Jin B, Wang C, Miao B, et al. (2009) Limiting spectral distribution of large-dimensional sample covariance matrices generated by VARMA. *J Multivariate Anal* 100: 2112–2125.

Jobst AA (2013) Multivariate dependence of implied volatilities from equity options as measure of systemic risk. *International Review of Financial Analysis* 28: 112–129.

Johnstone IM (2001) On the distribution of the largest eigenvalue in principal component analysis. *Ann Stat* 29: 295–327.

Kawahara Y, Yairi T, Machida K (2007) Change-point detection in time-series data based on subspace identification. *Proceedings of the 7th IEEE International Conference on Data Mining*, 559–564.

Kritzman M, Li Y, Page S, et al. (2011) Principal components as a measure of systemic risk. *J Portf Manage* 37: 112–126.

Kwiatkowski D, Phillips P, Schmidt P, et al. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series has a unit root? *J Econometrics* 54: 159–178.

Laloux L, Cizeau P, Bouchaud JP (1999) Noise Dressing of Financial Correlation Matrices. *Phys Rev Lett* 83: 1467–1469.

Laloux L, Cizeau P, Potters M, et al. (2000) Random matrix theory and financial correlations. *Int J Theor Appl Financ* 3: 391–397.

Lequeux P, Menon M (2010) An eigenvalue approach to risk regimes in currency markets. *J Deriv Hedge Funds* 16: 123–135.

Lewis M (2010) *The Big Short: Inside the Doomsday Machine*. W. W. Norton & Company Inc., New York, NY.

Liu H, Aue A, Debashis P (2015) On the Marčenko-Pastur law for linear time series. *Ann Stat* 43: 675–712.

Liu S, Yamada M, Collier N, et al. (2013) Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks* 43: 72–83.

Longin F, Solnik B (2001) Extreme correlation of international equity markets. *J Financ* 5: 649–676.

Lorden G (1971) Procedures for reacting to a change in distribution, *Ann Math Stat* 42: 1897–1908.

Marčenko VA, Pastur LA (1967) Distribution for some sets of random matrices. *Math USSR-Sbornik* 1: 457–483.

Mayya KBK, Amritkar RE (2006) Analysis of delay correlation matrices. *Quant Financ*.

Meng H, Xie WJ, Jiang ZQ, et al. (2014) Systemic risk and spatiotemporal dynamics of the US housing market. *Sci Rep-UK* 4: 3655.

Meric I, Kim S, Kim JH, et al. (2001) Co-movements of U.S., U.K., and Asian stock markets before and after September 11, 2001. *J Money Invest Bank* 3: 47–57.

Moustakides GV (1986) Optimal stopping times for detecting changes in distributions. *Ann Stat* 14: 1379–1387.

Muirhead RJ (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Murphy KM, Topel RH (1985) Estimation and inference in two-step econometric models. *J Bus Econ Stat* 34: 370–379.

Newey WK, West KD (1987) A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55: 703–708.

NYSE Financial Index (2014) *NYSE Euronex*. Available from: http://www.nyse.com/about/listed/nykid.shtml.

Page ES (1954) Continuous inspection schemes. *Biometrika* 41: 100–115.

Pan RK, Sinha S (2007) Collective behavior of stock price movements in an emerging market. *Phys Rev E* 76: 1–9.

Petterson M (1998) Monitoring a freshwater fish population: Statistical surveillance of biodiversity. *Environmetrics* 9: 139–150.

Petzold M, Sonesson C, Bergman E, et al. (2004) Surveillance in longitudinal models: Detection of intrauterine growth restriction. *Biometrics* 60: 1025–1033.

Phillips P, Perron P (1988) Time series regression with a unit root. *Biometrika* 75: 335–346.

Pillai KCS (1976a) Distribution of characteristic roots in multivariate analysis. Part I: Null distributions. *Can J Stat* 4: 157–184.

Pillai KCS (1976b) Distribution of characteristic roots in multivariate analysis. Part II: Non-null distributions. *Can J Stat* 5: 1–62.

Plerou V, Gopikrishnan P, Rosenow B, et al. (2002) Random matrix approach to cross correlations in financial data. *Phy Rev E* 65: 066126.

Poor V, Hadjiliadis O (2009) *Quickest Detection*, Cambridge University Press, New York, NY.

Preisendorfer RW (1988) *Principal component analysis in meteorology and oceanography*. North Holland, Amsterdam.

Pukthuanthong K, Roll R (2009) Global market integration: An alternative measure and its application. *J Financ Econ* 94: 214–232.

Pukthuanthong K, Berger D (2012) Market Fragility and International Market Crashes. *J Financ Econ* 105: 565–580.

Reinhart C, Rogoff K (2011) *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press, Princeton, New Jersey.

Fitch cuts Greece's issuer default ratings to 'RD'. (2012, March 9). *Reuters*. Available from: http://www.reuters.com/article/2012/03/09/idUSL2E8E97FN20120309.

Shiryaev AN (1978) *Optimal Stopping Rules*. Springer-Werlag, New York.

Shiryaev AN (2002) Quickest detection problems in the technical analysis of financial data. *Mathematical Finance - Bachelier Congress, 2000 (Paris)*. Springer, Berlin, 487–521.

Silverstein JW (1985) The smallest eigenvalue of large dimensional Wishart matrix. *Ann Probab* 13: 1364–1368.

Silverstein JW (1995) Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J Multivariate Anal* 55: 331–339.

Smith R, Sidel R (2010) Banks keep failing, no end in sight. *Wall Street J*. Available from: http://online.wsj.com/news/articles/SB20001424052748704760704575516272337762044.

Solnik B, Boucrelle C, Le Fu Y (1996) International market correlation and volatility. *Financ Anal J* 52: 17–34.

Sugiyama M, Suzuki T, Nakajima S, et al. (2008) Direct density ratio estimation in high-dimensional spaces, *Ann I Stat Math* 60: 699–746.

Tartakovsky AG, Rozovskii BL, Blazek RB, et al. (2006) A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE T Signal Proces* 54: 3372–3382.

Thottan M, Ji C (2003) Anomaly detection in IP networks. *IEEE T Signal Proces* 15: 2191–2204.

Thurner S, Biely C (2007) The eigenvalue spectrum of lagged correlation matrices. *Acta Phys Pol B* 38: 4111–4122.

Tracy CA, Widom H (1996) On orthogonal and symplectic matrix ensembles. *Commun Math Phys* 177: 727–754.

Trivedi R, Chandramouli R (2005) Secret key estimation in sequential steganography, *IEEE T Signal Proces* 53: 746–757. bibitemTulino2004 Tulino AM, Verd S (2004) Random Matrix Theory and Wireless Communications. *Found Trend Commun Inf Theory* 1: 1–182.

Wetherhill GB, Brown DW (1991) *Statistical Process Control*. Chapman and Hall, London.

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–838.

Wigner EP (1955) Characteristic vectors of bordered matrices with infinite dimensions. *Ann Math* 62: 548–564.

Wishart J (1928) The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* 20: 32–52.

Yamada M, Kimura A, Naya F, et al. (2013) Change-point detection with feature selection in high-dimensional time-series data. *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* 171: 1827–1833.

Yao JF (2012) A note on a Marčenko-Pastur type theorem for time series. *Stat Probab Letter* 82: 20–28.

