Systemic Risk from Asset Concentration and Common Holdings among Banks

Celso Brunetti[◊], Jeffrey H. Harris[○] and Shawn Mankad[□]

Abstract

The recent financial crisis has focused attention on identifying and measuring systemic risk. In this paper, we propose a novel approach to estimate the portfolio composition of banks as function of daily interbank trades and stock returns. While banks' assets are reported to regulators and/or the public at relatively low frequencies (e.g. quarterly or monthly), our approach is able to derive bank assets holdings at higher frequencies. From asset holdings, we are able to derive precise estimates of (i) portfolios' concentration within each bank—a measure of the degree of diversification—and of (ii) common holdings across banks—a measure of shock propagation. We find evidence that systemic risk measures derived from our approach lead, in a forecasting sense, several commonly used systemic risk indicators.

Key words:

JEL:

[◊]Federal Reserve Board (celso.brunetti@frb.gov); [○]American University and SEC (jharris@american.edu); [□]Cornell University (spm263@cornell.edu) corresponding author, this material is based upon work supported by the National Science Foundation under Grant No. 1633158 (Mankad).

1 Introduction

The recent financial crisis has accentuated the need for effective monitoring, oversight and regulation of financial markets and institutions. In response, governments around the world have created new regulatory frameworks. For instance, in the U.S., the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank) enhanced oversight and regulation and similarly the European Union introduced several supervisory bodies (European Banking Authority, European Securities and Markets Authority, etc.) to do the same. Internationally, the Financial Stability Board has been created with the mandate of promoting financial stability across multiple regulatory jurisdictions.

One consequence of launching these new regulatory regimes in the Digital Age is a vast and increasing amount of data that is available to regulators on the behavior of market participants. For instance, a main thrust of Dodd-Frank was to create transparency and accountability in markets through a myriad of data recording and reporting requirements that apply to exchanges and their market participants. Yet, despite increased reporting requirements, most regulators have access only to data directly related to their legal purview. Given the significant number of agencies and fragmented nature of the regulatory oversight ecosystem, it remains a major challenge for regulators to gain deep insight into the balance sheet of complex financial institutions which trade products across various regulated and unregulated markets. Thus, a common and important problem faced by regulatory bodies around the world is how to integrate financial data streams together in a principled manner to yield new and informative insights.

The aim of this work is to help address this key issue in a setting where regulators have access to two informationally-linked data sources available at the daily frequency: (i) equities (stock movements) and (ii) interbank lending data. Stock returns are of course widely and publicly available, while interbank lending data is accessible to central banks. Using these two sources, we build on the accounting frameworks of Shin [2009, 2010], Elliott et al. [2014] and Brunetti et al. [2015] to extract daily estimates of the composition of the unobserved portfolios held by individual banks. [1] More specifically, we can estimate the weights of each individual bank's underlying portfolio, thus giving regulators meaningful and timely information about the systemic holdings across the banking system.

Our methods involve solving a matrix factorization problem within a novel Bayesian estimation framework (details in Section 2 below). The solution provides estimates of each individual bank's underlying portfolio which we then use to characterize risk within and among banks. In particular, we derive an index of portfolio concentration for each individual bank which captures the diversification, or lack thereof, of each bank portfolio, and an index of portfolios similarity across banks. The concentration of an individual bank's portfolio on a particular investment has well-known risk implications [Klein and Bawa, 1977; Santis and Gerard, 1997; Gale and Gottardi, 2017; Pastor, Stambaugh and Taylor, 2017]. By estimating the daily holdings, our methods allow regulators to better assess, in a timely manner, the probability of concentrated risk within a bank

[1] We partition quarterly balance sheets of the banking sector to isolate the underlying (and unobserved) portfolios held by banks. This partitioning, when combined with balance sheet identities, implies a variant of the non-negative matrix factorization problem [extensively studied in other domains, e.g. Lee and Seung, 1999]. In particular, our matrix factorization problem requires solving for one factor that is subject to probability constraints.

without having to directly examine bank balance sheets on a more low frequency basis (typically a quarter). Moreover, the similarity between banks in terms of holdings indicates how interconnected banks portfolios are. This measure of common asset holdings across banks is very important for the propagation of shocks [see e.g. Greenwood et al., 2015, Caccioli et al., 2014, 2015].

While our methods indirectly estimate banks' holdings, we provide evidence that our estimates are close approximations of real balance sheet data. In fact, we validate our estimates rigorously in two ways. First, since the statistical model and estimation framework are novel,, our first validation relies on Monte Carlo simulations and demonstrates that our estimation approach produces reliable results. In the simulation analysis we compare different estimation techniques and we show that our Bayesian approach is the most reliable. Second, we validate the model itself from an accounting point of view by showing that the estimates from the model match closely real accounting data.

Once we know the asset composition of each bank portfolio, we derive two indices. The first index looks at the *concentration* of individual portfolios while the second looks at the degree of *common holdings* across different banks. Both indices convey important information in a forecasting sense. We show that these two variables one-way Granger cause risk measures that are published by the European Central Bank[2] (ECB) during the most recent financial crisis. That is, we find that a more similar and concentrated banking sector is a leading factor and harbinger of market stress at monthly horizons. In this respect our paper contributes to a growing literature, prompted by the recent crisis, on measures of

[2] http://sdw.ecb.europa.eu/home.do

systemic risk. Acharya, Pedersen, Philippon and Richardson [2017] develop a systemic risk measure based on the (expected) amount a bank is undercapitalized in a crisis event. Brownlees and Engle [2017], measure systemic risk as the contribution of each firm in terms capital shortfall in severe market movements. Adrian and Brunnermeier [2016] used quantile regressions to compute *CoVar*, the *Var* for the financial sector conditional on a bank having had a *VaR* loss. Huang, Zhou and Zhu [2009] propose a measure of systemic risk based on default probabilities of individual banks (extracted from CDS data) and forecasted asset return correlations. In a related paper, Giudici, Sarlin and Spelta [forthcoming], introduce a measure of systemic risk which combines direct exposures with common exposures—i.e. they combine what Brunetti et al [2015] refer to as correlation network and physical network.[3] Our measures are different from what the literature has proposed so far. We propose a measure that captures concentration in a bank portfolio (i.e. diversification or lack thereof) and hence is specific to a bank risk profile, and a measure of similarities across bank's portfolios which links the riskiness of each bank to other banks in the system. A shock to a bank which is not well diversified could have a larger impact on the bank's balance sheet and can propagate to other banks if banks in the system hold similar portfolios. While many of the available risk measures relate to capital adequacy and hence are more concerned with the liability side of a bank balance sheet, our measures provide insights on the riskiness of the asset side of the balance sheet. In the last part of the paper we compare our systemic risk indicators to those

[3] See also Tarashev, Borio and Tsasaronis [2010], de Jonghe [2010], Segoviano and Goodhart [2009]. For a survey on systemic risk measures, see Biasis, Flood, Lo and Valavanis [2012].

developed by Acharya et al [2017] and Brownlees and Engle [2017] and find evidence that our risk indicators lead, in a forecasting sense, those other measures of systemic risk.

Finally, we show that the moments of our measures (standard deviation, skewness and kurtosis) convey information. Higher moments may be capturing heterogeneity in banks' risk behavior and may be linked to uncertainty in such behaviors. We also add to the growing literature that shows common asset holdings by banks can create dangerous market conditions, wherein shocks to bank equity can result in financial contagion through fire sales [e.g. Greenwood et al., 2015; Caccioli et al., 2014, 2015]. We provide a novel empirical methodology for regulators to employ in monitoring the banking sector. Using interbank lending and information gleaned from stock market price changes, our methods yield insight into the balance sheets of banks at a higher frequency than more cumbersome and less timely quarterly or annual disclosures or audits allow. Moreover, our methods complement other approaches to assess and monitor systemic risk that build on network science techniques [Billio et al., 2012; Diebold and Yılmaz, 2014; Brunetti et al., 2015; Giudici et al., forthcoming].

2 An Accounting Framework

As a starting point for building our systemic risk measures, we first employ the same framework as in Shin [2009, 2010], Elliott et al. [2014] and Brunetti et al. [2015] wherein individual bank balance sheets can be connected via interbank lending and common holdings, and then aggregated to the industry level. Let there be n banks under consideration and X be the vector of interbank debt (the total value of liabilities held by other banks). Π_{ij} is the share of bank i's liabilities held by bank j, W_{ik} is the weight invested in each of the K assets by bank i ($\sum_k W_{ik} = 1$), Y_{ik} denotes the market value of bank i's assets, e_i indicates bank i's equity (which we proxy for with the market value of equity), and d_i is the total value of liabilities of bank i held by non-banks.

Consider a financial system in which banks connect lenders to borrowers as intermediaries, collecting deposits from households and firms and investing the deposits in a portfolio of assets, including loans to the household sector (via mortgages and consumer debt) and firms. The balance sheet for any individual bank i can be partitioned as follows.

Assets	Liabilities
$\sum_k W_{ik} Y_{ik}$	e_i
	x_i
$\sum_j x_j \Pi_{ij}$	d_i

We obtain the balance sheet identity as

$$\sum_k W_{ik} Y_{ik} + \sum_j x_j \Pi_{ij} = e_i + x_i + d_i$$

or, using matrix notation, as

$$\Pi X + (W \odot Y)u = E + X + D$$

where u is a vector of ones of length K. We can therefore express the portfolio of assets held by each bank as follows

$$(W \odot Y)u = E + (I - \Pi)X + D$$

where I is the $n \times n$ identity matrix; $\odot$ denotes the Schur product (element wise multiplication), so that $C = (A \odot B)$ and $C_{ij} = A_{ij}B_{ij}$.

Recall that D represents debt claims on the banking sector by households, mutual and pension funds and other non-bank institutions. Following Shin [2009], we assume that the debt liabilities to non-banks evolve slowly. We also assume that W, the weight invested in each of the k assets, evolves slowly, whereas the value of the corresponding holdings fluctuates more rapidly over time (e.g. from day to day, or week to week). Thus, over appropriately short intervals, changes to D and to W are negligible and the following identity applies to the balance sheet

$$(W \odot Y)u = E + (I - \Pi)X + D$$

$$\big(W \odot (Y_t - Y_{t-1})\big)u = E_t + D + (I - \Pi_t)X_t - (E_{t-1} + D + (I - \Pi_{t-1})X_{t-1})$$

$$\underbrace{\big(W \odot (Y_t - Y_{t-1})\big)u}_{Unobserved} = \underbrace{E_t - E_{t-1} + (I - \Pi_t)X_t - (I - \Pi_{t-1})X_{t-1}}_{Observed}. \qquad (1)$$

Note that we assume that changes to the equity account $(E_t - E_{t-1})$ can be readily measured for public banks from public stock prices, wherein the market incorporates important, but largely non-specific information about the status of the bank into daily stock price changes. In this light, stock returns might reflect information about the assets held by the bank, the loan portfolio held by the bank, or any other asset or liability on the bank's balance sheet. Conceptually, changes in these other accounts are reflected in the bank's equity accounts and so we proxy for these underlying changes through changes in bank stock prices.

