ESTIMATION OF TIME-VARYING KERNEL DENSITIES AND CHRONOLOGY OF THE IMPACT OF COVID-19 ON FINANCIAL MARKETS

A PREPRINT

Hugo Soleau
Université Paris Dauphine-PSL
hugo.soleau@dauphine.eu

Paul-Antoine Charbit
Université Paris Dauphine-PSL
paul-antoine.charbit@dauphine.eu

Elio Mario Sawaya
Université Paris Dauphine-PSL
elio-mario.sawaya@dauphine.eu

Ali Ahmed Bachir CHACHA
Université Paris Dauphine-PSL
ali-ahmed-bachir.chacha@dauphine.eu

February 1, 2024

ABSTRACT

This article is a reproduction of the study carried out by Garcin et ali [1] on the estimation of the return distributions of various indices.
We study estimation methods based on approximation by time-varying kernel densities. The aim of these methods is to estimate dynamic distribution laws for series of data from non-parametric methods. The main challenge is to build a criterion so as to minimize dependence statistics over the filtering parameters (bandwidth & discount factor) rather than carrying out a series of statistical and graphical tests.
To do this, we first look at estimating the filtering parameters of our data series. We then apply diverse divergence statistics on our estimated densities/distributions, in order to find a financial chronology of the covid crisis. Finally, we analyse the difficulties encountered in reproducing the paper and present the various extensions we have considered or implemented.

1 Introduction

Predicting returns is a broad topic in quantitative finance that has been studied by many researchers and professionals. In particular, the various players in the research field have gradually taken an interest in methods for estimating the distribution of returns on different stock prices. Over the last few years, some of these methods have demonstrated their robustness and plausibility: kernel density estimation methods.
In particular, dynamic kernel density estimation methods enable the modelling of distribution laws that fluctuate over time using a non-parametric approach and allowing certain assumptions to be made through the relationships between returns and distributions described in the dynamics.
For this reason, in their article "Estimation of time-varying kernel densities and chronology of the impact of COVID-19 on financial markets", Garcin et al look at these methods in order to determine the best possible distributions for series of returns through the estimation of filtering parameters h and ω which will respectively denote the bandwidth and discount factor.
In turn, we attempt to retrace these methods by looking at the various statistics defined and used, which we apply to the case of the COVID-19 crisis.

First, we recall that in non-parametric statistics, Kernel Density Estimation (KDE) can be viewed as an extension of histograms in a continuous world, where a kernel is assigned to each data point. In the classical framework,

the estimation is static, and the density is simply given by:

$$\hat{f}(x) = \frac{1}{th} \sum_{i=1}^{t} K(\frac{x - X_i}{h}) \qquad (1)$$

In the paper and in our reproduction, the used kernel function is an Epanechnikov kernel given by :

$$K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{(|u| \leq 1)}$$

In this framework, the parameter h can be viewed as a regularization parameter: the larger it is, the smoother the density and the better its generalization to new data. Conversely, as $h \to 0$, each kernel becomes a Dirac delta function centered at its data point.

Parameter selection in time-varying kernel densities is even more challenging since it introduces another free parameter: the smoothing between densities at different dates. In fact, the new method introduced and its variant are CDF-based while traditional MLE method are PDF-based. The aim here is not to obtain the parameters which best fit the data, but to obtain the parameters that give the best description of the true densities, i.e. we are interested in the parameters that guarantee the best uniformity and independence of the PIT.

The methodology of our study is organized into three distinct sections. The first section dives into the innovative technique of dynamic kernel density estimation and its various iterations, contrasting these with the more traditional maximum likelihood estimation (MLE) based methods. The second section is dedicated to exploring different measures of divergence, providing detailed explanations for each. Finally, the third section presents an empirical application of our methodology, specifically focusing on stock markets during the COVID-19 pandemic. This section culminates in a comprehensive presentation of our findings, highlighting the significant contributions of our approach to understanding the intricate dynamics influenced by the financial crisis.

Notations

- X is the distribution of the return it could be parametric or non-parametric
- X_i is a scalar representing the i-th observation return
- X_t is a scalar representing the t-th daily price
- h is is a scalar representing the bandwidth parameter
- w is is a scalar representing the discount parameter
- $Z_t^{h,w}$ is PIT
- $\hat{F}_t^{h,w}$ is the estimated cumulative distribution function of returns
- $\hat{f}(x)$ is the estimated static density
- $\hat{f}_t(x)$ is the estimated dynamic density at the t-th period

2 Model and estimation methods

The use of kernel densities is a common feature of estimation in non-parametric cases, since it makes it possible to estimate the probabilistic behaviour of observed and measured phenomena independently of any factor linked to the distribution.

This is why, following the example of the work carried out by Garcin et al[1], we have reproduced estimates based on dynamic kernel densities. The densities we use are exponentially weighted such that recent data have more weights than other, this statement is true either for the first estimation (from date 0 to date t_0) and for the subsequent ones ($t_0 + 1$ to T). The rational behind this is quite straightforward since it is well-known that the market is not static (succession of bullish, bearish and neutral phases).

Therefore we give the various equations we used that are given in [1]:

$$\hat{f}_{t+1}^{h,w}(x) = w\hat{f}_t^{h,w} + \frac{1 - w}{h} K(\frac{x - X_{t+1}}{h}) \qquad (2)$$

$$\hat{F}_{t_0}^{h,w}(x) = \frac{1-w}{1-w^{t_0}} \sum_{i=1}^{t} w^{t_0-i} \mathcal{K}\left(\frac{x-X_i}{h}\right) \qquad (3)$$

with $\mathcal{K}$ the primitive of K and its limit at infinity worth 1.

$$W_{i,t} = \begin{cases} \frac{1-w}{1-w^{t_0}} w^{t-i} & \text{if } i \leq t_0, \\ (1-w)w^{t-i} & \text{else.} \end{cases} \qquad (4)$$

Also, by integrating equation 2, we obtain the cdf function corresponding to time t+1:

$$\hat{F}_{t+1}^{h,w}(x) = w\hat{F}_t^{h,w}(x) + (1-w)\mathcal{K}\left(\frac{x-X_{t+1}}{h}\right)$$

The purpose of a Kernel is then to perform a continuous generalization of a histogram within which we count the number of occurrences in given intervals. The thinner the intervals, the more precise the density estimate will be. The proximity function K in (2) thus reaches its maximum at 0 and gradually decreases as $|x \breve{} X_t|$ moves away from 0. The impact of X_t on the density will therefore be maximal for x = X_t. Also, the larger h is, the wider the Kernel will be, thus widening the interval on which each observation has an impact.