Zhang M, Kolkiewicz AW, Wirjanto TS, et al. (2015) The impacts of financial crisis on sovereign credit risk analysis in Asia and Europe. *Int J Financ Eng* 2: 143–152.

Appendix

A. Spectral analysis of correlation

This section outlines the concept of correlation structure of a multivariate time series that is referred to throughout the paper. Specifically we are interested in investigating the strength of contemporaneous comovement between components in time series.

A.1. Eigenspectrum analysis

Consider a d-dimensional random vector $\boldsymbol{X}$, with

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_d \end{bmatrix} \sim F \tag{35}$$

for some distribution F and linearly independent $X_i, i = 1, \ldots, d$. Let $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ be the population covariance matrix of $\boldsymbol{X}$. Since $\boldsymbol{\Sigma}$ is a square symmetric matrix, it can be diagonalized to obtain its eigendecomposition

$$\boldsymbol{\Sigma} = \sum_{i=1}^{d} \lambda_i \boldsymbol{\upsilon}_i \boldsymbol{\upsilon}_i' \tag{36}$$

$$= \boldsymbol{\Upsilon} \boldsymbol{\Lambda} \boldsymbol{\Upsilon}'. \tag{37}$$

The eigendecomposition of $\boldsymbol{\Sigma}$ is deterministic up to a scaling factor and in reference to any eigenvectors hereafter, we will mean the normalized scale, i.e., $\boldsymbol{\upsilon}'\boldsymbol{\upsilon} = 1$. It is also unique if $\lambda_i \neq \lambda_j \, \forall i, j$ and we will assume this property to be true.

Computing the eigendecomposition of a covariance matrix amounts to performing a PCA, which is one of the most frequently used techniques in multivariate statistical analysis. The population PCA is defined as a linear transformation $\boldsymbol{X} \mapsto \boldsymbol{Y}$, such that the (principal) components (PCs) of $\boldsymbol{Y}$ are linearly uncorrelated and have successively decreasing variances. In particular the map $\boldsymbol{Y} = \boldsymbol{\Upsilon}'\boldsymbol{X}$ achieves this, if we choose to index the eigendecomposition, such that $\lambda_1 > \ldots > \lambda_d$. Then, an interpretation of the eigenvectors is that $\boldsymbol{\upsilon}_1$ is the direction in $\mathbb{R}^d$ of largest variance in the data; $\boldsymbol{\upsilon}_2$ is the direction of largest variance that is orthogonal to $\boldsymbol{\upsilon}_1$; and each subsequent $\boldsymbol{\upsilon}_i$ points in the direction of greatest variance such that it is orthogonal to the previous eigenvector. This means that the first PC contains the most information about the structure of the system and every subsequent PC contains less information than the previous one.

In addition, if $\boldsymbol{Y}_i = \boldsymbol{\upsilon}_i'\boldsymbol{X}$ is the i-th PC, then

$$\mathrm{Var}[\boldsymbol{Y}_i] = \mathrm{Cov}[\boldsymbol{\upsilon}_i'\boldsymbol{X}, \boldsymbol{\upsilon}_i'\boldsymbol{X}] \tag{38}$$

$$= \boldsymbol{\upsilon}_i' \mathrm{Cov}[\boldsymbol{X}, \boldsymbol{X}] \boldsymbol{\upsilon}_i \tag{39}$$

$$= \boldsymbol{\upsilon}_i' \boldsymbol{\Sigma} \boldsymbol{\upsilon}_i \tag{40}$$

$$= \boldsymbol{\upsilon}_i' \boldsymbol{\upsilon}_i \lambda_i \tag{41}$$

$$= \lambda_i, \tag{42}$$

so λ_i is the variance of the i-th PC. The eigenvalues of $\boldsymbol{\Sigma}$ are all positive because it is symmetric and positive definite. Since the PCs are uncorrelated and the variance of each is given by the corresponding eigenvalue, we can define the total variance in the system as

$$\Omega = \sum_{i=1}^{d} \lambda_i, \tag{43}$$

and the variance accounted for by the first k PCs as

$$\omega_i = \frac{\lambda_i}{\Omega}. \tag{44}$$

Then the proportion of variance explained by the first k PCs is

$$\phi_k = \sum_{i=1}^{k} \omega_i \qquad k \leq d. \tag{45}$$

This framework translates the concentration of cumulative eigenvalue magnitude, or energy, to the strength of co-movement in the components of $\boldsymbol{X}$. The value of ϕ_k can be interpreted as the extent to which the variates in $\boldsymbol{X}$ move together due to some k uncorrelated factors. If a high proportion of total variance is explained by relatively few PCs, then this would manifest in an eigenspectrum with the few corresponding eigenvalues being much larger than the remaining bulk. Large eigenvalues are associated with a structure in the data whereas small eigenvalues are associated with noise. Therefore highly correlated random variables will result in high values of ϕ_k for a relatively small k.

Consider, for an illustrative purpose, an extreme case with a bivariate random vector $\boldsymbol{X} = [X_1 \; X_2]'$ such that $X_1 \sim N(0, 1)$ and $X_2 = 2X_1$. The population covariance matrix for this system is

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \tag{46}$$

with an eigendecomposition

$$\boldsymbol{\Upsilon} = \begin{bmatrix} 0.45 & -0.90 \\ 0.90 & 0.45 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Lambda} = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}. \tag{47}$$

The first PC is proportional to $[1, 2]'$, capturing the full structure of the system, namely variance on the line $X_1 = 2X_2$. The eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = 0$. Therefore, for a sample of n observations, ϕ_1 will be close to one. On the other hand, if X_1 and X_2 are independently distributed standard Normal random variables, then $\lambda_1 = \lambda_2 = 1$ and, statistically, the eigenvectors of a sample covariance matrix from this system will explain an approximately equal amount of variance.[‡]

A.2. Sample eigenspectrum analysis

In most practical cases the population covariance matrix is unknown, so the true eigenvectors and eigenvalues cannot be determined with certainty. Instead the available information is a sample of n realizations of $\boldsymbol{X}$, which we write in matrix form as

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix} \in \mathbb{R}^{d \times n}. \tag{48}$$

The analysis in Section A.1 has a sample analog. First $\boldsymbol{\Sigma}$ is estimated by the sample covariance matrix $\mathbf{S}$, which has an eigendecomposition

$$\mathbf{S} = \sum_{i=1}^{d} l_i \boldsymbol{u}_i \boldsymbol{u}_i' \tag{49}$$

[‡]It will be shown in Section B that the eigenvalues are systematically unequal.

$$= \mathbf{U}\mathbf{L}\mathbf{U}', \tag{50}$$

where the $\boldsymbol{u}_i$'s are eigenvectors with associated eigenvalues l_i. Unlike (37), this decomposition is stochastic and, in fact, $\boldsymbol{u}_i$ and l_i are observation-dependent estimates of $\boldsymbol{v}_i$ and λ_i, respectively. Our estimators of (43)–(45) are then given by

$$\hat{\Omega} = \sum_{i=1}^{d} l_i, \tag{51}$$

$$\hat{\omega}_i = \frac{l_i}{\hat{\Omega}}, \tag{52}$$

$$\hat{\phi}_k = \sum_{i=1}^{k} \hat{\omega}_i \qquad k \leq d. \tag{53}$$

Remark 1. Note that the rows of $\mathbf{X}$ are typically pre-processed to have zero mean. If they were also standardized to have unit variance, then we would have $\mathbf{S} = \mathbf{R}$, where $\mathbf{R}$ is the sample correlation matrix and $\hat{\Omega} = \text{tr}(\mathbf{R}) = d$.