Similarly, intraday transaction-level interbank lending data can be used to construct Π_t, the adjacency matrix of interbank transactions, and X_t, the vector of debt held by other banks. We also note that X_t can only be partially observed—banks can lend each other money through various (often unobservable) mechanisms. For instance, the European banks which we study can trade across the e-MID electronic system (which we observe—see data description below), bilaterally via phone brokers in the over-the-counter (OTC) market, and with the ECB directly.

Yet, in spite of utilizing partial information on interbank debt, the factorization discussed in the next section is able to produce estimates of W that are able to forecast financial market stress.

3 Balance Sheet Driven Bayesian Factorization

Given the accounting identity that links banks together through interbank lending arrangements and holdings of common assets, we aim to quantify the portfolio composition of each bank. Using our notation from above, let

$$Z_t = E_t - E_{t-1} + (I - \Pi_t)X_t - (I - \Pi_{t-1})X_{t-1}$$

and

$$V_t = Y_t - Y_{t-1}.$$

Written in element form, Equation (1) implies that the i-th bank's balance sheet satisfies

$$(Z_t)_i = \sum_k (W)_{ik}(V_t)_{ki}$$

Assuming that the investment opportunity set is the same for all banks, we can express the same equation in matrix form

$$Z = WV \qquad (2)$$

subject to $\sum_{k=1}^{K} W_{ik} = 1$ for all i and $W_{ij} \geq 0$ for all i, j, where $Z = [Z_1, Z_2, \ldots, Z_T]$ is an $n \times T$ matrix, W is an $n \times k$ matrix with non-negativity constraints on the rows, and $V = [V_1, V_2, \ldots, V_T]$ is an $k \times T$ matrix.

Equation (2) can be readily seen as a probability matrix factorization problem, where Z is given and the objective is to estimate W and V. The main contributions in non-negative matrix factorization typically pose an optimization problem based on minimizing the Frobenius norm of the difference between Z and the estimated factors to estimate W and V in (2) [Lee and Seung, 1999, Lin, 2007, Mankad et al., 2016]. However, due to the non-negativity and sum-to-one constraints on W, the resultant problem is challenging to solve.[4]

In this light, we develop a novel Bayesian estimation framework, capturing the non-negativity constraints using appropriate distributional assumptions.[5] Specifically, we assume that each row of W, denoted by w_i, is distributed according to a Dirichlet distribution with common concentration parameter α.

$$p(w_i) = Dir(\alpha) \qquad (3)$$

The Dirichlet distribution, whose range is all discrete probability distributions of length K, is commonly utilized in nonparametric Bayesian statistics to model unknown probability distributions [Antoniak, 1974, Sethuraman, 1994].[6] α is the

[4] The usual approach (see, for example Huck et al. [2010], Heinz et al. [2001]) to deal with the su-to-one constraint is to find approximate solutions (i.e., continuous relaxation of constraint via a Lagrangian penalty) or to ignore the constraint in the estimation and normalize the factors ex-post in a second stage. Both conventions have computational advantages, but do not guarantee robust solutions. Indeed, we find that conventional optimization methods can provide qualitatively different solutions depending on the random seed, reducing the practical application of these methods.

[5] This contrasts with previous Bayesian factorization models that have a similar framework but solve for non-negativity without the probability constraint [Schmidt et al., 2009, Psorakis et al., 2011].

[6] More recently, it has been popularized in the Latent Dirichlet Allocation model of Blei et al. [2003] and applied extensively for summarizing unstructured text data with so-called topic modeling analysis. We use this distribution for the rows of W to capture the probability constraint.

common concentrated parameter of the Dirichlet distribution, which can take any value greater than zero. As α gets larger, the probabilities are closer to uniform, meaning that the asset holding factors within each bank portfolio and across banks are approximately equal. As α approaches zero, the distribution features greater amounts of sparsity (zeros) and concentrates on one or more components within each bank portfolio – though the component that is concentrated upon can vary between banks.

Since V represents changes in asset values at the daily level, over long enough intervals we expect its distribution be unimodal and centered on a small constant capturing market trends. We may also expect the true distribution of V to have heavier tails as has been established for stock returns [Upton and Shannon, 1979]. As our results will show the Gaussian distribution offers a suitable approximation with computational advantages. As such, elements of V are assumed to be independently Normally distributed with mean μ and variance σ_V^2

$$p(V) = \prod_{k,j} \mathbb{N}(\mu, \sigma_V^2)\,. \qquad (4)$$

We note that while the prior distribution assumes that the daily returns between asset classes are independent, the posterior distribution of V will in general exhibit correlations between asset classes.[7] Thus, in effect, the correlation structure between asset returns is learned implicitly through the estimation that is described later in this section.

[7] See the last sub-section of Appendix 1 for mathematical details.

We introduce one last random variable, σ^2, that controls the variance of additive Gaussian noise on each element of the matrix Z and is modeled with an inverse gamma density with shape η and scale θ.

$$p(\sigma^2) = IG(\eta, \theta) \qquad (5)$$

The inverse gamma density as a prior distribution for the noise variance σ^2 is a natural choice and has been extensively utilized [e.g. Brav 2000; Cremers 2002; Jones 2003; Korteweg and Sorensen 2010].

To complete the Bayesian specification, we assume that Z has the following conditional likelihood

$$p(Z|W, V, \sigma^2) = \prod_{ij} \mathbb{N}\big((WV)_{ij}, \sigma^2\big). \qquad (6)$$

The Normal distribution is again used for tractability and computational reasons, though this does not necessarily sacrifice the overall accuracy of the factorization even when Z follows a non-Gaussian distribution. Specifically, Equations (2) and (6) can be viewed as mixture models, where Gaussian means encoded within the columns of V are added together using weights in W. The number of components in the mixture are determined by the rank of the factors W and V, which is set by the analyst. With such mixture models, scholars have shown that with enough components the resultant mixture distribution given by WV has sufficient flexibility to approximate any continuous distribution for Z (subject to regularity conditions) to an arbitrary degree of accuracy (see Norets and Pelenis, [2012]; Rossi [2014]).[8] This also provides intuition for when our model will not work well. We expect our Bayesian specification to struggle if the true distribution of Z is

[8] The intuition for this result is that any density can be well approximated using multiple small variance normal components with different means to position the components appropriately.

discontinuous, truncated, or having other boundary effects. In our real data, we see no evidence for this concern.

We briefly discuss estimation of the Bayesian model next, with full derivations provided in the Appendix. By Bayes rule, the joint posterior is proportional to

$$\mathrm{p}(\mathrm{W},\mathrm{V},\sigma^2|\mathrm{Z}) \propto \mathrm{p}(\mathrm{Z}|\mathrm{W},\mathrm{V},\sigma^2)\mathrm{p}(\mathrm{W})\mathrm{p}(\mathrm{V})\mathrm{p}(\sigma^2), \quad (7)$$

where we utilize the fact that W, V, and σ^2 are assumed to be independently distributed as in Equations (3)-(6).

Computing the posteriors densities p(W|Z) and p(V|Z) requires solving an intractable integral of the joint posterior distribution in Equation (7). To overcome this challenge, we utilize a combination of standard Markov Chain Monte Carlo (MCMC) methods. The basic idea behind MCMC methods is to construct a Markov chain that has the desired distribution as its limiting distribution. Thus, once the Markov chain has converged to its equilibrium, repeatedly sampling states of the chain provides an empirical estimate of the desired distribution that is accurate to an arbitrarily high degree. From this empirical distribution, the expectation can be readily calculated.[9]

Since we can apply conjugate distributional properties[10] to derive explicit, closed forms of the posterior distributions for $V$and σ conditional on the data (Z) and the current state of each of the parameters (W, V, σ^2), Gibbs sampling is used to estimate the marginal distributions $p(V|Z)$ and $p(\sigma|Z)$. In other words, the

---

[9] See Casella and George (1992) and Chib and Greenberg (1995) for further information on MCMC methods, including how to determine whether the Markov chain has converged to its limiting distribution and other best practices.

[10] In Bayesian probability theory, if the posterior distributions (e.g., p(V | Z)) are in the same family as the prior probability distribution (e.g., p(V)), the prior and posterior are called conjugate distributions. Such distributions have a closed-form expression for the posterior distribution.

Markov chain is defined by the conditional posterior distributions and iterated until convergence as in any MCMC method, after which samples are drawn and averaged to derive point estimates.

A more general version of Gibbs Sampling, the Metropolis Hastings algorithm, is used to estimate $p(W|Z)$, because the conditional posterior distribution of W is not composed of conjugate distributions and thus cannot be characterized analytically. The estimation procedure exploits the fact that we are still able to compute the value of a function (shown explicitly in the Appendix) that is proportional to the desired distribution. This proportion is used to generate Markovian samples iteratively that converge to the desired distribution as the number of samples grows.

4 Validating the Model

In this section, we validate the model and Bayesian estimation framework from both a statistical perspective through a simulation exercise, and from an accounting perspective by comparing the estimates of W (the vector of weights invested in each asset class) against actual balance sheet data from each bank. In both tasks, we utilize non-parametric hypothesis tests to compare the distribution of the true W with its estimate. Specifically, we utilize four well known non-parametric tests. The Brown-Mood median test [Brown et al., 1951] and the Fisher-Pitman permutation test [Boik, 1987] assess whether two samples have identical medians and means, respectively. The third test is the more general Mann-Whitney U test [Mann and Whitney, 1947] which compares the full distributions of the estimated and true W to assess whether our estimate is stochastically smaller (or larger) than its true value.[11] And lastly we utilize the Two Sample Anderson Darling Test [Scholz and Stephens, 1987 following Anderson and Darling, 1954] to assess whether there are differences between the two samples with particular sensitivity at the tails of the distributions.[12]

Note that element-wise accuracy comparisons for W (like mean squared error) are not possible given the large number of asset classes and that the columns of the estimated W can be ordered arbitrarily, a common property of such factorization models. As such, in addition to the distributional tests, we also report the Rand Index, a classic accuracy measure for this setting [Rand 1971], which varies from zero to one, with larger values indicating more accurate estimates. To understand the accuracy of the overall factorization we report a Pseudo R^2, which is analogous to the R^2 in linear regression.[13]

Lastly, we also compare the accuracy of the proposed model relative to competing methods. As mentioned above, there are several optimization based techniques in the matrix factorization and Machine Learning literature that can be used to solve Equation (2). The methods we compare against are as follows:

1. The Semi Non-Negative Matrix Factorization model of Ding et al. [2010] with probability constraints enforced ex-post, which is state of the art in the matrix factorization field;

[11] To understand the precise hypothesis tested by the Mann-Whitney U test, let x and y be two random variables with cumulative distribution functions f and g respectively. The hypotheses for the test $H_0: f(\cdot) = g(\cdot)$ *versus* $H_1: f(a) < g(a)$ *or* $f(a) > g(a), \forall a$.