The authors presented two PIT based models, each one with different versions for the parameters bounds and the PITs taken into account. Both involves the minimization of an adapted Kolmogorv-Smirnov statistic.

The first one is:

$$d_v(Z_{t_0+1}^{h,w}, ..., Z_T^{h,w}) = \underset{0\leq\tau\leq v}{max} (\sqrt{T-\tau-t_0} \times k'_\tau(Z_{t_0+1}^{h,w}, ..., Z_T^{h,w}) \qquad (5)$$

Where k is

$$k(Z_{t_0+1}^{h,w}, ..., Z_T^{h,w}) = \begin{cases} \underset{t_0+1\leq s\leq T}{max} |Z_s^{h,w} - \frac{1}{T-t_0+1} \sum_{u=t_0+1}^{T} \mathbf{1}_{[0,Z_s^{h,w}]}(Z_u^{h,w})| & \text{if } \tau = 0 \\ \underset{t_0+1\leq s\leq T-\tau}{max} |Z_s^{h,w} Z_{s+\tau}^{h,w} - \frac{1}{T-\tau-t_0+1} \sum_{u=t_0+1}^{T-\tau} \mathbf{1}_{[0,Z_s^{h,w}]}(Z_u^{h,w}) \mathbf{1}_{[0,Z_{s+\tau}^{h,w}]}(Z_{u+\tau}^{h,w})| & \text{else.} \end{cases} \qquad (6)$$

The reason behind this is quite simple: the PITs have to be independent in order to avoid to have globally uniform PITs, but low in bearish markets and high in bullish ones. In fact, this approach only test the pairwise independence of the PITs, and this is for computational comfort.

This approach is divided into three distinct sub-methods.
The first sub-method constrains the bandwidth parameter h to be greater than 0, while the discount factor w is bounded between 0 and 1. The optimal bandwidth and discount factor will then be the parameters which minimize the uniformity and divergence statistics:

$$(h^*, w^*) = \underset{h>0,0<w\leq 1}{argmin}\ d_v(Z_{t_0+1}^{h,w}, ..., Z_T^{h,w})$$

The second sub-method modifies the lower bound of h to be $1 - v^{-1}$ (the rationale being that a density cannot change too rapidly; in essence, with a minimum $w = 1 - v^{-1}$, therefore $1 - w = v^{-1}$. If $v = 22$, as utilized in the article, this implies that at a minimum, a density will be completely modified after 22 days, thereby smoothing the transition. This constrained version is as such :

$$(h_c^*, w_c^*) = \underset{h>0,1-v^{-1}<w\leq 1}{argmin}\ d_v(Z_{t_0+1}^{h,w}, ..., Z_T^{h,w})$$

The final method, known as the 'censored' approach, aims to accentuate the extreme Probability Integral Transforms (PITs) in order to better fit the fat tails of the distribution. Here, the divergence statistic will be calculated only for the PITs that are lower than ρ or higher than $1 - \rho$. In the paper, ρ is set to 0.05.

$$D_\nu(Z_{h,\omega}^{t_{0+1}}, \ldots, Z_{h,\omega}^{T}) = \max_{t_{0+1}\leq s<s+\nu-1\leq t\leq T} \left(\sqrt{t-s+1} \times k(Z_{h,\omega}^{s}, \ldots, Z_{h,\omega}^{t})\right)$$

The PIT based approach involves testing the uniformity over each sub-interval greater than ν. Consequently, the statistic D_ν is designed to calculate the uniformity of each sub-interval within the range $[t_0 + 1, T]$ with sizes exceeding ν. Therefore, it does not directly test the independence of the PITs, but indirectly does so by assessing the uniformity of

each sub-interval greater than ν. The various constraints can also be applied here.

Harvey and Oryshchenko pioneered the application of maximum likelihood for selecting the parameters h and ω. The primary objective of this approach is to identify parameters that maximize the likelihood of observing a given outcome, given the data at hand. It is essential to acknowledge that this method incorporates the dynamic nature of the density, and thus of its estimation.

$$(h^{HO}, \omega^{HO}) = \underset{h>0, 0<\omega\leq 1}{argmax} \sum_{s=t_0}^{T-1} \log\left(\hat{f}_s^{h,\omega}(X_{s+1})\right)$$

It is interesting to have a look at this new approach in order to monitor whether the initial recommended method using PITs is efficient enough to asses a good forecast quality.

To evaluate the relevance of the proposed criterion, the authors suggest generating 2000 random variables following a Cauchy distribution with a time varying location parameter ($t/100$) and scale of 1. Subsequently, they estimate the densities on these generated data. The statistical analysis is then conducted by comparing these estimated densities with the true distribution, which is known in this context (since data are simulated).

$$f_t(x) = \frac{1}{\pi\left[1 + (x - \frac{t}{100})^2\right]}$$

This law lies in the high frequency of large values, thus moving away from observations made upstream. The likelihood method must therefore be smoothed to reconcile observations with high values while avoiding excessively low likelihoods.

