

Riskcasting

Arnaud de Servigny & Timothée Sohm-Quéron
Bramham Gardens, Paris, France

March 30, 2020

The Riskcasting[®] methodology is protected by a global patent

TECHNICAL PAPER



Contents

1	Data Processing from Option Prices to the Discount Factor	3
1.1	Background to get the Discount Factor	3
1.2	A quick overview of Estimation Techniques	5
1.3	The Risk Neutral Density of returns: Q	6
1.4	The Empirical/Historical Density of returns: P	12
2	From the Discount Factor to the Relative Risk Aversion	16
2.1	Relative Risk Aversion	16
2.2	Denoising the Relative Risk Aversion	17
2.3	Extracting the daily Moments as Input Factors	18
3	The Asset Allocation Phase	19
3.1	Description of the Classification Methodology	19
3.2	Mathematical formulation	19
3.3	Retrieving the Probability to be in a given Class, given the Information Acquired	21
3.4	From the Probability to be in a given Class to the Allocation Rule itself	21

1 Data Processing from Option Prices to the Discount Factor

1.1 Background to get the Discount Factor

In this first section we define what the Discount Factor is and on which assumptions it relies. The evolution of prices of financial securities is tied to two engines: one is the evolution of fundamental data, the other one is the evolution of behavioural preferences. We focus here on the second motor. We show that when there is no arbitrage opportunity, the price of a security is the discounted value of its possible payoff. The more risk averse the investor, the more he will prefer the present and thus the more aggressively he will discount future payoffs.

The Discount Factor is a quantity that summarizes risk preferences of investors. If markets are complete and there is no opportunity of arbitrage, then the Discount Factor exists and is unique [1]. To present how the Discount Factor is used in asset pricing, we use the example detailed in Barone-Adesi [1]. Let's use a representative agent with a utility function U , in a two-period economy. The utility function $U()$ represents the utility of the investor over these two periods, while $u()$ is the utility related to a single period. This investor displays a non-satiation property, as more consumption is always preferred to less consumption. His utility increases with his consumption (his wealth). As a consequence, the first derivative of his utility function is always positive. Moreover, our investor is risk averse: his utility is an increasingly concave function. Let's write the utility of this investor in a two-period economy, with β a subjective discount factor and c_t the consumption at time t :

$$U(c_t; c_{t+1}) = u(c_t) + \beta u(c_{t+1}) \quad (1)$$

The investor chooses at t the amount of asset he wants to purchase, but does not know yet the payoff of the asset at $t+1$. As a consequence, he wishes to maximize his global expected utility $E_t([U(c_t; c_{t+1})])$. We deduce a maximization program. ξ being the amount of the asset the investor chooses to buy at time t , e the original endowment of the agent, P_t the price of the asset at time t and x_{t+1} the future payoff of the asset, the objective becomes:

$$\begin{aligned} & \underset{\xi}{\text{maximize}} && u(c_t) + \beta E_t[u(c_{t+1})] \\ & \text{subject to} && c_t = e_t - P_t \xi \\ & && c_{t+1} = e_{t+1} + x_{t+1} \xi \end{aligned}$$

The first constraint is a budget constraint property, as the investor cannot consume more than he owns. The second constraint means that the investor will consume his entire endowment at the end of the second period. Substituting the two constraints into the objective function we get:

$$\underset{\xi}{\text{maximize}} \quad u(e_t - P_t \xi) + \beta E_t[u(e_{t+1} + x_{t+1} \xi)] \quad (2)$$

In order to obtain the optimum, we now set the derivative with respect to ξ equal to 0, using the chain rule:

$$0 = -P_t u'(e_t - P_t \xi) + \beta E_t [u'(e_{t+1} + x_{t+1} \xi) x_{t+1}] \quad (3)$$

We obtain:

$$P_t u'(c_t) = \beta E_t [u'(c_{t+1}) x_{t+1}] \quad (4)$$

We finally get the traditional fundamental asset pricing equation:

$$P_t = E_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right] \quad (5)$$

And we can define the *marginal rate of substitution* (*MRS*) and the *discount factor* $m_{t,t+1}$:

$$m_{t,t+1} \stackrel{def}{=} \beta E_t \left[\frac{u'(c_{t+1})}{u'(c_t)} \right] \stackrel{def}{=} MRS \quad (6)$$

The *MRS* is also defined as the *discount factor*. We can therefore price any security as:

$$P_t = E_t [m_{t,t+1} x_{t+1}] \quad (7)$$

Or in the continuous case (under the historical measure P), with m_t the discount factor, x_t the expected payoff and $p_{t,T}$ is the historical probability of state S_T :

$$P_t^P = \int_{\mathbb{R}} \underbrace{m_{t,T}(S_T)}_{\text{Discounting including risk preferences}} \underbrace{x_T(S_T)}_{\text{Payoff in T}} \underbrace{p_{t,T}(S_T)}_{\text{Historical probability of the payoff}} dS_T \quad (8)$$

We now present the version of this equation with a change in measure, under the risk neutral measure Q . We can have both an economic perspective of this subject, and a mathematical view. As seen in (5), the way a risk averse agent discounts future payoffs (the Discount Factor) depends on his marginal utility (risk preference). On the contrary, the risk-neutral agent always has the same marginal utility, leading to a constant discount factor. For him, in (5), we have $\beta = e^{-rT}$. Mathematically speaking, this is due to the fact that under the risk-neutral measure, discounted prices follow a martingale with an expected rate of return equal to the risk free rate. If not, we would have arbitrage opportunity. The main point is that under the risk neutral measure, risk neutral agents discount future payoffs at the risk free rate r :

$$P_t^q = e^{-rT} \int_{\mathbb{R}} x_T(S_T) q_{t,T}(S_T) dS_T \quad (9)$$

Combining (8) and (9) we can obtain that the *discount factor* is a function of the relation between the risk neutral measure q and the historical density p , and also includes the risk free rate r :

$$m_{t,T}(S_T) = e^{-rT} \frac{q_{t,T}}{p_{t,T}} \quad (10)$$

Figure 1.1 describes a typical evolution of the Discount Factor for the S&P 500.