A.3. Time series eigenspectrum analysis

Consider a d-dimensional time series $\{\boldsymbol{X}_t\}$ whose population covariance is $\boldsymbol{\Sigma}$. We are interested in performing a sequential analysis of the eigenspectrum of this random process by using the above framework. Of particular interest is the quantity in (53), which indicates the strength of contemporaneous comovement between the components of $\{\boldsymbol{X}_t\}$. This can be obtained by applying the PCA on a fixed-size rolling sample of observations from $\{\boldsymbol{X}\}_t$. Consider a data matrix with the most recent n observation available at time t, which we denote by

$$\mathbf{X}_t^{(n)} = \begin{bmatrix} \boldsymbol{x}_{t-n+1} & \boldsymbol{x}_{t-n+2} & \cdots & \boldsymbol{x}_t \end{bmatrix}. \tag{54}$$

Following the analysis in Section A.1, at time t, the sample covariance matrix

$$\mathbf{S}_t^{(n)} = \frac{1}{n-1}\mathbf{X}_t\left(\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}_t' \tag{55}$$

has an eigenvalue decomposition as in equation (49). This yields an empirical eigenspectrum $\boldsymbol{l}_t = [l_{1t}, \cdots, l_{dt}]'$ and replacing l_i with l_{it} in (51)–(52) we get $\hat{\Omega}_t$ and $\hat{\omega}_{it}$ respectively. Finally

$$\hat{\phi}_{kt} = \frac{\hat{\omega}_{kt}}{\hat{\Omega}_t}. \tag{56}$$

A.4. Eigenspectrum analysis versus average Pearson correlation

For the purpose of assessing the integration between two financial assets we may naturally question the difference between the framework above and a simple averaging of the elements of the correlation matrix. Obviously strong correlations of opposite signs can result in a low average correlation coefficient. However even computing the average absolute value of correlation coefficients can be disadvantageous. The PCA approach provides a more granular view of the structure of asset

co-movements. Specifically we can see the concentration of variance captured by eigenvalues of different ranks and this can be meaningful information, especially when different eigenvectors have different economic interpretations. For example the dominant eigenvector has been identified as influence of the entire market and subsequent eigenvectors can explain clusters of stocks with similar return dynamics, which have been interpreted as industry effects, for example (Allez and Bouchaud, 2011; Laloux et al., 1999; Pan and Sinha, 2007; Plerou et al., 2002). In addition we may also be interested in a measure of co-movement that takes into account the different return volatilities of assets. In this case the relative magnitude of eigenvalues of the covariance matrix will be a better measure than the average correlation coefficient (Kritzman et al., 2011).

B. The distribution of eigenvalues

We will now discuss the distributional properties of correlation matrix spectra. Some work on the joint distribution of the eigenvalues has been done in the multivariate statistical analysis. However most results on the limiting behaviour of eigenvalues come from the Random Matrix Theory (RMT). In this section we present theoretical results on the distribution of eigenvalues.

B.1. Random matrices

A random matrix is a matrix whose elements are random variables. We shall denote random matrices by calligraphic letters to distinguish them from those with deterministic elements. Thus a random matrix, $\mathcal{X}$, has $X_{ij} \sim F$ as its (i, j)-th element, for some distribution F. The corresponding realizations are $\mathbf{X}$ and x_{ij}. Random matrices are studied as ensembles, each of which includes matrices that usually form an algebraic group (Bejan, 2005). We list some of the most studied ensembles and their construction (Edelman and Rao, 2005) below.

- **Gaussian orthogonal ensemble**: symmetric matrices that can be written as $\frac{(\mathcal{X}+\mathcal{X}')}{2}$ where $\mathcal{X}$ is $n \times n$ with i.i.d. standard real Normal elements.
- **Gaussian unitary ensemble**: Hermitian matrices that can be written as $\frac{(\mathcal{X}+\mathcal{X}^H)}{2}$ where $\mathcal{X}$ is $n \times n$ with i.i.d. standard complex Normal elements. We use $\mathcal{X}^H$ to denote the Hermitian transpose of $\mathcal{X}$.
- **Gaussian symplectic ensemble**: self-dual matrices that can be written as $\frac{(\mathcal{X}+\mathcal{X}^D)}{2}$ where $\mathcal{X}$ is $n \times n$ with i.i.d. standard quaternion Normal elements. We use $\mathcal{X}^D$ to denote the dual transpose of $\mathcal{X}$.
- **Wishart ensemble**: symmetric matrices that can be written as $\mathcal{X}\mathcal{X}^T$, where $\mathcal{X}$ is $d \times n$ with i.i.d. standard Normal elements that can be real, complex or quaternion. We let $\mathcal{X}^T$ denote $\mathcal{X}'$, $\mathcal{X}^H$ or $\mathcal{X}^D$ appropriately.

Of considerable importance to the multivariate statistical analysis is the study of Wishart matrices. Next we will review the Wishart distribution and results relating to the eigenspectrum of Wishart matrices.

B.2. Wishart distribution

Wishart matrices arise most prominently in the analysis of sample covariance matrices. Named after John Wishart, who first computed its joint element density, the Wishart distribution is a generalization of the chi-squared distribution to a higher dimension.

Definition 1 (Wishart Distribution)**.** Let $\mathcal{M} \in \mathbb{R}^{d\times d}$ be a symmetric random matrix such that

$$\mathcal{M} = \mathcal{X}\mathcal{X}' \tag{57}$$

where

$$\mathcal{X} = \begin{bmatrix} \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_n \end{bmatrix} \tag{58}$$

and $\boldsymbol{X}_i \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ independently for $i = 1, \ldots, n$. Then $\mathcal{M}$ has a *Wishart distribution with covariance* $\boldsymbol{\Sigma}$ *and n degrees of freedom*, which we denote by

$$\mathcal{M} \sim W_d(n, \boldsymbol{\Sigma}). \tag{59}$$

Recall that $\mathcal{X}'\mathcal{X}$ is positive semi-definite and thus $\boldsymbol{c}'\mathcal{M}\boldsymbol{c} \geq \mathbf{0}, \forall \boldsymbol{c} \in \mathbb{R}^d$. Now a Wishart matrix is nonsingular, if and only if $n \geq d$ (Muirhead, 1982). Under this condition $\mathcal{M}\boldsymbol{c} \neq \mathbf{0}$, so that $\boldsymbol{c}'\mathcal{M}\boldsymbol{c} > \mathbf{0}$ for every nonzero $\boldsymbol{c}$. In other words a Wishart matrix is positive definite, if and only if $n \geq d$. Only when this condition holds the Wishart matrix will have a density function as formalized in the following theorem from Muirhead (1982).

