Borradores de ECONOMÍA

Detecting anomalous payments networks: A dimensionality reduction approach

By: Carlos León

No. 1098
2019

Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

Detecting anomalous payments networks: A dimensionality reduction approach[1]

Carlos León[2]

The opinions contained in this document are the sole responsibility of the authors and do not commit Banco de la República or its Board of Directors.

Abstract

Anomaly detection methods aim at identifying observations that deviate manifestly from what is expected. Such methods are usually run on low dimensional data, such as time series. However, the increasing importance of high dimensional payments and exposures data for financial oversight requires methods able to detect anomalous networks. To detect an anomalous network, dimensionality reduction allows measuring to what extent its main connective features (i.e. the structure) deviate from those regarded as typical or expected. The key to such measure resides in the ability of dimensionality reduction methods to reconstruct data with an error; this reconstruction error serves as a yardstick for deviation from what is expected. Principal component analysis (PCA) is used as dimensionality reduction method, and a clustering algorithm is used to classify reconstruction errors into normal and anomalous. Based on data from Colombia's large-value payments system and a set of synthetic anomalous networks created by means of intraday payments simulations, results suggest that detecting anomalous payments networks is feasible and promising for financial oversight purposes.

Classification: E42, C38, C53

Keywords: anomaly, payments, network, dimensionality, clustering.

---

[1] The opinions and statements in this article are the sole responsibility of the author, and do not represent neither those of Banco de la República nor of its Board of Directors. I thank Oscar Acero for his assistance on intraday payments data processing. I thank Oscar Acero, Freddy Cepeda, Pamela Cardozo, Clara Machado and Fabio Ortega for their comments and suggestions. Any remaining errors are my own.

[2] Senior Researcher, Financial Infrastructure Oversight Department, Banco de la República; Extramural Fellow, CentER, Tilburg University. E-mail: cleonrin@banrep.gov.co / carlosleonr@hotmail.com.

Detección de redes de pagos anómalas:
Una aproximación desde la reducción de dimensiones

Carlos León

Las opiniones contenidas en el presente documento son responsabilidad exclusiva de los autores y no comprometen al Banco de la República ni a su Junta Directiva

Resumen

Los métodos para detección de anomalías buscan identificar observaciones que se desvían ostensiblemente de lo esperado. Esos métodos suelen utilizarse con datos de baja dimensionalidad, tales como las series de tiempo. Sin embargo, la creciente importancia de las series de redes de pagos y exposiciones –series de alta dimensionalidad- en el seguimiento de los mercados financieros exige métodos aptos para detectar redes anómalas. Para detectar una red anómala, la reducción de dimensiones permite cuantificar qué tan diferentes son las características conectivas de una red (i.e. su estructura) con respecto a aquellas que pueden ser consideradas como normales. Esto se consigue gracias a que la reducción de dimensiones permite reconstruir los datos con un error; ese error sirve de parámetro para determinar qué tan diferentes son las características conectivas de las redes. La descomposición por componentes principales es utilizada como método para reducir dimensionalidad, y un algoritmo de agrupamiento clasifica los errores de reconstrucción en normales o anómalos. Con base en datos del sistema de pagos de alto valor colombiano y un conjunto de redes de pagos anómalas creadas artificialmente a partir de métodos de simulación de pagos intradía, los resultados sugieren que la detección de redes de pagos anómalas es posible y prometedor para propósitos de seguimiento de los mercados financieros.

Clasificación: E42, C38, C53

Palabras clave: anomalías, pagos, redes, dimensionalidad, agrupamiento.

1 Introduction

Anomaly detection is about identifying observations that do not conform to what is typical. A generic approach to anomaly detection is to find observations that may be considered of low probability of occurrence as they lie beyond a threshold distance from actual data or a chosen statistical distribution for the data. In the case of a low dimensional system, say, a vector of observations (e.g. time series of an asset's price), this approach is rather straightforward –although by no means precise or unfailing. Yet, for high dimensional systems, such as a time series of matrices representing interactions or links between agents, this approach may be inconvenient. Discovering network-wide anomalies at the link level is particularly difficult because link anomalies may not be indicative of network-level anomalies (Lakhina, et al., 2004, Huang, et al., 2006, Ding & Tian, 2016)[3]. As a single observation of a network contains multiple variables (i.e. data is multidimensional), not any particular one but rather a combination of dimension values may signal an anomaly (Han & Kamber, 2006).

Payments systems are one of such high dimensional systems. In payments systems, a non-small number of agents send and receive funds to and from each other through time, creating a time series of matrices. In this vein, detecting an anomalous payments matrix out of a time series of payments matrices is challenging because of the dimensionality problem.

To detect anomalous payments networks, reducing the dimensionality of payments system data is convenient. Dimensionality reduction is the process of finding a suitable lower-dimensional space in which to represent the original data (Martínez, et al., 2011). From several alternatives to reduce the dimensionality of payments system matrices for anomaly detection[4], I choose Principal Component Analysis (PCA). PCA is a well-known and computationally efficient technique that permits the reduction of the number of variables being analyzed without losing *too much* information. From an anomaly detection viewpoint, PCA-based dimensionality reduction attains an intuitive separation of the space into normal and anomalous subspaces (Lakhina, et al., 2004). Alternatively, from a predictive perspective (see Shmueli, 2010), PCA compresses matrices' main features, and allows using them as predictors of what an expected payments matrix is.

[3] Furthermore, working on link-level anomalies may turn computationally hard as payments networks of n participants may contain up to n^2 links; when no self-connections are allowed, $n(n-1)$ links.
[4] Network analysis statistics (e.g. density, clustering, average distance, reciprocity) may be interesting choices for anomaly detection. However, by definition, they focus on a particular feature or dimension of the network (e.g. connectedness, transitivity, closeness, reciprocity), hence they are inconvenient for network-level anomaly detection purposes. Hence, as suggested by Lakhina, et al. (2004), it is difficult to extract meaningful information about network anomalies from any kind of statistics.

Reducing the dimensionality of payments networks for anomaly detection is a novel approach spurred by the increasing prevalence of network analysis for financial oversight purposes.[5] To the best of my knowledge, the work of Triepels, et al. (2017) is the first and only to undertake this approach. They employ an autoencoder (i.e. a feed-forward neural network) to reduce the dimensionality of the Dutch part of the Eurosystem payments network extracted from TARGET2.[6]

The result of the dimensionality reduction is a *lossy compression* of payments data. It is lossy because only the essential features of data are preserved, thus it is useful to detect anomalous observations (see Han & Kamber, 2006, Triepels, et al., 2017). Afterwards, based on lossy compression of typical or normal payments data, observed payments data is reconstructed with an error. If the reconstruction error is small, the observed payments matrix conforms to what is typical; on the other hand, if reconstruction error is large, the observed payments matrix is to be considered an anomaly –presumably, produced by a distinct (i.e. atypical) generating process. Instead of an autoencoder, I employ PCA, which is a simpler and more tractable method –at the expense of neglecting non-linear features in payments networks. I use a clustering algorithm to classify reconstruction errors into normal and anomalous.

Anomaly detection is challenging because there are very few observations that can be labeled as outliers and they do not fit a consistent pattern (Alpaydin, 2014). Consequently, Triepels, et al. (2017) use artificial bank runs (i.e. additional liquidity outflows) to evaluate the performance of the autoencoder approach. Instead, I use synthetic anomalous networks derived from payments simulation methods.

Overall, results show that the reconstruction of those synthetic anomalous payments networks from PCA-based lossy compression yields reconstruction errors that deviate from what is expected. An agglomerative clustering algorithm uses the difference in reconstruction errors to identify those payments networks that may be classified as anomalies because of their divergence in connective structure. When the assumptions behind the synthetic payments network are severe enough, the identification of anomalies reaches accuracy levels around eighty percent. Hence, it is fair to say that the suggested approach is promising for the oversight of payments networks and –possibly- of other financial networks (e.g. exposures, ownership).

To be able to identify anomalous networks is relevant for financial oversight purposes. Network-wide anomalies may signal forthcoming events or be evidence of a previous

[5] Literature about reducing the dimensionality of other types of networks exist. Most of it is dedicated to internet traffic networks (see Lakhina, et al., 2004, Huang, et al., 2006, Ding & Tian, 2016).

[6] TARGET2 is the large-value payments system that enables central banks and commercial banks in the Eurosystem to transfer money between each other under real-time gross settlement. It is owned and operated by the Eurosystem (i.e. European Central Bank and national central banks of all European Union member states that have adopted the Euro).

incidence. Both are key for financial oversight. Moreover, based on the identification of network-wide anomalies, financial authorities may further examine their source, which may be participant-driven (i.e. caused by a financial institution) or market-driven (i.e. a system-wide change). In this vein, the herein suggested approach provides a simple and feasible method for detecting anomalous networks, which add to existing monitoring tools based on lower-dimension data.