The parameters estimated through the PIT method provide a better overall fit to the true distribution, yet the estimated density is less smooth, exhibiting bumps in the tails, attributable to an excessively low discount factor. In comparison to the MLE method, we get these values for the parameters h and w :

$h^* = 0.488 vs h^{HO} = 0.596 and w^* = 0.902 vs w^{HO} = 0.989$

The MLE method yields a smoother probability density function (pdf), but its center is shifted, a phenomenon that can be explained by its tendency to smooth out the left-side bump. Conversely, the censored version, which only considers extreme PITs in the minimization of d_ν, enhances the smoothing of the estimates, but it too introduces a leftward shift in its center. Moreover, it features even flatter shape than the density estimated via MLE parameters, which is already more flattened compared to the true density.

In conclusion, the PIT method holds a significant advantage over the MLE approach due to its greater flexibility. For instance, we could negotiate a trade-off between the center fit and tail smoothing by varying the exclusion criterion of the central PITs (ρ).

However, the initial approach with the PITs method gives us a better estimate of dynamic densities than the maximum likelihood method, but only for the Kolmogorov-Smirnov statistic. The PITs method would therefore benefit from being adapted to other statistics for greater efficiency.

3 The different divergence statistics

The research focus on 4 divergence statistics, half of them being based on CDFs while the 2 others on Pdfs.

The Kolmogorov-Smirnov statistic is given by:

$$k(\hat{F}_t, \hat{F}_{t_0}) = \underset{x\in\mathbb{R}}{sup} |\hat{F}_t(x) - \hat{F}_{t_0}(x)|$$

And is defined as the maximum divergence between the empirical CDF and the theoretical oen (here drawn form uniform distribution).

The p-Wasserstein distance (can be seen as a generalization of the Kolmogorov-Smirnov statistic) which for $p \geq 1$ represents the minimum cost to reconfigure one pdf into another. Where the Kolmogorov-Sirnov statistic consists of the distance L^∞ between the distribution functions, the Wasserstein distance is the distance L^p between the quantiles of the CDFs. In our case, we will take p=1 (i.e. distance L^1 between the CDFs).

$$W_p(\hat{F}_t, \hat{F}_{t_0}) = \left(\int_0^1 |\hat{F}_t^{-1}(\alpha) - \hat{F}_{t_0}^{-1}(\alpha)|^p d\alpha \right)^{\frac{1}{p}}$$

Thus the 1-Wasserstein statistic will resumed to the cumulative absolute divergence between two CDFs as follows :

$$W_1(\hat{F}_t, \hat{F}_{t_0}) = \left(\int_{\mathbb{R}} |\hat{F}_t(x) - \hat{F}_{t_0}(x)| dx \right)$$

The Hellinger distance which represents the cumulative difference between densities as opposed to distribution functions :

$$H(\hat{f}_t, \hat{f}_{t_0}) = \sqrt{\frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{\hat{f}_t(x)} - \sqrt{\hat{f}_{t_0}(x)} \right)^2 dx}$$

And finally, we can focus on the Kullback-Leibler divergence which is linked to the Shannon entropy and which is not a distance function like the previous ones since it is non-symmetric.

$$D_{KL}(\hat{f}_t || \hat{f}_{t_0}) = \int_{\mathbb{R}} \hat{f}_t(x) log \left(\frac{\hat{f}_t(x)}{\hat{f}_{t_0}(x)} \right) dx$$

The Kullback-Leibler (KL) divergence and Hellinger distance measures dissimilarity between two probability distributions differently. KL divergence quantifies information loss when using the distribution measured in t0 to approximate the one measured in t. It is the expectation (under f_t) of the log-difference between the two distributions. In contrast, the Hellinger distance provides a symmetric metric of the geometrical (it respect mathematical properties of a distance: triangular inequality and symmetry) gap between distributions. It is bounded between 0 and 1, where 0 indicates identical distributions and 1 denotes total disjointness.

In fine, while KL divergence is suited for information-centric problems due to its focus on entropy (since it is the cost in terms of information, of using $f_t o$ in place of f_t) the Hellinger distance is preferred for evaluating dissimilarity in a geometric context because of its adherence to metric space axioms.

4 Results

In this section, we look first at the robustness of our various estimation methods with respect to the statistics defined above using simulated data, then at the estimation of our filter coefficients and the study of our estimates with respect to critical data corresponding to those of various indices during the COVID-19 crisis.

Initially, we reproduced the study by simulating our various statistics (see Figure 1). We find that, like Garcin et ali, the Kolmogorov-Smirnov statistic fits our data well for the various approaches. But we also observe distribution tails testifying to the balancing carried out in the choice of the discount parameter (here $\omega^* = 0.902$, $\omega^{HO} = 0.989$ and $\omega_c^* = 0.979$) and that of the bandwidth parameter (here $h^* = 0.488$, $h^{HO} = 0.596$ and $h_c^* = 1.375$).

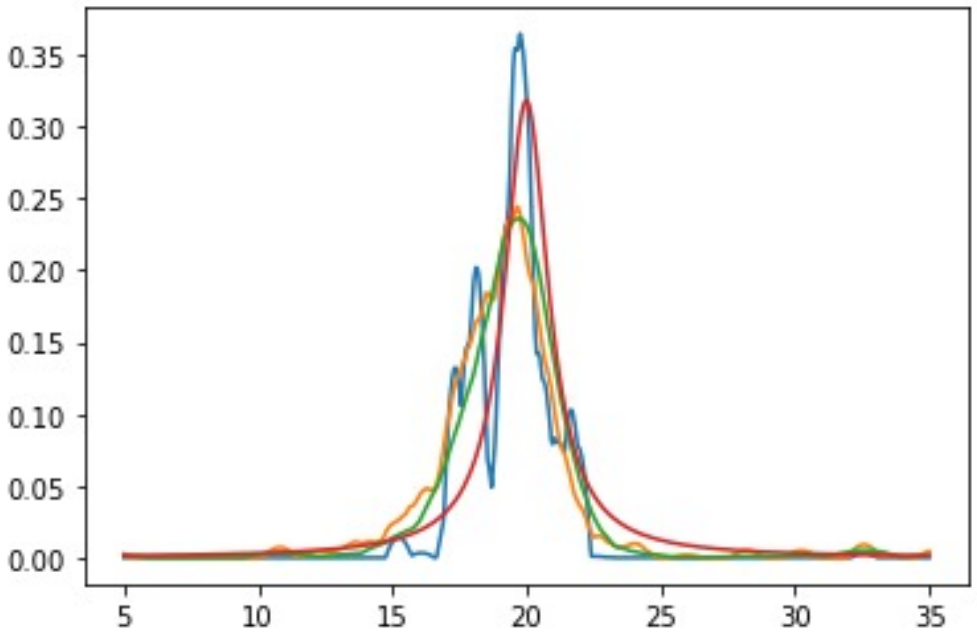

Figure 1: For the time-varying Cauchy distribution, true pdf (red curve) and estimates at time $t + 1 = 2000$. The three solid lines represent the estimates with the three competing vectors of parameters: (h^*, ω^*) (blue curve),(h^{HO}, ω^{HO}) (yellow curve), and (h_c^*, ω_c^*) (green curve).