Theorem 1. *Assume that* $\mathcal{M} \sim W_d(n, \boldsymbol{\Sigma})$ *with* $n \geq d$*. Then* $\mathcal{M}$ *has the density function*

$$f_{\mathcal{M}}(\mathbf{M}; n, d, \boldsymbol{\Sigma}) = \frac{1}{2^{\frac{dn}{2}}\Gamma_d(\frac{n}{2})\,(\det \boldsymbol{\Sigma})^{\frac{n}{2}}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\mathbf{M}\right)(\det \mathbf{M})^{\frac{n-d-1}{2}}, \tag{60}$$

where $\Gamma_d(\cdot)$ *denotes the multivariate gamma function.*

Having introduced the Wishart distribution we can view the sample covariance matrix as a random matrix $\mathcal{S}$. Recall that, unlike the usual definition of the sample covariance matrix, the elements of $\mathcal{S}$ are treated as random variables. Then we have that $\mathcal{S} \sim W_d\left(n-1, \dfrac{\boldsymbol{\Sigma}}{n-1}\right)$. Thus, the joint distribution of its elements exists if and only if $n \geq d$, in which case it is given by

$$f_{\mathcal{S}}(\mathbf{S}; n, d, \boldsymbol{\Sigma}) = \frac{\left(\frac{n}{2}\right)^{\frac{dn}{2}}}{\Gamma_d(\frac{n}{2})\,(\det \boldsymbol{\Sigma})^{\frac{n}{2}}} \operatorname{etr}\left(-\frac{1}{2}n\boldsymbol{\Sigma}^{-1}\mathbf{S}\right)(\det \mathbf{S})^{\frac{n-d-1}{2}}. \tag{61}$$

The eigenspectrum of $\mathcal{S}$ is of interest both in the general study of random matrices, not least due to its importance in multivariate statistics, and, as per Section A.1, in our study of comovement in multivariate series. Modeling the sample covariance matrix as a random matrix allows for analysis that is based on tools from the RMT. However we shall first present what is known about the distribution of eigenvalues from the multivariate analysis. We begin with the following theorem from James (1960).

Theorem 2 (Joint Density of Eigenvalues)**.** *Let $\mathcal{S}$ be a random matrix distributed as $W_d(N, \frac{\mathbf{\Sigma}}{N})$, with $N = n - 1 \geq d$. Define $\mathcal{L} = \operatorname{diag}(l_1, \ldots, l_d)$ to be a diagonal matrix of the eigenvalues of $\mathcal{S}$ and $\mathcal{U} \in O(d)$ such that $\mathcal{S} = \mathcal{U}\mathcal{L}\mathcal{U}'$. Then the joint density of the eigenvalues $l_1 > \ldots > l_d > 0$ is*

$$f_{\mathcal{L}}(l_1 = x_1, \ldots, l_d = x_d; N, d, \mathbf{\Sigma}) = \left(\frac{N}{2}\right)^{\frac{dN}{2}} \frac{\pi^{\frac{d}{2}^2} (\det \mathbf{\Sigma})^{-\frac{N}{2}}}{\Gamma_d\left(\frac{N}{2}\right)\Gamma_d\left(\frac{d}{2}\right)} \prod_{i=1}^{d} x_i^{\frac{N-d-1}{2}} \prod_{i<j}^{d} \left(x_i - x_j\right) \cdot \int_{O(d)} \operatorname{etr}\left(-\frac{N}{2}\mathbf{\Sigma}^{-1}\mathbf{U}\mathbf{X}\mathbf{U}'\right) d\mathbf{U}, \tag{62}$$

where $\mathbf{X} = \operatorname{diag}(x_1, \ldots, x_d)$ and $(d\mathbf{U})$ is the normalized invariant measure on $O(d)$.

The integral term in (62) is in general difficult to evaluate analytically. In a null case $\mathbf{\Sigma} = \lambda \mathbf{I}_d$, using the facts that

$$\int_{O(d)} d\mathbf{U} = 1 \tag{63}$$

and

$$\operatorname{tr}(\mathbf{U}\mathbf{X}\mathbf{U}') = \operatorname{tr}(\mathbf{X}), \tag{64}$$

we have

$$\int_{O(d)} \operatorname{etr}\left(-\frac{N}{2}\mathbf{\Sigma}^{-1}\mathbf{U}\mathbf{X}\mathbf{U}'\right) d\mathbf{U} = \int_{O(d)} \operatorname{etr}\left(-\frac{N}{2\lambda}\mathbf{U}\mathbf{X}\mathbf{U}'\right) d\mathbf{U} \tag{65}$$

$$= \operatorname{etr}\left(-\frac{N}{2\lambda}\mathbf{X}\right) \int_{O(d)} d\mathbf{U} \tag{66}$$

$$= \exp\left(-\frac{N}{2\lambda} \sum_{i=1}^{d} x_i\right). \tag{67}$$

Thus the joint density of the eigenvalues becomes

$$f_{\mathcal{L}}(l_1 = x_1, \ldots, l_d = x_d; N, d, \lambda \mathbf{I}_d) = \left(\frac{N}{2\lambda}\right)^{\frac{dN}{2}} \frac{\pi^{\frac{d}{2}^2}}{\Gamma_d\left(\frac{N}{2}\right)\Gamma_d\left(\frac{d}{2}\right)} \prod_{i=1}^{d} x_i^{\frac{N-d-1}{2}} \prod_{i<j}^{d} \left(x_i - x_j\right) \cdot \exp\left(-\frac{N}{2\lambda} \sum_{i=1}^{d} x_i\right). \tag{68}$$

However the non-null case is much more involved. James (1960) obtains an infinite series representation for a general $\mathbf{\Sigma}$ in terms of a hypergeometric function of matrix arguments:

$$\int_{O(d)} \operatorname{etr}\left(-\frac{1}{2}N\mathbf{\Sigma}^{-1}\mathbf{U}\mathbf{X}\mathbf{U}'\right) d\mathbf{U} = {}_0F_0{}^{(d)}(-\frac{N\mathbf{X}}{2}, \mathbf{\Sigma}^{-1}). \tag{69}$$

$$\tag{70}$$

See Muirhead (1982) for a thorough introduction.

Using similar tools an exact expression of distribution of the largest eigenvalue, l_1, is (see Constantine, 1963; Muirhead, 1982)

$$P(l_1 \leq x) = \frac{\Gamma_d\left(\frac{d+1}{2}\right)}{\Gamma_d\left(\frac{d+N}{2}\right)} \det\left(\frac{N}{2} x \boldsymbol{\Sigma}^{-1}\right)^{\frac{N}{2}} {}_1F_1{}^{(d)}\left(\frac{N}{2}; \frac{N+d}{2}; -\frac{N}{2} x \boldsymbol{\Sigma}^{-1}\right). \tag{71}$$

The distribution of the largest eigenvalue is useful in various applications. For example, in the estimation of a sparse mean vector, the maximum of d i.i.d Gaussian noise variables is of interest (Johnstone, 2011). In addition, the PCA is often used as a dimensionality reduction tool in which extreme eigenvalues play a role in the choice of how many PCs to retain. Most importantly for our study the relative energy of the largest eigenvalue produced from asset returns can shed light on the extent of interconnectedness of a financial market, and, hence, its propensity to experience systemic shocks.

There are issues in practical applications of the density representations mentioned so far. Firstly the true covariance matrix, $\boldsymbol{\Sigma}$, is a required parameter that is usually unknown. Secondly the non-null case of $\boldsymbol{\Sigma}$ results in expressions that are difficult to evaluate analytically. Hypergeometric functions involve infinite series that converge very slowly for large d, and so even numerical solutions are typically inefficient. We proceed with reviewing some results from random matrix theory that are more useful in this sense.

