



---

*Research article*

## Trajectorial asset models with operational assumptions

Sebastian Ferrando<sup>1,\*</sup>, Andrew Fleck<sup>2</sup>, Alfredo Gonzalez<sup>3</sup> and Alexey Rubtsov<sup>1</sup>

<sup>1</sup> Department of Mathematics, Ryerson University, 350 Victoria St., Toronto, Ontario M5B 2K3, Canada

<sup>2</sup> Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada

<sup>3</sup> Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Funes 3350, Mar del Plata 7600, Argentina

\* **Correspondence:** Email: [ferrando@ryerson.ca](mailto:ferrando@ryerson.ca); Tel: +1416979500x7415.

**Abstract:** The paper addresses the problem of providing a framework and an algorithm to evaluate super and sub replicating prices, for European options, having interesting risk-reward characteristics. A general operational framework is put forward and illustrated by an algorithmic construction of one-dimensional models for option pricing. Asset models are defined based on a class of investors characterized by how they operate on financial data leading to potential portfolio rebalances. Once observable variables are selected for modeling, necessary conditions constraining these variables and resulting from the operational setup are derived. Future uncertainty is then reflected in the construction of combinatorial trajectory spaces satisfying such constraints. As the risky asset unfolds, it can be tested dynamically for the validity of observable sufficient conditions that rigorously imply the validity of the models. The paper describes the resulting algorithmic construction of such trajectory spaces and, in the absence of probability assumptions, a minmax algorithm that is available to evaluate the super and sub replicating prices.

**Keywords:** trajectorial asset models; superhedging; option pricing; operational; Agent Based Models

**JEL Codes:** C61, C63, G11, G12, G13

---

### 1. Introduction

It is well known (see Eberlein and Jacod (1997) and Carassus et al. (2006)) that a large class of stochastic models put forward in the literature, and in order to obtain financial options' prices, provide non informative super and sub replicating bounds. More specifically, these bounds degenerate into *absolute bounds*. The latter refers to bounds that hold for all possible no arbitrage models and hence, are non informative (Merton (1973)).

Here, we address the following problem: can we justify simple models for asset evolution where we can evaluate sub and super replication prices and, in such a way that these quantities have useful risk-rewards characteristics?

An algorithmic construction, and its mathematical justification, that evaluates such bounds has already been provided in Degano et al. (2018). The said algorithm operates in a general probability-free context. To motivate such a setting, we point out that the support of a stochastic process is an offshoot of its assumed probability distribution and implicitly incorporates several analytical constraints. As such, it can include paths which can never appear as a rebalancing stock value (e.g. Brownian Motion has infinite speed). One speculates, from our perspective, that this latter feature is the cause for worst case bounds to degenerate into absolute bounds.

The issue at stake is the meaning of a price provided by a stochastic model (e.g. the Black and Scholes model): can you superhedge pathwise (in particular perfectly hedge) with such a price? The answer is no and this is so for several reasons: the price assumes continuous time rebalancing, market conditions for the applicability of the model are not spelled out, it is not clear how and when unfolding chart paths will correspond to a process's path, etc. The same issues are also present when superhedging is replaced by risk-taking investment, the above list of reasons will make the meaning of the model's profit and loss profile unclear. In this paper, we propose models that are robust, the superhedging property is then expected to be upheld for future unfolding trajectories. Superhedging is what you aim for as the cheaper alternative of exact hedging is not available given that real markets are incomplete (we are not allowing to use options for hedging). So it boils to a question on how robust and realistic the proposed models are and how tight the computed prices are. The whole paper is an attempt to address these issues.

The above reasoning suggests to place emphasis in, well justified, scenario construction as a prior step to imposing a probability distribution. Proceeding with this motivation, we show, by example, models with realistic bounds and useful risk-reward characteristics. The paper Carassus and Vargiolu (2018), using a completely different approach but emphasizing properties of the support of the process, does provide examples of useful price bounds. On the other hand, Kahalé (2017) employs convex programming for the estimation of bounds in a context of trading with a larger pool of assets.

We provide an implementation of the general algorithm from Degano et al. (2018). Most importantly, we put forward a general methodology, based on financial operations, observations and worst case estimation, to algorithmically construct a non-probabilistic asset model evolution. We show, via simple examples, how an operational approach based on observable variables can be put to work to price an option written on a single risky asset. The paper could be considered to provide a proof of concept by introducing an alternative approach to asset model construction and it is here illustrated in specific models. The methodology is general as it is based on constructing scenarios (trajectories), from a worst case perspective, and superhedging a future obligation. The mathematical framework is presented in Ferrando et al. (2019) and Degano et al. (2018). The financial setting is the most basic one, namely a riskless bank account paying zero interest rates and a risky asset with no transaction costs (in particular there is no bid-ask spread). The goal is to define models based on observable quantities that relate directly to a class of investors interested in objectively gauging the market price of a European option  $Z$  written on the asset. For simplicity we restrict ourselves to path independent European options. No analytical or probabilistic assumptions are required; the basic modeling object, to be constructed, is a set of trajectories  $\mathcal{S}$ . The algorithmic construction and

justification for the trajectory sets is the main thrust of the paper. It is known (Degano et al. (2018)) that, under some general conditions, knowledge of  $\mathcal{S}$  allows the evaluation of lower and upper bounds for the option's price. In the absence of a probabilistic structure, these bounds are defined via a minmax optimization giving sublinear/superlinear operators acting on general functions. The upper bound represents the minimal investment required for the existence of a portfolio that superhedges the option for each trajectory  $\mathbf{S} \in \mathcal{S}$  (with a dual meaning for the lower bound). Section 2 summarizes the notions and results that we use from Degano et al. (2018). We recall that worst option bounds in a stochastic setting can be obtained by considering all equivalent martingale measures (Föllmer and Schied (2013)).

The paper constructs trajectory sets  $\mathcal{S}$  via combinatorial definitions, this is in contrast to the modeling with random paths from a stochastic process. The trajectory sets are obtained by imposing constraints that in turn correspond to how a class of investors sample a financial chart and how they rebalance their portfolios as a response to changes of their data summaries. We allow for all possible trajectories satisfying the said constraints. The constructions naturally rely on operational constraints and observable quantities, the latter are estimated via a worst case methodology; on this topic we refer to Sniedovich (2016) for a general discussion (we do not rely on any specifics of that paper).

Expressed in other words: we describe an operational setup that provides a model for chart values  $x(t_i)$  and times  $t_i$  at which investors rebalance their portfolios. Such setting is based on observable variables that result from sampling a given chart and the ensuing portfolio rebalances. Trajectory sets model the times  $t_i$  and the values  $x(t_i)$  plus any other relevant variables that restrict the future manifold of trajectories (in this paper we concentrate on the variation but other possibilities could be incorporated as well). The trajectory construction is described by an algorithm.

There could be at least two quick criticisms to the setting outlined above; we deal with them briefly now and offer the rest of the paper as additional arguments. *i)* the models are subjective as they are investor dependent and so, in this way, are quite idiosyncratic. The proposed models are not subjective as they depend on objective actions, namely: specific sampling of charts, accumulation of quantities and portfolio rebalances as a response to values of observed variables. These actions define, implicitly, a nonempty class of investors. Our assumptions only involve results from measurements and operations, both allow for a precise discussion on how and when an unfolding chart  $x(t)$  may fail to be associated to some  $\mathbf{S} \in \mathcal{S}$ . In particular, the models do not intend to reproduce option's market prices (which may require considering multiple classes of investors and their bets on the risky asset) but to provide a particular investor class with an objective tool to assess risk-reward alternatives. *ii)* the combinatorial definition (i.e. considering all possible cases), used to construct trajectory sets based on worst case estimation, will produce unrealistic option prices. Clearly, the obtained option bounds could be too wide but, we argue, this feature will depend on the relevance of the proposed trajectory set. We show, by example, that the bounds we obtain for some models are realistic when compared to market prices. In fact, in the proposed type II models, we isolate a specific feature, based on a general property that can naturally be incorporated in any model construction, that restricts the effect of a worst case estimated parameter propagating into the price bounds' spread (see Section 11). Moreover, the fundamental risk-reward alternative, in the proposed setting, can be analyzed objectively given that we only rely on observable quantities. That is, one can know the empirical implications on risk and reward due to specific observable conditions not being satisfied by the unfolding chart. This is not the case in the usual modeling with stochastic processes in finance. As an example, we indicate the difficulty

in the analysis of parameter estimation and the unknown logical repercussions of calibrating to market values (e.g. the use of implied volatility, see Section 9.3.5 in J.Y. Campbell and MacKinlay (1997) as well as Schoutens et al. (2006)).

