Application of isobars to stock market indices

Kristýna Ivanková[1]

Abstract. Isobar surfaces, a method for describing the overall shape of multidimensional data, are estimated by nonparametric regression and used to evaluate the efficiency of selected markets based on returns of their stock market indices.

Keywords: Isobars, Efficient market hypothesis, Nonparametric regression, Extreme value theory

1 Introduction

This article describes the isobar surfaces approach, which finds a one-dimensional ordering of multidimesional data, and surfaces in $\mathbb{R}^d$ that enclose the u-th quantile of the data distribution according to the one-dimensional ordering.

An isobar maps every direction to a particular distance from a center (specified by the quantile function). The resultant surface for a fixed quantile u is also called an isobar. The article [4] contains the first definition of isobars. Both authors of [4] extend their work in further articles in different ways.

M. F. Barme-Delcroix continued to study their theoretical properties and their connections with single-dimensional extreme value theory. In [1] the stability of multivariate intermediate order statistics is discussed. The article [2] studies multidimensional outlier-prone and outlier-resistant distributions and [7] extends the theory of outlier-proneness. In [3], limit laws for multidimensional extremes are studied with the usage of the one-dimensional Fisher-Tippett theorem. P. Jacob, the second autor of [4], focused on practical estimation of the isobar shapes. The article [5] focuses on estimation of the edge of the bounded support using nonparametric regression and [6] extends this method for unbounded support using asymptotical location and isobars.

In our article, we use the shape of the isobar surfaces to test the *efficient market hypothesis* stating that returns of efficient stock market indices have the behaviour of Brownian motion. The hypothesis is stated e.g. in [8]. Various reasons for deviations from this hypothesis for otherwise efficient markets were discussed in the literature.

The first part is concerned with theory and estimation of isobars. In the second part we perform a simulation study for Gaussian distribution with different parameters and an assessment of the shape and stability of the resulting isobars. In the third part we'll estimate isobars for returns of stock market indices (the NASDAQ Composite Index and the PX Index) and their lagged values and evaluate the efficient market hypothesis for each index. Finally, we summarize the results and outline future progress.

2 Isobar surfaces and their estimation

Isobars are defined in generalized polar coordinates, so a coordinate transformation of data is required beforehand. The transformation of a non-zero vector $\mathbf{x} \in \mathbb{R}^d$ to generalized polar coordinates is

$$r = \|\mathbf{x}\|_2 , \quad \theta = \frac{\mathbf{x}}{\|\mathbf{x}\|_2},$$

where $\|\mathbf{x}\|_2$ is the Euclidean norm of the vector $\mathbf{x}$. Observe that the generalized angle θ lies on S^{d-1}, the sphere of unit radius in $\mathbb{R}^d$.

We'll use the definition of isobar as it appears in [4], pp. 2. For every $u \in (0, 1)$, the u-level isobar is defined as a mapping of a fixed θ to the value of the inverse distribution function of the Euclidean distance

[1]Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08, Prague 8, Czech Republic

from the origin: $\theta \to F^{-1}_{R\,|\,\Theta}(u)$. The name "$u$-level isobar" will also be used interchangeably for the surface $S_u = F^{-1}_{R\,|\,\Theta}(u)$ determined by each θ with a fixed quantile u in the inverse of the conditional distribution function $F^{-1}_{R\,|\,\Theta}$. For this definition and the estimation method described later to be valid, the random variable $X = (R, \Theta)$ whose multidimensional realizations comprise our sample needs to satify certain requirements. We assume continuity of the mariginal density $f_\Theta(\theta)$, conditional density $f_{R\,|\,\Theta}(r\,|\,\theta)$ and the conditional distribution function $F_{R\,|\,\Theta}(r\,|\,\theta)$. We also need the distribution function to be a bijection so that its inverse exists. The introduced mapping is assumed to be continuous and strictly positive.