We then compare our different approaches for each of our statistics. In these graphs, we see that the censoring method converges better than the other two methods for all our measurements. The results for the likelihood method confirm its superiority over the PIT-based approach, except in the case of the Kolmogorov-Smirnov statistic. Our results are consistent with those of the original article.

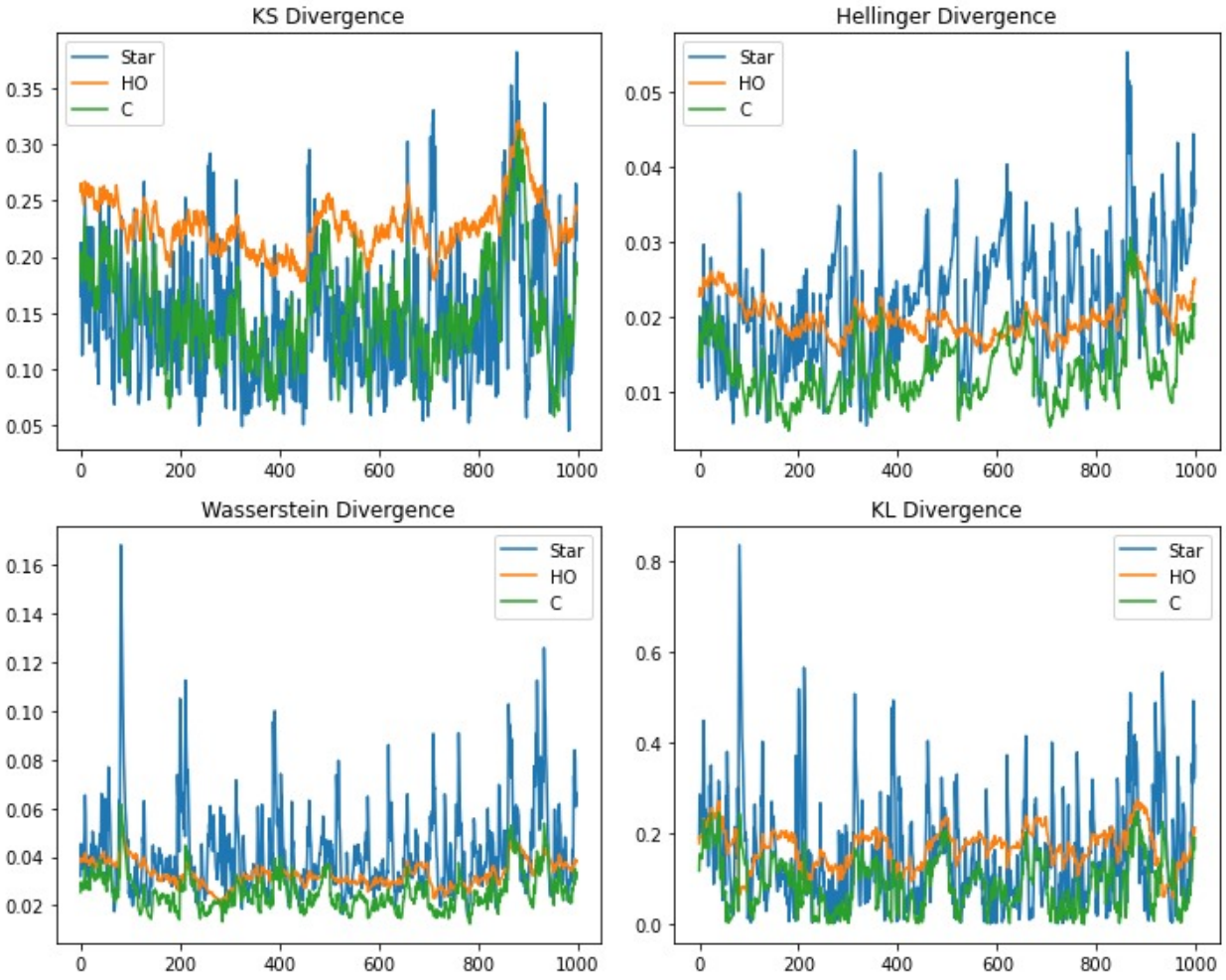

Figure 2: Divergence statistics with respect to f_t, as afunction of the instant t,when ft is the time-varying Cauchy density defined in Equation (9): the Kolmogorov–Smirnov statistic (top left), the Hellinger distance (top right), the Wasserstein distance (bottom left), and the Kullback-Leibler divergence (bottom right). The three curves represent the three competing vectors of parameters: (h^*, ω^*) (blue curve), (h^{HO}, ω^{HO}) (orange curve), and (h_c^*, ω_c^*) (green curve). The more erratic aspect of the black curves is to be explained by the lower selected free parameters.

In order to apply the method that was presented using the PITs, we considered several stock market indices, namely the American indices NASDAQ, S&P 500; the European indiex EURO STOXX 50; and the Asian indices Nikkei 225 & KOSPI.

The data were extracted from Yahoo Finance between 17 April 2015 and 28 May 2020, shortly after the first major spike in COVID-19. The advantage of choosing three different markets to conduct the study is to compare the impact that the pandemic may have had on three distinct zones.
We then estimated the best filter parameters h, ω with the different methods presented below for our data (see Table 1).