A third criticism relates to robustness of worst case estimation; this is an empirical concern that crucially depends on the relevance of the chosen model variables. We make specific comments on this issue at appropriate times along the paper, see for example Sections 6.3 and 9.2.

The approach does not preclude adding probabilities to the models; the methodology introduces possible future events prior to probabilities. Once a trajectory space is in place, if deemed appropriate, probabilities can be added a posteriori. In the case of a continuous time stochastic process model  $\{X_t\}$ , one could set  $x(t) = X_t(w)$  and consider that the samples  $x(t_i)$  are given by a sequence of, investor dependent, stopping times  $\{\tau_i\}$ ; that is,  $x(t_i) = X_{\tau_i(w)}(w)$ . Our approach dispenses with any modeling of  $x(t)$ , in particular, it avoids unnecessary implicit assumptions related to the stochastic process  $\{X_t\}$ .

Chart paths are observable but are not uniquely defined, e.g. their values will depend on the sampling frequency. In our models, a sampled trajectory uniquely emerges as a result of restricting ourselves to a class of investors whose operations act as preconditions resulting on the observed samples. This is analogous to a measurement setup in physics. The basic modeling ingredients are trajectory coordinates values  $S_i$  and times  $T_i$  associated to portfolio rebalances. The paper shows by example that there are interesting ways to construct trajectory sets with relevant financial features that produce models supporting a rigorous risk-reward analysis.

Noise driven stochastic models provide access to a powerful stochastic calculus and are put to good use to model an array of important financial situations (e.g. Brigo and Mercurio (2006)). From our point of view, a deficiency in the foundations of that modeling approach is that the sources of uncertainty (noise) appearing in the modeling equations are not, in the majority of cases, empirically linked to an actual phenomena. Therefore, repercussions on the stochastic model due to changing market conditions (e.g. non stationary effects) are difficult to detect. This foundational deficiency undermines the meaning of financial implications dependent on the underlying model. One speculates that the proliferation of asset models in finance indicates the lack of general, external, criteria to assess their suitability. Justification of some of the proposed models is then sought in statistics but, without external criteria of validity, a moderately large number of modeling assumptions provides unreliable conclusions. We refer to Pfliederer (2014) for a general (critical) discussion of models in finance and economics. We are aware of Feng et al. (2012), and related literature, where attempts are made to link investment behaviors in agent based models with stochastic emerging phenomena. We also remark on the economic literature on rule-based decisions (see Gilboa (2009)).

In contrast to the above described situation, we use the proposed methodology to establish rigorous results (see Section 6) showing when the models will succeed. The results allow to detect dynamically, i.e. as the chart unfolds, when hypothesis, required for validity of the said results, could fail. In this way we establish objective conditions implying falsifiable models.

The paper uses some overlapping of explanations across different, but related, sections. In this way the formal definitions of quantities, being them observable features or mathematical variables, is given an empirical meaning allowing to appreciate their roles in the constructions. In order to follow several of our arguments, it is crucial to keep in mind the empirical meaning of the variables used in the trajectory set construction.

The presentation is organized as follows, Section 2 introduces some notation and basic notions

needed from the general framework from Ferrando et al. (2019) and Degano et al. (2018). Section 3 provides an overview of the characteristics of charts that will be modeled by trajectory sets. Section 4 describes the basic operations performed on charts by our class of investors. Section 5 provides the paper basic construction given by a recursive algorithm. Section 5.1 is a first introduction to some of the parameters that are used in the construction of models. Section 5.2 describes an algorithm for the recursive step completing the construction introduced in Section 5. Section 6 establishes rigorous results giving observable conditions implying the validity of the models introduced. Section 7 presents an analysis of trajectory approximation error accumulation. Section 8 sets model parameters to worst case observed estimates. Section 9 completes the specifications from Section 8 by calibrating some remaining parameters, enumerating characteristics of the data employed and providing a simple description of the meaning of worst case. The section also illustrates several characteristics of estimated parameters. Section 9.5 displays price bounds for the different models introduced. Section 10 explains and illustrates the risk-reward tension of the models. Section 11 explains how a particular feature of type II models blocks the effect of worst case parameter estimation propagating into every possible node and, in this way, providing tighter option bounds. Section 12 provides an extended discussion of what we believe to be the most important take away from the paper.

The supporting material in Appendix 1 links to a file providing details of Matlab software implementing an algorithm that constructs the trajectory models as well as an algorithm that evaluates the price bounds and optimal hedging. The supporting material in Appendix 2 links to a bundle of files containing Matlab software, data and further output from the software.

## 2. Option pricing in trajectory models

This introduces some notation and notions from Degano et al. (2018) (that reference should be consulted for details).

A trajectory set model  $\mathcal{S}$ , in this article, is given by a set of sequences  $\mathbf{S} = \{\mathbf{S}_i \equiv (S_i, T_i, W_i)\}_{i \geq 0}$  with  $S_0 = s_0, T_0 = 0, W_0 = 0, \forall \mathbf{S} \in \mathcal{S}$ . Of course, other variables/coordinates are also possible.  $S_i$  models the values of the tradable risky asset at times  $T_i$  when an investor rebalances her portfolio and  $W_i$  models a variable representing a financial observable that presumably restricts the future possible manifold of asset values.

The financial context is of a risky asset  $x(t)$  and a zero interest bank account. The financial goal is to price a European option with time to expiration  $T$ . Given this context, our constructions will imply that for each  $\mathbf{S} \in \mathcal{S}$  there exists  $N(\mathbf{S})$  satisfying  $T_{N(\mathbf{S})} = T$ . The market model is given by  $\mathcal{M} = \mathcal{S} \times \mathcal{H}$  where  $H = \{H_i\} \in \mathcal{H}$  are sequences of functions  $H_i : \mathcal{S} \rightarrow \mathbb{R}$  representing trading strategies. Precise definitions are given below. The following conditional spaces will play a key role. For given  $k \geq 0$  and  $\mathbf{S} \in \mathcal{S}$ , set:

$$\mathcal{S}_{(\mathbf{S},k)} \equiv \{\mathbf{S}' \in \mathcal{S} : \mathbf{S}_i = \mathbf{S}'_i \forall 0 \leq i \leq k\}. \quad (1)$$

Notice  $\mathcal{S}_{(\mathbf{S},0)} = \mathcal{S}$  and that if  $\mathbf{S}' \in \mathcal{S}_{(\mathbf{S},k)}$ , then  $\mathcal{S}_{(\mathbf{S}',k)} = \mathcal{S}_{(\mathbf{S},k)}$ . The tuple  $(\mathbf{S}, k)$  will be referred to as a *node* (equivalently, we may refer to a node by  $\mathbf{S}_k$  as well).

**Definition 2.1** (Portfolio Set). A portfolio  $H$  is a sequence of (pairs of) functions  $H = \{\Phi_i = (B_i, H_i)\}_{i \geq 0}$  with  $B_i, H_i : \mathcal{S} \rightarrow \mathbb{R}$ .

- A portfolio  $H$  is said to be *admissible* for  $\mathcal{S}$  if for each  $\mathbf{S} \in \mathcal{S}$  there exists a nonnegative integer  $N_H(\mathbf{S}) \leq N(\mathbf{S})$  such that  $H_i(\mathbf{S}) = 0$  for all  $i \geq N_H(\mathbf{S})$ .
- A portfolio  $H$  is said to be *self-financing* at  $\mathbf{S} \in \mathcal{S}$  if for all  $i \geq 0$ ,

$$H_i(\mathbf{S}) S_{i+1} + B_i(\mathbf{S}) = H_{i+1}(\mathbf{S}) S_{i+1} + B_{i+1}(\mathbf{S}). \quad (2)$$

- A portfolio  $H$  is called *non-anticipative* if for all  $\mathbf{S}, \mathbf{S}' \in \mathcal{S}$ , satisfying  $\mathbf{S}'_k = \mathbf{S}_k$  for all  $0 \leq k \leq i$ , it then follows that  $\Phi_i(\mathbf{S}) = \Phi_i(\mathbf{S}')$ .