A description of the ordering of multidimensional data by quantiles follows. Consider a sample of n independent realizations of the random variable X, e.g. $X_i = (R_i, \Theta_i)$, $1 \le i \le n$. For every i there exists an unique u_i-level isobar containing the point X_i. Denoting $X_{i,n}$ the realizations ordered by their respective quantile values u_i, the maximum value is given by the point $X_{n,n}$ which belongs to the upper-level isobar with level $\max_{1\le i\le n} u_i$. In practice, we'll assess the 1-level isobar on the grounds of the *asymptotical location property* as described in [6]. For large n, the furthest points from the origin lie near the $\frac{n-1}{n}$-level isobar. The 1-level isobar is then simply the edge of the bounded support. Citing Definition 3 from [6], pp. 175:

> The distribution of r.v. X on $\mathbb{R}^d$ is said to have the asymptotical location property if a.s. for each $\epsilon > 0$ and each sample X_i, $1 \le i \le n$, with the same distribution as X and with size $n \ge n_0 = n_0(\epsilon, \omega)$: $\inf_{x\in S_{(n-1)/n}} \mathrm{dist}(x, X_{n,n}) \le \epsilon$ and $\sup_{x\in S_{(n-1)/n}} \min_{1\le i\le n} \mathrm{dist}(x, X_i) \le \epsilon$.

Isobar estimation is performed by the non-parametric regression of [5, 6]. For the estimation we'll assume *homotheticity* of isobars, e.g. for some strictly positive continuous function $v(\theta)$ and a distribution function G,

$$F_{R\,|\,\Theta}(r\,|\,\theta) = G\left(\frac{r}{v(\theta)}\right) \quad \text{for } r \in [0, v(\theta)].$$

The function $v(\theta)$ corresponds to the 1-level isobar and unambiguously describes the shape of all isobars. The distribution of $\frac{\mathbf{x}}{v(\theta)}$ is spherically symmetric and it can be fully described by G on $[0, 1]$.

We estimate $v(\theta)$ using radial regression:

$$w(\theta) = E(R\,|\,\Theta = \theta) = \int_0^{v(\theta)} 1 - G\left(\frac{r}{v(\theta)}\right) dr = c\, v(\theta),$$

where c is the expected value of G. The estimate of the expected value of R given $\Theta = \theta$ describes the shape of 1-level isobar up to a multiplicative constant. This constant is chosen in a way that the estimated expected value shape $\hat{w}(\theta)$ contains the whole data after scaling:

$$\hat{v}(\theta) = \frac{\hat{w}(\theta)}{\hat{c}}, \quad \text{where } 1/\hat{c} = \max_{1\le i\le n} \tfrac{R_i}{\hat{w}(\Theta_i)}.$$

The original method in [5] performs non-parametric regression on data transformed into hyperspherical coordinates (r, φ), resulting in the estimate of $w(\varphi_1, \ldots, \varphi_{d-1})$. This parametrization, however, suffers from pole singularities in higher dimensions ($d > 2$), which hurts non-parametric regression. Therefore we propose to estimate $w(\theta)$ in the domain $(r, \mathbf{x})$ after projecting the data on the unit sphere S^{d-1} and adding r as an extra coordinate. The estimate of $w(\theta)$ then corresponds to the estimate of $w(\mathbf{x})$ constrained to $\mathbf{x} \in S^{d-1}$. This method is a little slower due to the extra coordinate, but doesn't suffer from degeneracies. In the following, we'll use and evaluate both methods.

3 Simulation study – normal distribution

Since we assume Brownian motion for data obtained from efficient markets, we need to know how non-parametric isobar shape estimation behaves in the ideal case of normal distribution. Computations were performed in the `R` environment for statistical computing. For multidimensional non-parametric regression the `np` package was used. The best results were obtained by choosing locally-weighted linear regression, Gaussian kernels and k-nearest-neighbour kernel bandwidth estimation.