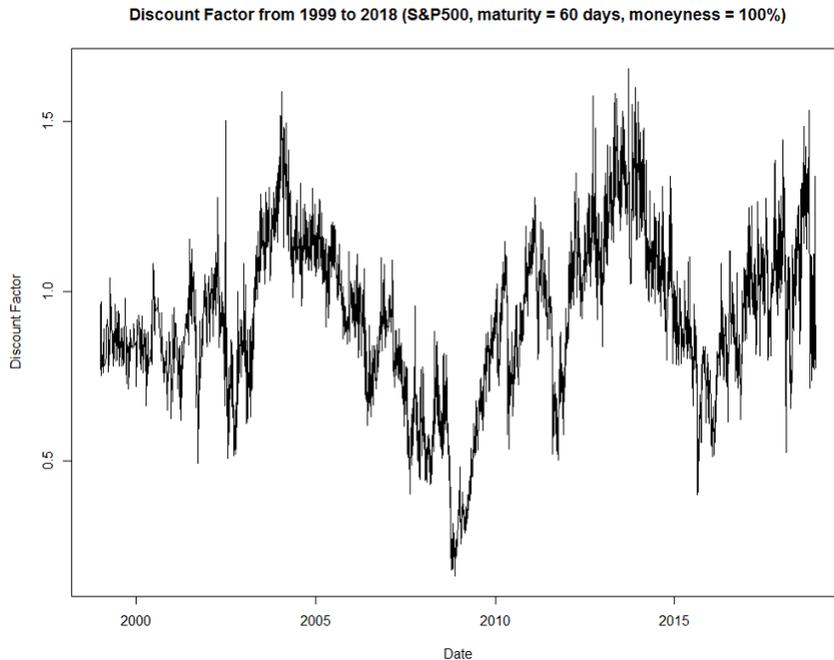


Figure 1: S&P 500, Discount Factor from 1999 to 2018, maturity = 60 days, moneyness = 100%

1.2 A quick overview of Estimation Techniques

The initial approaches to estimating the Discount Factor were purely economically driven, as the Discount Factor is the marginal rate of substitution. Despite some refinements (like in Campbell [4] with a time-varying relative risk aversion), most of these models suffer from two drawbacks. First they struggle to adequately specify the utility function of investors. Second, consumption data is extremely difficult to retrieve. To skip this difficulty one standard practice is to use market returns as a proxy.

Thanks to this proxy a new type of model can be introduced, with a parametric approach. This idea is quite intuitive, as we wish to express the Discount Factor as a simple function of market returns. The main paper about this approach is written by Engle and Rosenberg [8]. Their idea is to use two specifications for the Discount Factor. In the first version, the Discount Factor is a power function of the underlying asset gross returns. In their second version, they use a polynomial function with $N+1$ terms. Despite some encouraging results, these parametric approaches tend to over-simplify the problem and having

to specify the shape of the relationship between returns and the Discount Factor causes a major difficulty about which shape to use.

Most of the approaches rely on the ratio between the historical state density of returns and the risk-neutral one. Our methodology to estimate the Risk Neutral Density is quite similar in spirit to the one described in Figlewski [7]. We showed in (10) that:

$$m_{t,T}(S_T) = e^{-rT} \frac{q_{t,T}}{p_{t,T}}$$

With $q_{t,T}(S_T)$ being the state price density of returns (also called risk-neutral density) and $p_{t,T}(S_T)$ the empirical distribution of returns. The next challenge is to find a way to estimate both these densities.

1.3 The Risk Neutral Density of returns: Q

The Risk Neutral Density is not directly observable and we need to find a way to estimate it. To do so, we first recall the definition of the Leibniz integral rule, and then start with the pricing formula of an European call option.

Definition of the Leibniz integral rule We start with the definition of the Leibniz integral rule. As a reminder, this rule is used to derive an integral with a parameter as limit. We first define the integral:

$$\int_{a(x)}^{b(x)} h(x, t) dt \tag{11}$$

The Leibniz integral rule states that:

$$\frac{\partial}{\partial x} \left(\int_{a(x)}^{b(x)} h(x, t) dt \right) = h(x, b(x)) \frac{\partial}{\partial x} b(x) - h(x, a(x)) \frac{\partial}{\partial x} a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} h(x, t) dt \tag{12}$$

First derivative: Call Price (under the Risk Neutral Measure) One of the most fundamental theorem of asset pricing is that the price of an asset at time t is equal to its discounted expected payoff at time T . The terminal payoff of a call option is $\max(0, S_T - X)$, leading to:

$$C = e^{-rT} \int_X^{+\infty} (S_T - X) f(S_T) dS_T \tag{13}$$

With r the risk-free rate, T the remaining time until maturity, X the strike, S_T the price of the underlying at time T , and $f(S_T)$ the probability density for the terminal value of the underlying, known as the *Risk Neutral Density*. Using the notation in (11), we define:

$$h(X, S_T) = (S_T - X) f(S_T) \tag{14}$$

We can now have the first derivative of the call price, with respect to the strike X :

$$\frac{\partial C}{\partial X} = \frac{\partial}{\partial X} e^{-rT} \int_X^{+\infty} (S_T - X) f(S_T) dS_T$$

As e^{-rT} does not depend on X :

$$\frac{\partial C}{\partial X} = e^{-rT} \frac{\partial}{\partial X} \int_X^{+\infty} (S_T - X) f(S_T) dS_T$$

We can use the Leibniz rule (12) and the defined function (14):

$$\frac{\partial C}{\partial X} = e^{-rT} \left[h(X, +\infty) * 0 - h(X, X) * 1 + \int_X^{+\infty} \frac{\partial}{\partial X} h(X, S_T) dS_T \right]$$

Leading to:

$$\frac{\partial C}{\partial X} = e^{-rT} \left[-(X - X) f(X) + \int_X^{+\infty} -f(S_T) dS_T \right]$$

Finally:

$$\frac{\partial C}{\partial X} = -e^{-rT} \int_X^{+\infty} f(S_T) dS_T \quad (15)$$

Second derivative: Call Price First we begin by rearranging the last expression (15), since $f(S_T)$ is the risk neutral density, and its integral between 0 and $+\infty$ is equal to 1:

$$\frac{\partial C}{\partial X} = -e^{-rT} \left(1 - \int_0^X f(S_T) dS_T \right)$$

And:

$$-e^{rT} \frac{\partial C}{\partial X} - 1 = - \int_0^X f(S_T) dS_T$$

Finally:

$$e^{rT} \frac{\partial C}{\partial X} + 1 = \int_0^X f(S_T) dS_T \quad (16)$$

Let's work with these two expressions, and derive them with respect to the strike. Let's start with the right part:

$$\int_0^X f(S_T) dS_T$$

We can easily apply again the Leibniz integral rule:

$$\frac{\partial}{\partial X} \int_0^X f(S_T) dS_T = f(S_T) \frac{\partial}{\partial X} X - 0 + \int_0^X \frac{\partial}{\partial X} f(S_T) dS_T = f(S_T)$$