$H_i(\mathbf{S})$  and  $B_i(\mathbf{S})$  represent the number of shares and the funds in the bank account respectively, held at time  $T_i$ . We abuse the notation by writing  $H = \{H_i\}_{i \geq 0}$  instead of  $H = \{(B_i, H_i)\}_{i \geq 0}$ . Given  $\mathbf{S} \in \mathcal{S}$  and a self-financing portfolio  $H$ , the portfolio value defined by  $V_H(i, \mathbf{S}) = B_i(\mathbf{S}) + H_i(\mathbf{S})S_i$  is equal to  $V_H(i, \mathbf{S}) = V_H(0, S_0) + \sum_{k=0}^{i-1} H_k(\mathbf{S}) \Delta_k S$ , during the period  $[i, i + 1)$  for  $i = 0, \dots, N_H(\mathbf{S}) - 1$  and  $\Delta_k S \equiv (S_{k+1} - S_k)$ ; notice that  $V_H(0, S_0) = B_0(\mathbf{S}) + H_0(\mathbf{S})S_0$ . Clearly, to specify self-financing portfolios, it is enough to provide sequences  $H = \{H_i\}$  of non-anticipative functions and an associated real number  $V_0 = V_H(0, S_0)$ , since the corresponding  $B_i$  are obtained from (2).

We call  $\mathcal{M}$  *n-bounded* (or just simply *bounded*) if there exists a constant  $n$  so that :  $N_H(\mathbf{S}) \leq N(\mathbf{S}) \leq n$  for all  $\mathbf{S} \in \mathcal{S}$  and  $H \in \mathcal{H}$ .

The models are discrete in the sense that we index potential portfolio rebalances by integer numbers. Otherwise, stock charts and investment amounts can take values in general subsets of the real numbers and time can flow continuously.

Let  $Z : \mathcal{S} \rightarrow \mathbb{R}$  denote a general function, from time to time, we will refer to such function informally as the *derivative* or *payoff function*. Results in the paper mainly apply to path dependent European options but at some points we restrict to path independent options. This choice is to keep developments simple and so avoid getting diverted by inessential complications. Similarly, one could handle American options but doing so will distract from our main focus. The usual dynamic programming approach to pricing American options in discrete time (see for example Cutland and Roux (2012) section 7.6) can be easily adapted to superhedging such type of options.

The quantity  $\bar{V}_k(\mathbf{S}, Z, \mathcal{M})$  introduced in the next definition is the minimum capital needed, conditional on a node  $(S, k)$ , to superhedge  $Z$ .

**Definition 2.2** (Conditional Minmax Bounds). Given a market  $\mathcal{M} = \mathcal{S} \times \mathcal{H}$ ,  $k \geq 0$  and  $\mathbf{S} \in \mathcal{S}$ . Let  $Z$  be a function defined on  $\mathcal{S}$ , define:

$$\bar{V}_k(\mathbf{S}, Z, \mathcal{M}) \equiv \inf_{H \in \mathcal{H}} \sup_{\mathbf{S}' \in \mathcal{S}(\mathbf{S}, k)} [Z(\mathbf{S}') - \sum_{i=k}^{N_H(\mathbf{S}')-1} H_i(\mathbf{S}') \Delta_i S']. \quad (3)$$

Also set  $\underline{V}_k(\mathbf{S}, Z, \mathcal{M}) \equiv -\bar{V}_k(\mathbf{S}, -Z, \mathcal{M})$ . Since  $\bar{V}_0(\mathbf{S}, Z, \mathcal{M})$  and  $\underline{V}_0(\mathbf{S}, Z, \mathcal{M})$  depend on  $\mathbf{S}$  only through  $S_0$ , we adopt the notation  $\bar{V}(S_0, Z, \mathcal{M})$  and  $\underline{V}(S_0, Z, \mathcal{M})$  respectively. These quantities are called *price bounds*.

Stochastic models imply existence of a fair price interval for the derivative  $Z$  if the market does not contain arbitrage strategies. In our context, the arbitrage-free condition is replaced by the following notion of 0-neutral market.

**Definition 2.3** (0-Neutral Market). The market  $\mathcal{M}$  is *conditionally 0-neutral* at node  $(\mathbf{S}, k)$  if

$$\bar{V}_k(\mathbf{S}, Z = \mathbf{0}, \mathcal{M}) = 0.$$

For  $k = 0$ , we will just refer to  $\mathcal{M}$  as 0-neutral.

The notion of 0-neutral market, taken from Degano et al. (2018), was originally introduced in Britten-Jones and Neuberger (1996) and was considered equivalent to arbitrage-free in the latter reference. In our general setting, it is only a necessary condition for a market to be arbitrage-free while simultaneously allowing for arbitrage opportunities and a well defined theory of option pricing. 0-neutrality is the key condition to obtain a well defined fair price interval (Degano et al. (2018)).

Theorem 2.1 below is stated for a bounded market and  $\mathcal{H}$  assumed closed under addition. This is done to avoid introducing further notions, the result holds in more generality as indicated in Ferrando et al. (2019).

**Theorem 2.1** (Price Interval). Consider a bounded market  $\mathcal{M} = \mathcal{S} \times \mathcal{H}$  and a function  $Z$  defined on  $\mathcal{S}$ ; fix  $\mathbf{S} \in \mathcal{S}$  and  $k \geq 0$ . If  $\mathcal{H}$  is closed under addition and  $\mathcal{M}$  is conditionally 0-neutral at  $(\mathbf{S}, k)$ , then:  $\underline{V}_k(\mathbf{S}, Z, \mathcal{M}) \leq \bar{V}_k(\mathbf{S}, Z, \mathcal{M})$ .

Upper bounds allow to superhedge, namely, for a given  $\epsilon > 0$ , there exists  $H \in \mathcal{H}$ , an *optimal portfolio*, with initial capital  $\bar{V}(S_0, Z, \mathcal{M}) + \epsilon$  such that

$$\bar{V}(S_0, Z, \mathcal{M}) + \epsilon + \sum_{i=0}^{N(\mathbf{S})-1} H_i(\mathbf{S}) \Delta_i S \geq Z(\mathbf{S}) \quad \forall S \in \mathcal{S}. \quad (4)$$

We emphasize that the superhedging property is the crucial property in applications. An analogous property holds for the lower bound.

It then remains to address the issue of how to guarantee, for a given trajectory set model  $\mathcal{S}$ , that the conditional 0-neutral property holds at any given node  $(\mathbf{S}, k)$ . For our simple case of zero interest rates and a single risky asset this just means that for any given node  $(\mathbf{S}, k)$ , there exists  $\mathbf{S}'$  and  $\mathbf{S}''$  in  $\mathcal{S}_{(\mathbf{S}, k)}$  satisfying:

$$0 \leq (S'_{i+1} - S_i) \quad \& \quad (S''_{i+1} - S_i) \leq 0, \quad (5)$$

details are found in Degano et al. (2018). Appendix 1 complements the presentation of trajectory based markets describing an algorithm to evaluate option bounds and optimal portfolios.

In the paper we consistently differentiate between *chart* values (these being observable on a market) and *trajectory coordinates* (given by the models).

### 3. Operational setting

We refer to *charts* as the market quoted price  $x(t)$  at time  $t$  of a risky asset  $x$ . We formalize this by considering a chart to be a map  $x : \mathcal{T} \rightarrow \hat{\delta}_0 \mathbb{N}$  where  $\hat{\delta}_0 \equiv 0.01$  (of the monetary unit under consideration) and  $\mathcal{T}$  a time interval. One could naturally accommodate assets that take negative values but, for simplicity, we do not do so. Unless explicitly indicated otherwise, an interval of real numbers  $[a, b]$  is understood to mean  $[a, b] \cap \Delta \mathbb{Z}$ , for a given  $\Delta > 0$ , when the *variable represents time*. So,  $\mathcal{T} \subseteq \Delta \mathbb{Z}$ . We will use  $T$ , to denote the time length to maturity of a European option contract active

in a given interval  $[t_0, t_0 + T]$  and written on the asset  $x$ . Instead of replacing  $\hat{\delta}_0$  by 0.01 in the rest of the paper, we use  $\hat{\delta}_0$  for readability.