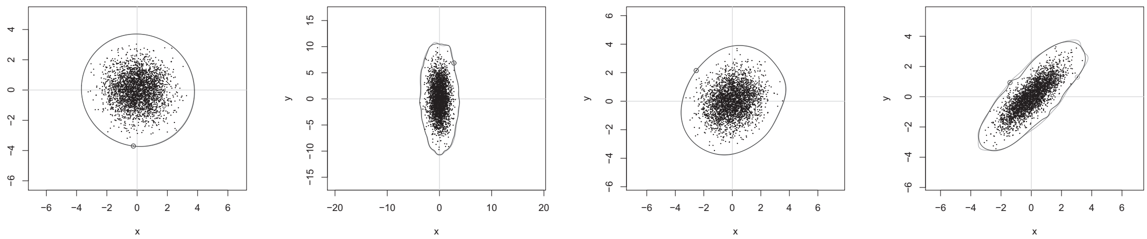

Figure 1: Isobars for Gaussian distibution samples with $n = 3000$ and covariance matrices $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}$, $\begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$.

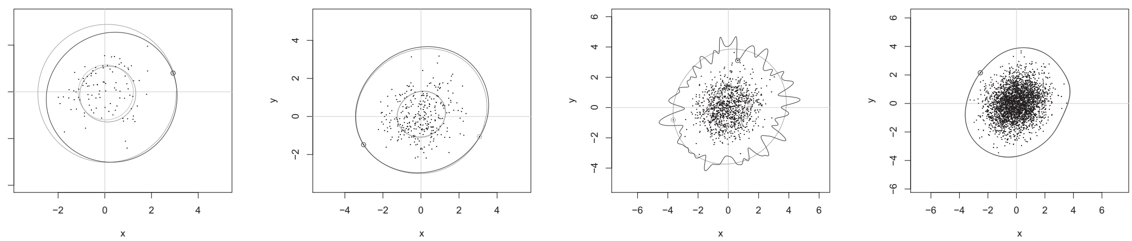

Figure 2: Isobars for Gaussian distibution samples with covariance matrix $\begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$ and $n \in \{100, 300, 1000, 3000\}$.

We've estimated isobar shapes of two-dimensional Gaussian distribution samples with zero mean and several covariance matrices. We've tested various sample sizes (n: 100, 300, 1000 or 3000) from distributions with diagonal covariance matrices (variances (1, 1) or (1, 9)) and from distributions with non-zero covariances (variances: (1, 1), covariances: 0.2 or 0.8). An excerpt of the results is shown in Figures 1 and 2. Inner shapes represent the expected value estimation $\hat{w}(\theta)$, outer shapes represent the estimation of the 1-level isobar $\hat{v}(\theta)$. The result for the hyperspherical parametrization of [5] is gray while the result of the proposed projection approach is black. The farthest point for each parametrization is highlighted. The obtained isobar shapes started resembling circles (uncorellated marginals with equal variances) and ellipsoids (unequal variances or corellated marginals) around sample sizes 300 and 1000.

The most time-consuming part of estimation is bandwidth selection. Both parametrizations get stuck in local minima – the hyperspherical parametrization often averages all data ($k = n$), while the projection approach has better details, but may overfit (k too small). We solve the problem by allowing the algorithm to take multiple restarts of the randomized bandwidth search, but for larger n, the speed of this approach can become prohibitive. We've chosen 5 and 10 restarts for the hyperspherical and projection approach, respectively. Since the software doesn't support periodic domains needed in the hyperspherical parametrization, we had to emulate this functionality by placing copies of the data along multiples of 2π. This has slowed the estimation to the level of the second parametrization, which needs an extra coordinate. The results have shown that both parametrizations perform equally well, at least in the case of $d = 2$.

4 Application – stock market indices

After the choice of methods of isobar shape assessment we'll proceed to their application to stock market index returns. The efficient market hypothesis states that returns (closing−opening price) of market indices in efficient markets follow Brownian motion (see e.g. [8]). In practice, this assumption is mostly violated by the periodic structure (day, week, quarter, year) of agent behaviour. Further bias mostly reveals non-rational behaviour, non-zero information costs or delayed reactions. Our goal is to measure the efficiency of a market using isobar shapes.