[12] This is comparable to the Kolmogorov-Smirnov test, which is not as appropriate in our setting since the real balance sheet data has multiple zero values. Moreover, the Anderson Darling test has been shown in Monte-Carlo studies to have comparatively greater statistical power [Razali et al., 2011].

[13] The Pseudo $R^2 = 1 - \|Z - \widehat{W}\widehat{V}\|_F^2 / \|Z - \bar{Z}\|_F^2$.

2. Fuzzy K-means, a classic machine learning algorithm, that produces estimates of W based on a Gaussian mixture model [Bezdek et al., 1984].

4.1 Simulation

We test the accuracy and validity of the proposed model under different simulation settings. The first simulation establishes self-consistency of the proposed factorization, that is, we generate the matrix Z from the model implied by the factorization. Then we perform the estimation with perfect knowledge of the true underlying parameterizations. In practice, this information would of course not be known at the start of the estimation, but since this is the first time that we test this methodology, we need to establish whether the estimation procedure is valid. The second simulation misspecifies the hyper-parameters and initial values to help us gain insight into the sensitivity and validity of the estimation under the more realistic condition that the underlying parameterizations are unknown.

First, we generate W, V, and σ according to their distributions, where the concentration value and initializations are set to be equal to the values used in our real data analysis ($\alpha = 0.2$, $\mu = 0$, $\sigma_V = 0.45$, $k = 100$, and $\theta = 0.5$) with W and V of dimension 49×8 and 8×23, respectively, also chosen to match the real data.[14] Second, under the hyper-parameter misspecification scenario, we perform the estimation with the value of the prior parameter α ranging from 0.15 to 0.35 (the true value always remains 0.20). Studying the performance of the estimation for

[14] The number of banks in our sample is not fixed since during our sample period there have been mergers, acquisitions and bankruptcies. The choice of 49 banks in the simulation is only indicative of the number of banks in our sample. Our data on the interbank market are confidential and we cannot be more precise. The number of investment assets, 8, is the result of the banks' balance sheet analysis. Details are discussed in the data section.

varying levels of α misspecification is of particular importance, since α is also the main parameter for the distribution of W. For completeness we also incorrectly initialize $\mu_{init} = 1$, $\sigma_{V,init} = 1$ and $\sigma_{init} = 1$, so that every hyper-parameter is misspecified.

Panel A in Table 1 presents the results of the self-consistency scenario. Each of our four non-parametric statistical tests indicate that the model and estimation is self-consistent. Note that the average p-values shown in parentheses in Table 1 are above 5% for all tests, so in this setting, we fail to reject the null hypotheses from each of the four tests. The Pseudo R^2 and Rand Index also provide evidence that the estimation overall and of W specifically are high quality, with values close to one. Thus, our estimates of W, which are particularly important within the systemic risk context, match well the true distribution when using the ex-post correct parameterization.

Panel B in Table 1 presents results for values of α ranging from 0.15 to 0.35 (straddling the actual 0.2 value) in various hyper-parameter misspecification scenarios. The Pseudo R^2 values show that the overall quality of the estimation remains high and is essentially unaffected by misspecified hyper-parameters, which is perhaps expected given that in Bayesian analysis the posterior distribution uses the observed data to recalibrate prior assumptions. Focusing on W, for all α, our tests largely show that the estimated and true distributions are statistically identical (excepting the most conservative Anderson-Darling test). We regard this as evidence that our estimation procedure is robust to mild misspecification. In fact, while the Rand Index decreases when the hyper-parameters are misspecified, the performance under misspecification is still superior to competing methods in as shown in Panel C.

Since we have numerical evidence that the estimation is statistically valid and with performance that is favorable compared to alternative techniques, we now turn our attention to whether the model can be validated in terms of actual financial accounting data.

4.2 Validation with Balance Sheet Data

To estimate *W,* we use daily interbank trading data coupled with daily stock returns of publicly-traded European banks spanning January 2006 through December 2012.[15] We obtain interbank trading data from e-MID, the only electronic market for interbank deposits in the Euro region, which offers interbank loans ranging from overnight (one day) to two years in duration, with overnight contracts representing 90% of total volume during our sample period (see Brunetti et al. [2011]). Our e-MID trading data includes a large number of banks, but we analyze only those which are publicly-traded. We integrate the e-MID data through the balance sheet, coupling assets traded overnight with the corresponding daily stock returns (as a proxy for equity changes). To preserve confidentiality we only say that the number of publicly traded banks in our data is between 50 and 60 and that these banks are all located in several European countries. Our sample includes large, medium and small size banks. Summary statistics for the publicly traded banks in our sample are shown in Table 2, where

[15] We stop in December 2012 because liquidity in the e-MID market dried up as shown in Brunetti et al. [2015].

we see that their average daily interest rate in the e-MID dropped post-Lehman (starting September 12, 2008), while daily volume started to decline much earlier when the ECB noted worldwide liquidity shortages (August 7, 2007). Daily volume for the publicly traded banks declined from over 1 billion euros prior to the start of the financial crisis to 100 million euros post-Lehman. As the crises unfolded, the banks averaged negative stock returns with increasing volatility.

We assume that banks rebalance their overall portfolios monthly, so we estimate W each month using the Bayesian framework. Recall, however, that $Z = [Z_1, \ldots Z_T]$ is constructed with daily stock returns and daily e-MID activity as proxies for equity changes and interbank activity, respectively.[16]

Using public data from annual reports, we construct the true vector of weights held across eight investment types, W, by first partitioning the balance sheets of each bank on December 31 each year into eight categories: Cash, Commercial Loans, Intangible Assets, Interbank Assets, Residential Loans, Investments, Other Holdings, and Remainder (total assets minus all other categories). Then we arrange the balance sheet into a matrix, so that each row is a bank, i.e., W is a $N \times 8$ matrix, where N is the number of banks in our sample Lastly, we normalize W, by dividing each entry by its row sum (each bank's total assets).

We compare the density of our estimated W as of December 31 of each year in our sample to the observed W constructed using real balance sheet data,[17] see Figure 1. Note that our estimates closely approximate actual values except for the tails

[16] When beginning the MCMC estimation each month, W, V, and σ^2 are initialized to their estimates from the previous month. 20,000 samples are drawn using a burn-in of 10,000 iterations.
[17] We performed a grid search to find the best α value. This was done both every month and once simultaneously for all time points. Results were qualitatively similar between both approaches.

of the distribution—in the actual data approximately 44% of values are less than 1% (i.e. an individual bank holds a very small amount of assets in a particular investment), while in our estimated W 44% of values are less than 4%. Whereas this difference is not practically meaningful (as shown in the next section), it does impact the non-parametric statistical tests as reported in Table 3. We consistently fail to reject the null hypothesis for the permutation test (same mean) and median test (same median). Interestingly, we the Mann-Whitney U test, which compares the full distributions of the estimated and true W, indicates that at the onset of the crisis, 2006-2008, the estimates of W are statistically different from the true data while during and after the crisis, 2009-2012, our procedure is able to produce accurate estimates of the distribution of W. This is confirmed by the pseudo R^2 which indicates that our factorization explains less variation in Z at the onset of the crisis.

The Anderson-Darling test, the most conservative of the four, rejects the null due to the differences highlighted above on the lower tail. Consistent with the hypothesis testing, the Pseudo R^2 and Rand Index values show that the estimation quality is generally good, with values clearly bounded away from zero.[18] More generally, these results resemble the misspecification scenario in the simulation study, where the Anderson-Darling test is strongly rejected but all other metrics indicate accurate estimates. Thus, we believe that the bulk of this evidence

[18] Tables A1 and A2 in the Appendix show results for estimation under competing methods. The competing approaches tend to achieve slightly higher Rand Index values, but massively underperform with the hypothesis testing compared to our proposed approach.

supports that our estimates of W are accurate and drawn from nearly identical distributions to the true W.[19]

5. Compiling Systemic Risk Measures

Having established the validity of our approach in estimating portfolio weights across the spectrum of the banks in our sample, we now turn our attention to two key questions: i) Are bank portfolios well diversified? ii) How similar are portfolio holdings across banks? To answer the first question we develop a concentration index which captures the degree of diversification of each bank's portfolio. To answer the second question, we develop a similarity index which captures how similar portfolio holdings are across banks. We view these two metrics as indicators of systemic risk within the banking system. First, concentrated holdings on a small number of assets within an individual bank exposes the bank to asset-specific risk. Ceteris paribus, should a bank with a portfolio concentrated only in one or two assets be forced to sell those assets, the price impact could be higher when compared to a bank liquidating a more diversified portfolio. This measure of risk is specific to the bank. Second, the similarity of asset holdings across banks suggests that shocks to any particular asset class will be borne across the entire banking system. The similarity of portfolio holdings is the theoretical justification of network analysis such as Diebold and Yilmaz [2014] and Billio et al. [2012], and is based on a simple consideration: If two banks, A and B, hold the same asset and an exogenous shock forces A to liquidate the asset, the price of the

[19] A strict interpretation of these various statistical tests seems to be an excessively high standard for validating the model. For example, the data might fail a test because we use a normal approximation as opposed to a t-distribution.

asset will decline and therefore change the value of B's portfolio potentially leading to B also selling the asset at an unfavorable price. Braverman and Minca [2014] adopt this argument to describe how common asset holdings can transmit financial distress among banks. [20] Our similarity measure is capturing the network effect in that is describing a propagation mechanism. In case of a shock to a bank, the concentration measure tells us how risky a bank's portfolio is while the similarity measure assesses the likelihood that the shock will propagate. Note that an advantage of our approach is to allow estimation of portfolio weights at a higher frequency than usually reported in bank's official filings.

After obtaining an estimate for W in a given month, we calculate a Herfindahl index [Rhoades, 1993] of diversification/concentration across our eight asset classes. Specifically, let the superscript (t) index time in months, then

$$H_i^{(t)} = \sum_k W_{ik}^{(t)2} \quad (7)$$

To measure the similarity in assets held across banks, we define the pairwise portfolio similarity between bank i and bank j as

$$Sim_{ij}^{(t)} = \sum_{k=1}^{K} min\left(W_{ik}^{(t)}, W_{jk}^{(t)}\right).$$

The similarity[21] index is bounded between 0 and 1, with zero values indicating each pair of banks hold different assets while values equal to one indicating both

[20] Others have obtained similar theoretical findings establishing that overlapping portfolios and common assets holdings can serve to amplify economic shocks, thus raising the chances of simultaneous failures [Wagner, 2010; Beale et al., 2011; Haldane and May, 2011; Raestin, 2014; Caccioli et al., 2014, 2015; Greenwood et al., 2015].