The following informal enumeration indicates the ingredients that we will use to construct trajectory sets in the paper. In this way, trajectory coordinates will result from a class of investors that rebalance their portfolios based on prescribed operations on chart data. For fixed real numbers  $\delta$ ,  $\delta_0$ ,  $\delta \geq \delta_0$ , and  $\Delta > 0$ :

1. Charts  $x$  are sampled at dynamic times  $r_l$  according to  $\delta_0$ .
2. Time intervals have a lower bound resolution  $\Delta$  in the sense that  $(r_{l+1} - r_l) \geq \Delta > 0$ .
3. Times  $t_i$ , for the  $i$ -th portfolio rebalance, satisfy  $t_i \in \{r_l\}$  and are given by an investor prescribed threshold  $\delta$ .
4. Sampled variation  $w(x, t)$  is accumulated for the samples  $x(r_l)$ .
5. There is a dynamical number of portfolio rebalances  $N(x, [t_0, t_0 + T])$  that take place in  $[t_0, t_0 + T]$ .

Hence, we deal with a class of investors that will react to chart changes of size  $\delta$  and observe chart values at a certain time resolution  $\Delta$  and spatial resolution  $\delta_0$ . From this perspective, these portfolio managers want to evaluate investment opportunities related to a European option. They want to compare market prices relative to their operational portfolio rebalancing.

Given this setup, the operations of sampling and portfolio re-balancing performed on chart samples will create the associations:

$$x(t_i) \rightarrow S_i, t_i \rightarrow t_0 + T_i, \text{ which we write compactly as: } x(t_i) \rightarrow \mathbf{S}_i \text{ or } x \rightarrow \mathbf{S}.$$

For the precise definition of the notion of *association*, we refer to Definition 6.1. Notice that  $w(x, t_i)$  is neglected above; a specific  $w(x, t_i)$  is introduced in Section 4 and its role is explained elsewhere in the paper. Historical, worst case, values of  $w(x, t)$  and the operational setup, will limit the minimum and maximum possible number of samples as well as their ranges. In this way, we will obtain necessary constraints restricting the possible values of  $\mathbf{S}_{i+1}$  for given  $\mathbf{S}_0, \dots, \mathbf{S}_i$ . Such restrictions are used to define  $\mathcal{S}$ ; therefore, the meaning of  $\mathbf{S}_{i+1}$  is that of possible values at which an investor will re-balance her portfolio given a prescribed way of processing the data. Section 6 discusses the reasons why the association  $x(t_i) \rightarrow \mathbf{S}_i$  can fail. More precisely, we provide hypothesis so that the values of an unfolding chart will be associated to a trajectory in  $\mathcal{S}$ . We also analyze the related issue of the error build up whenever the association  $x \rightarrow \mathbf{S}$  is given by an approximation (this is done in Section 7).

An inequality of the form  $\delta \leq |x(t') - x(t)|$ ,  $t < t'$ , as well as analogous expressions, will be referred to as a  $\delta$ -increment (notice the abuse of language as the value of  $x(t')$  could be smaller than  $x(t)$ ).

#### 4. Charts and investors' operations

Set

$$T = M_T \Delta, \delta = F \delta_0, \delta_0 = G \hat{\delta}_0 \text{ so if } E \equiv F G, \delta = E \hat{\delta}_0, \quad (6)$$

where  $M_T, F, G$  are positive integers. Our investor will fix  $\delta$ , by a process of calibration, and will not rebalance her portfolio for changes smaller than  $\delta$ .  $\Delta$  is the smallest time resolution at which the investor will observe charts. The parameter  $\delta_0$  will provide the grid  $\delta_0 \mathbb{N}$ , we then require, by construction, that coordinates  $S_i \in \delta_0 \mathbb{N}$  and  $W_i \in \delta_0 \mathbb{N}$ . Taking  $\delta_0 = \hat{\delta}_0$  is the natural choice and



simplifies the developments. Our only reason to consider  $\delta_0 > \hat{\delta}_0$  is to save on computations, something we need to do to produce output for the values of  $T$  and  $\Delta$  we have chosen to illustrate the models.

**Definition 4.1.** [Dynamic sampling times] Given  $\delta_0 > 0$ , a chart  $x$  and interval  $[t_0, t_0 + T]$ ; a sequence of increasing dynamic sampled times is given by  $r = r(x) = \{r_l\}_{l=0}^L \subseteq [t_0, t_0 + T]$  (so  $r_l \in \Delta \mathbb{Z}$ ),  $L = L(x, [t_0, t_0 + T])$ ,  $r_0 \equiv t_0$ , satisfying:

$$\delta_0 \leq |x(r_{l+1}) - x(r_l)|, \quad 0 \leq l < L - 1, \quad r_L \equiv t_0 + T, \quad (7)$$

notice that  $\delta_0 \leq |x(r_L) - x(r_{L-1})|$  may or may not hold. Consistently with these properties, we also require the times  $r_l$  to satisfy: if  $r_l < t < r_{l+1}$ ,  $t \in \Delta \mathbb{Z}$ , then  $\delta_0 > |x(t) - x(r_l)|$ .

The collection  $\{r_l\}$  as above is referred as *sampling times*. The measured times  $\{r_l\}$  are independent of a monetary unit re-scaling, i.e. after a change of the form  $\delta_0 \$ \rightarrow \delta'_0 = c\delta_0 \phi$ .

From a financial point of view, and when dealing with a long period of historical data, it is more reasonable to define the sampling times in terms of proportional changes  $(x(r_{l+1}) - x(r_l))/x(r_l)$  (i.e. *returns*). Similarly for the rebalancing times  $t_i$ .

**Definition 4.2.** [Dynamic rebalancing times] Given  $\delta > 0$ , a chart  $x$  and interval  $[t_0, t_0 + T]$ , a sequence of increasing dynamic rebalancing times is given by  $t = t(x) = \{t_i\}_{i=0}^N \subseteq \{r_l\}_{l=0}^L$ ,  $N = N(x, [t_0, t_0 + T])$ ,  $t_0(x) \equiv t_0$ , satisfying:

$$\delta \leq |x(t_{i+1}) - x(t_i)|, \quad 0 \leq i < N - 1, \quad t_N \equiv t_0 + T. \quad (8)$$

The times  $t_i$  are chosen in a *greedy* way, namely given  $t_i = r_{l_i}$ ,  $t_{i+1}$  is the smallest element  $r_{l_{i+1}}$  in  $\{r_l\}$  satisfying (8) if such element exists, otherwise  $r_{l_{i+1}} = T$  and  $N = l_{i+1}$ .

Notice that  $\delta \leq |x(t_N) - x(t_{N-1})|$  may or may not hold. The latter ambiguity will imply that the trajectory set will need to account for a larger set of possible values for  $x(t_N)$ , conditional on given values  $S_0, \dots, S_{N-1}$  (see extended discussion in Section 5.1).  $N = N(x, [t_0, t_0 + T])$  is the number of times a portfolio is rebalanced in the interval  $[t_0, t_0 + T]$ , it follows that  $N(x, [t_0, t_0 + T]) \leq L(x, [t_0, t_0 + T])$ . Portfolio rebalancing times  $t_i = t_i(x)$  satisfy  $t_0(x) = t_0 = r_0 < t_1 < \dots < t_{N-1} \leq r_{L-1}$ . Setting  $t_N = r_L = t_0 + T$  as a rebalancing time, as we do, is inconsequential in the absence of transaction costs (as is our standing assumption in the paper).