Our data consists of weekly closing and opening prices for the past ten years (sample size around 500) obtained from the Reuters Wealth Manager service. The y-axis denotes the current value of stock market index returns, the x-axis denotes their lagged values. Under the efficient market hypothesis, the isobar shape for this configuration should be close to a circle (since Brownian motion is independent with itself when lagged). From all the studied stock market indices we've selected two. The first one is the NASDAQ Composite Index, comprised of 2742 stocks of the NASDAQ Stock Market. The other one is

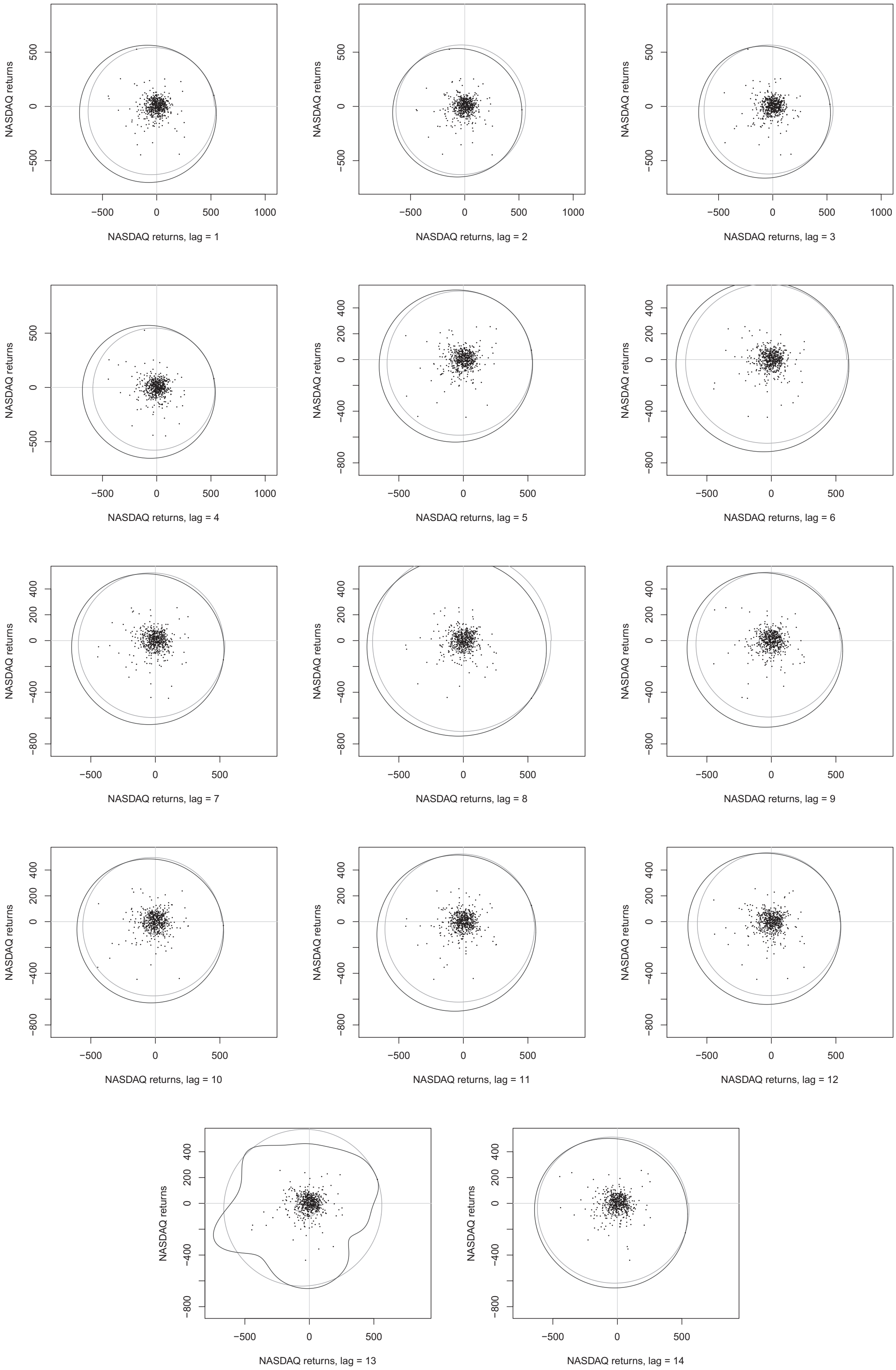

Figure 3: Lags of 1–14 weeks for the NASDAQ Composite Index.