B.3. Results from the RMT

RMT addresses the properties of large matrices whose entries are random variables. The study of random matrices has focused on their eigenvalues as early as the 1920s with the work of Wishart (1928). A couple of decades later, the field has become prominent in nuclear physics, where dynamic systems were approximated by discretization, leading to matrices of very large dimensions. This motivated interest in the limiting behaviour of the eigenvalues and, indeed, the pioneering work of Wigner (1955) and Marčenko and Pastur (1967) is concerned with applications in physics. Significant improvements in computing have, since, proliferated studies of high dimensional data and, nowadays, large matrices are often seen in many fields. Results from the RMT have been used in such areas as:

1. Wireless communication: d input channels and n output channels. See Wetherhill and Brown (1991) , for example.

2. Climate studies: d measurement locations and n time points. See Preisendorfer (1988), for example.

3. Financial data: d is the number of securities and n is the number of return observations. Examples are Gopikrishnan et al. (2001) and Laloux et al. (1999).

Much of the RMT is concerned with the spectral distribution of infinitely large matrices. Specifically, as the size of the matrix grows infinitely large, the eigenvalue behaviour is described in terms of the limiting empirical spectral distribution.

Definition 2 (Empirical Spectral Distribution)**.** Let $\mathbb{1}_{[A]}$ be the indicator function for event A. The

empirical spectral distribution (*ESD*) of a square matrix $\mathbf{X}$ is defined by

$$F^{\mathbf{X}}(x) = \frac{1}{d}\sum_{i=1}^{d} \mathbb{1}_{[l_i \leq x]}. \tag{72}$$

Put otherwise, $F^{\mathbf{X}}(x)$ equals the proportion of eigenvalues not greater than x.

If the ESD of a matrix converges to a non-random distribution F as its dimension increases to infinity, then F is said to be the *limiting spectral distribution* (LSD). Analysis of the LSD of large dimensional random matrices is a central theme in the RMT and is the subject of the most famous results, such as Wigner's *semi-circle law* (Wigner, 1955) and its analog for other random matrix ensembles. Bai and Silverstein (2010) provide a recent comprehensive reference.

Suppose, for instance, that $\mathcal{X}$ is such that its elements are independently distributed random variables $X_{ij} \sim N(0,1)$. The population covariance is given by $\mathbf{I}_d$ and we have that $\lambda_1 = \lambda_2 = \ldots = \lambda_d = 1$. However, a Monte Carlo simulation of randomly drawn $\mathcal{X}$ with $n = d = 10$ produces[§] the ordered eigenvalues $l_1 > l_2 > \ldots > l_{10}$ with systematic errors, as summarized in Table 13. The implication of this argument is that we cannot expect the empirical eigenvalues to have equal energy even if the population eigenvalues are in fact equal. This phenomenon is sometimes described as greater spacing between sample eigenvalues than their population counterparts.

Table 13. Ordered eigenvalues of realizations of $\mathbf{S}$ with $n = d = 10$, averaged over 1000 simulations.

Statistic	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
Observed	3.05	2.25	1.67	1.21	0.84	0.53	0.29	0.13	0.03	0.00
Error	2.05	1.25	0.67	0.21	-0.16	-0.47	-0.71	-0.87	-0.97	-1.00

The next result demonstrates that the distribution of the eigenvalues as $n, d \to \infty$ converges to a non-random distribution. For the case when $\mathcal{X}$ has real elements the LSD is known as the Marčenko-Pastur (MP) law (Marčenko and Pastur, 1967).

Theorem 3 (Marčenko-Pastur law)**.** *Let $\mathcal{X} \in \mathbb{R}^{d\times n}$ be a matrix whose columns are d-dimensional random variables $\boldsymbol{X}_j = [X_{1j} \ \ldots \ X_{dj}]'$ satisfying:*

1. *Independence: the elements X_{ij} are i.i.d. $\forall i, j$.*
2. *Zero mean: $\mathbb{E}(X_{ij}) = 0 \quad \forall i, j$.*
3. *Constant and uniform variance: $\mathbb{E}(X_{ij}^2) = \sigma^2 \quad \forall i, j$.*
4. *Finite fourth moment: $\mathbb{E}(|X_{ij}|^4) < \infty \quad \forall i, j$.*
5. *Asymptotic aspect ratio: $n, d \to \infty$ such that $\dfrac{d}{n} \to y \leq 1$.*

[§]After subtracting the mean and standardizing.

Then the ESD of the sample covariance matrix, $\mathcal{S}$, converges in probability to the Marčenko-Pastur distribution with probability density function given by

$$f_y(x) = \frac{\sqrt{(b-x)(a-x)}}{2\pi\sigma^2 yx}\mathbb{1}_{[a,b]}(x), \tag{73}$$

where $a = \sigma^2(1 - \sqrt{y})^2$ and $b = \sigma^2(1 + \sqrt{y})^2$.

The assumption of independence in the Marčenko-Pastur law regards both the *row-wise* and *column-wise* structure of $\mathcal{X}$. In the case of a d-dimensional time series, $\{\boldsymbol{X}_t\}$, this is called *contemporaneous* and *temporal* structure respectively. Multivariate data without any structure, such as white noise, will have a LSD with the probability density function given in equation (73). However in many practical setting the dependence structure between the entries of $\boldsymbol{X}_t$ is not as trivial and the LSD in such cases will be different. Some extensions of the Marčenko-Pastur Law have addressed contemporaneous dependence, i.e. dependence between the components $X_i, i = 1, \ldots, d$. Silverstein (Silverstein, 1995) considers matrices of the form $\mathcal{Y} = \mathcal{T}_p^{\frac{1}{2}}\mathcal{X}$ where $\{\mathcal{T}_p\}$ is a sequence of Hermitian non-negative definite matrices. A strong LSD is established and a characteristic equation for its Stieltjes transform is given for $\frac{1}{n}\mathcal{Y}\mathcal{Y}'$ under the assumption that $\{\mathcal{T}_p\}$ is bounded in spectral norm and its spectral density converges in distribution to a non-random density. More recently Bai and Zhou (2008) extended the result to a more general $\{\mathcal{T}_p\}$ that satisfies a mild moment condition. There have also been studies that consider temporal dependence using linear models. Jin et al. (2009) establish the existence of a LSD for large sample covariance matrices generated by a vector autoregressive moving average (VARMA) process. An explicit form for the probability density function is provided for VAR(1) and VMA(1) models. Yao (2012) studied a similar problem and determined the LSD and its Stieltjes transform assuming a general linear process. Most recently Liu et al. (2013) extend the Marčenko-Pastur law to multivariate MA(∞) models and find the LSD of the symmetrized lagged autocovariance matrix in terms of its Stieltjes transform. Davis et al. (2014) also consider sample covariance matrices resulting when the rows of $\boldsymbol{X}_t$ are copies of some time-dependent linear process. Their main result states that the joint distribution of the first k eigenvalues converges to a Poisson point process.