Table 1: Table of optimized estimated bandwith and discount factor minimizing criterion d_ν

	h^*	ω^*	h_c^*	ω_c^*	h^{HO}	ω^{HO}
NASDAQ Composite	6.5e-3	0.947	6.5e-3	0.980		
S&P 500	1.0e-5	0.946				
EURO STOXX 50	7.4e-5	0.911	1.1e-4	0.946		
Nikkei 225	1.0e-2	0.935	1.1e-2	0.958		
KOSPI	4.4e-3	0.893	7.2e-2	0.953		

It can be seen that our parameters are relatively close or similar to within an order of magnitude of those obtained by Garcin et al. The reasons for these differences are explained below. But for the moment we note that the filtering parameters h and ω have roughly similar values between the different indices studied.
Furthermore, we note a difference between the methods with and without censoring. Censored estimation performs much more selective filtering, with bandiwdth and discount factor values well above those of the PIT-based approach.

Subsequently, and in order to be consistent with the results of the Garcin et ali study, we take the same filtering values $h^m = 0.012$ and $\omega^m = 0.955$, which correspond to the largest values obtained on all the data, thus ensuring greater robustness in the estimation of our distributions. We then plot the estimated pdf for the returns of the S&P 500, EUROSTOXX 50 and KOSPI. We find that the distribution tails are thick, reflecting the events of COVID-19 (see Figure 3).

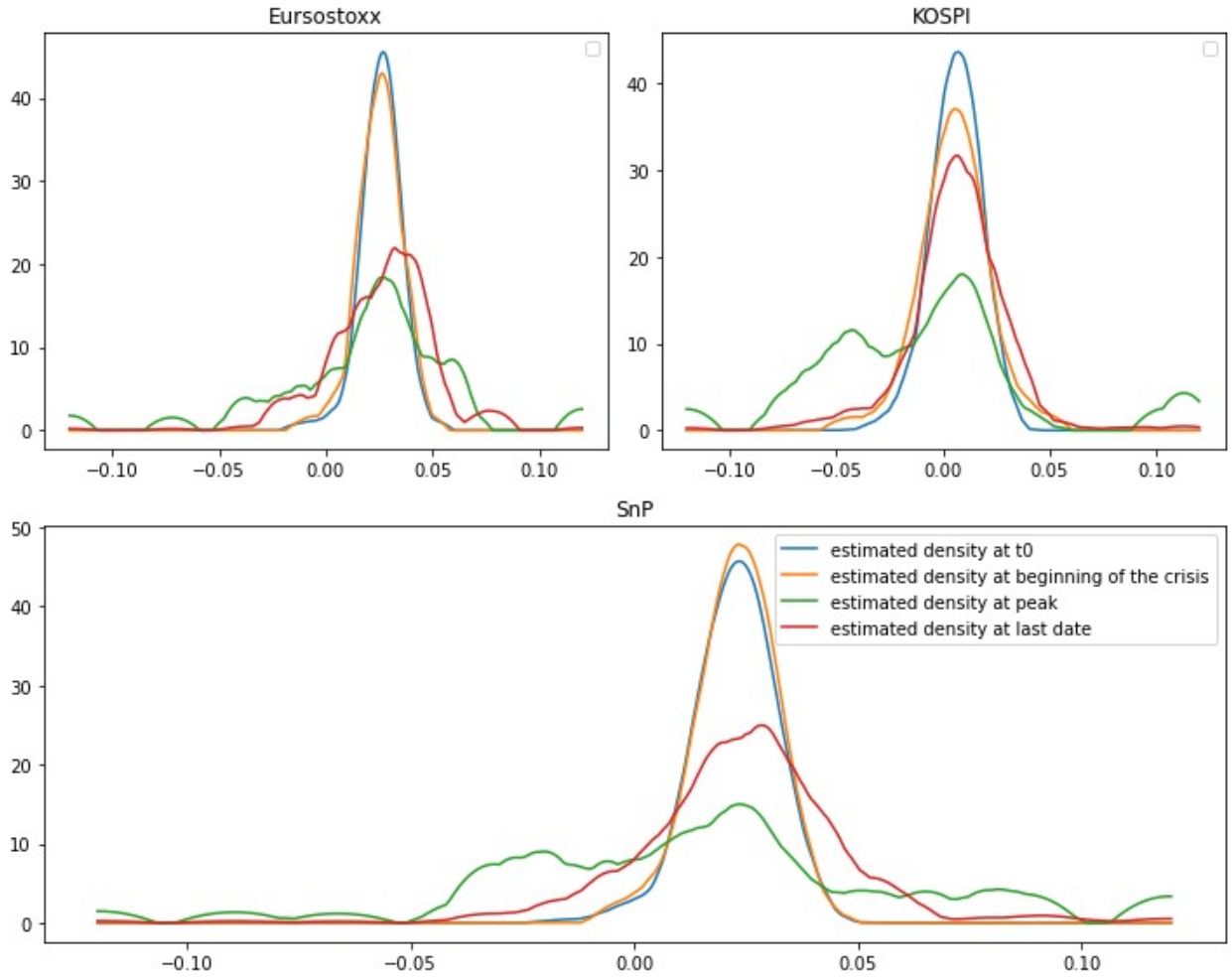

Figure 3: Daily evolution through time of four divergence statistics: the Kolmogorov–Smirnov statistic (top left), the Hellinger distance (top right), the Wasserstein distance (bottom left), and the Kullback-Leibler divergence (bottom right). The curves correspond to SP 500, EURO STOXX 50, and KOSPI indices. The dotted lines are simulated confidence intervals, with confidence levels, from the bottom to the top: 95%, 99% and 99.9%.

These observations are confirmed by the plot of daily divergence measurements in Figure 4, which shows the successive events of our crisis, with markets at equilibrium at the end of 2019 and the start of turmoil around the beginning of February 2020.