2 Dimensionality reduction with Principal Component Analysis

Principal Component Analysis (PCA) is a customary dimensionality reduction technique introduced by Pearson (1901).[7] PCA aims to reduce the dimensionality of correlated data by finding a small number of orthogonal (i.e. uncorrelated) linear combinations that account for most of the variability of the original data (McNeil, et al., 2005). Similarly, Campbell, et al. (1997) describe PCA as a technique that permits the reduction of the number of variables analyzed without losing too much information in the covariance matrix.

2.1 Compressing and decompressing data with PCA

Let X be a $d \times t$ matrix containing the original d-dimension and t-observation data, and $A = XX^{\mathrm{T}}$ the $d \times d$ covariance matrix of X, PCA is based on the eigenvector or spectral decomposition of the covariance matrix A, which states that any matrix $A \in \mathbb{R}^{d \times d}$ can be written as

$$A = \Gamma\Lambda\Gamma^{\mathrm{T}} \qquad [1]$$

In this setting, Λ is a diagonal $d \times d$ matrix in which diagonal entries correspond to the eigenvalues of A, $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$, such that $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_d$, and Γ is a $d \times d$ matrix containing the eigenvectors as columns paired to the eigenvalues, such that the j-th column contains the j-th eigenvector. By construction, the first principal component, corresponding to the first column in Γ, lies in the direction of maximum variance of the samples, whereas the second column corresponds to the direction of maximum variance in the remaining data (except for the variance represented by the first component), and so on (Ding & Tian, 2016).

When used as a dimensionality reduction technique, the main objective of PCA is to reduce the dimensionality from d to $k \ll d$ while retaining most variance in the original data

[7] In Economics and Finance, PCA is commonly used to build factor models (see Brooks, 2008) and to build indexes based on the contribution of the constituents to overall variance (see McNeil, et al, 2005).

(i.e. without losing *too much* information). As it is most likely that the first dimensions retain most of the variance in correlated datasets, discarding the last dimensions should not result in a significant loss of information.[8] If this is the case, the data effectively resides in an k-dimensional subspace out of the original d-dimensional space.

The fraction of variance retained when reducing the dimensionality from d to k is given by the cumulative percentage contribution of the first k-th eigenvalues to the sum of eigenvalues,

$$w_k = \sum_{j=1}^{k} \lambda_j \Big/ \sum_{j=1}^{d} \lambda_j, \text{ where } 0 < w_k \leq 1 \qquad [2]$$

Reducing dimensions from d to k yields a $d \times k$ matrix $\overline{\Gamma}$. This matrix contains the top k eigenvectors (in columns) corresponding to the first k eigenvalues from [1]. If $k = 1$, $\overline{\Gamma}$ is a $d \times 1$ matrix known as the first eigenvector or the first principal component, which is the one that captures the most variance of the data. From $\overline{\Gamma}$, a lower dimension reconstructed version of X may be obtained by calculating $\hat{X}$ (see Lakhina, et al., 2004, Ghodsi, 2006, Ng, 2017).

$$\hat{X} = \overline{\Gamma}\overline{\Gamma}^{\mathrm{T}} X \qquad [3]$$

$\hat{X}$ is a reconstructed version of X from $\overline{\Gamma}$, with its dimension matching that of the original data ($d \times t$). As it is a reconstructed version of X, based on a lower dimension encoding, there is a reconstruction or decompression error. Let z represent the number of elements in X (i.e. $z = d \times t$), the squared reconstruction error (e) is calculated as

$$e = \sum_{i=1}^{z} (x_i - \hat{x}_i)^2 \qquad [4]$$

Overall, the widespread use of PCA for dimensionality reduction is due to the following features (see Sree & Venkata, 2014, Ding & Tian, 2016): First, it is the optimal linear scheme for compressing and decompressing a set of highly dimensional vectors into a set of lower dimensional vectors. Second, the square error of the reconstructed data is inversely proportional to the dimension. Third, the model is stable without adjusting parameters in the process. Fourth, compression and decompression are easy to conduct –only matrix multiplication is required. Nonetheless, there is a caveat: in its standard form, PCA is a linear dimensionality reduction method that may overlook non-linear features in data.

[8] On the other hand, as it is unlikely to find prominent eigenvalues when variables are uncorrelated (e.g. if the dataset is random), discarding dimensions could compromise the integrity of information.

2.2 An illustrative example: image compression

A straightforward application of PCA for dimensionality reduction is image compression (see Ng, 2017). In this case, as an image is a matrix of values corresponding to its pixels, it is possible to compress and reconstruct any image with PCA.

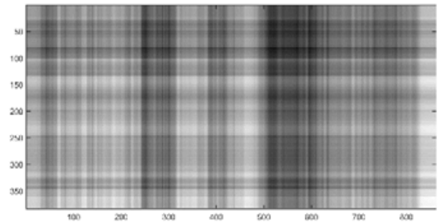
Panel a. in Figure 1 exhibits a grayscale version of the Colombian fifty thousand Peso bill in a 382-row by 862-column image or matrix. Panel b. exhibits a reconstruction based on the first eigenvector of the original image ($k = 1$); as expected, the reconstruction based on dropping 861 (out of 862) dimensions offers a modest version of the original, with a reconstruction error (e) about 898.29. As the number of dimensions increases (i.e. as k increases), the quality of the reconstruction rises (i.e. the image becomes intelligible) and the reconstruction error decreases. When k reaches about one tenth (panel e.) or half (panel f.) of the original dimensions, the image starts to be analogous to the original.

A naïve method for detecting whether the image of a bill corresponds to that of the Colombian fifty thousand Peso bill could be designed based on image compression.[9] This could comprise the following steps. First, choosing a k corresponding to a lossy compression that retains the essential features only. Second, for a non-small number of different Colombian fifty thousand Peso bills (e.g. newly printed, slightly deteriorated, somewhat deteriorated), calculate $\overline{\Gamma}$ (as in [1]) and the corresponding reconstruction error e (as in [4]). Third, use the typical $\overline{\Gamma}$ to obtain the reconstructed version of the bill (as in [3]) and its corresponding reconstruction error (as in [4]). Fourth, based on the empirical distribution of e and a threshold for what is to be considered an anomalous reconstruction error, determine whether the bill corresponds to the Colombian fifty thousand Peso bill or not.

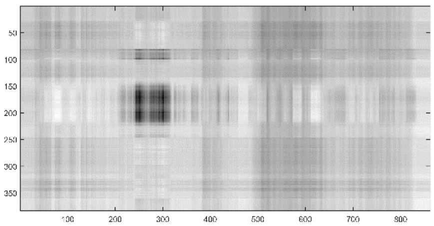
This naïve method may be adjusted to detect whether a payments matrix corresponds to what is expected as a typical payments matrix. Instead of an image, a payments matrix. Instead of pixels' values, the matrices would contain the value of payments between financial institutions.

[9] It is *naïve* because it could serve the purpose of identifying whether the image of a given bill deviates *too much* from the average Colombian fifty thousand Peso bill, but it will fail to detect counterfeit bills. Moreover, it is *naïve* as we are not considering some desirable features of image recognition methods, such as invariance to translation, rotation, and size.

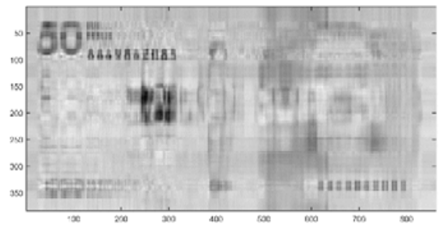
a. $k = d$, $w_k = 1.00$, $e = 0.00$, size 543KB



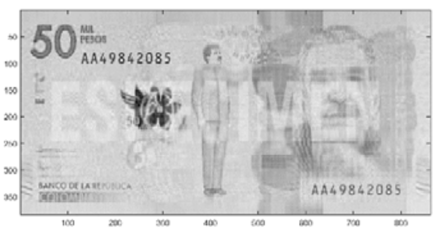
b. $k = 1$, $w_k = .9779$, $e = 898.29$, size 261KB



c. $k = 2$, $w_k = .9825$, $e = 711.79$, size 219KB



d. $k = 10$, $w_k = .9894$, $e = 429.58$, size 275KB



e. $k = 38$, $w_k = .9949$, $e = 208.34$, size 367KB

f. $k = 191$, $w_k = .9993$ $e = 26.99$, size 529KB

Figure 1. Image compression based on PCA dimensionality reduction. Source: author's calculations, based on original PNG format image downloaded from http://www.banrep.gov.co/es/billete-50mil-pesos.