The distribution of the largest eigenvalue in the Wishart ensemble is of particular importance in applications such as hypothesis tesing and signal detection. While an exact expression for this had been known, and is given in (71), its hypergeometric function factor renders it intractable in many practical settings. Surveys by Pillai (1976a, b) contain additional discussions on challenges in obtaining marginal distributions from the joint distribution of eigenvalues of Wishart matrices. In some cases the RMT techniques have been useful in this context. Johnstone (2001) establishes that the limiting distribution of the largest eigenvalue obeys the *Tracy-Widom law* in the following theorem.

Theorem 4 (Johnstone's Theorem)**.** *Let $\mathcal{X}$ be a $d \times n$ random matrix with i.i.d. elements $X_{ij} \sim N(0,1)$. Let l_1 be the largest sample eigenvalue of $\mathcal{M} = \mathcal{X}\mathcal{X}'$ and define the centre and scaling constants*

$$\mu_{dn} = \left(\sqrt{n-1} + \sqrt{d}\right)^2 \tag{74}$$

$$\sigma_{dn} = \left(\sqrt{n-1} + \sqrt{d}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{d}}\right)^{\frac{1}{3}}. \tag{75}$$

If $d, n \to \infty$ *such that* $\frac{d}{n} \to y \leq 1$*, then*

$$\frac{l_1 - \mu_{dn}}{\sigma_{dn}} \xrightarrow{\mathfrak{D}} W_1 \sim F_{TW}^{(1)} \tag{76}$$

where the convergence is to the Tracy-Widom law of order 1, which has the distribution function

$$F_{TW}^{(1)}(y) = \exp\left(-\frac{1}{2}\int_y^\infty q(x) + (x-y)q^2(x)dx\right) \qquad y \in \mathbb{R} \tag{77}$$

where $q(x)$ *solves the Painlevé II differential equation*

$$\begin{aligned} q''(x) &= xq(x) + 2q^3(x), \\ q(x) &\sim \mathrm{Ai}(x) \text{ as } x \to \infty, \end{aligned} \tag{78}$$

and $\mathrm{Ai}(x)$ *denotes the Airy function.*

The distribution in (77) was obtained by Tracy and Widom (1996) as the limiting law of the largest eigenvalue of an $n \times n$ Gaussian symmetric matrix. However Johnstone's result states that it also applies to Wishart matrices if l_1 is standardized by the constants in (74) and (75). Evaluating the Painlevé II differential equation, and hence the Tracy-Widom distribution, requires numerical approximation techniques such as in Edelman and Persson (2005). Bejan (2005) provides tables with evaluations and p-values that are analogous to the traditional tables for the Student's t-distribution for example.

C. Stylized properties of eigenvalue-based time series

Our objective here is to expand on the notion of a spectrum-based time series as presented in Section A.3. Ultimately we are interested in performing a sequential analysis of the strength of contemporaneous dependence in the multivariate time series. We recap the setting: consider, at time $t \geq n$, a data matrix $\mathbf{X}_t^{(n)}$ of the most recent n observations $\boldsymbol{x}_{t-n+1}, \ldots, \boldsymbol{x}_t$ from a d-variate time series, $\{\boldsymbol{X}_t\}$. The corresponding sample covariance matrix is $\mathbf{S}_t^{(n)}$ and its eigenvalues, $\boldsymbol{l}_t$, can be used to construct a single observation of the statistics $\hat{\omega}_{it}$ or $\hat{\phi}_{kt}$ as defined in equations (52)-(53). We can also generate series $\{\boldsymbol{l}_t\}$, $\{\hat{\omega}_{it}\}$ and $\{\hat{\phi}_{kt}\}$ by applying the PCA sequentially on a fixed-size rolling window of size n.

To clarify matters let us outline this procedure for an i.i.d series of $\boldsymbol{X}_t \sim N_4(\mathbf{0}, \mathbf{I}_4)$. We generate n arbitrarily ordered (index-wise) observations which we assume is information available at the start of sampling. We compute an initial $\mathbf{S}_0^{(n)}$ and its eigenvalues, $\boldsymbol{l}_0$. Upon arrival of the next observation, we truncate the oldest observation, leaving the most recent n points, with which $\mathbf{S}_1^{(n)}$ and $\boldsymbol{l}_1$ are computed. Let this process continue and the series of interest will be generated. Figure 25 shows the relative magnitude series of eigenvalues produced from a Monte Carlo simulation of 1000 sequential spectra using $n = 100$. Each eigenvalue accounts for approximately a quarter of the variance. In Figure 26 we plot the cumulative explained variance. Recall that $\hat{\phi}_k$ is simply the sum of the first k eigenvalues, normalized by the sum of all eigenvalues. If we choose to standardize the data to have unit variance then the eigenvalues would simply be scaled down by d, the number of variates.

It is useful to compare the empirical eigenvalues to the theoretical ones. In the above experiment, we obtain a mean value of 1.225 for l_1, whereas the theoretical upper support of the Marčenko-Pastur

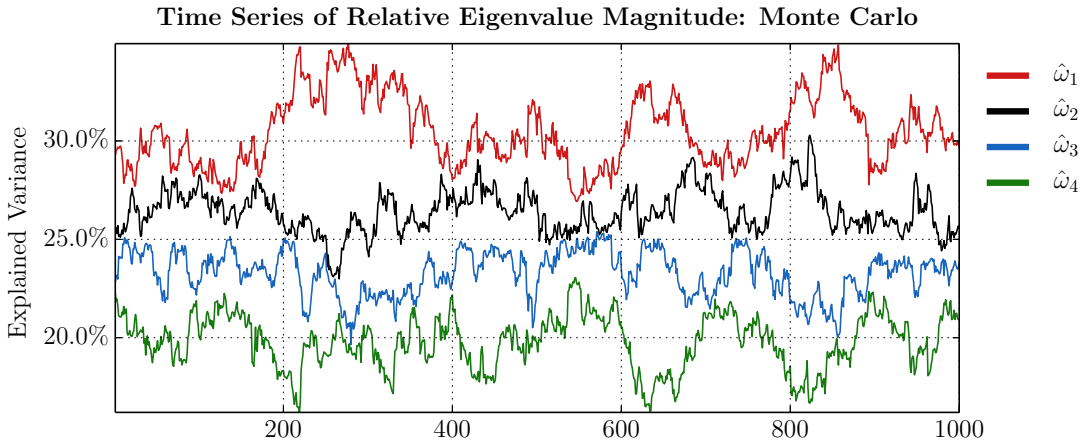

Figure 25. Normalized eigenvalues of a sample covariance matrix $\mathbf{S}_t^{(n)}$ of $\boldsymbol{X}_t \sim N_4(\mathbf{0}, \mathbf{I_4})$, with an initial sample of $n = 100$ rolled over 1000 sequential observations.