We now work on the left side of (16):

$$\frac{\partial}{\partial X} \left(e^{rT} \frac{\partial C}{\partial X} + 1 \right) = e^{rT} \frac{\partial^2 C}{\partial X^2}$$

We get the final expression of the risk neutral density:

$$f(S_T) = e^{rT} \frac{\partial^2 C}{\partial X^2} \quad (17)$$

Using Put Price We can prove that starting with put prices, we can obtain the exact same result. First, let's recall the pricing equation of the put price:

$$P = e^{-rT} \int_0^{+\infty} (X - S_T) f(S_T) dS_T \quad (18)$$

We define the function:

$$h(X, S_T) = (X - S_T) f(S_T) \quad (19)$$

We use the Leibniz integral rule:

$$\frac{\partial P}{\partial X} = e^{-rT} \frac{\partial}{\partial X} \int_0^X (X - S_T) f(S_T) dS_T$$

Leading to:

$$\begin{aligned} \frac{\partial P}{\partial X} &= e^{-rT} \left[h(X, X) * 1 - h(X, 0) * 0 + \int_0^X \frac{\partial}{\partial X} h(X, S_T) dS_T \right] \\ \frac{\partial P}{\partial X} &= e^{-rT} \left[\int_0^X f(S_T) dS_T \right] \end{aligned}$$

We rearrange the last expression:

$$e^{rT} \frac{\partial P}{\partial X} = \int_0^X f(S_T) dS_T$$

We now derive again with respect to X and use the same result as in (1.3):

$$f(S_T) = e^{rT} \frac{\partial^2 P}{\partial X^2} \quad (20)$$

In conclusion, starting with calls or puts, we always end up with the same expression for the risk neutral density.

Volatility Surface - Literature review In most articles ([7] and [1] for example), authors do not necessarily pay attention to having fixed maturity for options across time. As a consequence, time series of the Discount Factor tend to blend information that is not completely comparable. We found one article by Vial [10] which uses fixed maturities. Unfortunately the author uses a volatility surface directly from the *OptionMetrics* database, which uses options with little or no volume. Thus, the calculated Risk Neutral Densities incorporate information which is not relevant.

Filters The database used in our study is the usual OptionMetrics database. It contains the closing price of all options on single stock and indices in the US market from January 1996 to May 2016. We first apply a filter on volumes (at least 40 traded options per day), and on prices (which need to quote at least 0.05\$). We then exclude options with less than 7 days to maturity. Finally, we add some liquidity constraints necessary to the construction of the volatility surface: we need at least three different maturities for each stock, and at least 5 traded puts and 5 calls for each of these maturities.

Volatility Surface - Construction We then move on to the construction of the volatility surface itself. The idea is quite simple. First, for each available maturity, we interpolate the implied volatility across all available strikes, thanks to splines. We then obtain, for each maturity, the implied volatility for different levels of moneyness (defined as K/S_t , with K the strike of the option and S_t the current price of the underlying). We then interpolate, for each of these moneynesses, the implied volatility across all maturities. The final output is the implied volatility level for any fixed level of moneyness and maturity. We first eliminate the outliers, based on the idea that the price of a call (put) option should monotonically decrease (increase) with the strike. The second possible source of abnormal behaviour is the concavity in the volatility smile. First, let's keep in mind the expression of the Risk Neutral Density (17):

$$f(S_T) = e^{rT} \frac{\partial^2 C}{\partial X^2}$$

The Risk Neutral Density depends on the second derivative of the price of the option with respect to the strike. If we observe some concavity at any point, then we should have $\frac{\partial^2 C}{\partial X^2} < 0$, resulting in a negative probability. To check whether we have concavity, we need to use an approximation of the second derivative at available discrete points. We emphasise the fact that these points are not always symmetrically distributed. Let's start with the general Taylor-Young formula:

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 \quad (21)$$

We consider three options (possibly puts or calls) with respective strikes $K_1 < K_2 < K_3$. We want to evaluate the second derivative at the K_2 point. We define $h_1 = K_3 - K_2$ and $h_2 = K_2 - K_1$. Using (21) we engineer two changes of variable:

$$f(a+h_1) \approx f(a) + h_1 f'(a) + \frac{1}{2} h_1^2 f''(a) \quad (22)$$

$$f(a-h_2) \approx f(a) - h_2 f'(a) + \frac{1}{2} h_2^2 f''(a) \quad (23)$$

Ultimately, combining (22) and (23):

$$f(a+h_1) + f(a-h_2) \approx 2f(a) + (h_1 - h_2)f'(a) + (0.5(h_1^2 + h_2^2))f''(a)$$

Rearranging:

$$(0.5(h_1^2 + h_2^2))f''(a) \approx f(a + h_1) + f(a - h_2) - 2f(a) - (h_1 - h_2)f'(a)$$

Yielding:

$$f''(a) \approx \frac{f(a + h_1) + f(a - h_2) - 2f(a)}{0.5(h_1^2 + h_2^2)} - \frac{(h_1 - h_2)f'(a)}{0.5(h_1^2 + h_2^2)} \quad (24)$$

This is the final formula to approximate the second derivative at any point. We can check that it boils down to the usual formula when we have the traditional case $h_1 = h_2$:

$$f''(a) \approx \frac{f(a + h_1) + f(a - h_1) - 2f(a)}{0.5(h_1^2 + h_1^2)} - \frac{0 * f'(a)}{0.5(h_1^2 + h_1^2)}$$

Yielding to the the usual formula for the discrete approximation of the second derivative:

$$f''(a) \approx \frac{f(a + h_1) + f(a - h_1) - 2f(a)}{h_1^2}$$

In conclusion, we can estimate the first and second derivative at each point. For calls, if the first derivative is positive and/or the second derivative is negative, we then remove the point. For puts, we do it if the first derivative is negative and/or the second derivative is negative. We need to keep in mind that this method eliminates most of the points leading to concavity, but not all of them. To interpolate the implied volatility, the standard practice is to use splines. The main objective of splines is to fit discrete points with a set of polynomial functions. The points used to fit the splines are called *knots*. Even if all of these polynomials have the same form, the coefficient in the polynomials will change between two consecutive sets of two knots. This kind of fit takes into account multiple behaviours across the range of strikes. One of the standard practices is to use natural cubic splines. In this case the set of polynomial functions are of the form (with K_i the i^{th} strike):

$$\begin{cases} S_0(K) = a_0 + b_0K + c_0K^2 + d_0K^3, K_0 \leq K \leq K_1 \\ \dots \\ S_{n-1}(K) = a_{n-1} + b_{n-1}K + c_{n-1}K^2 + d_{n-1}K^3, K_{n-1} \leq K \leq K_n \end{cases}$$