We can write for  $0 \leq i \leq N - 1$ :

$$\Delta_i x \equiv x(t_{i+1}) - x(t_i) \equiv (\hat{k}_{i+1} - \hat{k}_i) \hat{\delta}_0 \equiv \hat{m}_i \hat{\delta}_0. \quad (9)$$

The relevance of using a  $\hat{\cdot}$  notation in  $\hat{m}_i$ , and in related variables, will be made clear later on in the paper.

$$\Delta_i w \equiv w(x, t_{i+1}) - w(x, t_i) \equiv \sum_{j=0}^{s_i-1} |x(r_{l_i+j+1}) - x(r_{l_i+j})| \equiv (\hat{j}_{i+1} - \hat{j}_i) \hat{\delta}_0 \equiv \hat{\delta}_0 \sum_{j=0}^{s_i-1} |\hat{p}_j|, \quad (10)$$

where  $t_i = r_{l_i} < \dots < r_{l_i+s_i} = t_{i+1}$ , i.e.  $l_{i+1} = l_i + s_i$  and  $0 \leq s_i$ . Notice the possibility of  $s_i = 0$ , this could happen only for the case  $i = N - 1$  and there are no  $\hat{\delta}_0$ -increments in between  $t_i$  and  $t_{i+1} = T$  and so  $x(t_{i+1}) - x(t_i) = 0$ . Setting  $w(x, t_0) = 0$  it follows that  $w(x, t_i) = \sum_{l=0}^{l_i-1} |x(r_{l+1}) - x(r_l)|$ .

Also,

$$\Delta_i t \equiv t_{i+1} - t_i = r_{l_i+s_i} - r_{l_i} \equiv (n_{i+1} - n_i) \Delta \equiv q_i \Delta. \quad (11)$$

Notice that,

$$s_i \Delta \leq \sum_{l=l_i}^{l_i+s_i} (r_{l+1} - r_l) = q_i \Delta, \text{ i.e. } 0 \leq s_i \leq q_i \leq M_T. \quad (12)$$

Given the above definitions and notation, we can write:

$$\Delta_i x = \sum_{j=0}^{s_i-1} (x(r_{l_i+j+1}) - x(r_{l_i+j})) = \hat{\delta}_0 \sum_{j=0}^{s_i-1} \hat{p}_j = \hat{m}_i \hat{\delta}_0$$

so

$$\sum_{j=0}^{s_i-1} \hat{p}_j = \hat{m}_i. \quad (13)$$

Moreover, given that times  $t_i$  are chosen in a greedy fashion we also have for any  $1 \leq R < s_i$ ,

$$|\sum_{j=0}^{R-1} \hat{p}_j| < E, \quad G \leq |\hat{p}_j| < 2E, \quad 0 \leq j \leq s_i - 2, \quad (\text{and so } |\hat{p}_{s_i-1}| \leq (|\hat{m}_i| + E)), \quad (14)$$

where  $G \leq |\hat{p}_j|$  follows from  $\hat{\delta}_0 |\hat{p}_j| = |x(r_{l_i+j+1}) - x(r_{l_i+j})| \geq \delta_0$ . Given that  $\hat{m}_i \hat{\delta}_0 \equiv x(t_{i+1}) - x(t_i)$ , hence  $|\hat{m}_i| \hat{\delta}_0 \geq \delta$  is guaranteed for  $i < N - 1$  and so

$$|\hat{m}_i| \geq E, \quad \text{for } 0 \leq i < N - 1. \quad (15)$$

We have introduced the dichotomy between  $\hat{\delta}_0$  and  $\delta$ , the former being an absolute minimum value for chart changes and the latter being investor dependent.

Analogous to the use of  $\delta_0$ , and also to reduce the execution time of a software implementation, one could introduce  $\Delta_0 \geq \hat{\Delta}_0$  where  $\hat{\Delta}_0$  is the smallest time resolution supported by the data (i.e.  $t \in \hat{\Delta}_0 \mathbb{Z}$ ). In that case one then will require  $(r_{l+1} - r_l) \geq \Delta_0$  and  $r_l \in \hat{\Delta}_0 \mathbb{Z}$ .

To summarize the above:

- $r_l$  are the sampling times.
- $t_i \in \{r_l\}$  will denote the portfolio rebalancing times.
- $s_i$  is the number of chart samples in between two portfolio rebalances.
- $q_i$  the possible number of time intervals of size  $\Delta$  in between two consecutive portfolio rebalances.
- $\hat{m}_i$  the possible number of  $\hat{\delta}_0$  chart units in between two consecutive portfolio rebalances.
- $\hat{p}_j$  the possible number of  $\hat{\delta}_0$  chart units in between two consecutive samples.
- $N(x, [t_0, t_0 + T])$  total number of portfolio rebalances in  $[t_0, t_0 + T]$ .

We will rely on  $s_i \leq q_i$  given by (12), that is, we are taking an upperbound. This does not create any drawbacks as we will be replacing the role of  $s_i$  by  $J$  with  $J \leq q_i$ .

Through the process of calibration, the time and space intervals reflect actions of realistic investors, therefore, we expect the models to reflect features of options' market values. This observation is confirmed for type II models, we refer to Figures 13, 14, 16 and 17 for examples.

#### 4.1. Unfolding chart parameters

The definitions in this section are only first required in Section 6; we place the definitions here as they refer to a given chart  $x$  and time interval  $[t_0, t_0 + T]$ . Moreover, they provide a preliminary meaning to the corresponding parameters used for model building in Section 5.2.

**Definition 4.3** (Unfolding  $S$ - $T$  Nodes).

$$\mathcal{N}_E(x, [t_0, t_0 + T]) \equiv \left\{ \left( \frac{x(t_{i+1}) - x(t_i)}{\hat{\delta}_0}, \frac{t_{i+1} - t_i}{\Delta} \right) : 0 \leq i \leq P - 1 \right\}, \quad (16)$$

where the  $\{t_i\}$  are as in Definition (4.2) and  $P \equiv P(x, [t_0, t_0 + T]) \equiv N \equiv N(x, [t_0, t_0 + T])$  (the latter as introduced in Definition 4.2) if  $\delta \leq |x(t_N) - x(t_{N-1})|$  or  $P \equiv P(x, [t_0, t_0 + T]) \equiv N - 1$  otherwise. Notice the possibility that  $\mathcal{N}_E(x, [t_0, t_0 + T]) = \emptyset$ . We also set

$$\tau(x, [t_0, t_0 + T]) \equiv (t_P - t_0), \quad (17)$$

notice the possibility  $\tau(x, [t_0, t_0 + T]) = (t_0 - t_0) = 0$  whenever  $N = 1$ .

**Definition 4.4** (Minimal wait for a  $\delta$ -increment).

$$\Gamma(x, [t_0, t_0 + T]) \equiv \min_{0 \leq \gamma \leq T, \gamma \in \Delta \mathbb{N}} \{ \gamma : \forall [t', t' + \gamma] \subseteq [t_0, t_0 + T], \exists s \in (t', t' + \gamma], \quad (18)$$

$$|x(s) - x(t')| \geq \delta \}.$$

That is,  $\Gamma(x, [t_0, t_0 + T])$  is the smallest length such that for *any*  $t$  satisfying  $[t, t + \Gamma(x, [t_0, t_0 + T])] \subseteq [t_0, t_0 + T]$  there exists  $s \in (t, t + \Gamma(x, [t_0, t_0 + T]))$  such that  $\delta \leq |x(s) - x(t)|$ . We adopt the usual definition that the minimum over an empty set is equal to  $T$ . Clearly, if  $x$  has  $n$   $\delta$ -increments in  $[t_0, t_0 + T]$  then  $\Gamma(x, [t_0, t_0 + T]) \leq (T - (n - 1)\Delta)$  and if  $n = 1$  we may have  $\Gamma(x, [t_0, t_0 + T]) = T$ .

### 5. Trajectory sets: recursive construction

The previous section assumed the availability of an observable chart  $x(t)$  and introduced sequences of integers  $s_i, q_i, \hat{m}_i, \hat{p}_j$ ,  $0 \leq j < s_i$  and  $N$ . The integers  $\hat{k}_i, n_i$  and  $\hat{j}_i$  are *derived* quantities, in the sense that each set  $s_i, q_i, \hat{m}_i, \hat{p}_j$ ,  $0 \leq j < s_i$  gives a triple  $(\hat{k}_i, n_i, \hat{j}_i)$ . Necessarily, the meaning of the variables impose some *constraints* obeyed by these sequences. We will construct trajectory models by allowing trajectories to be jointly specified by the mentioned constraints. The end result being the definition of a possible set of chart samples, time rebalances and sampled variation values.