[21] In unreported results, we measure similarity using the Euclidean distance, KL divergence, correlation, and cosine similarity (as in Getmansky et al. [2017] for the insurance industry) and several other criteria. Results are consistent across these different similarity measures.

banks hold identical asset portfolios. For example, assume two banks and three assets and the following holdings:

	Bank A	Bank B	Similarity Index
Investment 1	50%	30%	30%
Investment 2	10%	60%	10%
Investment 3	40%	10%	10%

The example shows that the index has the desirable property of been bounded between zero (no common holdings) and 1 (same holdings and same and weights), and captures portfolio similarities adequately. However, it may underestimate contagion since, if Bank A is hit by a shock and liquidates most if its assets, the value of the entire B's portfolio will be affected.

Within the banking system, asset concentration and portfolio similarity are related. If an individual bank holds a concentration of assets that perform poorly *ex-post*, the poor performance might stress other institutions. Likewise, if each bank in the system is fully diversified across asset classes, then, by definition, all bank balance sheets will be similar and highly interconnected. In this regard, we posit that the interdependence between asset concentration and bank similarity reflects systemic risk.

Figure 2 displays the first four cross-sectional moments of the Herfindahl distribution of bank asset concentration/diversification over time. As the figure shows, European banks grew more concentrated (on average) from 2006 through September 2008 when Lehman Brothers failed. During the following year, however, the average concentration fell dramatically before gradually rising again to near pre-Lehman levels through 2012. The standard deviation of asset concentration has risen alongside the rise in concentration since mid-2009.

Notably, the skewnesss and kurtosis of asset concentration was falling pre-Lehman, but rose dramatically in the subsequent four months, suggesting that individual bank holdings were less concentrated and banks were investing in different portfolios. Both skewness and kurtosis of asset concentration remained high through mid-2011 before falling dramatically through 2012.

Figure 3 displays the first four moments of pairwise portfolio similarity over time. The average similarity among European banks generally rose from 2008 through mid-2010 before falling through 2012. Overall, the standard deviation of similarity increased from 2006-2008 and remained relatively high through 2012. The skewness in similarity is negative and, after becoming more negative from 2008-2010, has reverted toward zero through 2012. The kurtosis of similarity appears to follow the opposite pattern from skewness.

While the patterns in moments of asset concentration and portfolio similarity suggest that these metrics may reflect real economic conditions, our aim is to test whether these metrics are useful for forecasting purposes.

5.1 Comparison with ECB risk and macro indicators

Recall that our metrics are constructed using daily interbank trades and daily equity price change, so that we impound both expected future performance as well as current liquidity demands for each bank. As a benchmark test for the usefulness of these metrics in forecasting, we relate moments of concentration and similarity to three measures of systemic risk published by the ECB—the

Composite Systemic Risk Index, the Simultaneous Default Probability, and the Sovereign Composite Systemic Risk Index. The composite systemic risk indicator is a weighted average of several measures of financial stress that focus on different aspects of the financial system including money, bond, equity and forex markets as well as financial intermediaries. The Simultaneous Default Probability, and the Sovereign Composite Systemic Risk Index are essentially CDS implied probabilities.[22]

Figure 4 displays these three ECB metrics from 2006 through 2012.[23] Note that the ECB metrics all rise from early 2007 through early 2009 before abating slightly until early 2010. All three metrics then rise through late 2011 before generally falling off through 2012. These metrics have been extensively used—e.g. Hollo et al [2012].

Table 4 reports the contemporaneous correlation between the moments of our concentration and similarity indices and the three measures of systemic risk published by the ECB—these measures are in first difference to guarantee stationarity.[24]

Interestingly, the higher is the concentration in the banking sector (mean), the higher is the systemic risk indicator, which captures stress conditions in European markets. Similarly, high level of variation (standard deviation) in concentration and similarities across bank portfolios, are associated to high levels of stress

---

[22] ECB Financial Stability Review, June 2012, available at: https://www.ecb.europa.eu/pub/pdf/other/financialstabilityreview201206en.pdf?6b3b7eb08f53f6ad069f5b6dd15275c8

[23] The Simultaneous Default Probability metric is only available since 2007.

[24] We perform two stationarity tests, the generalized least squares Dickey–Fuller (DF) test proposed by Elliott, Rothenberg, and Stock (1996) and the Augmented Dickey-Fuller (ADF) test.

conditions. The systemic risk indicator is also linked to higher moments of the similarity index. The Probability of Simultaneous Default is positively correlated to the similarity index indicating that when the similarity index is high, the probability of default is increasing (recall that the ECB systemic indicators are in first difference and hence they indicate a change). Overall, we find strong contemporaneous linkages between the ECB systemic risk measures and our indices.

The second part of Table 4 reports contemporaneous correlations between major EU macroeconomic indicators—the Consumer Confidence Index (CCI), Industrial Production (IP), the Purchasing Managers' Index (PMI) and Retail Sales—and the moments of the concentration and the similarity indices (also in this case, all variables have been differenced to accommodate for non-stationarity). Table 4 indicates strong contemporaneous linkages between macro variables and our indices. Recall that our indices are computed using data on the interbank market, e-MID, which is only in part capturing interbank activities, and stock market returns of a relatively small number of publicly traded banks. Nevertheless, the correlation between our indices and macro variables is statistically significant.

We further investigate the lead-lag relations among the moments of our indices and the ECB systemic risk and macro indicators by estimating bivariate VARs. and testing for Granger-non-causality.[25] Figure 5 presents the results of our analysis. $A \rightarrow B$ indicates that A Granger-causes B at the 5 percent significance level, while $A \leftrightarrow B$ indicates feedback effect, i.e. A Granger-causes B and B Granger-causes A at the 5 percent significance level. Figure 5 shows that, with

[25] VARs optimal lag specification is based on AIC.

only a few exceptions of feedback effects, our indices are able to cause, in a forecasting sense, most of the ECB systemic risk and macro indicators. Interestingly, the first three moments of the concentration index seem to contain valuable information for forecasting.

5.2 Comparison with SRISK and MES

Acharya et al [2017] developed a model of systemic risk in which the level of capitalization of the financial sector has implication on the real economy. In the model, the contribution of each financial institution to systemic risk is measured by the systemic expected shortfall which is a function of how much the institution is undercapitalized conditional on the entire financial system being undercapitalized. The systemic expected shortfall depends on the institution's leverage and on the institution's marginal expected shortfall (MES). Acharya et al [2017] show that MES is able to predict systemic risk in the recent financial crisis.

Brownlees and Engle [2017] introduce a conditional capital shortfall measure of systemic risk, named SRISK. This measure captures the contribution of a financial institution to systemic risk and, similar to Acharya et al [2017], is based on the capital shortfall of the institution conditional on a severe market downturn. SRISK is able to capture the riskiness of US financial institutions leading to the 2007-2009 crisis. Aggregating SRISK across institutions, the authors also propose an early warning index of distress.

Both MES and SRISK combine balance sheet information and asset price information for publicly traded financial institutions. We took both measures and compared them to our indices.[26] The measures were provided to us on a company-by-company basis. We selected all European companies in the countries where the e-MID banks are based.[27] In total we construct MES and SRISK measure using 313 financial institutions including banks, insurance companies, brokers/dealers. To aggregate MES and SRISK measures across institutions, we normalize these measures by market capitalization.

It is important to note that our concentration and similarities indices are computed by looking at 50-60 banks while MES and SRISK have been computed by looking at a much larger set of institutions.

Figure 6 depicts MES and SRISK together with the concentration index and the similarity index. MES and SRISK show similar behaviors after the crisis with the concentration index. In fact, the correlation coefficient between the concentration index and MES and SRISK is 43% and 39%, respectively. There seems to be no much co-movements between the similarity index and MES and SRISK.

We further investigate the lead-lag relationship between MES and SRISK, on one hand, and the concentration and the similarity indices, on the other, through the lens of bivariate VARs and Granger-causality tests. We find that the concentration index causes, in a forecasting sense, both the MES (p-value = 0.073) and SRISK (p-value = 0.76). The reverse is not true—i.e. SRISK and MES do not Granger-cause the concentration index. As evidenced in Figure 6, we do not find any linkage in terms of Granger-causality between the similarity index and MES and SRISK.

---

[26] We are thankful to Rob Capellini director of the V-Lab at NYU for sharing the data.

[27] For confidentiality reasons we cannot match the two datasets exactly — i.e., we cannot select the companies in V-Lab with the banks in our e-MID dataset.

6 Discussion and Conclusion

Panel A: Self-Consistency Scenario						
$\hat{\alpha}$	Pseudo R^2	RI	Permutation Test	Median Test	MW Test	AD Test
0.20	0.936	0.824	0.000 (1.000)	-1.147 (0.369)	0.323 (0.682)	2.249 (0.091)
Panel B: Hyper-Parameter Misspecification Scenario						
$\hat{\alpha}$	Pseudo R^2	RI	Permutation Test	Median Test	MW Test	AD Test
0.10	0.929	0.757	0.000 (1.000)	0.247 (0.610)	0.862 (0.425)	4.698 (0.009)
0.15	0.938	0.762	0.000 (1.000)	0.428 (0.587)	1.049 (0.352)	4.362 (0.041)
0.20	0.938	0.776	0.000 (1.000)	0.566 (0.595)	1.272 (0.256)	3.332 (0.045)
0.25	0.936	0.775	0.000 (1.000)	0.761 (0.518)	1.576 (0.148)	6.939 (0.005)
0.30	0.941	0.782	0.000 (1.000)	0.609 (0.521)	1.711 (0.119)	9.212 (0.001)
Panel C: Competing Methods						
Method	Pseudo R^2	RI	Permutation Test	Median Test	MW Test	AD Test
Semi-NMF	0.800	0.740	0.000 (1.000)	0.623 (0.587)	-0.135 (0.779)	4.248 (0.015)
Fuzzy K-Means	NA	0.633	0.000 (1.000)	4.112 (0.004)	3.078 (0.010)	106.406 (0.000)

Table 1: Simulation results averaged over 100 iterations. Pseudo R^2 is defined analogously to the linear regression setting; RI is the Rand Index of W (values closer to 1 indicate more accurate estimates). The permutation test refers to the Fisher-Pitman test [Boik, 1987] while the median test refers to the Brown-Mood test [Brown et al., 1951] and assess whether two samples have identical means and medians, respectively. MW test refers to the Mann-Whitney U test [Mann and Whitney, 1947] which compares the full distributions of the estimated and true W. The AD test refers to the Two Sample Anderson Darling Test [Scholz and Stephens, 1987] to assess whether there are differences between the two samples. The statistical tests compare the estimated and true distribution of W; average test statistics are reported with p-value in parentheses. Failing to reject the null hypothesis provides evidence in support of the estimation for all statistical tests. Note that Pseudo R^2 is not reported for the Fuzzy K-Means algorithm, because it only estimates W and an estimate of both W and V is required.