3 Data

Akin to Triepels, et al. (2017), I use payments data from the large-value payments system. In Colombia, the large-value payments system (CUD, Spanish acronym for the Deposit Accounts System) works under a real-time gross settlement system owned and managed by Banco de la República –the central bank of Colombia. CUD is a non-tiered payments system, in which financial institutions of several types (e.g. commercial banks, investment banks, securities brokers-dealers, pension funds, mutual funds, insurance companies, and financial cooperatives) are eligible to participate directly.[10] Based on previous research, Colombia's large-value payments system networks' connective structure conforms to stylized facts of large-value payments networks (see Cepeda, 2008, León & Berndsen, 2014, Bernsdsen, et al., 2018).[11]

The sample starts in January 2, 2017 and ends in December 28, 2018. There are 484 days and 125 participants in this sample. Therefore, I work with a hypermatrix (i.e. a cube) of size [125, 125, 484]. Each matrix is a weighted matrix, with elements calculated as the contribution to the sum of all weights; hence, elements in each layer of the hypermatrix are values between 0 and 1, and the sum of all elements in each layer is 1. In order to make matrices comparable through time, I keep constant the number of participating financial institutions along the selected period.[12]

From this dataset, I extract the expected or typical network. In the sample, there are no episodes of turmoil or failure that could be labeled as outright outliers. However, from a low dimensional perspective, it is common to regard high- and low-value days as atypical.[13] Therefore, in order to extract a time series of typical networks, I build a *filtered* dataset by excluding those observations in which the value of transactions exceeds a two-tail 90% confidence interval. That is, I discard days with total value of payments below the fifth percentile or above the 95th percentile, as in Figure 2.

The resulting filtered hypermatrix' size is size [125, 125, 436]. As expected, excluding those (48) observations enables to work with a distribution of payments that resembles a Gaussian distribution; for instance, the kurtosis of the distribution decreased from 7.9 to 2.9. It is worth emphasizing that this exclusion does not compromise the exercise: the aim is to identify anomalies arising from atypical changes in the connective features (i.e. the structure) of payments networks –in which the absolute value of payments is not considered. That is, detecting anomalous payments networks by identifying atypical overall transaction values is the simplest (i.e. low dimensional) method available, and it should be attempted before incurring in a high dimensional method such as the one herein described.

[10] Other institutions participate directly in CUD too, such as the Central Bank, the Ministry of Finance, information processors, and financial infrastructures. As the aim is detecting anomalies in payments between financial institutions, these other types of institutions are excluded. Yet, under some circumstances, it may be convenient to add these participants to the sample.

[11] For instance, CUD networks are inhomogeneous and hierarchical, presumably approaching a modular scale-free connective structure. They display typical features of other large-value payments networks, such as sparseness, clustering, small by mean geodesic distance, and with right-skewed (i.e. heterogeneous) distributions of connections and their intensity.

[12] In this case, this is a judicious choice. However, in practice, the first two dimensions (i.e. rows and columns) of the cube could change as financial institutions enter or exit the payments system.

[13] This is a standard procedure that is able to identify the holiday season (e.g. from the days before Christmas until mid-January), days circa national holidays and other days in which market activity deviates from its norm.

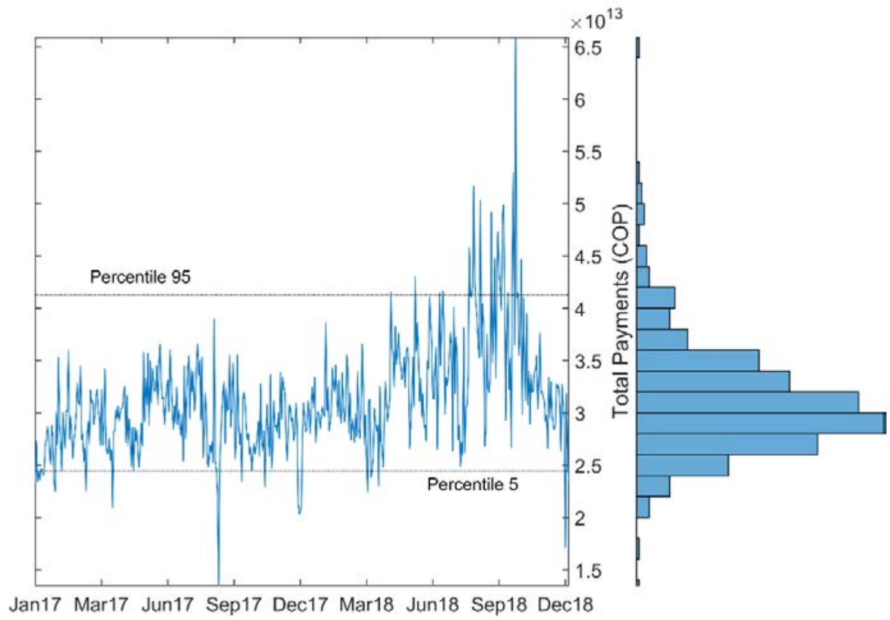


Figure 2. Time series of CUD payments (January 2, 2017-December 28, 2018). Dotted lines correspond to the 90% confidence interval. Discarding the observations below the fifth percentile or above the 95th percentile reduces the kurtosis of the distribution of total payments from 7.9 to 2.9. Source: author's calculations.

As some of those excluded observations correspond to seasonal effects that are rather straightforward to anticipate or that are easy to identify by monitoring the value of transactions, they are not to be considered as network-wide anomalies. Consequently, there is a pending challenge: to get a dataset of payments networks that may be labeled as anomalous. This challenge is well-documented in related literature (see Alpaydin, 2014, Triepels, et al., 2017). Triepels, et al. (2017) use artificial bank runs to circumvent this challenge[14]. In my case, I take advantage of payments simulation methods to create synthetic abnormal payments networks.

Synthetic abnormal payments networks are created as follows. First, I select three participating financial institutions that are interesting due to their importance to the large-value payments system –based on their centrality[15] in payments networks. Due to

[14] The bank runs designed by Triepels, et al. (2017) are based on a probabilistic (i.e. exponentially increasing binomial probability) model of additional liquidity outflows for a given bank. To the best of my understanding, the outflows and inflows of all other banks are unaffected by these additional outflows.

[15] All references to centrality in this article correspond to overall payments network centrality. This is calculated as a PCA-weighted average comprising six centrality indexes (i.e. in and out degree; in and out strength; hub and authority). See Newman (2010) for a comprehensive review of centrality measures.

confidentiality reasons, I name them Bank A, Bank B and Bank C. Second, based on the intraday payments corresponding to the filtered dataset (i.e. observations within the two-tail 90% confidence interval), for each selected financial institution I simulate two sets of abnormal intraday payments sequences.[16] The first set corresponds to a simulation of intraday payments under the assumption that the selected financial institution has a balance equal to zero at the start of the day. The second set corresponds to a simulation of intraday payments under the assumption that the selected financial institution is unable to make any payments throughout the day. Intuitively, I expect to find greater anomalies under the second because setting financial institution's initial balance equal to zero should have a weaker impact as funds received throughout the day will allow making payments. Third, for each one of the two sets of abnormal intraday payments sequences, I build its corresponding payments hypermatrix. As three financial institutions are selected, I have six abnormal payments hypermatrices to work with.

Overall, the anomaly of synthetic networks arises from two sources that should affect the connective structure of the payments network. First, from the reduced ability of the selected participant to fulfil its payments to its counterparties (i.e. direct impact). Second, as an outcome of the latter, from the reduced ability of its counterparties to fulfil their payments (i.e. indirect impact). A simpler approach, say, changing the payments from a participant to its counterparties alone (i.e. without the simulation of subsequent payments among all participants) may fail to produce a structural change in the payments network –which is what I expect an anomaly detection model to capture.

4 In-sample results

The method I designed for in-sample assessing whether a payments matrix is abnormal (or not) comprises several steps. First, choosing a k corresponding to a lossy compression that retains the essential features only. Second, based on k, calculating $\overline{\overline{\Gamma}}$, which is a typical or expected $\overline{\Gamma}$ for the filtered hypermatrix. Third, using $\overline{\overline{\Gamma}}$ to obtain the reconstructed versions of the filtered hypermatrix and the synthetic abnormal payments hypermatrix, and their corresponding reconstruction errors. Fourth, comparing the distribution of the two sets of reconstruction errors; that is, test the null hypothesis of both sets pertaining to the same continuous distribution. This section describes each of these steps in detail.

[16] Payments simulation methods allow replicating the functioning of a payments system. In my case, there are two general assumptions in the simulation. First, following the observed sequence of intraday payments (i.e. no reaction by participants). Second, rejecting payments that could not be processed due to insufficient funds (i.e. no overdraft facilities, no reprocessing of payments). These are conservative assumptions that aim at producing anomalies in the form of network-wide structural changes.

4.1 Selecting k

There is no standard or accepted method for selecting the number of dimensions to drop from the original data. Lossy compression is about discarding most of the dimensions in order to preserve the essential features only. Figure 3 shows the variance retained and the reconstruction error for $k = 1,2, \ldots, 10$, for each observation in the filtered hypermatrix. It is rather clear that a small number of dimensions retains a large portion of variance in the dataset, and that –as expected- there is an inverse relation between reconstruction error and retained variance.