In order to avoid misunderstandings, one should keep in mind that the previously introduced variables:  $r_i, t_i, \hat{m}_i, q_i, s_i, N(x, [t_0, t_0 + T])$  and  $\hat{p}_j$  are either observable or operationally prescribed and, in that way, are a result of acting on unfolding charts. In order to avoid a confusing proliferation of notation, the same symbols will be used to denote the corresponding modeling variables used to define trajectory sets but with the following exemptions:  $N(x, [t_0, t_0 + T])$  will be replaced by  $N(\mathbf{S})$ ,  $\hat{m}_i$  for  $m_i$ ,  $\hat{k}_i$  for  $k_i$ ,  $\hat{j}_i$  for  $j_i$  and  $\hat{p}_j$  for  $p_j$ . In the last four cases the reason to drop the  $\hat{\cdot}$  notation is due to the fact that  $x(t) \in \hat{\delta}_0 \mathbb{N}$  while trajectory coordinates  $S_i, W_i \in \delta_0 \mathbb{N}$ .

The two related, but different, uses of the same notation for charts and trajectories should not be identified of course, doing so will preclude the analysis of how well the model approximates the

observational setting. Namely, how relevant are the proposed trajectories to fit the values of the unfolding chart at the re-balancing times.

Trajectory coordinates  $S_i$  and  $W_i$  will take values in  $\delta_0 \mathbb{N}$ . As already indicated, the possibility to take  $\delta_0 > \hat{\delta}_0$  will allow us to save on computations and this is the reason why we introduce  $\delta_0$ . Conceptually, the reader could consider  $\delta_0 = \hat{\delta}_0$  but in practical terms we noticed the need for  $\delta_0 > \hat{\delta}_0$  due to the time horizons and sampling times we have considered (these choices were dictated by the data we got access to). The choice  $\delta_0 > \hat{\delta}_0$  means we will be *approximating* the values  $x(t_i)$  which introduces the issue of error accumulation, we deal with this issue in Section 7.

We describe now a formal recursive definition of a generic (abstract) trajectory set  $\mathbf{S}$ . Section 5.2 completes the formal definition by providing the construction needed in the recursive step. Such constructions are driven by a set of parameters providing a class of trajectory sets. We assume given an interval  $[t_0, t_0 + T]$ , where a European option of interest is active. We set the notation  $\mathbf{S} = \{S_i \equiv (S_i, t_0 + T_i, W_i)\} \equiv \{(k_i \delta_0 + s_0, t_0 + n_i \Delta, j_i \delta_0)\}$ ,  $k_i \in \mathbb{Z}$ ,  $n_i < n_{i+1}$ ,  $0 \leq j_i < j_{i+1}$ ,  $n_i, j_i \in \mathbb{N}$ . The recursive step will split into type I and type II models, in the former case, as already indicated, the variable  $W_i$  will be ignored something we remind readers from time to time.

Trajectories  $\mathbf{S}$  will then be specified by setting  $s_0 = x(t_0) + k_0 \delta_0 = x(t_0)$ , for  $x(t_0)$  given,  $t_0 + T_0 = t_0 + n_0 \Delta = t_0$ ,  $W_0 = j_0 \delta_0 = 0$  and proceeding recursively: given a triple  $(k_i, n_i, j_i)$  we should provide a set of *admissible* values  $(k_{i+1}, n_{i+1}, j_{i+1})$  (these are specified in Section 5.2); for each such admissible triple we set:

$$\Delta_i S \equiv (S_{i+1} - S_i) = (k_{i+1} - k_i) \delta_0 = m_i \delta_0, \quad (19)$$

$$\Delta_i W \equiv (W_{i+1} - W_i) = (j_{i+1} - j_i) \delta_0 = \sum_{j=0}^{s_i-1} |p_j| \delta_0, \quad (20)$$

$$\Delta_i T \equiv (T_{i+1} - T_i) = (n_{i+1} - n_i) \Delta = q_i \Delta, \text{ where } m_i, p_j \in \mathbb{Z}, 1 \leq |p_j| \text{ (as } \delta_0 \leq \delta_0 |p_j|), \quad (21)$$

$$1 \leq q_i \leq M_T, 0 \leq s_i \leq q_i, s_i, q_i \in \mathbb{N}. \quad (22)$$

Moreover, define

$$N(\mathbf{S}) = i \text{ where } T_i = T,$$

notice that  $N(\mathbf{S})$  exists as  $\Delta_i T \geq \Delta$  for any  $0 \leq i < N(\mathbf{S})$ .

For the case  $i < N(\mathbf{S}) - 1$  we will also require  $|m_i| \geq \frac{\delta}{\delta_0} = F$ . Also, for any  $1 \leq R < s_i$  and  $i \leq N(\mathbf{S}) - 1$ ,

$$\sum_{j=0}^{s_i-1} p_j = m_i, \quad \left| \sum_{j=0}^{R-1} p_j \right| < F, \quad 1 \leq |p_j| < 2F, \quad 0 \leq j \leq s_i - 2, \quad (23)$$

$$\text{(and so } 1 \leq |p_{s_i-1}| \leq (|m_i| + F)).$$

Given  $(S_i, T_i, W_i)$ , we introduce the convenient notation for type II models:

$$\mathcal{N}_A(S_i, t_0 + T_i, W_i) = \{(S_{i+1} = k_{i+1} \delta_0 + s_0, t_0 + T_{i+1} = t_0 + n_{i+1} \Delta, \quad (24)$$

$$W_{i+1} = j_{i+1} \delta_0), \text{ with } (k_{i+1}, n_{i+1}, j_{i+1}) \text{ admissible}\}, \quad (25)$$

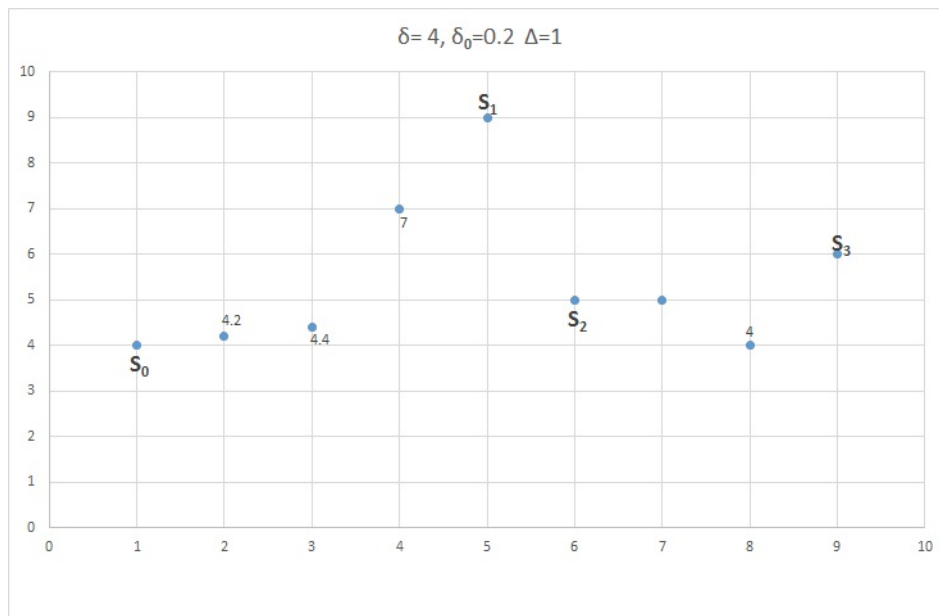
and for type I models:

$$\mathcal{N}_A(S_i, t_0 + T_i) = \{(S_{i+1} = k_{i+1} \delta_0 + x(t_0), t_0 + T_{i+1} = t_0 + n_{i+1} \Delta), \quad (26)$$

with  $(k_{i+1}, n_{i+1})$  admissible). (27)

Notice that the values  $(p_j)_{j=0}^{s_i-1}$ ,  $m_i$ ,  $q_i$  need to be prescribed in order to obtain the set of admissible triples  $(k_{i+1}, n_{i+1}, j_{i+1})$  (or tuples  $(k_{i+1}, n_{i+1})$  for type I models). This will be done in Section 5.2 for each of the two types of models. Subsequent sections indicate how to estimate the models' parameters. Estimation in this paper involves worst case estimation on historical data. The precise meaning of the latter, we expect, to be quite evident in each instance; specifics are provided in Section 9. It is a simple minded methodology that requires that informative variables have been chosen to define trajectories. *Informative* here means that the parameters that confine the model variables are contained in worst case ranges that reflect observable and possible financial situations that impact the values of the risky asset and the class of investors under consideration.