the PX Index comprised of 14 stocks of the Prague Stock Exchange (only five of which are Czech). We've studied isobar shapes for lags between one and fifteen weeks.

Isobar shapes for the NASDAQ Composite Index (Figure 3) are very close to circles except for the 13-week lag, which can be explained by the expected quarterly periodicity of agent behaviour. Based on visual examination, the market of NASDAQ may follow the efficient market hypothesis. On the other hand, the isobar shapes of the PX Index (Figure 4) deviate from circles for longer lags (of 4, 7, and 10–14 weeks). This shows that the efficient market hypothesis doesn't apply to the market described by the PX Index. The wild shape of the PX Index two-week lag might also be due to a local minimum of the bandwidth – the five-fold increase of the number of restarts might not be sufficient yet (indices not shown here exhibited similar shapes also for the hyperspherical parametrization). This can be resolved by a further increase of the number of restarts, or by a change of the search algorithm. Another possibility is to switch to a different software package.

Our future goal is to create a measure of market efficiency. Since this can be formulated as similarity of the isobar shape to a circle, this measure can be based on the number of neighbors included in the kernel during bandwidth selection.

5 Conclusion

We've investigated the possibilities of using the isobar surfaces approach with homothetic isobars for both simulated and real data. In the simulation study, we've found suitable methods for estimating nonparametric regression, the sample size needed for the application of our methods, and isobar shapes for Gaussian distrubusions with varying parameters. This knowledge was applied during the assessment of the efficient market hypothesis using isobar shapes. We've assessed the isobar shape for stock market index returns. For the NASDAQ Composite Index the shapes supported the efficiency hypothesis, for the PX Index we can reject the hypothesis. During the estimation of the isobar shape by nonparametric regression we've encountered problems during the automated bandwidth selection – even with an increased number of parameter search restarts, the search still stays in local minima. The problem can be solved by changing to a different implementation of nonparametric estimation. This will allow us to focus on finding an objective measure of market efficiency in our future work.

Acknowledgements

This work has been supported by the Czech grant agency grant 402/09/H045.

References

[1] Barme-Delcroix, M. F., and Brito, M.: Multivariate stability and strong limiting behaviour of intermediate order statistics. *J. Multivariate Anal.* **79** (2001), 157–170.

[2] Barme-Delcroix, M. F., and Gather, U.: An isobar surfaces approach to multidimensional outlier-proneness. *Extremes* **5** (2002), 131–144.

[3] Barme-Delcroix, M. F., and Gather, U.: Limit laws for multidimensional extremes. *Stat. Prob. Lett.* **77** (2007), 1750–1755.

[4] Delcroix, M. F., and Jacob, P.: Stability of extreme value for a multidimensional sample. *Stat. Analyse Donées* **16** (1991), 1–21.

[5] Jacob, P., and Suquet, C.: Regression and edge estimation. *Stat. Prob. Lett.* **27** (1996), 11–15.

[6] Jacob, P., and Suquet, C.: Regression and asymptotical location of a multivariate sample. *Stat. Prob. Lett.* **35** (1997), 173–179.

[7] Martins, A. P., Ferreira, H., and Pereira, L.: Multidimensional outlier-proneness of dependent data and the extremal index. *Statistical Methodology* **5** (2008), 72–82.

[8] Osborne, M. F. M.: Brownian motion in the stock market. *Operations research* **March-April 1959**, 145–173.