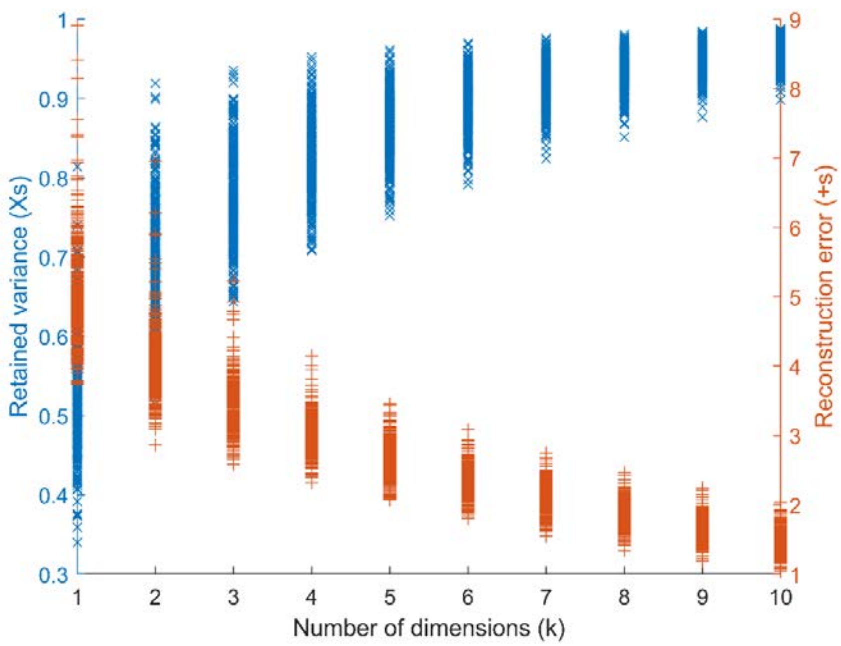


Figure 3. Variance retained (×) and reconstruction error (+), for each observation in the filtered hypermatrix, by k-dimensions. Source: author's calculations.

For brevity, next section uses $k = 1$. On average, with $k = 1$ the retained variance is about 0.54 and the reconstruction error is about 5.04.

4.2 Compressing and decompressing the expected payments network

Based on my choice of $k = 1$, I calculate the expected matrix of $\overline{\Gamma}$. This expected matrix ($\overline{\overline{\Gamma}}$) may be estimated based on a single expected payments network, which may be calculated as the average of networks in the filtered hypermatrix. However, this would force us to use a single matrix to compress and reconstruct all the networks in the filtered and synthetic hypermatrices –with an overly averaged network, and a limited number of reconstruction errors to work with.

Consequently, I undertake a bootstrapping approach to obtaining the expected matrix ($\bar{\bar{\Gamma}}$). Bootstrapping is a simple and useful method for assessing uncertainty in estimation procedures, and for obtaining a more accurate estimation than that obtained with a "raw" simple estimate (Dowd, 2005). In bootstrapping, a uniform random number generator is used to resample (with replacement) from a given dataset. This enables to obtain an entire set of expected matrices ($\bar{\bar{\Gamma}}$) –instead of one. In turn, I obtain a larger set of reconstruction errors to test whether the synthetic payments networks pertain to the same continuous distribution of the filtered payments networks.

In my case, the bootstrapping approach consists of calculating $\overline{\Gamma}$ for each one of the 1,000 random samples extracted from the filtered hypermatrix. I calculate each sample as the average network for 20-randomly selected (with replacement) networks from the filtered hypermatrix. As a result, after compressing and decompressing, I obtain the cumulative distribution of errors for the sample comprising 1,000 20-network averages (Figure 4).[17]

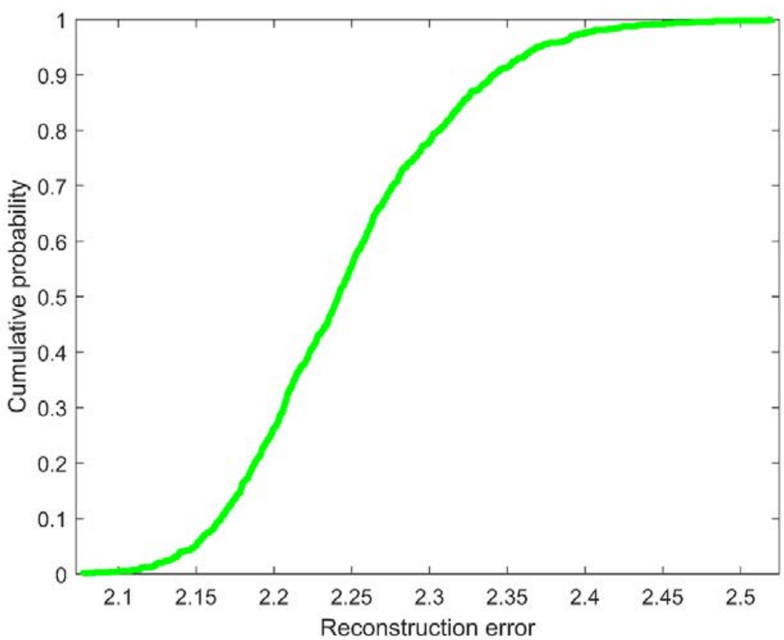


Figure 4. Cumulative distribution of reconstruction errors from bootstrapping. Source: author's calculations.

This distribution of errors in Figure 4 represents the loss caused by compressing and decompressing the sample average networks with $k = 1$. This distribution will serve as a

[17] As the networks that result from decompression may contain elements inconsistent with payments data (e.g. negative payments), the reconstruction procedure contains a lower bound to what a feasible payment is. The procedure converts all payments with values lower than the minimum registered in the sample into a null payment.

benchmark for what the loss is when compressing and decompressing each observation in the filtered and the synthetic hypermatrices.

4.3 Compressing and decompressing filtered and synthetic payments networks

From the bootstrap, I obtain a set of 1,000 estimations of $\bar{\bar{\Gamma}}$, each one with a $[125, k]$ dimension. I use each one of these 1,000 estimations to compress and decompress the filtered and synthetic payments networks.[18] Intuitively, because synthetic payments networks are constructed by means of simulating an abnormal sequence of payments, I expect to find that their reconstruction error is greater than that of filtered payments networks. Moreover, I expect that a test of the equality of probability distributions will reject the null hypothesis that synthetic payments networks are drawn from the filtered payments networks distribution; that is, I expect to verify that decompressing synthetic network payments yields abnormal reconstruction errors that may be used to signal anomalous networks.

Figure 5 compares the distribution of reconstruction errors from the bootstrap method (Figure 4) to those obtained when reconstructing each observation in the filtered hypermatrix and to those obtained when reconstructing each observation in the synthetic hypermatrices built under the assumption of inability to make payments throughout the day for Bank A, Bank B and Bank C. The reconstruction of the filtered and synthetic hypermatrices is based on the 1,000 estimations of $\bar{\bar{\Gamma}}$ from the bootstrap.

The distribution of reconstruction errors of the three synthetic hypermatrices corresponding to the inability to pay of banks A, B and C differs from that of the filtered hypermatrix. From Figure 5, those three distributions are to the right of the filtered hypermatrix, and their right tails are visibly fatter. Thus, as expected, reconstruction errors of the three synthetic hypermatrices are greater than those of the filtered hypermatrix.

[18] The flowchart in the Appendix (Figure A1) illustrates the process designed to obtain reconstruction errors for filtered and synthetic payments networks.

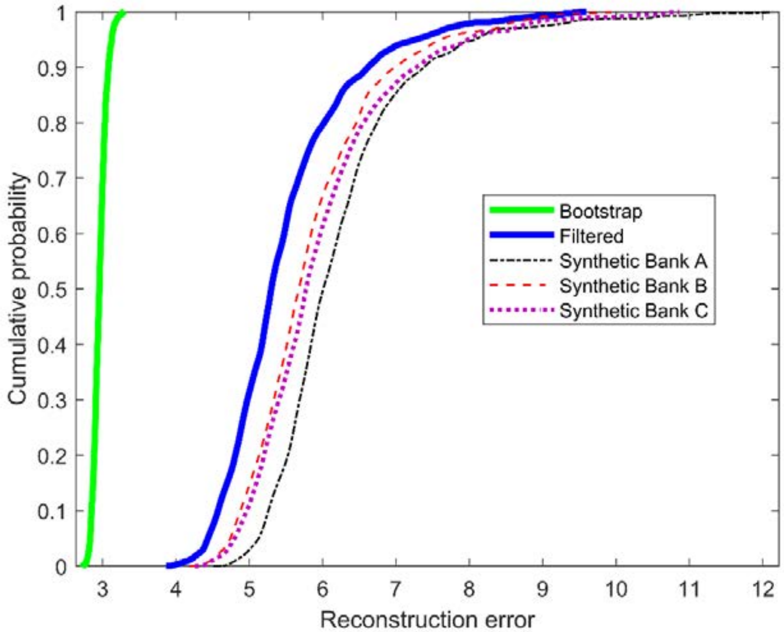


Figure 5. Cumulative distribution of reconstruction errors from bootstrapping, filtered and synthetic payments networks, with $k = 1$. Synthetic payments networks correspond to those built under the assumption of inability to make payments for Bank A, Bank B and Bank C. Source: author's calculations.