This type of splines is extremely useful as they impose smoothness on derivatives between two consecutive knots. This is important for traders as they cannot afford any discontinuity of greeks. However, as pointed by Figlewski [7], this kind of method suffers from one major drawback, which is that the spline needs to go through all points, and as a consequence, can lead to local concavity, which may not be annoying for traders purpose, but is not acceptable for us. One of the other drawbacks of the cubic spline is that despite displaying smoothness for the two first derivatives, it is not smooth with respect to the third derivative, and it can cause some kinks when we go back in the strike-price space [7]. Bliss

and Panigirtzoglou [3] are the first to propose an alternative smoothing splines. These kinds of splines do not have to go through each point, and contain a penalty function. As a consequence, Figlewski [7] uses a 4-th order spline, with one single knot at the at-the-money point (meaning two polynomials to estimate, one between K_1 and $K_i = S_t$, the other one between $K_i = S_t$ and K_N). However, as soon as we add more complexity, the fourth order polynomial may display inflection points, leading to rare cases of concavity. We therefore choose a different approach: we use quadratic splines, with a convexity constraint added. This method is different from the one presented by Shimko [9], as he only uses the parabola of the best least-squares fit and no assumption about convexity (which should however arise from the general form of the data). We should however emphasise that our approach suffers from one major drawback: we don't obtain smoothness on the second derivative of our splines, and will have to add another polynomial fit on the Risk Neutral Density as explained later.

Construction of the Risk Neutral Density In the previous subsection we presented how we generated the volatility surface. We can now extract the Risk Neutral Density in a similar manner. Starting from this input, we simply renew the filter on the first and second derivative, for each of the fixed studied maturities (here 30, 60 and 90 days), to get rid of most outliers. We then interpolate the implied volatility thanks to quadratic splines with a convexity constraint. Having the implied volatility associated with each strike in the strike range, we can translate it in the strike-price space. And we finally use (17) and (24) in order to get the estimation of the Risk Neutral Density at each point. It should be noted that we only use Out-of-The-Money (OTM) options to construct the Risk Neutral Density.

As discussed in 1.3, the major drawback with this approach is the discontinuity in the second derivative in splines. This leads to discontinuities in the strike-Risk Neutral Density space. To generate a smoother result, we use a polynomial fit (of order 10) on the calculated Risk Neutral Density, as in this example (Figure 2):

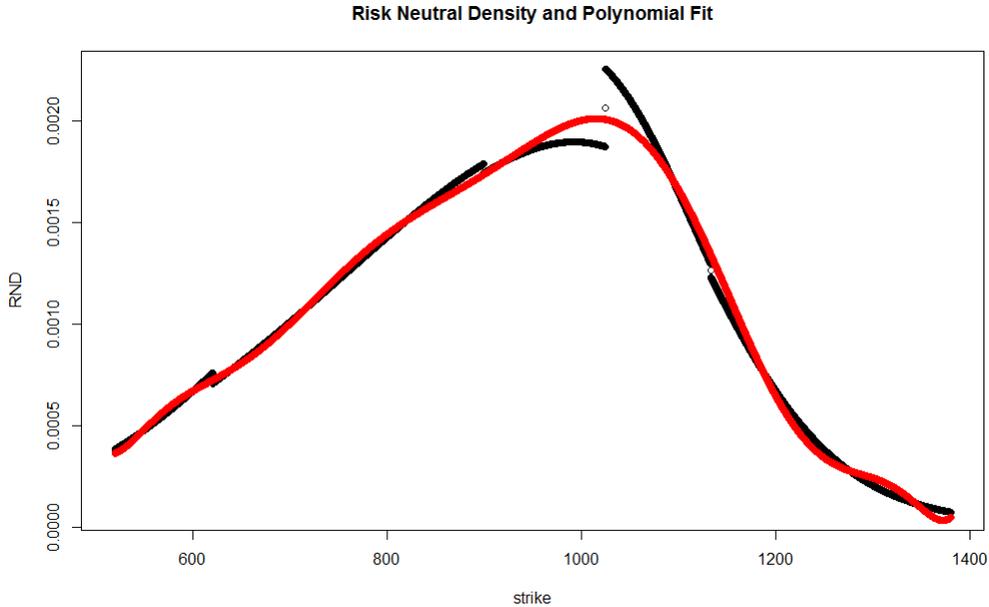


Figure 2: S&P 500, maturity = 60 days

1.4 The Empirical/Historical Density of returns: P

In the previous subsection we computed the risk neutral density. We now need to estimate the second part of (1.2), i.e. the Empirical Density (also known as the Historical Density). We first present the traditional approach, using GARCH-like models, such as the work of Barone-Adesi [1]. The main idea is to fit as well as possible past returns, and then simulate different paths (on a length corresponding to the expiry of the used options) in order to obtain a density. To do so, Barone-Adesi first fits historical daily returns with a GJR-GARCH (sometimes other types of GARCH models are used, as in Stork [6], in which an E-GARCH is used) on a long period of time (12 years). The idea is to use a long period of time to capture as much as possible the excess skewness, the excess kurtosis and other extreme returns. The GJR-GARCH is typically:

$$\log\left(\frac{S_t}{S_{t-1}}\right) = \mu + \epsilon_t$$

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \gamma I_{t-1}\epsilon_{t-1}^2$$

With S_t the underlying price, $\epsilon_t = \sigma_t z_t$ and $z_t \sim f(0, 1)$ and I_{t-1} when $\epsilon_{t-1} < 0$, and 0 otherwise. The parameters of this model are estimated by maximum likelihood. There are two main advantages with this kind of models. First, we use a GARCH model for a time-varying volatility, which is an important requirement

to fit the behaviour of financial asset. Second, thanks to $I_{t-1}\epsilon_{t-1}^2$, we can capture the asymmetry of innovations on volatility between positive and negative shocks. Once the parameters of the model are estimated, in order to perform Filtered Historical Simulations [2], we need to scale the return innovation. As we have the estimated residuals $\hat{\epsilon}_t$ and the estimated volatility $\hat{\sigma}_t$ of the fitted model, we can deduce a time series of scaled innovations:

$$\hat{z}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$$

Finally we can move on to the simulation part of the process. We can simulate N path with the estimated parameters of the model. But instead of drawing z_t from a Gaussian distribution, we draw it from the set of historical scaled innovations \hat{z}_t . Thus, Barone-Adesi simulate 35 000 paths using this, and construct a probability density according to the final values of S_T .