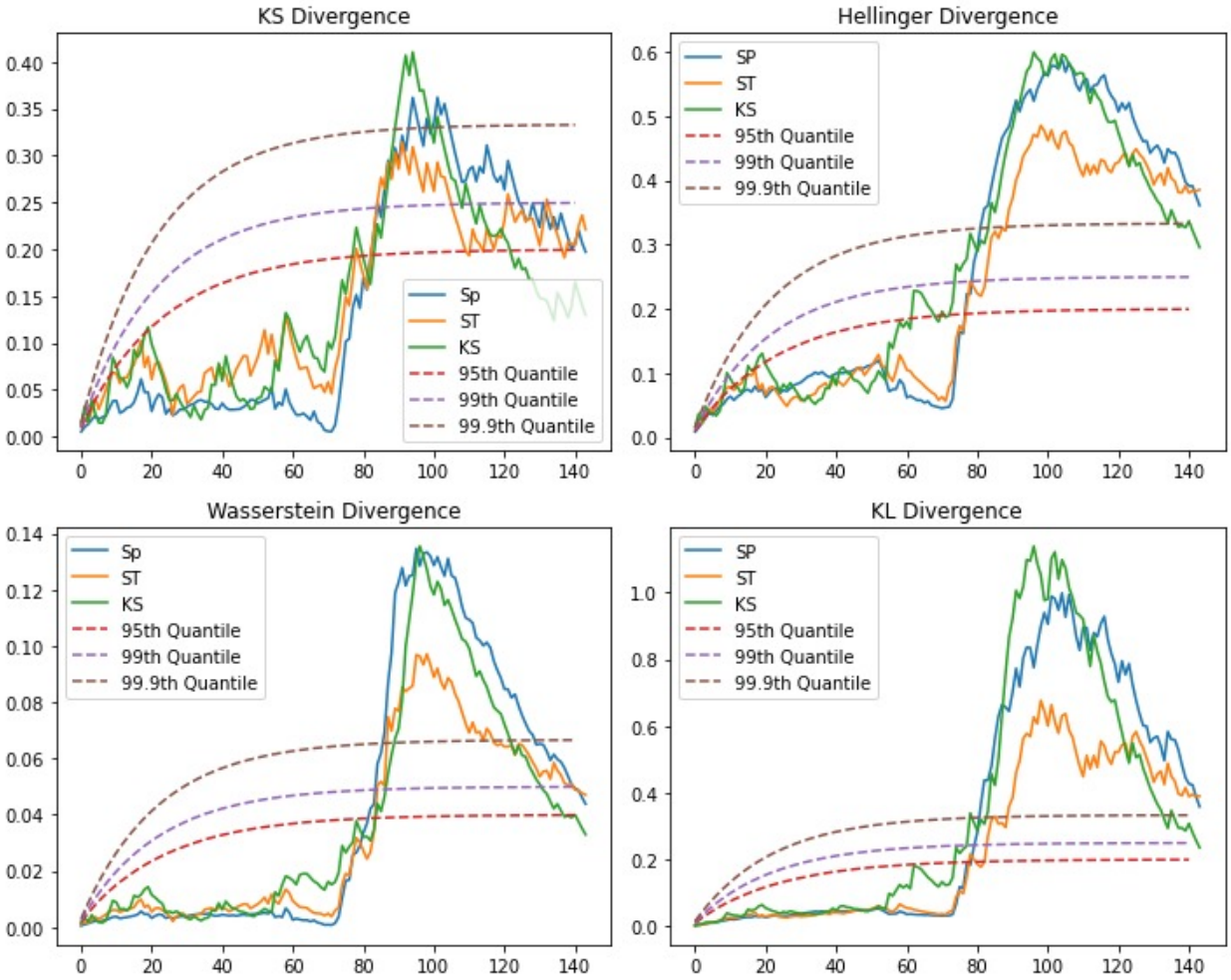

Figure 4: Estimated dynamic pdf of daily price returns for SP 500 (top left), EURO STOXX 50 (top right), and KOSPI (bottom) indices.

Finally, as regards the optimisation carried out to obtain the parameters minimising our dependency statistic, we have exploited various methods in order to converge on a solution identical to that of the authors. The methods used are: gradient descent, stochastic gradient descent, truncated Newton method (TNC), gradient descent coupled to TNC, annealing algorithm, genetic algorithm, gradient descent coupled to the annealing algorithm. After various trials varying the methods, the initial parameters and the path steps used, we arrived at parameters close to those obtained by the authors, but which were nevertheless different. We explain these differences in several ways.

First of all, we found that the complexity of our problem generates a loss of information that no longer allows us to determine variations below a certain parameter threshold. For the h parameters, our calculation of the d statistic remains constant for $h \leq 1e-5$. However, as some of our optimal parameters h are of the order of 1e-5, we had to neglect the part of the search plan defined by $h \leq 1e-5$.
We then noticed that our various algorithms were highly dependent on the initial parameters. We therefore chose to use stochastic methods to estimate our parameters. Unfortunately, these methods were unable to decide in favour of one parameter or another. We therefore had to revert to our original methods. However, in returning to our first methods, we decided to couple them together in an attempt to converge within a radius arbitrarily defined as a 5% disc around the parameters found by the first methods.
Finally, by plotting our parameter plane, we see that our plane is full of local minima. In particular, we see (cf. Figure 5) that a set of minima lie in the area of the $[0,1] \times [0.9,1]$ plane. Consequently, it appears that most of our methods converge in this zone, which seems to contain the global minimum, and our optimisation is therefore consistent for most of the parameters obtained.