Figure 6 compares the distribution of reconstruction errors from the bootstrap method (in Figure 4) to those obtained when reconstructing each observation in the filtered hypermatrix and to those obtained when reconstructing each observation in the synthetic hypermatrices built under the assumption of initial balance set to zero. The distributions of reconstruction errors of the three synthetic hypermatrices corresponding to initial balance set to zero are somewhat different from that of the filtered hypermatrix. The difference with respect to the filtered hypermatrices is lower than that under the alternative assumption of inability to make payments throughout the day (in Figure 5). As before, this is an expected outcome because the inability to pay is a stricter assumption than setting initial balances to zero.

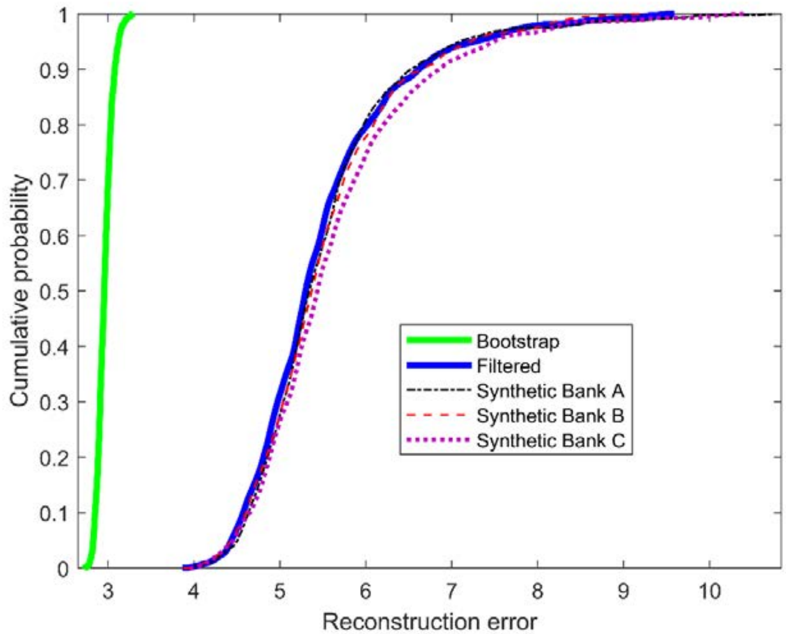


Figure 6. Cumulative distribution of reconstruction errors from bootstrapping, filtered and synthetic payments networks with $k = 1$. Synthetic payments networks correspond to those built under the assumption of initial balance set to zero for Bank A, Bank B and Bank C. Source: author's calculations.

4.4 Testing differences among reconstruction errors distributions

Table 1 shows the first four moments of each distribution of errors calculated in the previous section –under the assumption of $k = 1$. In addition, the p-value from the two-sample Kolmogorov-Smirnov tests the hypothesis of distribution equality with respect to reconstruction errors from the filtered hypermatrix. This non-parametric test rejects the null hypothesis that errors in the six synthetic hypermatrices are drawn from the same continuous distribution as the filtered hypermatrix.[19]

[19] Results are robust to other choices of k (see Table A1 and A2, in the Appendix).

Payments Network (hypermatrix)		Mean	Standard deviation	Skewness	Kurtosis	K-S test p-value[c]
Bootstrap[a]		2.95	0.09	0.55	3.30	0.00
Filtered[b]		**5.48**	**0.88**	**1.53**	**6.26**	**1.00**
Inability to make payments	Synthetic Bank A[b]	6.22	0.99	2.04	9.55	0.00
	Synthetic Bank B[b]	5.85	0.90	1.34	5.36	0.00
	Synthetic Bank C[b]	5.98	1.01	1.56	6.61	0.00
Zero initial balance	Synthetic Bank A[b]	5.52	0.89	1.97	9.29	0.00
	Synthetic Bank B[b]	5.52	0.85	1.22	4.91	0.00
	Synthetic Bank C[b]	5.62	0.96	1.52	6.55	0.00

Table 1. Reconstruction errors' distribution, with $k = 1$. [a] Estimated on the 1,000 samples obtained by bootstrap from the filtered hypermatrix. [b] Estimated on the 436,000 errors obtained by reconstructing the 436 observations in the corresponding hypermatrix with the 1,000 estimations of $\overline{\overline{\Gamma}}$. [c] Two-sample Kolmogorov-Smirnov test of distribution equality, with respect to the reconstruction errors from the filtered hypermatrix. Source: author's calculations.

5 Out-of-sample results

Rejecting the null hypothesis of distribution equality with respect to in-sample reconstruction errors is noteworthy. However, as anomaly detection is a *predictive modeling* problem (see Shmueli, 2010), testing whether reconstruction errors enable to identify anomalous networks on new (unseen, unprocessed) data is fundamental. Hence, to test out-of-sample performance, a new dataset comprising new filtered and synthetic data from February 1st to March 29th, 2019 is created. I train the model with (old) data from January 1st, 2017 to December 28th 2018 –the same dataset used for in-sample results.[20]

This new dataset has three different subsets that mix filtered and synthetic data. The first subset combines filtered and synthetic data under the assumption that the top-5 financial institutions by network centrality are either starting with zero balance in their deposit accounts or are unable to make payments during the day. The second mimics the first subset but excludes the zero initial balance assumption. Based on the second, the third subset considers the top-3 financial institutions by payments network centrality only. As the average severity of subsets is increasing (i.e. the second subset is stricter than the first, and the third is the strictest), I expect the out-of-sample results to be increasing in their accuracy as well.

[20] The flowchart in the Appendix (Figure A2) illustrates the process designed to obtain in-sample (i.e. training) and out-of-the sample reconstruction errors.

5.1 Out-of-sample classification

In order to test the out-of-sample performance, the method should be able to distinguish those payments networks built with filtered data from those built with synthetic data. That is, the reconstruction error obtained from compressing and decompressing a payments network should allow classifying it as normal or anomalous with fair accuracy.

However, the reconstruction error does not provide a straightforward classification of each payments network as normal or anomalous; it is a measure of how distant perfect reconstruction is for a payments network, but does not deliver its class or category. Therefore, it is necessary to implement an ancillary method to extract information out of reconstruction errors in order to classify them as anomalous.

I choose an agglomerative clustering algorithm to perform unsupervised anomaly detection (see Thottan et al., 2010). In this case, an agglomerative clustering algorithm uses reconstruction errors as distances to group them into two clusters, consisting of normal and anomalous payments networks.[21] The normal group will always be the one that contains the lowest average reconstruction error; this is a fair and judicious conjecture, as the main assumption is that compressing and decompressing normal payments networks yields low reconstruction errors when compared to those attained with anomalous ones. In this vein, I will use the lowest reconstruction errors as an *attractor*, with all payments networks pertaining to that cluster classified as normal. On the other hand, payments networks clustered in the other group (i.e. distant to the low reconstruction error cluster) are classified as anomalies[22].

Figure 7 displays a dendrogram that illustrates the classification by means of agglomerative clustering on reconstruction errors on the second subset. In this case, there are five different synthetic payments networks and one filtered payments network. The synthetic networks comprise the top-5 financial institutions in the sample by centrality (banks V, W, X, Y, and Z)[23], under the assumption of no payments during the day. In parenthesis, the name of each bank is accompanied by the reconstruction error attained. Distances between

[21] In agglomerative clustering methods, groups of elements are successively merged by their similarity (i.e. low distance) until a single group is formed. The result of agglomerative clustering methods is a hierarchical structure that represents how elements relate to each other based on their cross-section similarities. See Martínez et al. (2011) and Han and Kamber (2006) for a throughout explanation.

[22] As before, for this classification we employ the bootstrapping approach. Hence, the agglomerative clustering is not based on a single reconstruction error but on a 1,000 random set of reconstruction errors. The agglomerative clustering algorithm used Euclidean distances and single linkage method; changing the linkage method from single to Ward results in a minor increase in predictive performance.

[23] I use the term "bank" in a general manner –for brevity and confidentiality reasons. However, this top-5 includes two non-banking institutions.

payments networks or their clusters are displayed in the horizontal axis; the larger the horizontal distance, the more different the reconstruction error.