Even if this approach is considered as state of the art to construct the Historical Density, we are not comfortable with some of its features. First, the idea of using such a long frame (12 years) leads to little or no reactivity in the estimation of P. It is true that the Historical Density should offer an objective evaluation of past returns. But we also need to take into account the time-varying regimes of the return distribution. There is little to no sense evaluating a common distribution of returns related to such different periods. Moreover, Barone-Adesi works with asset returns and not excess-returns, which are not fully comparable over time.

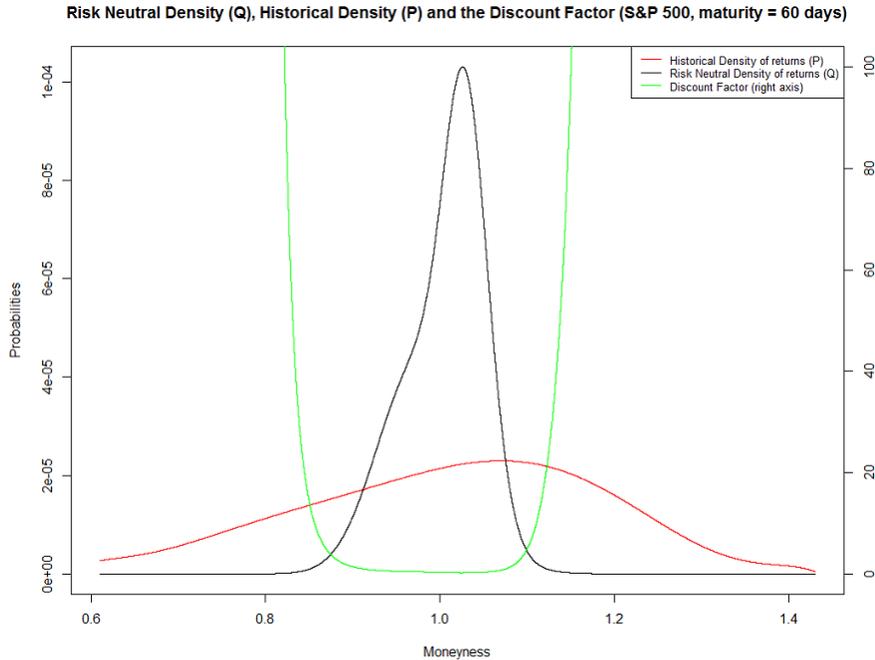


Figure 3: S&P 500, maturity = 60 days

Figure 3 displays the two distributions and the Discount Factor for a given date.

The approach we use starts with the observation made by Davies and Servigny [5]. They use returns of different asset classes (stock indexes, bonds, etc.) at different frequencies (from daily to quarterly), and run a test (Bontemps-Meddahi) to assess the "Gaussianity" on these time series. Their main finding is that the longer the time horizon, the more likely we are to accept the Gaussian hypothesis. As a consequence, our idea is to represent past returns with Gaussian Mixture Models (GMM). We do not simply rely on a Gaussian distribution as we still want to take into account the possible high kurtosis. Moreover, we want to be as dynamic as possible in order to get the current regime in the market (see Figures 4 and 5). As a consequence, the sample of data used to estimate the parameters of the GMM is proportional to the maturity of the options used to infer the Risk Neutral Density. The longer the maturity, the deeper we dig in history (the history used is 20 times the maturity, eg. 600 days for the 30 days maturity). The Gaussian Mixture contains two separate Gaussians, and is not calibrated on daily returns, but on rolling returns on periods similar to the used maturity. We find that the estimation is very stable as long as the starting values in the EM algorithm (used to calibrate the GMM) are not extreme. One interesting finding is that on average, most single stocks have one important "quiet" distribution with a small expected positive return, and conversely one

smaller "crash" distribution with an important expected negative return.
This density is estimated on the same maturities: 30, 60 and 90 days.