	Pre-Crisis Jan 1 2006 – Aug 7 2007				Crisis 1 Aug 8 2007 – Sept 12 2008			
	Mean	St. Dev.	Skew	Kurt.	Mean	St. Dev.	Skew	Kurt.
Stock Returns	0.001	0.141	1.070	507.643	-0.001	0.155	0.904	506.349
Volume (e-MID)	1637.923	1204.923	1.039	0.541	553.403	562.621	2.054	5.243
Rate (e-MID)	3.185	0.566	-0.039	-1.325	4.026	0.191	-0.597	1.719
	Crisis 2 Sept 16 – Apr 1 2009				**Crisis 3** Apr 2 2009 – Dec 31 2012			
	Mean	St. Dev.	Skew	Kurt.	Mean	St. Dev.	Skew	Kurt.
Stock Returns	-0.005	0.194	1.932	431.219	-0.002	0.200	0.211	560.613
Volume (e-MID)	158.231	103.638	1.152	0.658	158.333	158.640	1.232	-0.299
Rate (e-MID)	2.837	1.313	-0.190	-1.488	0.783	0.286	0.219	-1.913

Table 2: Summary statistics at the daily level for log stock returns, e-MID trading volume (millions of Euros), and e-MID interest rate. All e-MID statistics are computed using transactions that include at least one of the 54 banks as a counter-party in the overnight loan.

Year	Pseudo R^2	RI	Permutation Test	Median Test	MW Test	AD Test
2006	0.698	0.698	0.000 (1.000)	1.458 (0.174)	3.316 (0.001)	22.771 (0.000)
2007	0.882	0.598	0.000 (1.000)	0.583 (0.612)	3.326 (0.001)	47.608 (0.000)
2008	0.864	0.594	0.000 (1.000)	-0.729 (0.515)	2.249 (0.025)	48.401 (0.000)
2009	0.910	0.595	0.000 (1.000)	-1.442 (0.170)	1.202 (0.229)	38.931 (0.000)
2010	0.930	0.575	0.000 (1.000)	-1.010 (0.344)	0.979 (0.330)	39.619 (0.000)
2011	0.916	0.584	0.000 (1.000)	0.433 (0.719)	1.163 (0.245)	39.511 (0.000)
2012	0.940	0.521	0.000 (1.000)	0.433 (0.720)	0.922 (0.359)	35.221 (0.000)

Table 3: Validation results for estimation using the proposed method compared to actual European bank balance sheet data disclosed in annual reports. Pseudo R^2 is defined analogously to the linear regression setting; RI is the Rand Index of W (values closer to 1 indicate more accurate estimates). The permutation test refers to the Fisher-Pitman test [Boik, 1987] while the median test refers to the Brown-Mood test [Brown et al., 1951] and assess whether two samples have identical means and medians, respectively. MW test refers to the Mann-Whitney U test [Mann and Whitney, 1947] which compares the full distributions of the estimated and true W. The AD test refers to the Two Sample Anderson Darling Test [Scholz and Stephens, 1987] to assess whether there are differences between the two samples. The statistical tests compare the estimated and true distribution of W; average test statistics are reported with p-value in parentheses. Failing to reject the null hypothesis provides evidence in support of the estimation for all statistical tests.

	Herfindhal (Concentration index)				Similarity Index			
	Mean	St. Dev.	Skew	Kurt.	Mean	St. Dev.	Skew	Kurt.
Systemic Risk Ind.	0.206*	0.351*	-0.014	-0.082	0.146	0.501*	-0.338*	0.209*
Sim. Def. Prob.	-0.146	-0.232*	0.185*	0.193*	0.175*	0.077	-0.130	0.102
Sov. Syst. Risk Ind.	-0.141	-0.389*	0.049	0.092	-0.021	-0.133	0.151	-0.150
CCI	-0.525*	-0.382*	0.465*	0.484*	0.209*	-0.182*	-0.034	0.129
IP.	-0.057	-0.384*	0.004	0.122	-0.198*	-0.224*	0.347*	-0.246*
PMI	-0.598*	-0.228*	0.513*	0.481*	0.319*	-0.145	-0.178*	0.254*
Retail Sales	-0.109	0.037	0.134	0.141	0.035	0.047	-0.141	0.124

Table 4: Systemic Risk Indicator is in first difference (levels are non-stationary). Sim. Def. Prob. refers to the Probability of Simultaneous Default and is in first difference (levels are non-stationary). Sov. Syst. Risk Ind. refers to the Sovereign Systemic Risk Index and is in first difference (levels are non-stationary). CCI refers to the Consumer Confidence Index and is in first difference (levels are non-stationary). IP refers to Industrial Production and is in first difference (levels are non-stationary). PMI refers to the Purchasing Managers' Index and is in first difference (levels are non-stationary). Retail Sales is differenced twice to achieve stationarity.

* Indicates significance at 5 percent level.

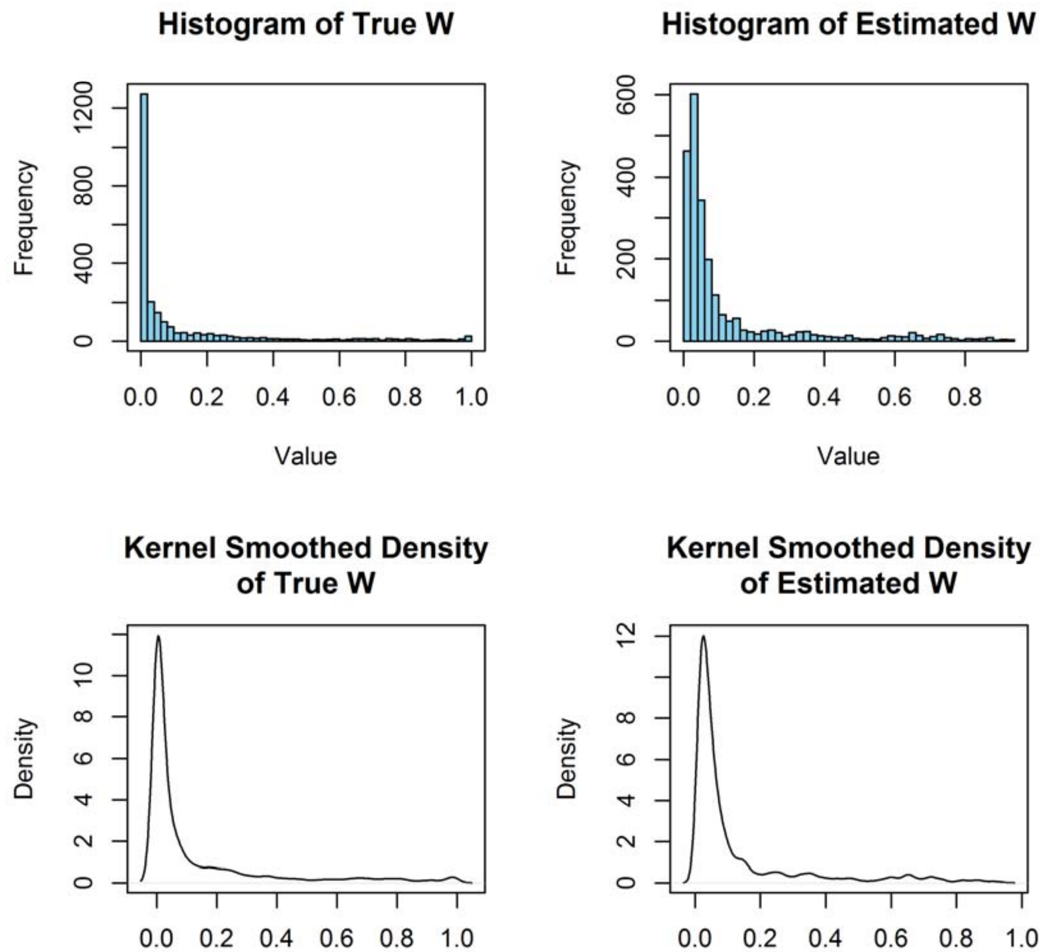

Figure 1: Distribution of the observed elements in W aggregated from all available years compared to the estimated W aggregated over the same times.

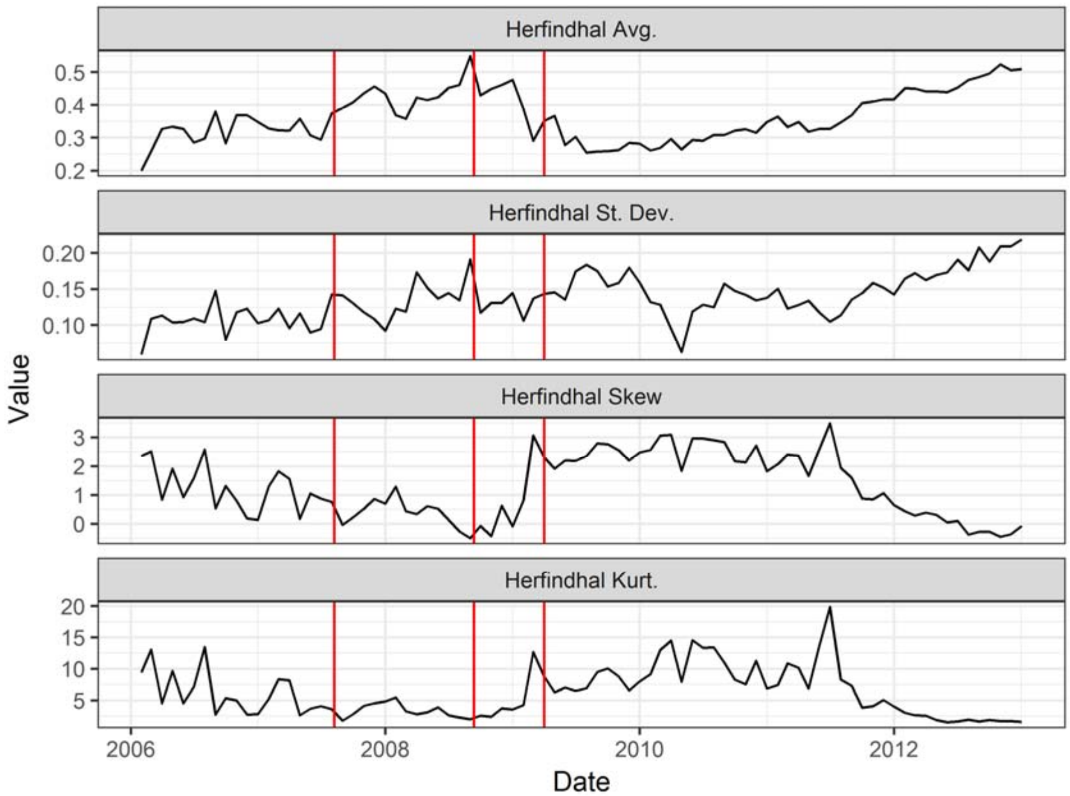

Figure 2: Herfindhal (Concentration index) summary statistics over time. The vertical lines denote three events: 1) August 7, 2007 when the ECB noted worldwide liquidity shortages; 2) September 12, 2008 (Lehman default); 3) April 1, 2009 when the ECB announced the end of the recession.