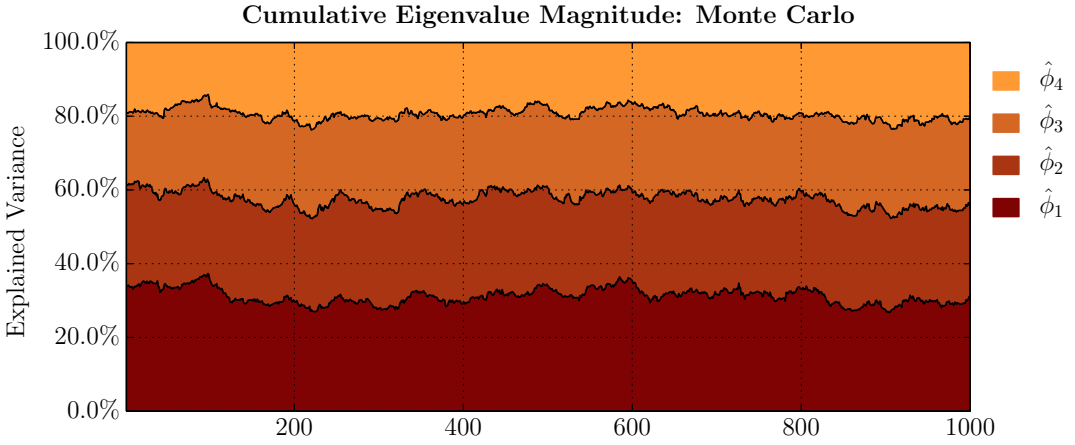

Figure 26. Cumulative normalized eigenvalues of a sample covariance matrix $\mathbf{S}_t^{(n)}$ of $\boldsymbol{X}_t \sim N_4(\mathbf{0}, \mathbf{I_4})$, with an initial sample of $n = 100$ rolled over 1000 sequential observations.

for an aspect ratio $y = \frac{4}{100}$ is 1.44. If we increase both d and n by a factor of 100, l_1 comes at a mean value of 1.39. In fact the limits of the smallest and largest eigenvalues have been shown (see Geman, 1980 and Silverstein, 1995) to converge to the lower and upper edges of the support of the Marčenko-Pastur density, denoted by a and b in Theorem 3 respectively. Formally, we have

$$l_1 \to (1 + \sqrt{y})^2, \; a.s. \tag{79}$$

$$l_d \to (1 - \sqrt{y})^2, \; a.s. \tag{80}$$

where the almost sure convergence is to be interpreted as occurring with probability one as $d, n \to \infty$. Thus comparing empirical extreme eigenvalues with their theoretical counterparts can give information about the structure in the data. If the data is white noise, as is the case underlying the Marčenko-Pastur law, then we would expect empirics to be close to theory. In contrast a relatively strong structure in the data would result in a significantly larger dominant eigenvalue than predicted by the RMT.

C.1. Autocorrelation

Clearly each $\{\hat{\omega}_{it}\}, i = 1, \ldots, d$, above is not serially independent because their values are obtained using a rolling window of observations. In our construction the last $n-1$ columns of $\mathbf{X}_t^{(n)}$ are the same as the first $n-1$ columns of $\mathbf{X}_{t+1}^{(n)}$. The window size, n, acts as a smoothing parameter in the sense that a larger n results in higher autocorrelations and smaller variance in the eigenvalue series. However the window size also affects the mean values through the dependence of the eigenvalue on the aspect ratio $y = \dfrac{d}{n}$. These dynamics are illustrated in Figure 27 where we plot $\{\hat{\omega}_{1t}\}$ and its autocorrelation functions for simulations using windows of varying length.

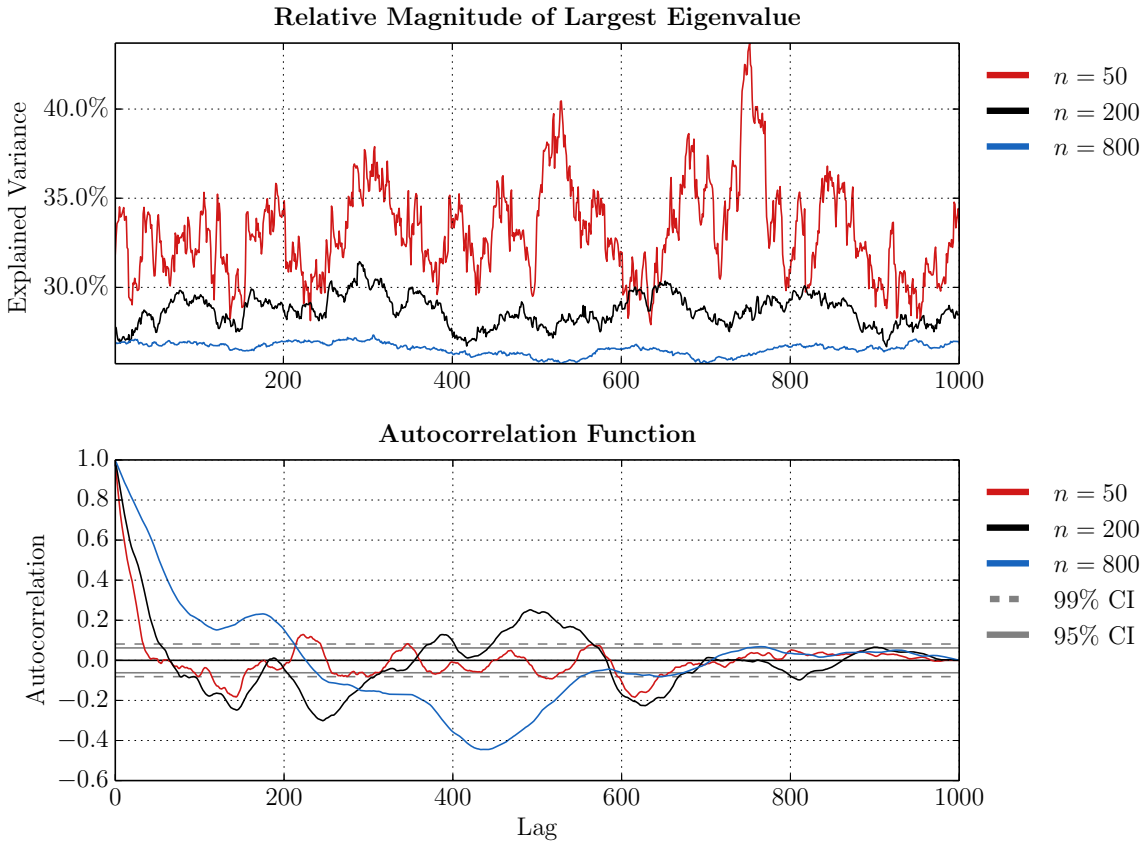

Figure 27. Top – cumulative normalized eigenvalues of a sample covariance matrix $\mathbf{S}_t^{(n)}$ of $\boldsymbol{X}_t \sim N_4(\mathbf{0}, \mathbf{I_4})$, with samples of varying window length. Bottom – autocorrelation functions and confidence intervals for each series in the top panel.

It can be seen from the figure that as n increases, the autocorrelation decays slower. Of course serial dependence could be statistically eliminated if we lagged our sampling, such that there is no overlap between successive data matrices $\boldsymbol{X}_t$. However a more common approach is to difference the series such that each observation represents increments $\Delta\hat{\omega}_{1t} = \hat{\omega}_{1t} - \hat{\omega}_{1t-1}$. In Figure 28, we plot the first difference of the resulting $\{\hat{\omega}_{1t}\}$ when $n = 50$. Note that the autocorrelation values now fall within the confidence intervals around a zero autocorrelation coefficient. This is rather intuitive as this differenced

series reflects changes in the amount of variance explained by the largest eigenvalues estimated from successive i.i.d. observations of the data.