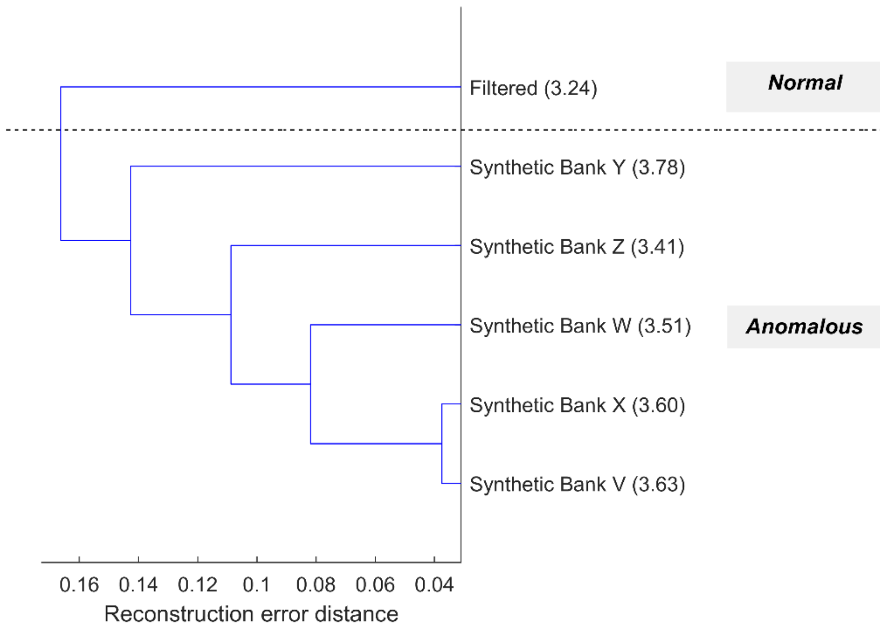


Figure 7. Dendrogram of reconstruction error distances. Distances between payments networks or their clusters are displayed in the horizontal axis; distances correspond to Euclidean distances between payments networks' reconstruction errors (in parenthesis). Source: author's calculations.

The dendrogram shows two main clusters –separated by the horizontal dashed line. The cluster that attained the lowest average reconstruction error (3.24) is classified as the normal cluster; as expected, this cluster contains the filtered payments network only. The other cluster, containing synthetic payments networks, displays higher reconstruction errors – hence, it is classified as the anomalous cluster. In this example, there is no misclassification: the filtered payments network is the only element in the normal cluster (i.e. a true negative), whereas all synthetic payments networks are contained in the anomalous cluster (i.e. true positives). A misclassification would occur if one of the synthetic payments networks locates in the normal cluster (i.e. a false negative) or if the filtered payments network does not pertain to the normal cluster (i.e. a false positive).

5.2 Out-of-sample performance

I implement two standard tools for assessing out-of-sample performance in classification problems, namely the *confusion matrix* and the *Receiver Operating Characteristic* curve. The confusion matrix is a table whose rows represent the target (i.e. actual) class and whose columns represent the output (i.e. predicted) class. In my case, as there are two classes (i.e. normal and anomalous), the confusion matrix is a two-by-two table (see Table 2). For a classifier to have good accuracy, most of the predictions are along the diagonal of the confusion matrix (i.e. predicted class matches actual class), with the rest of the entries (i.e. below or above the diagonal) being close to zero (Han & Kamber, 2006). Entries in the diagonal are either true negatives (TN) or true positives (TP). Those outside the diagonal correspond to either false positives (FP) or false negatives (FN); these two are also known as Type I and Type II errors, respectively.

		Output (predicted) class: Normal	*Output (predicted) class: Anomalous*
Target (actual) class	*Normal*	True negatives (TN)	False positives (FP)
	Anomalous	False negatives (FN)	True positives (TP)

Table 2. Confusion matrix. Source: author's design.

From the confusion matrix, three rates or ratios are usually reported. First, the True Positive Rate or *sensitivity*, which is the proportion of positives that are correctly classified (TPR= $\mathrm{TP}/(\mathrm{TP{+}FN})$). Second, the True Negative Rate or *specificity*, which is the proportion of negatives that are correctly classified (TNR= $\mathrm{TN}/(\mathrm{TN{+}FP})$). Third, *accuracy*, which is the ratio of true positives and true negatives to the sum of true and false positives and negatives (ACC= $(\mathrm{TP{+}TN})/(\mathrm{TP{+}FP{+}TN{+}FN})$).

As there are three different subsets, there are three different confusion matrices. In addition, there are several confusion matrices by the choices of k (i.e. number of dimensions to keep in compression). Table 3 displays the three corresponding confusion matrices for $k = 5$; other choices of k do not change the main outcome –as shown in the Appendix (Table A3).

Target (actual) class		Output (predicted) class: Normal	Output (predicted) class: Anomalous
	Normal	37,464 (0.08)	2,536 (0.01)
	Anomalous	250,442 (0.57)	149,558 (0.34)

a. Subset #1. (top-5 institutions, zero initial balance or inability to make payments; 440,000 simulations)

Target (actual) class		Output (predicted) class: Normal	Output (predicted) class: Anomalous
	Normal	38,214 (0.16)	1,786 (0.01)
	Anomalous	65,361 (0.27)	134,639 (0.56)

b. Subset #2. (top-5 institutions, inability to make payments; 240,000 simulations)

Target (actual) class		Output (predicted) class: Normal	Output (predicted) class: Anomalous
	Normal	38,145 (0.24)	1,292 (0.01)
	Anomalous	33,666 (0.21)	86,334 (0.54)

c. Subset #3. (top-3 institutions, inability to make payments; 160,000 simulations)

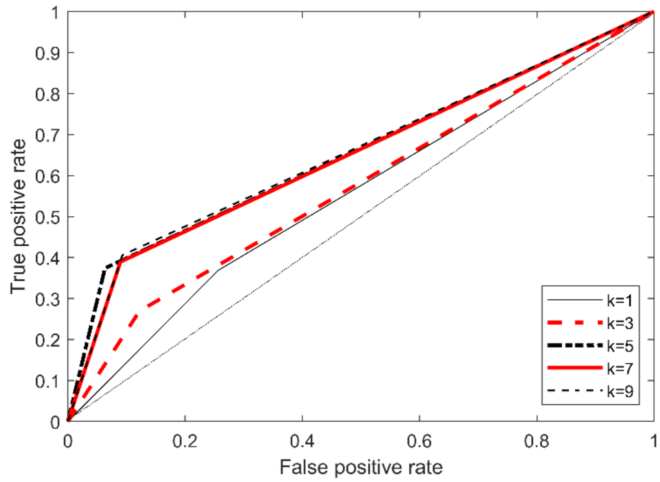
Table 3. Confusion matrices. Number of simulations and percentage of simulations (in parenthesis). Calculations based on $k = 5$. Source: author's calculations.

As expected, the first subset attains a poor classification. This subset considers the top-5 institutions by payments network centrality but comprises two different assumptions about the source of the anomaly, namely the inability to make payments or zero initial balance. For this subset, sensitivity, and accuracy are rather low, about 0.37 and 0.42, respectively. However, the specificity is high, 0.94. Overall, results show that, on the first subset, the method is successful at classifying filtered payments networks as normal but fails at classifying synthetic payments networks as anomalies.

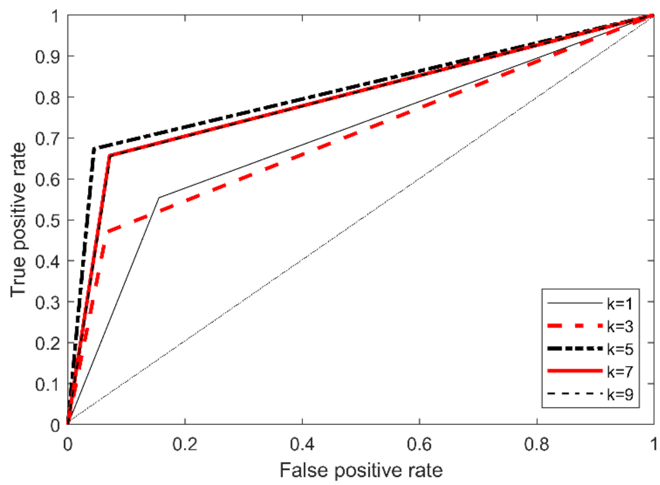
The second subset does not include the zero initial balance assumption, which is the mildest of the two included in the first subset. Correspondingly, classification improves. Specificity remains high, about 0.95. Sensitivity reaches 0.67, which means that about two thirds of synthetic payments networks are classified as such. Accuracy reaches 0.72. This notable improvement in classification is due to discarding those synthetic payments networks whose assumptions do not entail a potentially large effect (i.e. zero initial balance).

The third subset does not include the assumption of zero initial balance, and it is restricted to the top-3 financial institutions by payments network centrality. That is, the synthetic payments networks in the third subset correspond to those assumptions that are the most likely to cause a potentially large effect in the networks' connective structure. Specificity remains high at 0.95. Sensitivity reaches 0.72, which means that almost three fourths of synthetic payments are correctly classified as anomalous. Accuracy reaches 0.78.