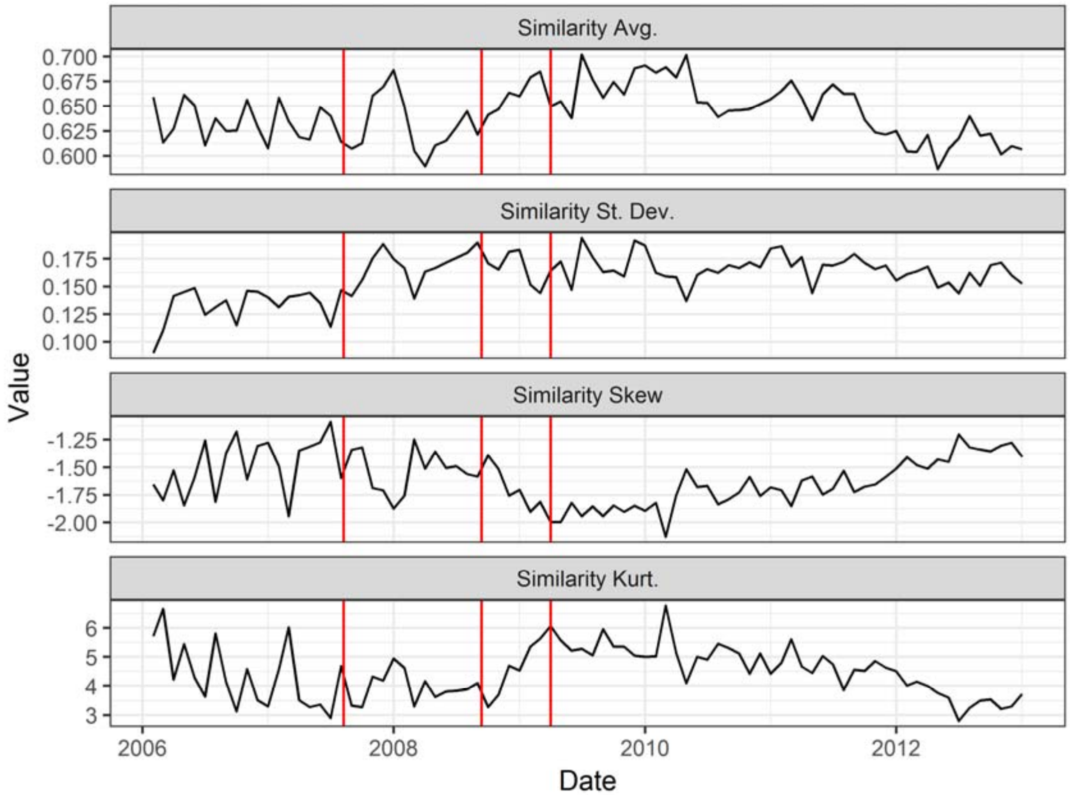

Figure 3: Similarity index summary statistics over time. The vertical lines denote three events: 1) August 7, 2007 when the ECB noted worldwide liquidity shortages; 2) September 12, 2008 (Lehman default); 3) April 1, 2009 when the ECB announced the end of the recession.

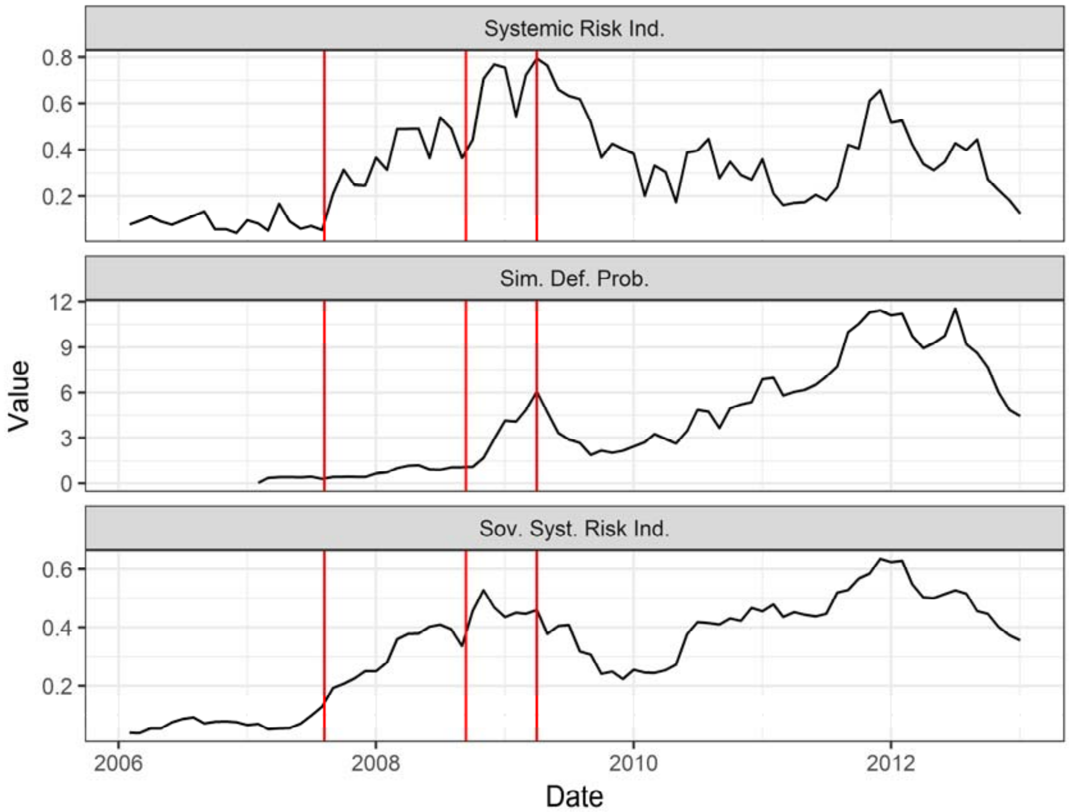

Figure 4: The top panel shows the raw time series for systemic risk measures published by the ECB. The vertical lines denote three events: 1) August 7, 2007 when the ECB noted worldwide liquidity shortages; 2) September 12, 2008 (Lehman default); 3) April 1, 2009 when the ECB announced the end of the recession.

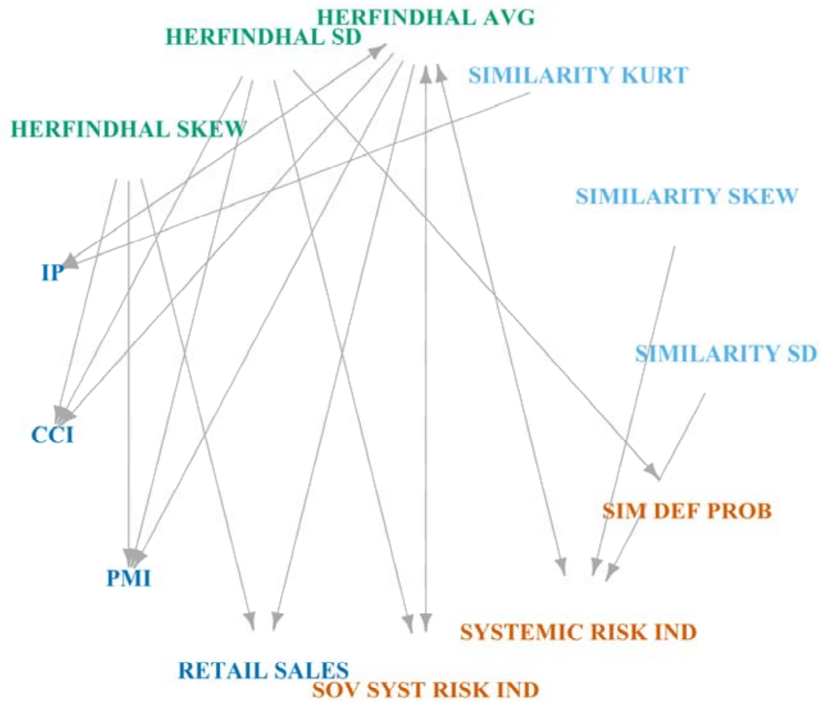

Figure 5: Granger Causality relationships at the 5% significance level amongst the derived variables, systemic risk measures, and macro-economic variables.