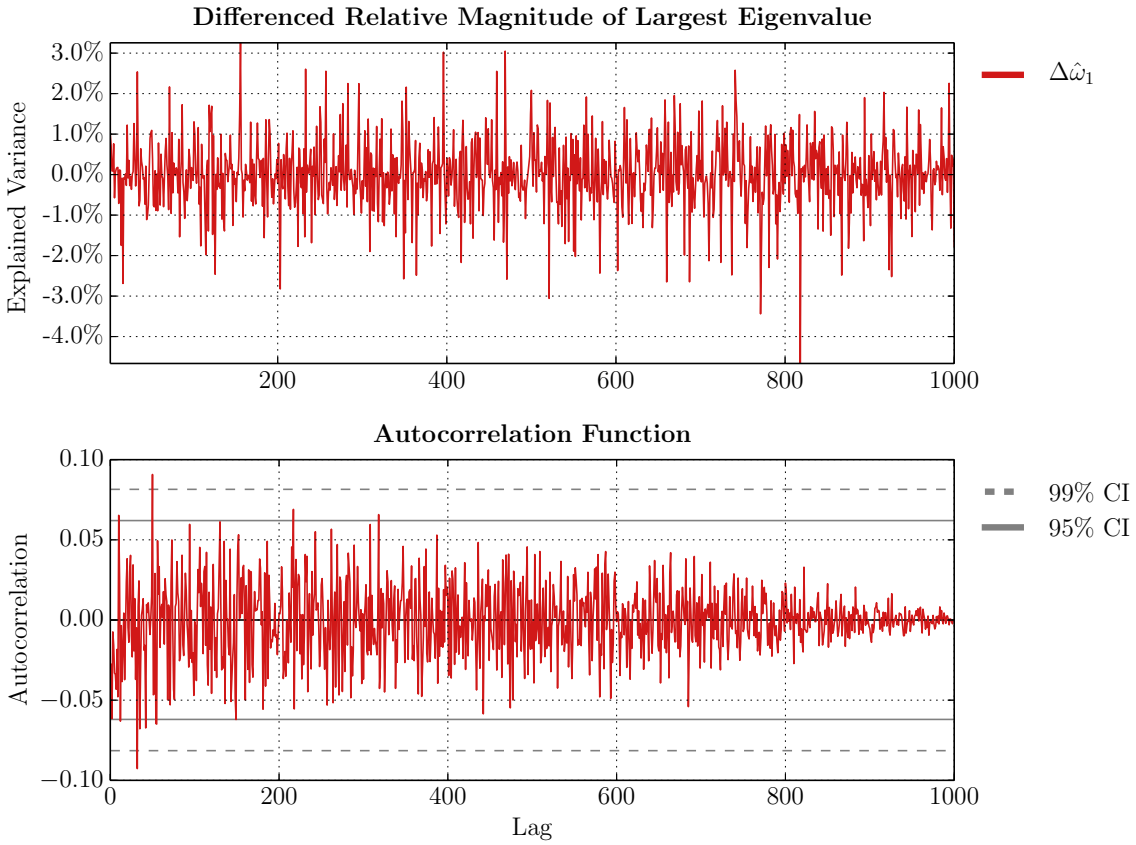

Figure 28. Top – first difference of $\hat{\omega}_{1t}$ generated with $n = 50$. Bottom – Autocorrelation function with confidence intervals for the series in the top panel.

C.2. Stationarity

Results from Section B on the distribution of eigenvalues provide justification for the claim of eigenvalue series stationarity under certain conditions. Consider first the relative magnitude of single eigenvalues $\hat{\omega}_{it}, i = 1, \ldots, d$. The stationarity of each depends on that of l_{it}. We know that in the case of a d-variate i.i.d. Gaussian process, the distribution of the largest eigenvalue exists and is given by equation (71). Asymptotic distributions are given by Anderson (1963) for fixed d and $n \to \infty$ in the form

$$\sqrt{\frac{n}{2}}\left(\frac{l_1}{n} - 1\right) \xrightarrow{\mathfrak{D}} N(0, 1), \tag{81}$$

and by Johnstone (2001) as in Theorem 4. Denote this by $F_{l_1}(l_1 \leq x; n, d, \mathbf{\Sigma})$ to emphasize the parameter set of the distribution. Clearly the conditions under which $\{l_{1t}\}$ is stationary depend on the underlying time series and the sampling procedure which produces each l_{1t}. Specifically, if the underlying process is at least second-order stationary, then taking disjoint samples $\mathbf{X}_t^{(n)}$ will result in i.i.d realizations l_1.

Table 14. Stationarity tests for 1000 simulated series with $n = 100$ and $d = 4$.

Test	$\hat{\omega}_i$				$\Delta\hat{\omega}_i$			
	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
Reject H_0^{ADF}	0%	0%	0%	0%	100%	100%	100%	100%
Reject H_0^{PP}	0%	0%	0%	0%	100%	100%	100%	100%
Do not reject H_0^{KPSS}	0%	0%	0%	0%	98%	100%	92%	94%

Table 15. Stationarity tests for 1000 simulated series with $n = 100$ and $d = 4$.

Test	$\hat{\phi}_k$			$\Delta\hat{\phi}_k$		
	$k = 1$	$k = 2$	$k = 3$	$k = 1$	$k = 2$	$k = 3$
Reject H_0^{ADF}	0%	0%	0%	100%	100%	100%
Reject H_0^{PP}	0%	0%	0%	100%	100%	100%
Do not reject H_0^{KPSS}	0%	0%	0%	92%	100%	96%

To the best of our knowledge there are no results on non-extreme eigenvalues.

However, in empirical applications, the rolling-window fashion by which the eigenvalues are calculated results in some dependency. Thus we conduct three tests on $\hat{\omega}_{it}$ and $\Delta\hat{\omega}_{it}$ to investigate their stationarity using the following unit-root tests: (1) Augmented Dickey-Fuller (ADF) parametric test (Dickey and Fuller, 1979); (2) Phillips-Perron (PP) nonparametric test (Phillips and Perron, 1988) and (3) Kwiatkowski, Phillips, Schmidt and Shin (KPSS) Lagrange multiplier test (Kwiatkowski et al., 1992). The null hypotheses for these tests are respectively given by

$$H_0^{ADF} = H_0^{PP} : \text{the series has a unit root,} \tag{82}$$

$$H_0^{KPSS} : \text{the series does not have a unit root.} \tag{83}$$

A unit root is a feature of a nonstationary time series, so that a rejection of the hypothesis in (82) for a given time series is to be interpreted that it is second-order stationary. On the contrary the opposite conclusion is drawn from the KPSS test if there is sufficient evidence to reject its null hypothesis, H_0^{KPSS}. Results for our tests are summarized in Table 14, where the relative magnitude and changes in relative magnitude for each eigenvalue are taken into consideration. The tests consistently conclude that $\hat{\omega}_{it}$ is not stationary since, in 1000 simulations, there are no rejections of the unit root hypothesis. However the first-differenced series is found to be stationary by all of the tests in the vast majority of simulations. Table 3 summarizes results from similar tests for the cumulative sums $\hat{\phi}_{kt} = \sum_{i=1}^{k} \hat{\omega}_{it}, k = 1, \ldots, d-1$. As before the time series are found to be stationary only after first differencing.