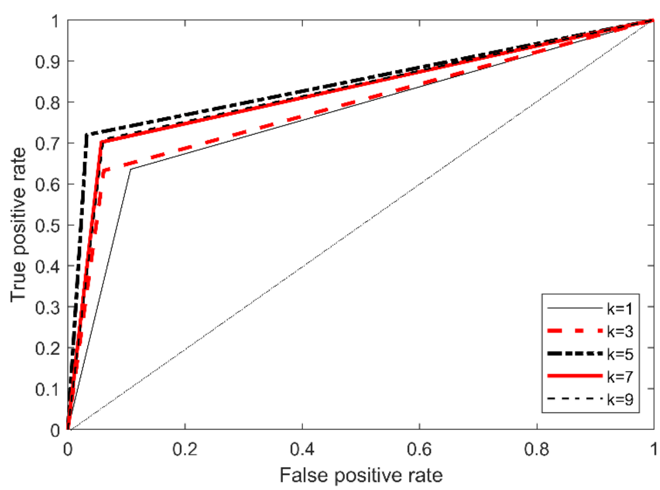
The second selected tool for assessing out-of-sample performance is the *Receiver Operating Characteristic* (ROC) curve. It is a tool that enables to visualize the trade-off between the rate at which the model can accurately recognize positive cases versus the rate at which it mistakenly identifies negative cases as positives (Han & Kamber, 2006). The vertical axis of the ROC curve represents the True Positive Rate (TPR), whereas the horizontal represents the False Positive Rate. Thus, the closer the curve to the vertical axis, the more accurate the model; the closer to the diagonal line (the 45 degree line), the less accurate –a perfect diagonal line corresponds to a random guess. Figure 8 displays the ROC curves corresponding to the three subsets under analysis, under different values for k.



a. Subset #1 (top-5 institutions, zero balance or inability to make payments)



b. Subset #2 (top-5 institutions, inability to make payments)



c. Subset #3 (top-3 institutions, inability to make payments)

Figure 8. ROC curves. Source: author's calculations.

Results portrayed in ROC curves are consistent with those attained with confusion matrices. Curves are closer to the diagonal when assumptions include those that may be considered as mild. Discarding those mild assumptions causes ROC curves to depart from the diagonal, and to achieve better trade-offs between true positive cases and false positives.

Finally, Table 4 reports the area under each ROC curve for different choices of k. The area under the ROC curve measures the accuracy of the model (Han & Kamber, 2006). Results in Table 4 are consistent with those reported before.

Subset	$k =$									
	1	2	3	4	**5**	6	7	8	9	10
#1	0.56	0.56	0.57	0.61	**0.66**	0.66	0.65	0.65	0.66	0.66
#2	0.70	0.67	0.70	0.76	**0.81**	0.81	0.79	0.79	0.79	0.79
#3	0.76	0.77	0.79	0.82	**0.84**	0.83	0.82	0.82	0.82	0.83

Table 4. Area under the ROC curve, for different choices of k. The best performing k for each subset (before rounding) is in bold. Source: author's calculations.

Therefore, based on the confusion matrices, their corresponding ratios and the ROC curves, it is fair to say that the method can classify synthetic payments networks. Nevertheless, the severity of the assumptions used to construct the synthetic payments networks determines the ability to classify. The stricter the assumptions, the better the out-of-the-sample performance of the classifier to detect anomalies.

6 Final remarks

Financial networks analysis has become a common tool for financial authorities dealing with markets' oversight. The identification of anomalous networks for oversight purposes is challenging as they are multidimensional. Reducing the dimensionality of financial networks is an obvious path to surmount such challenge.

In this article, I use principal component analysis (PCA) to reduce the dimensionality of Colombian large-value payments system's networks to achieve a method able to identify networks that deviate manifestly from what is considered an expected or typical network connective structure, i.e. anomalous networks. PCA is a well-known and computationally efficient technique for dimensionality reduction that has been used for network-wide anomaly detection in non-financial networks (see Lakhina, et al., 2004, Huang, et al., 2006, Ding & Tian, 2016).

Results suggest that detecting anomalous payments networks by means of PCA-based dimensionality reduction is feasible. Out-of-the-sample performance tests show that the method can identify synthetic payments networks whose connective structure deviates manifestly from what is expected for a typical payments network. The performance varies according to the severity of the assumptions used to construct synthetic payments networks. The stricter the severity, the better the performance. Therefore, it is fair to say that the suggested approach is judicious and promising for the oversight of payments networks and –possibly- of other financial networks (e.g. exposures, ownership).

The main contribution of this approach is to provide additional tools for financial oversight based on financial networks. First, anomalies may signal forthcoming events or be evidence from an incidence. Second, by identifying anomalous networks, financial authorities may decide further examining the cause of the anomaly, which may aid to find individual (i.e. at participant level) or system wide events. Consequently, financial authorities will find useful to implement an approach able to identify anomalous networks.

To be used as a practical financial oversight tool, computational and data collection costs should be low. Regarding the first, the core of the dimensionality reduction problem is equivalent to solving the eigenvalue for covariance matrix (as in [1]). As the number of participants in the Colombian large-value payments system is about 125 financial institutions (and it is expected to remain rather stable), the computational time required to attain that solution is low –about 0.003 seconds. For other countries, even if the number of participants in the large-value payments system is large, computational time remains low.[24] About the second, as the Colombian central bank owns and manages the large-value payments system, data is available on an end-of-the day basis, and its collection cost is nearly null. As large-value payments are usually owned and managed by central banks, or at least they have access to large-value payments data for financial oversight purposes, it is likely that data collection cost is also low for other jurisdictions.

Some shortcomings and challenges are worth stating. First, considering that anomalies are scarce and do not fit a consistent pattern (Alpaydin, 2014), results herein reported are dependent on the construction of synthetic anomalous networks. Observed anomalies and other synthetic datasets will serve to further test the approach. Second, it is advisable to examine whether implementing a non-linear approach to dimensionality reduction -such as that in Triepels, et al. (2017)- is necessary to increase the accuracy. Third, it is worth evaluating additional approaches to identify anomalies in payments networks. For instance, based on preliminary results (forthcoming), neural networks-based pattern recognition on large-value payments data has demonstrated promising results for detecting anomalies at the

[24] Computation time for a single payments network matrix of size [125, 125] is about 0.003 seconds in Matlab, on a workstation with a 6-core 3.6Ghz Intel Xeon processor and 128Gb RAM. Using random numbers from a uniform distribution, if size increases to [1000, 1000] and [10000, 10000], computation time reaches about 0.10 and 13 seconds, respectively.

participant level. Fourth, as large-value payments networks encompass a miscellaneous spectrum of payments, it is advisable to examine how focusing on specific types of payments (e.g. interbank lending, sovereign securities purchases) affects the feasibility of this approach.

7 References

Alpaydin, E. (2014). *Introduction to Machine Learning*. The MIT Press: Cambridge.

Berndsen, R., León, C., & Renneboog. L. (2018). Financial stability in networks of financial institutions and market infrastructures, *Journal of Financial Stability, 35*, 120-135.

Brooks, C. (2008). *Introductory Econometrics for Finance*. Cambridge University Press: Cambridge.

Campbell, J., Lo, A., & Mackinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton University Press: Princeton.

Cepeda, F. (2008). La topología de redes como herramienta de seguimiento en el sistema de pagos de alto valor en Colombia. *Borradores de Economía, 513*, Banco de la República.

Ding, M. & Tian, H. (2016). PCA-based network traffic anomaly detection. *Tsinghua Science and Technology, 21(5)*, 500-509.

Dowd, K. (2005). *Measuring Market Risk*. John Wiley & Sons: Chichester.

Ghodsi, A. (2006). *Dimensionality reduction: a short tutorial*. Department of Statistics and Actuarial Science, University of Waterloo.

Han, J. & Kamber, M. (2006). *Data Mining*. Morgan Kaufmann: San Francisco.

Huang, L., Nguyen, X., Garofalakis, M., Jordan, M.I., Joseph, A., & Taft, N. (2006). In-network PCA and anomaly detection. *NIPS'06 Proceedings of the 19th International Conference on Neural Information Processing Systems*, 617-624.

Lakhina, A., Crovella, M., & Diot, C. (2004). Diagnosing network-wide traffic anomalies. *Computer Communication Review, 34(4)*, 219-230.

León, C. & Berndsen, R. (2014). Rethinking financial stability: challenges arising from financial networks' modular scale-free architecture, *Journal of Financial Stability, 15*, 241-256.

Martínez, W.L., Martínez, A.R., & Solka, J. (2011). *Exploratory Data Analysis with Matlab*. CRC Press: Boca Ratón.

McNeil, A.J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press: Princeton.

Newman, M.E.J. (2010). Networks: An Introduction. Oxford University Press: New York.

Ng, S.C. (2017). Principal component analysis to reduce dimension on digital image. *Procedia, 111*, 113-119.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*, 559-572.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25(3)*, 289-310.