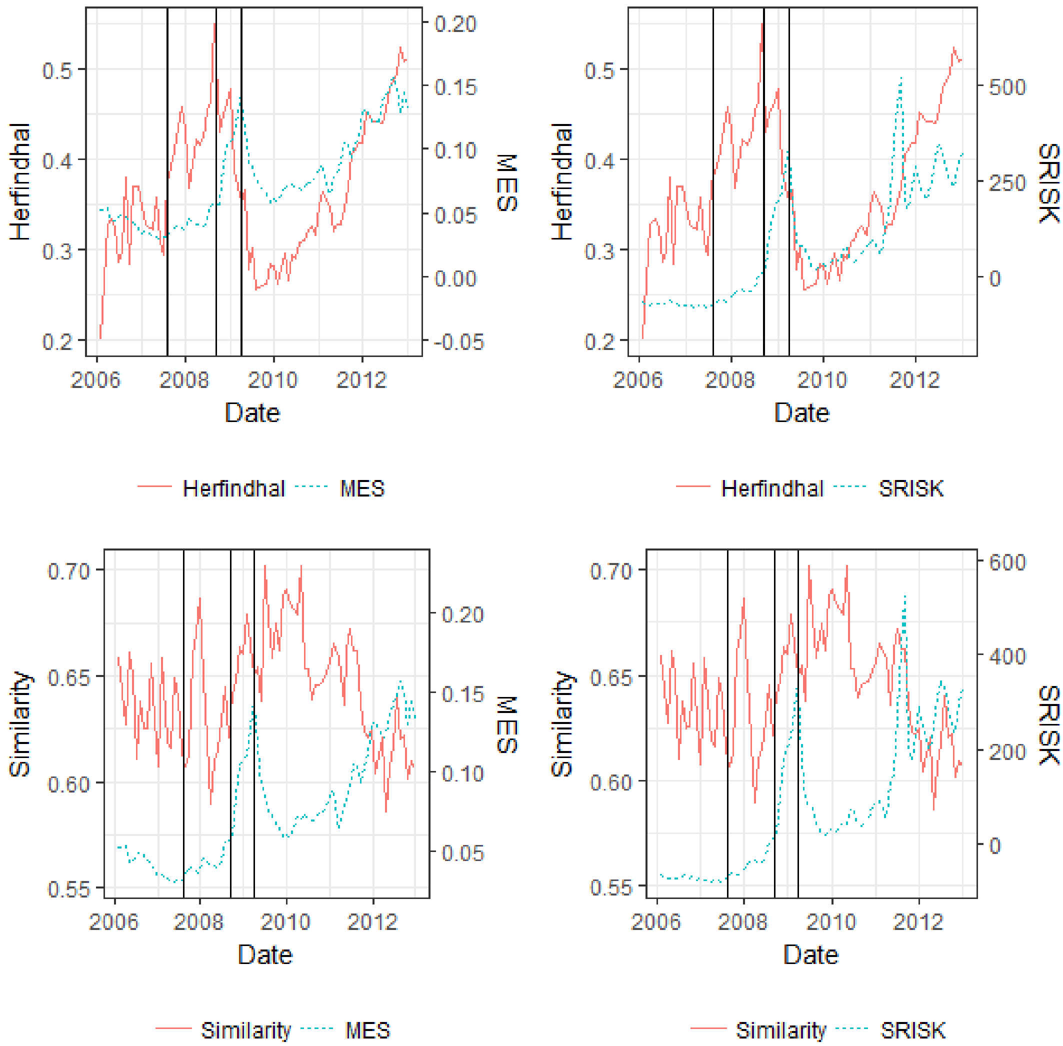

Figure 6: MES and SRISK indices and Concentration and Similarity indices. The vertical lines denote three events: 1) August 7, 2007 when the ECB noted worldwide liquidity shortages; 2) September 12, 2008 (Lehman default); 3) April 1, 2009 when the ECB announced the end of the recession.

References

V.V. Acharya, L.H. Pedersen, T. Philippon and M. Richardson. Measuring Systemic Risk. Review of Financial Studies, 30(1), 2–47, 2017.

T. Adrian and M.K. Brunnermeier. CoVar. American Economic Review, 106(7), 1705-1741, 2016.

T. W. Anderson and D. A. Darling. A test of goodness of fit. Journal of the American statistical association, 49(268):765–769, 1954.

C. E. Antoniak. Mixtures of Dirichlet processes with applications to bayesian nonparametric problems. The annals of statistics, pages 1152–1174, 1974.

Bezdek, J. C., R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. Computers & Geosciences 10(2-3), 191-203, 1984.

D. Biasis, M.D. Flood, A.W. Lo and S. Valavanis. A Survey of Systemic Risk Analytics. Annual Review of Financial Economics, 4, 255-296, 2012.

M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics, 104(3):535–559, 2012.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. Journal of Machine Learning Research, 3(Jan):993–1022, 2003.

R. J. Boik. The Fisher-Pitman permutation test: A non-robust alternative to the normal theory f test when variances are heterogeneous. British Journal of Mathematical and Statistical Psychology, 40(1):26–42, 1987.

A. Brav. Inference in long-horizon event studies: A bayesian approach with application to initial public offerings. The Journal of Finance, 55(5):1979–2016, 2000.

G. W. Brown, and A. M. Mood. On median tests for linear hypotheses.In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, volume 2, pages 159–166. University of California Press Englewood Cliffs, NJ, 1951.

C. Brownlees and R.F. Engle. SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. Review of Financial Studies, 30(1), 48–79, 2017.

C. Brunetti, M. Di Filippo, and J. H. Harris. Effects of central bank intervention on the interbank market during the subprime crisis. Review of Financial Studies, 24(6):2053–2083, 2011.

C. Brunetti, J. H. Harris, S. Mankad, and G. Michailidis. Interconnectedness in the Interbank Market. Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, Oct. 2015. URL http://ssrn.com/ abstract=2674602.

F. Caccioli, M. Shrestha, C. Moore, and J. D. Farmer. Stability analysis of financial contagion due to overlapping portfolios. Journal of Banking & Finance, 46:233– 245, 2014.

F. Caccioli, J. D. Farmer, N. Foti, and D. Rockmore. Overlapping portfolios, contagion, and financial stability. Journal of Economic Dynamics and Control, 51:50–63, 2015.

G. Casella, and E.I. George, 1992. Explaining the Gibbs sampler. The American Statistician, 46(3), pp.167-174.

S. Chib and E. Greenberg, 1995. Understanding the Metropolis-Hastings Algorithm. The American Statistician, 49(4), pp.327-335.

K. M. Cremers (2002). Stock return predictability: A Bayesian model selection perspective. The Review of Financial Studies, 15(4), 1223-1249.

O. de Jonghe. Back to the Basics in Banking? A Micro-analysis of Banking System Stability. Journal of Financial Intermediation, 19(3), 387-417, 2010.

F. X. Diebold and K. Yılmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics, 182(1): 119–134, 2014.

C. Ding, T. Li, and M. Jordan (2010). Convex and semi-nonnegative matrix factorizations. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (1), 45{55.

M. Elliott, B. Golub, and M. O. Jackson. Financial networks and contagion. American Economic Review, 104(10):3115–53, 2014.

B. Eraker. Mcmc analysis of diffusion models with application to finance. Journal of Business & Economic Statistics, 19(2):177–191, 2001.

D. Gale and P. Gottardi. Equilibrium Theory of Banks' Capital Structure. CESIFO Working Papers, 6580, 2017.

M. Getmansky, G. Girardi, K.W. Hanley, S. Nikolova, and L. Pelizzon, Portfolio Similarity and Asset Liquidation in the Insurance Industry (April 15, 2017). Available at SSRN: https://ssrn.com/abstract=3050561

J. Geweke and G. Zhou. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies, 9(2):557–587, 1996.

P. Giudici, P. Sarlin and A. Spelta. The Multivariate Nature of Systemic Risk: Direct and Common Exposures. Journal of Banking and Finance, Forthcoming.

R. Greenwood, A. Landier, and D. Thesmar. Vulnerable banks. Journal of Financial Economics, 115(3):471–485, 2015.

D. C. Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE transactions on geoscience and remote sensing, 39(3):529–545, 2001.

D. Hollo, M. Kremer, and M. Lo Duca, 2012. CISS-a composite indicator of systemic stress in the financial system.

A. Huck, M. Guillaume, and J. Blanc-Talon. Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 48(6):2590–2602, 2010.

X. Huang, H. Zhou and H. Zhu. A Framework for Assessing the Systemic Risk of Major Financial Institutions. Journal of Banking and Finance, 33, 2036–2049, 2009.

C. S. Jones. Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies, 16(3):793–843, 2003.

R. W. Klein and V. S. Bawa. The effect of limited information and estimation risk on optimal portfolio diversification. Journal of Financial Economics, 5(1): 89–111, 1977.

A. Korteweg and M. Sorensen. Risk and return characteristics of venture capitalbacked entrepreneurial companies. Review of Financial Studies, 23(10):3738– 3772, 2010.

G. Lamoureux and G. Zhou. Temporary components of stock returns: What dothe data tell us? Review of Financial Studies, 9(4):1033–1059, 1996.

D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrixfactorization. Nature, 401:788–791, 10 1999.

C.-b. Lin. Projected gradient methods for nonnegative matrix factorization. Neural computation, 19(10):2756–2779, 2007.

S. Mankad, S. Hu, and A. Gopal. Single Stage Prediction with Online Reviews for Mobile App Development and Management. ArXiv e-prints, July 2016.

B. Mann and D. R. Whitney. On a test of whether one of two random variablesis stochastically larger than the other. The annals of mathematical statistics, pages 50–60, 1947.

A. Norets, and J. Pelenis. Bayesian modeling of joint and conditional distributions. Journal of Econometrics, 168(2), 332-346, 2012.

L. Pastor, R.F. Stambaugh and L.A. Taylor. Portfolio Liquidity and Diversification: Theory and Evidence. NBER Working Paper No. w23670, 2017.

I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon. Overlapping community detection using Bayesian non-negative matrix factorization. Phys. Rev. E, 83:066114, Jun 2011.

N. M. Razali, Y. B. Wah. Power comparisons of shapiro-wilk, kolmogorovsmirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analytics, 2(1):21–33, 2011.

S. A. Rhoades. Herfindahl-hirschman index, the. Fed. Res. Bull., 79:188, 1993.

G. Santis and B. Gerard. International asset pricing and portfolio diversification with time-varying risk. The Journal of Finance, 52(5):1881–1912, 1997.

M. N. Schmidt, O. Winther, and L. K. Hansen. Bayesian non-negative matrix factorization. In International Conference on Independent Component Analysis and Signal Separation, pages 540–547. Springer, 2009.

F. W. Scholz and M. A. Stephens. K-sample anderson–darling tests. Journal of the American Statistical Association, 82(399):918–924, 1987.

M. Segoviano and C. Goodhart. Banking Stability Measures. IMF Working Paper 09/04, 2009.

J. Sethuraman. A constructive definition of dirichlet priors. Statistica sinica, pages 639–650, 1994.

H. S. Shin. Securitisation and financial stability*. The Economic Journal, 119(536): 309–332, 2009.

H. S. Shin. Financial intermediation and the post-crisis financial system. 2010.

N. Tarashev, C. Borio and K. Tsatsaronis. Attributing Systemic Risk to Individual Institutions: Methodology and Policy Applications. BIS Working Papers, 308, 2010.

D. E. Upton and D. S. Shannon. The stable Paretian distribution, subordinated stochastic processes, and asymptotic lognormality: an empirical investigation. The Journal of Finance 34 (4), 1031-1039 1979.

Appendix 1:Derivation of MCMC Sampling Algorithm

Detailed derivations are given below, followed by a summary of the main steps of the estimation. We will denote the rows of a matrix X as x_i or $x_{i.}$ and columns as $x_{.j}$. Also $X_{/x_i}$ denotes the matrix X excluding the i-th row.

Posterior of W

Since $p(W) = \prod_{i=1}^{n} p(w_i)$ (rows are i.i.d.) and w_i only affects z_i, it is easy to see that the posterior of W is a product of Gaussian likelihood and a Dirichlet prior

$$p(w_i|Z, W_{/w_i}, V, \sigma^2) \propto p(w_i)p\left(z_i \middle| W_{/w_i}, V, \sigma^2\right). \qquad \text{(A1)}$$

These are not conjugate distributions, which means that we can only compute the posterior distribution's value without characterizing the distribution analytically in closed form.