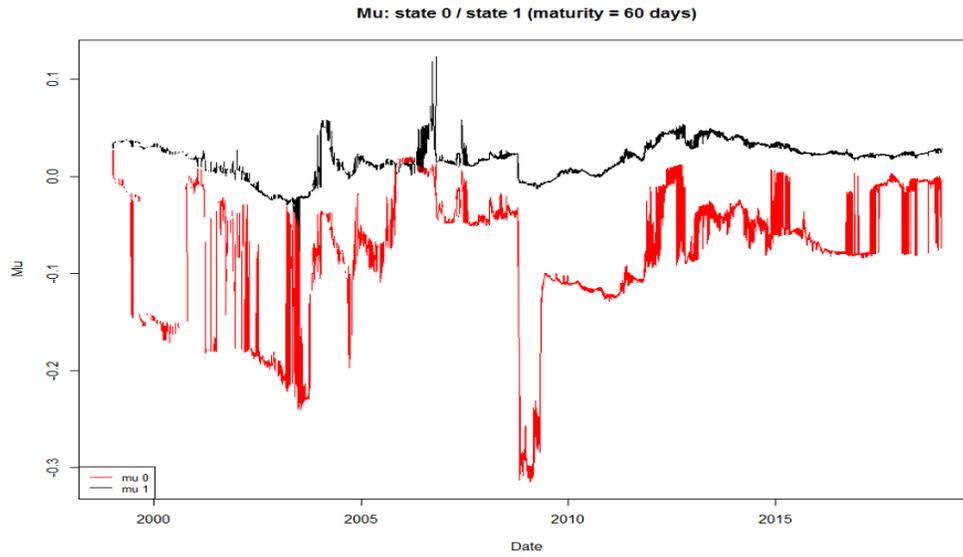


Figure 4: GMM: Mu, S&P 500, maturity = 60 days

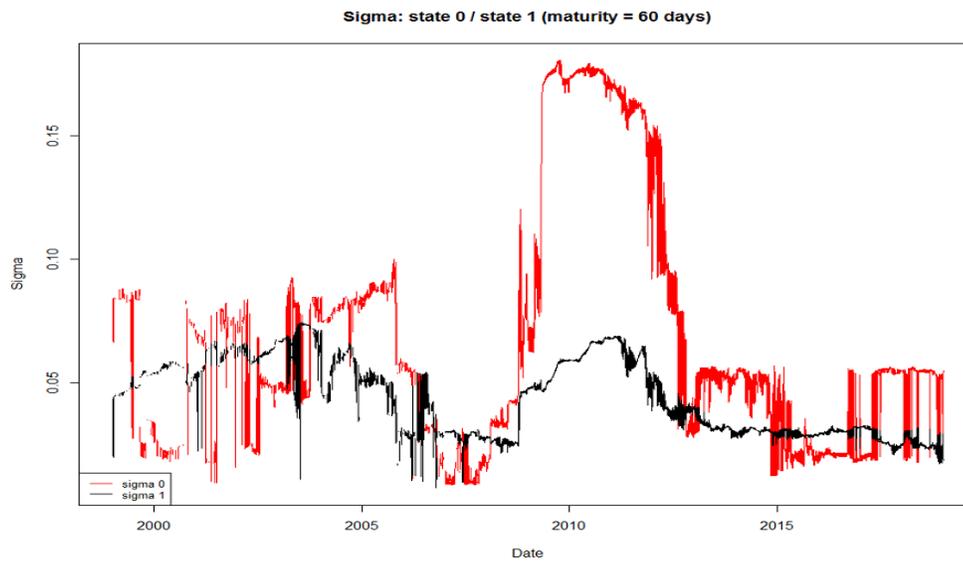


Figure 5: GMM: Sigma, S&P 500, maturity = 60 days

2 From the Discount Factor to the Relative Risk Aversion

2.1 Relative Risk Aversion

The previous part focused entirely on the estimation of the Discount Factor. We now want to link the Discount Factor with the preferences of investors. One of the advantages of the Discount Factor is that it allows us to compute the Arrow-Pratt measure of Relative Risk Aversion (RRA) with no assumption about the shape of the utility function. This methodology is robust for any type of utility function.

As a reminder, we link in 1.1 the Discount Factor and the utility function, with equation (6):

$$m_t(S_T) = \beta \frac{u'(S_T)}{u'(S_t)} \quad (25)$$

We can derive this expression, with respect to S_T :

$$\frac{\partial m}{\partial S_T} = m'_t(S_T) = \beta \frac{u''(S_T)}{u'(S_t)} \quad (26)$$

Leading to:

$$u'(S_T) = \frac{u'(S_t)m_t(S_T)}{\beta} \quad (27)$$

And:

$$u''(S_T) = \frac{u'(S_t)m'_t(S_T)}{\beta} \quad (28)$$

We can now define the Relative Risk Aversion coefficient:

$$RRA(S_T) \stackrel{def}{=} -S_T \frac{U''(S_T)}{U'(S_T)}$$

Using (27) and (28):

$$RRA = -S_T \frac{\beta U'(S_t)m'(S_T)}{\beta U'(S_t)m(S_T)} = -S_T \frac{m'(S_T)}{m(S_T)}$$

We also define the Discount Factor as the ratio between the Risk Neutral Density q and the Historical Density p . Thus:

$$RRA = -S_T \frac{(q/p)'}{(q/p)} = -S_T \frac{(q'p - p'q)/p^2}{(q/p)}$$

Where $q' = \frac{\partial q}{\partial S_T}$ and $p' = \frac{\partial p}{\partial S_T}$. Leading to the final expression of the RRA:

$$RRA = S_T \left(\frac{p'}{p} - \frac{q'}{q} \right) \quad (29)$$

As we already have the distribution of both q and p , we can simply approximate the first derivative of both, using Taylor-Young formula ("central-difference"):

$$f'(a) \approx \frac{f(a+h) - f(a-h)}{2h}$$

Thanks to this methodology we are able to find the RRA at each time t as a function of both the maturity and the level of moneyness. In particular, we estimate it for the 30, 60 and 90 days maturities, for various levels of moneyness, in order to maximize data availability:

Maturity	Min Moneyness	Max Moneyness
30 days	96%	102%
60 days	95%	103%
90 days	95%	103%

2.2 Denoising the Relative Risk Aversion

At this point, we start from time series of RRA (for specific levels of moneyness, as described above, and the three maturities: 30, 60 and 90 days). These time series are noised. When we wish to apply classification methods to raw spectra in general, the final classification model is usually mostly based on the highest frequencies. Therefore, we need to smooth the signal. Standard Fourier analysis is probably the most widely applied signal processing tool, but it fails to provide information for non-stationary signals. The wavelet transform has the advantage of being a local time-frequency analysis, which enables us to analyze non-stationary data at different resolutions.

Indeed, the basis for our approach is the concept of multi-resolution, which consists in decomposing our signal into nested subspaces V_m representing the approximation of the signal at increasing resolutions m . As $V_m \subset V_{m+1}$, the projection of the signal onto the subspace V_{m+1} can be represented in terms of V_m and its orthogonal complement W_m . V_m carries the filtered information, while W_m deals with the details (i.e, the differences between approximations at successive resolutions).

As the scale (resolution) gets finer, the approximations become more accurate. Thus, by using wavelets, it is possible to analyze the classification model at the scale which matters to us. In our case, we are interested in keeping only the signal which corresponds to the lowest possible frequency, given the size of the dataset we use. In other words, what matters to us is the structural change in risk aversion, beyond the short-term shifts and noise.

The number of decompositions (in order to get the smoothest signal as possible) depends on the size of the signal, and is defined in (2.2). The length of the signal is constant, as we always use the same wavelet (Symlet of length 16).

$$\text{max level of decomposition} = \log_2 \left[\frac{\text{length of the signal}}{\text{length of the filter} - 1} \right]$$

In our case, $m = 10$ at present and we only keep the most filtered level (level 1). Figure 6 shows how the filtered RRA series look like over time for different levels.