Sree, A. & Venkata, K. (2014). Anomaly detection using Principal Component Analysis. International. *Journal of Computer Science and Technology, 5(4)*, 124-126.

Thottan, M., Liu, G., Ji, C. (2010). Anomaly Detection Approaches for Communication Networks. In: Cormode, G. & Thottan, M. (eds.) *Algorithms for Next Generation Networks*. Computer Communications and Networks. Springer: London.

Triepels, R., Daniels, H., & Heijmans, R. (2017). Anomaly detection in real-time gross settlement systems. *Proceedings of the 19th International Conference on Enterprise Information Systems (ICEIS 2017), Vol.1*, 433-441. [doi:10.5220/0006333004330441]

Witten, I.H., Frank, E., & Hall, M.A. (2011). *Data Mining*. Morgan Kaufmann: Burlington.

8 Appendix

Payments Network (hypermatrix)		Mean	Standard deviation	Skewness	Kurtosis	K-S test *p*-value[c]
Bootstrap[a]		2.25	0.07	0.72	4.04	0.00
Filtered[b]		**4.70**	**0.73**	**1.89**	**8.84**	**1.00**
Inability to make payments	Synthetic Bank A[b]	5.55	0.92	2.09	10.49	0.00
	Synthetic Bank B[b]	4.96	0.67	1.47	6.34	0.00
	Synthetic Bank C[b]	5.10	0.81	2.07	10.11	0.00
Zero initial balance	Synthetic Bank A[b]	4.91	0.81	2.17	10.76	0.00
	Synthetic Bank B[b]	4.68	0.65	1.32	5.82	0.00
	Synthetic Bank C[b]	4.78	0.79	1.99	9.71	0.00

Table A1. Reconstruction errors' distribution, with $k = 3$. [a] Estimated on the 1,000 samples obtained by bootstrap from the filtered hypermatrix. [b] Estimated on the 436,000 errors obtained by reconstructing the 436 observations in the corresponding hypermatrix with the 1,000 estimations of $\overline{\overline{\Gamma}}$. [c] Two-sample Kolmogorov-Smirnov test of distribution equality, with respect to the reconstruction errors from the filtered hypermatrix.

Payments Network (hypermatrix)		Mean	Standard deviation	Skewness	Kurtosis	K-S test *p*-value[c]
Bootstrap[a]		1.81	0.06	0.80	5.07	0.00
Filtered[b]		**4.25**	**0.66**	**1.99**	**9.70**	**1.00**
Inability to make payments	Synthetic Bank A[b]	5.14	0.84	2.10	10.75	0.00
	Synthetic Bank B[b]	4.55	0.61	1.49	6.64	0.00
	Synthetic Bank C[b]	4.68	0.74	2.14	10.71	0.00
Zero initial balance	Synthetic Bank A[b]	4.52	0.75	2.19	11.17	0.00
	Synthetic Bank B[b]	4.29	0.59	1.36	6.14	0.00
	Synthetic Bank C[b]	4.38	0.72	2.05	10.20	0.00

Table A2. Reconstruction errors' distribution, with $k = 5$. [a] Estimated on the 1,000 samples obtained by bootstrap from the filtered hypermatrix. [b] Estimated on the 436,000 errors obtained by reconstructing the 436 observations in the corresponding hypermatrix with the 1,000 estimations of $\overline{\overline{\Gamma}}$. [c] Two-sample Kolmogorov-Smirnov test of distribution equality, with respect to the reconstruction errors from the filtered hypermatrix.

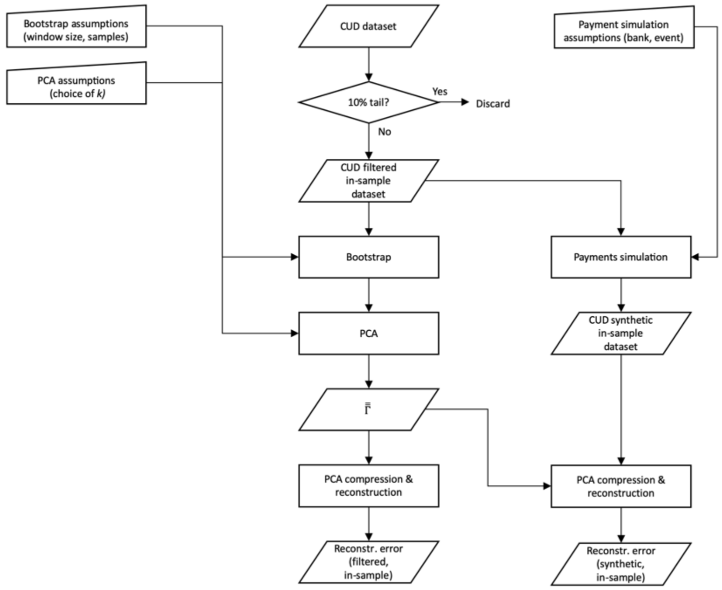


Figure A1. Obtaining reconstruction errors for filtered and synthetic payments networks. Source: author's design.

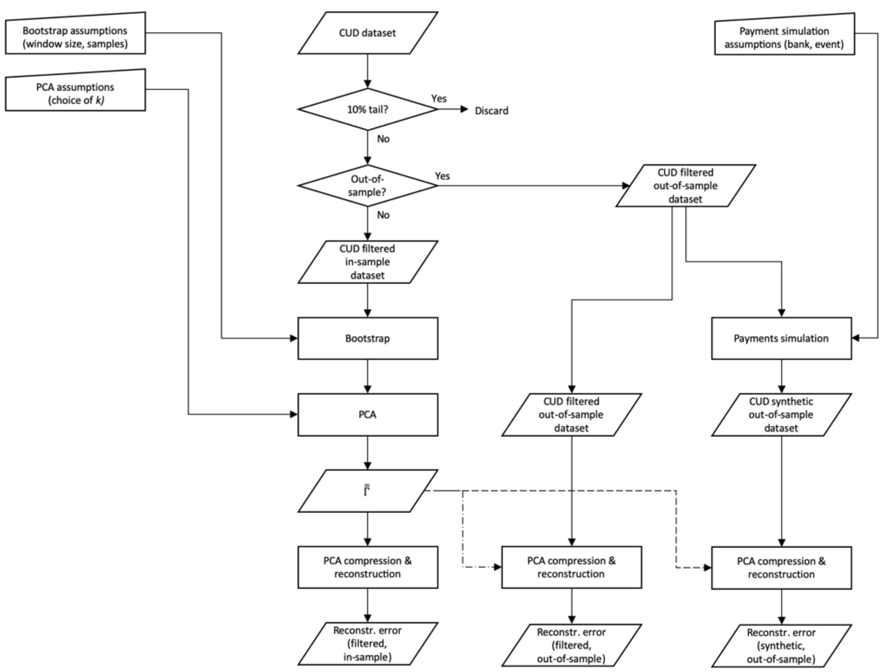


Figure A2. Obtaining in-sample and out-of-sample reconstruction errors for filtered and synthetic payments networks. Source: author's design.

Target (actual) class		Output (predicted) class: *Normal*	Output (predicted) class: *Anomalous*
	Normal	a. 0.07 b. 0.08 c. 0.08 d. 0.08 e. 0.08	a. 0.02 b. 0.01 c. 0.01 d. 0.01 e. 0.01
	Anomalous	a. 0.57 b. 0.66 c. 0.57 d. 0.56 e. 0.54	a. 0.34 b. 0.25 c. 0.34 d. 0.35 e. 0.37

a. Subset #1. (top-5 institutions, zero initial balance or inability to make payments; 440,000 simulations)

Target (actual) class		Output (predicted) class: *Normal*	Output (predicted) class: *Anomalous*
	Normal	a. 0.14 b. 0.16 c. 0.16 d. 0.16 e. 0.15	a. 0.03 b. 0.01 c. 0.01 d. 0.01 e. 0.01
	Anomalous	a. 0.37 b. 0.44 c. 0.27 d. 0.29 e. 0.29	a. 0.46 b. 0.39 c. 0.56 d. 0.55 e. 0.54

b. Subset #2. (top-5 institutions, inability to make payments; 240,000 simulations)

Target (actual) class		Output (predicted) class: *Normal*	Output (predicted) class: *Anomalous*
	Normal	a. 0.22 b. 0.24 c. 0.24 d. 0.24 e. 0.24	a. 0.03 b. 0.02 c. 0.01 d. 0.01 e. 0.01
	Anomalous	a. 0.27 b. 0.28 c. 0.21 d. 0.22 e. 0.22	a. 0.48 b. 0.47 c. 0.54 d. 0.53 e. 0.53

c. Subset #3. (top-3 institutions, inability to make payments; 160,000 simulations)

Table A3. Confusion matrices. Percentage of simulations. Calculations based on (a-) $k = 1$, (b.) $k = 3$, (c.) $k = 5$, (d.) $k = 7$, (e.) $k = 9$.


BANCO DE LA REPÚBLICA
COLOMBIA