As such, we use the Metropolis Hastings algorithms with a uniform proposal distribution, so that a candidate row $\breve{w}_i$ is generated by moving on the probability simplex randomly around the current state of w_i. Then the candidate row is accepted with probability $\min(1, \frac{p(\breve{w}_i|Z,W_{/\{w_i\}},V,\sigma^2)}{p(w_i|Z,W_{/\{w_i\}},V,\sigma^2)})$.

Posterior of V

We start by decomposing the posterior probability

$$p\left(v_{jk}|z, w, V_{/v_{jk}}, \sigma^2\right) \propto p(V)p(Z|W, V, \sigma^2) \qquad \text{(A2)}$$

$$\propto p\left(v_{jk}\right)p(Z|W, V, \sigma^2). \qquad \text{(A3)}$$

Recall that v_{jk} is i.i.d $N(\mu, \sigma_V^2)$. Therefore, the posterior of V is a product of a Gaussian prior and Gaussian distribution. By conjugacy, we have the posterior of v_{jk} to be

$$p\left(v_{jk}\middle|Z, W, V_{/v_{jk}}, \sigma^2\right) = N(\mu_p, \sigma_p^2), \quad \text{(A4)}$$

where

$$\sigma_p^2 = \left(\frac{\|w_{.j}\|_2^2}{\sigma^2} + \frac{1}{\sigma_v^2}\right)^{-1},$$

$$\mu_p = \sigma_p^2\left(\frac{\mu_{lik}\|w_{.j}\|_2^2}{\sigma^2} + \frac{\mu}{\sigma_V^2}\right),$$

$$\mu_{lik} = \frac{z_{.k}^T w_{.j} - (WV)_{.k}^T w_{.j} + \|w_{.j}\|_2^2 v_{jk}}{\|w_{.j}\|_2^2}.$$

Therefore we can sample directly in the Gibbs sampler from the posterior conditional distribution.

Posterior of σ^2

We follow standard arguments to exploit conjugacy properties of the inverse gamma and normal distributions.

$$p(\sigma^2|W,V,Z) \propto p(\sigma^2)p(W,V,Z|\sigma^2) \quad \text{(A5)}$$

$$\propto p(\sigma^2)p(Z|W,V,\sigma^2)p(W,V|Z,\sigma^2)$$

$$\propto p(\sigma^2)p(Z|W,V,\sigma^2)P(W,V)$$

$$\propto p(\sigma^2)p(Z|W,V,\sigma^2)$$

$$\propto IG(\eta \quad, \theta \quad)N(Z|W,V,\sigma^2).$$

Then by conjugacy, the posterior is

$$p(\sigma^2|W,V,Z) = IG(\eta', \theta') \quad \text{(A6)}$$

where

$$\eta' = \eta + \frac{NT}{2} + 1$$

$$\theta' = \frac{1}{2}\sum_{i,j}(Z - WV)_{ij}^2 + \theta.$$

Therefore we can sample directly in the Gibbs sampler from the posterior conditional distribution $IG(\eta', \theta')$.

Estimation Algorithm Summary

Let superscript (t) denotes the iteration number. Then using the definitions above, the following steps can be used to produce point estimate of $W, V,$ and σ^2.

1. Define the Dirichlet concentration parameter α and mean and variance of V (μ, σ_V). Randomly initialize $W^{(t)}, V^{(t)}, \sigma^{(t)}$.
2. For all i
   a) Using a uniform proposal distribution, form a candidate $\breve{w}_i$, i.e., $\breve{w}_i \sim U(w_{ij}^{(t)} - 0.01, w_{ij}^{(t)} + 0.01)$ with $\breve{w}_{iK} = 1 - \sum_{j=1}^{K-1} \breve{w}_{ij}$.
   b) Accept the candidate $w_i^{(t+1)} = \breve{w}_i$ with probability min(1, $\frac{p(\breve{w}_i | Z, W_{/w_i}^{(t)}, V^{(t)}, \sigma^{(t)})}{p(w_i^{(t)} | Z, W_{/w_i}^{(t)}, V^{(t)}, \sigma^{(t)})}$). Otherwise $w_i^{(t+1)} = w_i^{(t)}$.
3. For all j, k
   a) Sample $v_{jk}^{(t+1)} \sim N(\mu_p, \sigma_p^2)$.
4. Sample $\sigma^{(t+1)} \sim IG(\eta', \theta')$.
5. Repeat steps 2 through 4 until convergence.
6. Generate samples $t = T, T+1, \dots, T+N$ using steps 2 through 4.
7. Calculate point estimates $\widehat{W} = \frac{1}{N} \sum_{t=T}^{T+N} W^{(t)}, \widehat{V} = \frac{1}{N} \sum_{t=T}^{T+N} V^{(t)}, \widehat{\sigma} = \frac{1}{N} \sum_{t=T}^{T+N} \sigma^{(t)}$.

Correlation of Asset Returns in the Posterior of V

To show that two variables are conditionally independent, by definition we should show that

$$p(X, Y | Z) \propto u_1(X | Z) u_2(Y | Z),$$

i.e., we want to show that the posterior distribution (conditioning on data Z) can be factorized into a product of two appropriate functions.

With our model, the condition above with respect to V is

$$p\left(v_{jk}, v_{ik} | Z, W, V_{/v_{ik},v_{jk}}, \sigma^2\right) \propto u_1(v_{jk}) u_2(v_{ik}),$$

where v_{jk} represents the change in returns for asset class j on day k.

We will show this condition cannot be satisfied, i.e., that v_{jk} and v_{ik} are *dependent*. We start by decomposing the posterior probability

$$p\left(v_{jk}, v_{ik} | Z, W, V_{/v_{ik},v_{jk}}, \sigma^2\right) \propto p(v_{ik}) p(v_{jk}) p(Z|W, V, \sigma^2),$$

which is obtained through standard application of Bayes rule. Then it's easy to see that the independence condition above is satisfied only when $p(Z|W, V, \sigma^2)$ can itself be factorized into a product of two appropriate functions, like u_1 and u_2 above.

By Equation (6) of our paper,

$$p(Z|W, V, \sigma^2) = \prod_{ik} \mathbb{N}((WV)_{ik}, \sigma^2).$$

Then expanding the matrix product and plugging this into the Normal likelihood,

$$(WV)_{ik} = \sum_c w_{ic} v_{ck}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(z_{ik} - \sum_c w_{ic} v_{ck})^2}{-2\sigma^2}\right).$$

Without loss of generality, assume 2 asset classes so that $\sum_c w_{ic} v_{ck} = w_{i1} v_{1k} + w_{i2} v_{2k}$.

Then note that

$$\exp\left(\left(z_{ik} - \sum_c w_{ic} v_{ck}\right)^2\right) = \exp((z_{ik} - w_{i1} v_{1k} - w_{i2} v_{2k})^2)$$

$$= \exp\left(z_{ik}^2 + w_{i1}^2 v_{1k}^2 + w_{i2}^2 v_{2k}^2 - 2z_{ik} w_{i1} v_{1k} - 2z_{ik} w_{i2} v_{2k} - 2w_{i1} v_{1k} w_{i2} v_{2k}\right)$$

Since it is impossible to write $\exp(2w_{i1} v_{1k} w_{i2} v_{2k})$ as a product of two functions with arguments v_{1k} and v_{2k} respectively, the overall posterior likelihood for v_{1k} and v_{2k} also cannot be decomposed as such.

Thus, we have established that in general the posterior estimates for v_{ik} and v_{jk} will be correlated, i.e., the estimated returns for different asset classes contained in V are dependent.

Appendix 2: Validation with Balance Sheet Data: Competing Methods

Year	Pseudo R^2	RI	Permutation Test	Median Test	MW Test	AD Test
2006	0.850	0.696	0.000 (1.000)	6.663 (0.000)	3.483 (0.000)	14.986 (0.000)
2007	0.920	0.660	0.000 (1.000)	10.204 (0.000)	6.642 (0.000)	49.034 (0.000)
2008	0.907	0.655	0.000 (1.000)	10.204 (0.000)	7.013 (0.000)	51.946 (0.000)
2009	0.868	0.613	0.000 (1.000)	10.530 (0.000)	6.605 (0.000)	47.491 (0.000)
2010	0.903	0.654	0.000 (1.000)	10.241 (0.000)	6.362 (0.000)	45.888 (0.000)
2011	0.862	0.693	0.000 (1.000)	10.530 (0.000)	6.871 (0.000)	50.088 (0.000)
2012	0.424	0.650	0.000 (1.000)	10.386 (0.000)	6.637 (0.000)	49.440 (0.000)

Table A1: Validation results for estimation using the Semi-NMF model of Ding et al. [2010] with probability constraints enforced ex-post compared to actual European bank balance sheet data disclosed in annual reports. Pseudo R^2 is defined analogously to the linear regression setting; RI is the Rand Index of W (values closer to 1 indicate more accurate estimates). The statistical tests compare the estimated and true distribution of W; test statistics are reported with the p-value in parentheses. Failing to reject the null hypothesis provides evidence in support of the estimation for all statistical tests.

Year	Pseudo R^2	RI	Permutation Test	Median Test	MW Test	AD Test
2006	NA	0.652	0.000 (1.000)	3.540 (0.001)	3.704 (0.000)	22.176 (0.000)
2007	NA	0.601	0.000 (1.000)	3.498 (0.001)	4.655 (0.000)	37.496 (0.000)
2008	NA	0.518	0.000 (1.000)	6.851 (0.000)	6.230 (0.000)	59.457 (0.000)
2009	NA	0.538	0.000 (1.000)	1.731 (0.096)	3.021 (0.002)	30.826 (0.000)
2010	NA	0.632	0.000 (1.000)	0.577 (0.612)	2.189 (0.027)	25.552 (0.000)
2011	NA	0.665	0.000 (1.000)	4.183 (0.000)	4.394 (0.000)	32.243 (0.000)
2012	NA	0.572	0.000 (1.000)	2.019 (0.052)	3.269 (0.001)	28.306 (0.000)

Table A2: Validation results for estimation using Fuzzy K-means [Bezdek et al., 1984] compared to actual European bank balance sheet data disclosed in annual reports. Pseudo R^2 is defined analogously to the linear regression setting; RI is the Rand Index of W (values closer to 1 indicate more accurate estimates). The statistical tests compare the estimated and true distribution of W; test statistics are reported with the p-value in parentheses. Failing to reject the null hypothesis provides evidence in support of the estimation for all statistical tests.

Note that Pseudo R^2 is not reported for the Fuzzy K-Means algorithm, because it only estimates W, whereas other methods estimate both W and V.