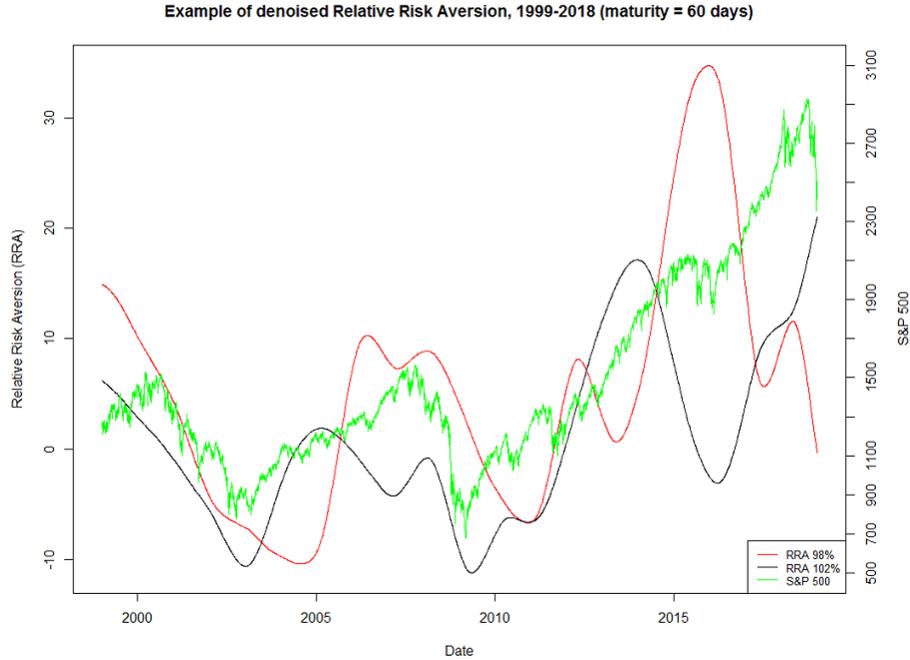


Figure 6: Relative Risk Aversion, S&P 500, maturity = 60 days

2.3 Extracting the daily Moments as Input Factors

Once we have generated the filtered time series of RRA per strike and maturity, using the wavelet methodology described above, we concentrate on reproducing three daily distributions of RRA, each of which corresponds to a specific maturity (respectively 30, 60 and 90 days).

In the 30 day-maturity case, we consider 7 series (from 96% to 102%, each being separated by 1% of moneyness), from which we compute the four moments (mean, variance, skew and kurtosis) for each date. In addition, we add a drift term, computing the first difference between the current mean level and the exponential moving average (and a decay factor of 0.2) of the mean. The same approach is followed in the 60 and 90-day cases, with both 9 series selected (from 95% to 103%, each being separated by 1% of moneyness).

The selection of the series has been performed carefully in order to deliver a good coverage of the distribution and a high degree of completeness of the underlying option data available.

3 The Asset Allocation Phase

We train three Naive Bayes Classifier with the 15 time series (moments of the three maturities, 30, 60 and 90 days) in order to forecast the direction of the S&P 500 over the next day. We use 5 years of data to train the models. Each of the models are retrained everyday in a rolling manner. The Naive Bayes Classifier is one of the simplest model in the "probability classifier" family. It needs few data to perform well and has relatively few parameters. The choice of training 3 different models, with 3 different targets (next daily return of the S&P 500), is made to enhance the robustness thanks to model averaging. The output is a probability of being in a specific return state over the next day. Because of the denoising part described in 2.2, the probability evolves smoothly over the time.

3.1 Description of the Classification Methodology

We split the in-sample dataset into two classes: 0's and 1's. The definition of these classes depends on the targeted model. A first model considers as class 1 occurrences of a daily return lower than -1% over the following day. The second model targets daily returns higher than +1% over the following day. The last model focuses on daily absolute returns higher than 1%.

For each of these three models, we share the same 15 input features. We note the i^{th} feature x_i . We note the class 1 as C_1 . The objective is to compute the probability to be in class 1, given today's set of 15 features:

$$p(C_1|x_1, \dots, x_{15}) \tag{30}$$

3.2 Mathematical formulation

The model we use is called *Gaussian Naive Bayes Classifier*. The two assumptions of the model are that features of the same class are independent and are distributed following a Gaussian distribution. The *Gaussian Naive Bayes Classifier* is not the only type of *Naive Bayes Classifier* but is the most used for two reasons. First it can handle continuous variables, and second it requires only two parameters to be fitted for each class (the sample mean and the sample standard deviation). We define:

Mathematical Notation	Definition
C_1	Class 1, defined depending on the model.
x_i	Particular feature i for a particular observation. Example: kurtosis 90 days for a particular observation = 0.32.
X	Set of 15 features for a particular observation. Corresponds to a vector: $X = \{x_1, \dots, x_{15}\}$.
$p(C_1 X)$	Probability to be in class 1, given today's set of 15 features.
$p(C_1)$	Probability of belonging in class 1 over the in-sample dataset. Also called the <i>prior</i> .
$p(X C_1)$	Probability that observation X occurs given the knowledge of belonging in class C_1 . Also called the <i>likelihood</i> .
$p(X)$	Called the <i>evidence</i> ie. the probability that observation X occurs. Does not change whether we compute the probability to be in class 0 or class 1. (see next equation)

Let's focus on the *evidence*, $p(X)$. The evidence can be written as, for both classes:

$$p(X) = \sum_{k=1}^2 p(C_k)p(X|C_k)$$

The evidence is said not to depend on any class. This is why it is considered as a pure scaling factor. However, we see that the evidence depends on X , the set of 15 features. This means that depending on the date, $p(X)$ will change.

Using Bayes' theorem, we can rewrite (30) as (given X the set of 15 features):

$$p(C_1|X) = \frac{p(C_1)p(X|C_1)}{p(X)} \quad (31)$$

$p(C_1)$ is called the prior, $p(X|C_1)$ the likelihood, and $p(X)$ the evidence. As X is the set of the 15 features, we can rewrite (31) as:

$$p(C_1|X) = \frac{p(C_1)p(x_1, \dots, x_{15}|C_1)}{p(X)} \quad (32)$$

The main assumption of the model is that features are independent conditionally on the class. Because of that, (32) becomes:

$$p(C_1|X) = \frac{p(C_1)p(x_1|C_1)p(x_2|C_1) \dots p(x_{15}|C_1)}{p(X)}$$

The evidence ($p(X)$) does not depend on the class, and is just a scaling factor. Finally, we can compute the probability to be in class 1:

$$p(C_1|X) = \frac{1}{p(X)}p(C_1)\prod_{i=1}^{15}p(x_i|C_1) \quad (33)$$

$p(C_1)$ is defined as the proportion of observations belonging in class 1, so it does not depend on the observation for which we try to predict $p(C_1|X)$. This means that it has no influence over the quintiles, as it is the same for all of the observations in the in-sample dataset.