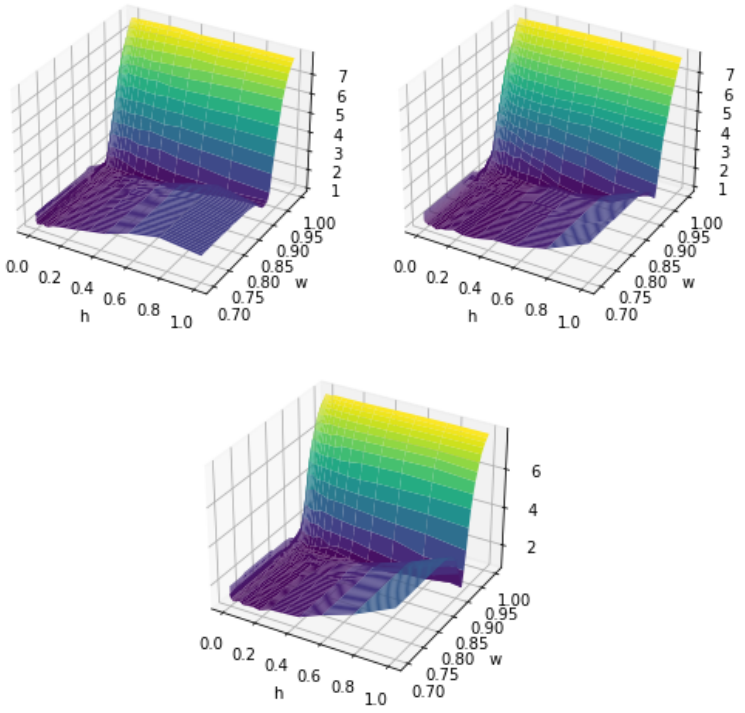

Figure 5: Projection of the dependency statistic according to h and ω in the space $[0, 1] \times [0.7, 1]$ for the EUROSTOXX 50, S&P 500 and KOSPI data series

5 Extension

In order to extend the study carried out by Garcin et ali, we decided to test our density estimates over the longer term. To do this, we have chosen to extend the range of data for the period up to the beginning of 2024 (01/01/2024). In addition, we have chosen not to re-evaluate the bandwidth and discount factor parameters in order to test the robustness of our estimates and the ability of our models to capture the information when our data ranges are extended.

We can see that the pdf of our various indices show distribution laws that correspond well to the different events that occurred over the time interval and to the different crises observed (see Figure 6).
Between the end of our first estimate and the beginning of 2024, numerous conflicts and anomalies occurred. These various phenomena are clearly shown in Figure 7, which plots the divergence measurements for the COVID-19 bounces and the various armed conflicts that occurred in the intervening period (Russian-Ukrainian war, Israeli-Palestinian conflict).

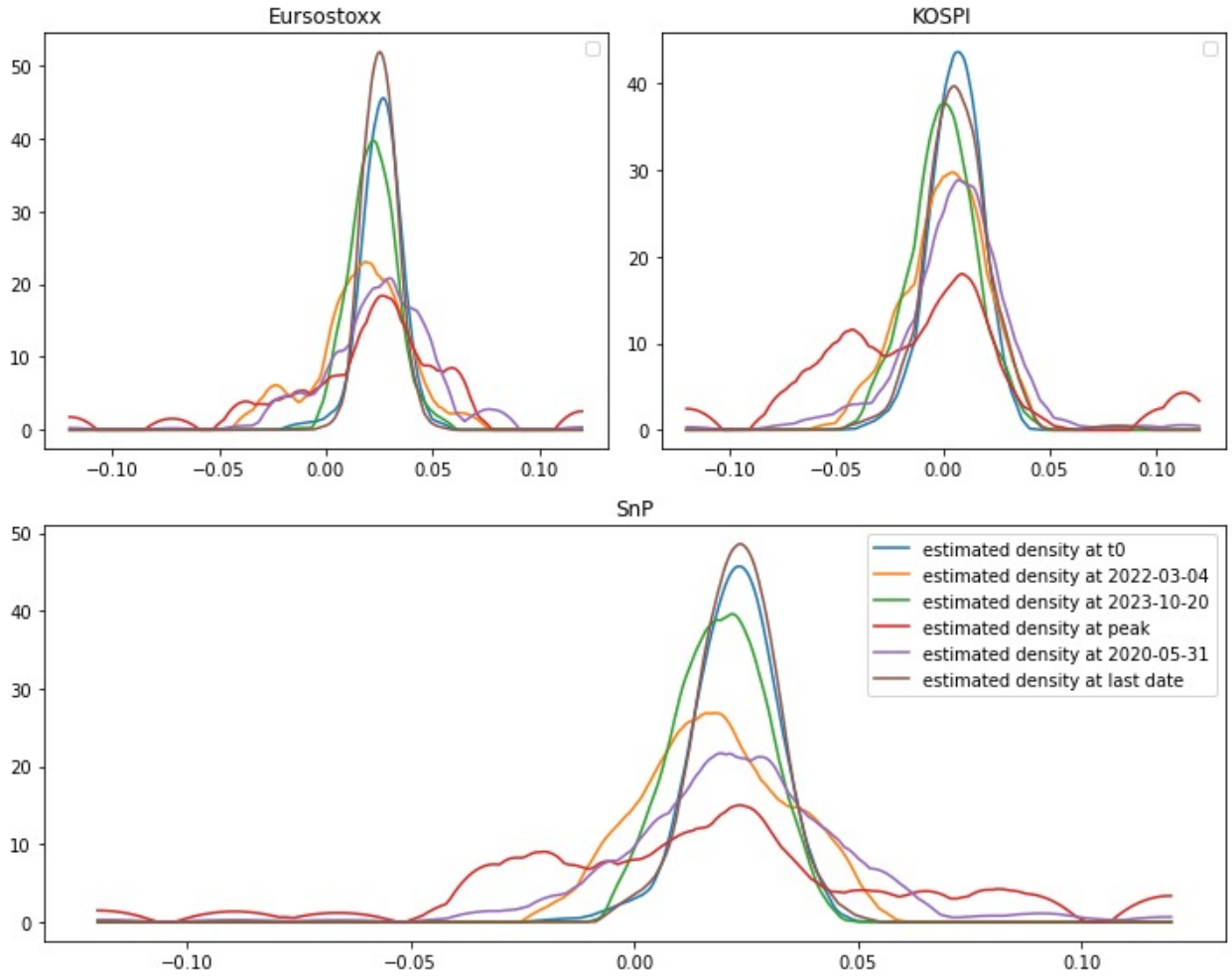

Figure 6: Estimated dynamic pdf of daily price returns for S&P 500 (top left), EURO STOXX 50 (top right), and KOSPI (bottom) indices.