**Remark:** Type II models will enforce (historical) constraints on  $W_i$  and, in this way, limit the number of possible future trajectories emanating from the node  $(S_i, t_0 + T_i, W_i)$ . For type I models, such historical constraints are not in effect, it follows that the coordinates  $W_i$  are irrelevant and so we have only the coordinates  $(S_i, t_0 + T_i)$ .



**Figure 1.** Illustration of chart values and portfolio rebalances.  $t_0 = r_0 = 1$ ,  $r_1 = 2$ ,  $r_2 = 3$ ,  $r_3 = 4$ ,  $r_4 = t_1 = 5$ ,  $r_5 = t_2 = 6$ ,  $r_6 = 8$ ,  $r_7 = t_3 = 9$ ,  $(m_1, q_1) = (25, 4)$ ,  $(m_2, q_2) = (-20, 1)$ ,  $(m_3, q_3) = (5, 3)$ . Where  $t_0 = 1\Delta$ ,  $T + t_0 = 9\Delta$ .

### 5.1. Model Parameters. Last Portfolio Rebalance (accounting for the possibility that $\delta > |x(t_N) - x(t_{N-1})|$ )

This section provides informal explanations of some of the parameters that will play a role in the algorithm constructing the two model types.

Recall the notation  $N = N(x, [t_0, t_0 + T])$  from Definition 4.2. As already hinted, the fact that the last rebalancing time  $t_N$  does not necessarily satisfy  $\delta \leq |x(t_N) - x(t_{N-1})|$  implies that a node  $(S_{N-1}, t_0 + T_{N-1}, W_{N-1})$  associated to  $(x(t_{N-1}), t_{N-1}, w(x, t_{N-1}))$  (such association also gives  $N(\mathbf{S}) = N(x, [t_0, t_0 + T])$ )

will be treated differently to other types of nodes. Parameters at nodes  $(S_i, t_0 + T_i, W_i)$  (leading to  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$ ), with  $i < N(\mathbf{S}) - 1$ , should prescribe that portfolio rebalances will take place only after the chart has moved by at least an amount  $\delta$  relative to its present value. At a node with  $i = N - 1$  it is not known which of the alternatives:  $\delta \leq |x(t_N) - x(t_{N-1})|$  or  $\delta > |x(t_N) - x(t_{N-1})|$ , will actually occur. Therefore, at a node  $(S_{N-1}, t_0 + T_{N-1}, W_{N-1})$  we will need to consider both possibilities. It follows that such a node will lead to a larger admissible set  $\mathcal{N}_A(S_{N-1}, t_0 + T_{N-1}, W_{N-1})$ .

In fact, the equality  $i = N - 1$  is uncertain. In order to deal with this uncertainty, we make use of additional information by means of relevant parameters set to worst case values. In this way we can partially discern if  $i < N - 1$  holds and in so doing we reduce the number of cases leading to the enlargement of  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$  just mentioned. We will make use of a parameter  $\Gamma$  that, when associated to a chart  $x$ , will have the following meaning: it is the smallest time length such that for *any*  $t$  (in the historical data set used for estimation),  $x$  moves up or down, from its current value  $x(t)$ , by at least  $\delta$  in the interval  $[t, t + \Gamma]$ .  $\Gamma$  corresponds to  $\Gamma(x, [t_0, t_0 + T])$  introduced in Section 4.1; the precise relationship is given by (75).

Another parameter,  $\tau$ , will also be used for the same purpose. The index  $i$  corresponding to  $(S_i, t_0 + T_i, W_i)$  is not being encoded in our data structure. This means that we will replace references to  $i$  by  $T_i$ . This implies the need to introduce  $\tau$  used to reflect the minimum time length after which  $\delta$ -increments may stop.  $\tau$  corresponds to  $\tau(x, [t_0, t_0 + T])$  introduced in Section 4.1; the precise relationship is given by (78).

In general, one would like to enforce on  $\mathcal{S}$  constraints that can be obtained empirically or as a result of further available information. Proceeding in this way will make  $\mathcal{S}$  smaller, and hence implying more accurate option's bounds, while still obeying our general philosophy of worst case construction.

## 5.2. Trajectory sets: recursive algorithmic step

This section states the recursive algorithmic construction leading to the two types of trajectory sets to be considered. The specifications complete the general recursive framework introduced in Section 5 where, we recall, an interval  $[t_0, t_0 + T]$  was given and only the sets  $\mathcal{N}_A(S_i, t_0 + T_i)$  and  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$  were missing in order to complete the specifications for type I and type II models respectively. All models introduced below assume a fixed  $T$  and  $\Delta$ .

The definitions of each of the models below use a list of parameters which enable the construction of each trajectory set. Section 5.3 provides an intuitive explanation of some features of the definitions. Later sections indicate how to set the values of these parameters being by the process of calibration, estimation or by using relationships that relate the parameters.

The notation  $\mathcal{N}_E$  ( $E$  for *empirical or estimated*) represents a given set of pairs  $(m, q)$ ,  $m \in \mathbb{Z}$ ,  $q \in \mathbb{N}$ . Such set will be estimated as described in Section 9. Define  $q_{min} \equiv \min\{q : (m, q) \in \mathcal{N}_E\}$ .

We point out to the appearance below of  $\sum_{j=0}^{J-1} p_j$  as a replacement for  $\sum_{j=0}^{s_i-1} p_j$  (similar comments apply to related quantities).  $J$  is needed given that we only know  $s_i \leq q_i$  (see, (12)); to account for this uncertainty we then consider all  $J$  satisfying  $J \leq q$  for a given  $(m, q)$ .

As an explanatory comment, we mention that all nodes for which we can ascertain  $T_i < \tau$  will have the same set of admissible nodes  $\mathcal{N}_{\mathcal{A}}(S_i, t_0 + T_i, W_i)$  for type II models (or  $\mathcal{N}_{\mathcal{A}}(S_i, t_0 + T_i)$  for type I models). In this case we foresee more  $\delta$ -increments are forthcoming and construct the trajectory set accordingly.

To repeat, we define possible values recursively, therefore, the cases below assume  $(S_i, t_0 + T_i, W_i)$

given (only  $(S_i, t_0 + T_i)$  is needed for the type I models); the initialization  $S_0 = s_0, T_0 = 0, W_0 = 0$  is in effect.

### 5.2.1. Type I models

#### Input Parameters

1.  $0 \leq \tau \leq T$ .
2.  $F, G \in \mathbb{N}$  ( $\delta \equiv F \delta_0, \hat{\delta}_0 \equiv G \hat{\delta}_0$ ).
3. A set  $\mathcal{N}_E$ , if  $(m, q) \in \mathcal{N}_E$ , then  $|m| \geq F, q \geq 1, m \in \mathbb{Z}, q \in \mathbb{N}$ .
4.  $\Gamma \in \Delta \mathbb{N}$  satisfying  $q_{min} \Delta \leq \Gamma \leq T$ .

#### Definition

1.  $\mathcal{N}_{S,T}^<(T_i) \equiv \{(m, q) \in \mathbb{Z} \times \mathbb{N} : -F < m < F, q \Delta + T_i = T, T_i < T\}$ .

Define the statements:

$$\begin{aligned} \mathbf{A} &\equiv [(\Gamma \leq (T - T_i) \ \& \ (\tau \geq \Delta)) \ \text{or} \ ((T_i < \tau) \ \& \ (\Delta q_{min} \leq (T - T_i))], \\ \mathbf{B} &\equiv [((T - T_i) < q_{min} \Delta)], \\ \mathbf{C} &\equiv \neg \mathbf{A} \ \& \ \neg \mathbf{B} = [(\tau \leq T_i) \ \& \ (q_{min} \Delta \leq (T - T_i)) \ \& \ (((T - T_i) < \Gamma) \ \text{or} \ (\tau = 0))]. \end{aligned}$$

If  $T_i = T$ , we terminate.