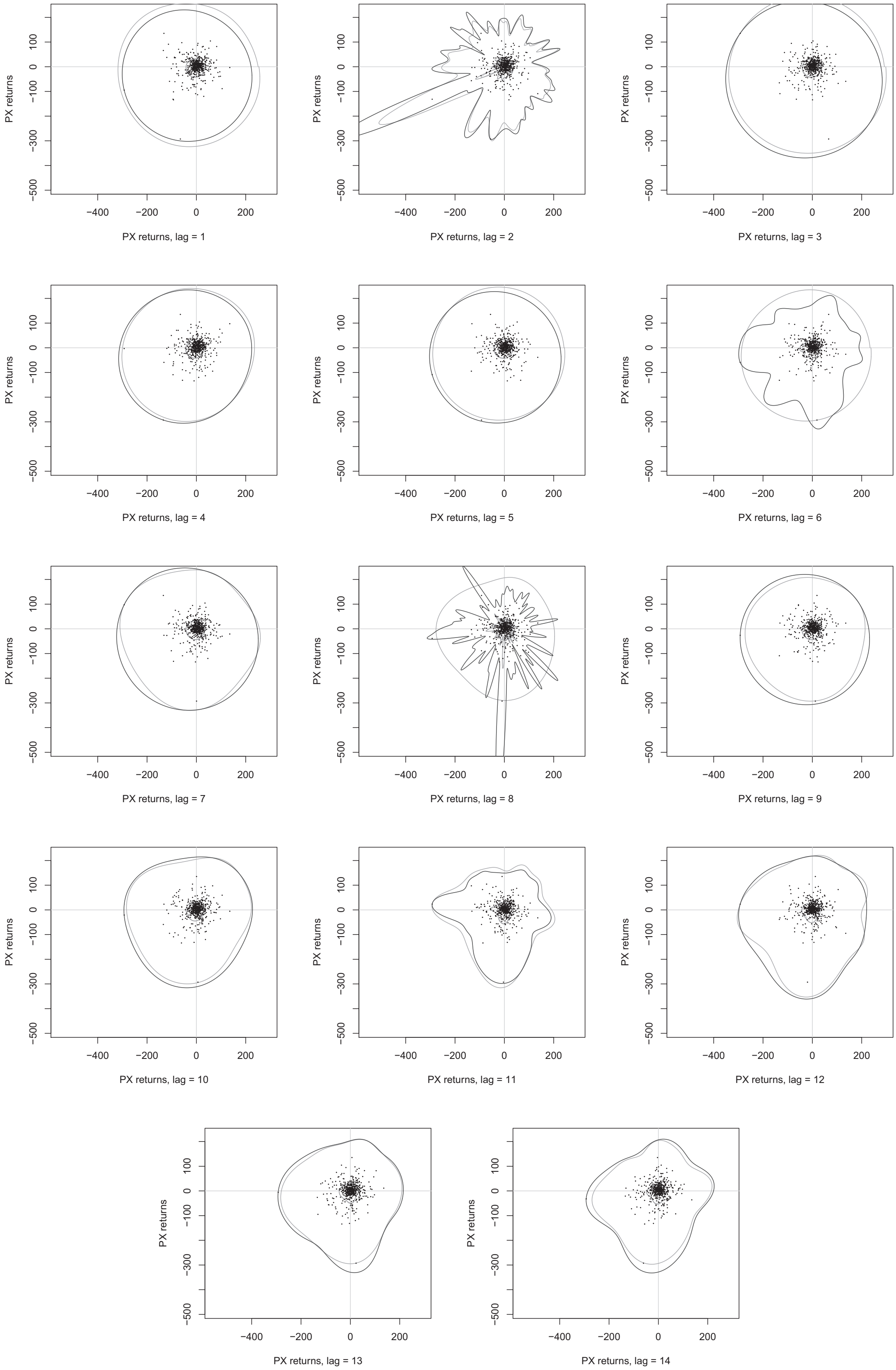

Figure 4: Lags of 1–14 weeks for the PX Index.