3.3 Retrieving the Probability to be in a given Class, given the Information Acquired

What we need to do next is to estimate $p(x_i|C_1)$, with i corresponding to the i^{th} feature. We assume that the x_i are distributed following a Gaussian distribution. We need to estimate 90 parameters: 3 models * 15 features * 2 parameters (μ and σ) per feature. To do so, we compute the sample mean and sample standard deviation for the dates labeled as 1, for the specific feature i . This is done using the in-sample dataset.

Algorithm 1 Compute $p(x_i|C_1)$

```

for  $i \in \{1, \dots, 15\}$  do
  keep  $x_{i,C_t=C_1}$ 
  compute  $\mu_{i,1} = \overline{x_{i,C_t=C_1}}$ 
  compute  $\sigma_{i,1} = sd(x_{i,C_t=C_1})$ 
end for

```

We know X , the set of 15 features of the out-of-sample date. Since $x_i \sim \mathcal{N}(\mu_{i,1}, \sigma_{i,1}^2)$, we are able to compute $p(x_i|C_1)$. We can calculate the probability to label the researched date as 1 using (33):

$$p(C_1|X) = \frac{1}{p(X)}p(C_1)\prod_{i=1}^{15}p(x_i|C_1)$$

3.4 From the Probability to be in a given Class to the Allocation Rule itself

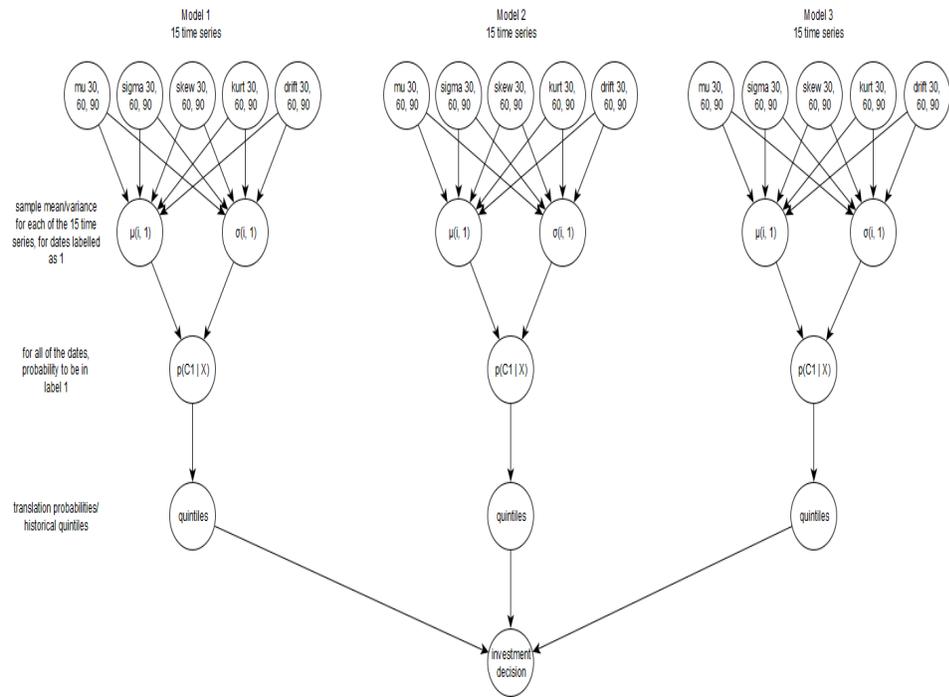
At this point, given any known set of 15 features, we are able to estimate the corresponding $p(C_1|X)$. Thanks to this, we are able to estimate the probability coming from the Naive Bayes Classifier, for each of the 3 models, trained on the past 5 years. From this, we consider the probability which applies to the day to come. Each of these three series of probabilities over the past five years are then clustered into quintiles, which enables us to define a triplet of three

quintiles related to the day to predict. The lower each quintile is, the higher the confidence to stay in the benign high return state.

Depending on this confidence level, we can cluster any market condition into three different broad states: a high risk of a drawdown category (0% allocation in equities, 100% in bonds), a neutral stance (classical 50/50 equity-bond allocation), and finally a highly probable positive return group (100% allocation in equities, 0% in bonds). The equity leg is the S&P 500, and the bond leg is the Bloomberg Barclays US Aggregate Index. The decision rule is as follows:

Allocation - Equities	Allocation - Bonds	Conditions (Quintiles)
100 %	0 %	All models: Quintile 1
0 %	100 %	At least one model: Quintile 5
50 %	50 %	All other combinations

We summarize below the methodology to go from 15 time series to the final investment decision using the following graph:



References

- [1] Giovanni Barone-Adesi and Hakim Dall'O. Is the price kernel monotone? Swiss Finance Institute Research Paper Series 10-03, Swiss Finance Institute, Jan 2010.
- [2] Giovanni Barone-Adesi and Kostas Giannopoulos. Backtesting derivative portfolios with filtered historical simulation. *European Financial Management*, 8:31–58, 2002.
- [3] Robert Bliss and Nikolaos Panigirtzoglou. Testing the stability of implied probability density functions. *Journal of Banking & Finance*, 26:381–422, 2002.
- [4] John Campbell. Consumption and the stock market: Interpreting experience. Technical Report 03, Swedish Economic Policy Review, 1996.
- [5] Greg B. Davies and Arnaud de Servigny. *Behavioral Investment Management*. McGraw-Hill Education, 2012.
- [6] Luiz F. Felix, Roman Kräussl, and Philip A. Stork. Implied volatility sentiment: A tale of two tails. Tinbergen Institute Discussion Paper 17-002, Tinbergen Institute Amsterdam, Jan 2017.
- [7] Stephen Figlewski and Justin Birru. Anatomy of a meltdown: The risk neutral density for the s&p 500 in the fall of 2008. *Journal of Financial Markets*, 15:151–180, 2012.
- [8] Joshua Rosenberg and Robert Engle. Empirical pricing kernel. *Journal of Financial Economics*, 64:341–372, 2002.
- [9] David Shimko. Bounds of probability. *Risk*, (6), 1993.
- [10] Christian Vial. *Forward-Looking Information in Portfolio Selection*. PhD thesis, University of St. Gallen, 2013.