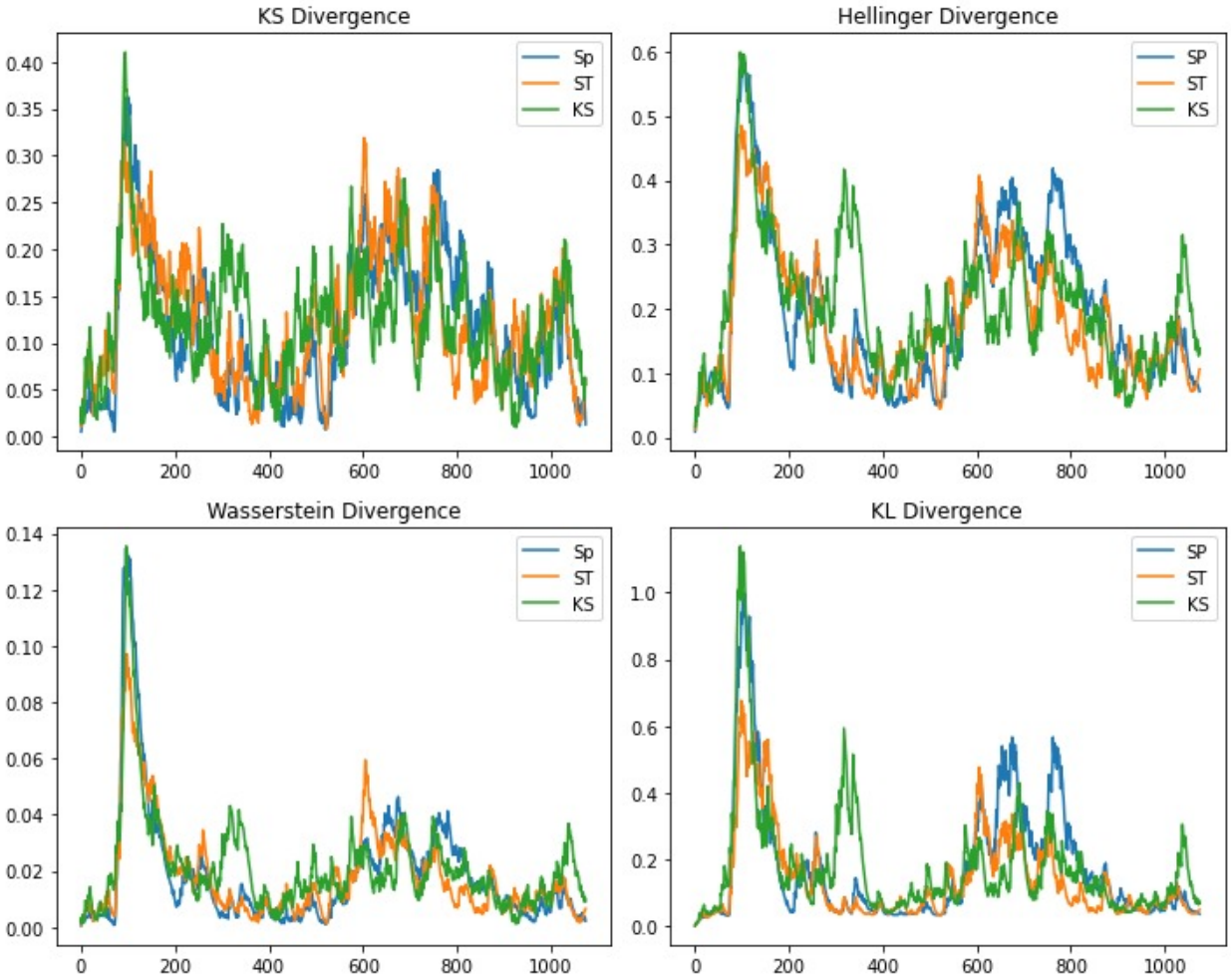

Figure 7: Daily evolution through time of four divergence statistics: the Kolmogorov–Smirnov statistic (top left), the Hellinger distance (top right), the Wasserstein distance (bottom left), and the Kullback-Leibler divergence (bottom right). The curves correspond to S&P 500, EURO STOXX 50, and KOSPI indices.

We conclude that our estimation methods have succeeded in capturing the various events that have occurred and that, despite the fact that the parameters have not been re-evaluated, our system has retained a high degree of accuracy in its estimates.

We had also considered enriching the method for estimating our distributions over longer time windows by considering sliding estimation windows, as proposed in the appendix to the original article, but this time with larger windows. However, the complexity of the calculations, relative to the power of our computers, meant that we were unable to carry out the study conclusively at this stage.
In addition, the time complexity of the calculations also limited our ability to test the robustness of our parameters and estimates by comparing the estimates of the different models for optimal filter parameters estimated from other index series.
Finally, we would have liked to study the robustness of our bandwidth and discount factor parameters on other market segments in order to compare the optimal parameters obtained. However, the great variability of our otpimisation algorithms does not allow us to determine the optimal parameters with any certainty.

6 Conclusion

In conclusion, time-varying kernel estimation methods demonstrate a high degree of robustness for similar data, with the ability to detect and plausibly estimate the distributions of our returns even in extreme cases of crisis, whatever their

source as long as we consider a very demanding filter criterion.
We also find that constraint methods have filtering requirements that enhance the accuracy of our density function estimates under the fixed conditions of PIT independence and uniformity.

7 References

References

[1] Garcin, Klein, Laaribi. *Estimation of time-varying kernel densities and chronology of the impact of COVID-19 on financial markets*, Journal of Applied Statistics, 2023.