#### A) [Move by an increment bigger than $\delta$ ]

- If condition **A** holds, let:  $\mathcal{N}_A(S_i, t_0 + T_i) \equiv \{(S_i + \Delta_i S, t_0 + T_i + \Delta_i T) : \exists (m, q) \in \mathcal{N}_E, q \Delta \leq (T - T_i), (\Delta_i S, \Delta_i T) = (m \delta_0, q \Delta)\}$ .

#### B) [Move to $T$ by an increment smaller than $\delta$ ]

- If condition **B** holds, let:  $\mathcal{N}_A(S_i, t_0 + T_i) \equiv \{(S_i + \Delta_i S, t_0 + T_i + \Delta_i T) : \exists (m, q) \in \mathcal{N}_{S,T}^<(T_i), (\Delta_i S, \Delta_i T) = (m \delta_0, q \Delta)\}$ . Note:  $T_{i+1} = T$ .

#### C) [Could move either by an increment smaller than $\delta$ or bigger than $\delta$ ]

- If condition **C** holds:  $\mathcal{N}_A(S_i, t_0 + T_i)$  is given by the union of the sets  $\mathcal{N}_A(S_i, t_0 + T_i)$  appearing in cases A) and B) above.

In case the statements **A** or **C** hold, we remark that the condition  $(m, q) \in \mathcal{N}_E$  satisfying  $q \Delta \leq (T - T_i)$  is clearly met and so  $\mathcal{N}_A(S_i, T_i) \neq \emptyset$ . Obviously, in case **B**,  $\mathcal{N}_A(S_i, t_0 + T_i) \neq \emptyset$  as well.

The following condition implies that each node  $(S_i, t_0 + T_i)$  is 0-neutral as required in (5):

$$\exists S'_{i+1}, S''_{i+1} \in \mathcal{N}_A(S_i, t_0 + T_i) \text{ such that } 0 \leq (S'_{i+1} - S_i) \ \& \ (S''_{i+1} - S_i) \leq 0. \quad (28)$$

We emphasize that (28) is not required for the construction of the models nor is needed in Proposition 6.0.1 below; (28) is needed to obtain meaningful price bounds. Condition (28) holds by construction when case *B*) executes and, therefore, also when case *C*) executes. Moreover, as it follows from the definition of type I model MIDP in Section 8.5, such models do satisfy (28) when case *A*) executes (and so MIDP models are 0-neutral).

Condition (28) is a very weak condition indeed, requiring that for any node  $(S_i, t_0 + T_i)$ , there are at least two children nodes emanating from it such that the stock moves in opposite directions. For general type I models one can easily check if the condition holds after the trajectory set has been constructed (as indicated, only case *A*) needs to be checked).

### 5.2.2. Type II models

#### Input Parameters

1.  $0 \leq \tau \leq T$ .
2.  $F, G \in \mathbb{N}$  ( $\delta \equiv F \delta_0, \delta_0 \equiv G \hat{\delta}_0, E \equiv F G$ ).
3. A set  $\mathcal{N}_E$ , if  $(m, q) \in \mathcal{N}_E$ , then  $|m| \geq F, q \geq 1, m \in \mathbb{Z}, q \in \mathbb{N}$ .
4.  $\{V_\rho\}, \rho \in (0, T], V_\rho \subseteq \mathbb{R}_+$ .
5.  $\Gamma \in \Delta \mathbb{N}$  satisfying  $q_{min} \Delta \leq \Gamma \leq T$ .

The collection  $\{V_\rho\}$  will be referred to as the *variation constraint sets*.

#### Definition

1.  $\mathcal{N}_{S,T}^<(T_i) \equiv \{(m, q) \in \mathbb{Z} \times \mathbb{N} : -F < m < F, q \Delta + T_i = T, T_i < T\}$ .
2.  $\mathcal{N}_W(m, q) \equiv \{(p_0, p_1, \dots, p_{J-1}) : p_j \in \mathbb{Z}, 1 \leq |p_j|, 1 \leq J \leq q, J \in \mathbb{N}; m = \sum_{j=0}^{J-1} p_j; |\sum_{j=0}^{R-1} p_j| < F, 0 \leq R \leq J-1, |p_j| < 2F, 0 \leq j \leq J-2\}$ , where  $(m, q) \in \mathbb{Z} \times \mathbb{N}$ .

Below we will make use of the statements **A**, **B** and **C** introduced for type I models.

For a given  $(m, q)$ , let  $z$  denote a number satisfying  $z = \sum_{j=0}^{J-1} |p_j| \delta_0$  where  $(p_j)_{j=0}^{J-1} \in \mathcal{N}_W(m, q)$ ; for any such  $z$  set:

$$N_A(m, q, z) \equiv \{(S_{i+1}, t_0 + T_{i+1}, W_{i+1}) \equiv (S_i + m\delta_0, t_0 + T_i + q \Delta, W_i + z)\}.$$

For a given  $(m, q)$  define:

$$N_A(m, q) \equiv \bigcup_{(W_i+z) \in V_{T_{i+1}}} N_A(m, q, z),$$

where  $T_{i+1} = T_i + q \Delta$ .

We assume that for any  $(S_i, T_i, W_i)$  and  $(m, q) \in \mathcal{N}_E \cup \mathcal{N}_{S,T}^<(T_i)$ :

$$N_A(m, q) \neq \emptyset; \tag{29}$$

equivalently, there exists  $z^*$  satisfying  $(W_i + z^*) \in V_{T_{i+1}}$  such that  $N_A(m, q, z^*) \neq \emptyset$ . In the construction that follows, and for the case  $T_{i+1} = T$ , it will be immaterial which such  $z^*$  is actually chosen.

If  $T_i = T$ , we terminate.



A) [Move by an increment bigger than  $\delta$ ]

- If condition **A** holds, let:

$$\begin{aligned} & \mathcal{N}_A(S_i, t_0 + T_i, W_i) \\ \equiv & \bigcup_{(m,q) \in \mathcal{N}_E, q\Delta < (T-T_i)} \mathcal{N}_A(m, q) \bigcup_{(m,q) \in \mathcal{N}_E, q\Delta = (T-T_i)} \mathcal{N}_A(m, q, z^*), \\ & \text{with } z^* \text{ satisfying } (W_i + z^*) \in V_T. \end{aligned}$$

B) [Move to  $T$  by an increment smaller than  $\delta$ ]

- If condition **B** holds, let:

$$\mathcal{N}_A(S_i, t_0 + T_i, W_i) \equiv \bigcup_{(m,q) \in \mathcal{N}_{S,T}^<(T_i)} \mathcal{N}_A(m, q, z^*), \quad (30)$$

with  $z^*$  satisfying  $(W_i + z^*) \in V_T$ .

C) [Could move either by an increment smaller than  $\delta$  or bigger than  $\delta$ ]

- If condition **C** holds:  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$  is given by the union of the sets  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$  appearing in cases A) and B) above.

If **A** or **C** hold, the condition  $(m, q) \in \mathcal{N}_E$  satisfying  $q\Delta \leq (T - T_i)$  is, clearly, non empty and so by (29)  $\mathcal{N}_A(S_i, t_0 + T_i, W_i) \neq \emptyset$ . This fact is also true when case B) holds due to (29). A requirement such as (29) is unavoidable as the sets  $\{V_\rho\}$  are independent of the sets  $\mathcal{N}_E$  and  $\mathcal{N}_W(m, q)$ . Therefore, validity of (29) does depend on the actual estimated values for these sets. Of course, one can also force (29) by construction (i.e. enlarge  $\mathcal{N}_W(m, q)$  or  $V_{T_{i+1}}$  as necessary) but we have chosen not to do so.

The coordinate  $W'_{i+1} \in V_{T_{i+1}}$  in  $(S_{i+1}, t_0 + T_{i+1}, W'_{i+1})$  is only used to prune future children nodes. Therefore, once we have arrived at final time  $t_0 + T$  there is no need to keep several of the possible values of  $W_{i+1}$  for future pruning. For this reason, under case A) with  $T_{i+1} = T$  or case B) (in the latter case  $T_{i+1} = T$  always holds), for a pair  $(S_{i+1}, t_0 + T)$ , as long as there exists at least one  $W_{i+1} \in V_T$  we then include  $(S_{i+1}, T, W_{i+1})$  in  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$ . That is, there is no need to consider any other  $\tilde{W}_{i+1} \in V_T$  for the given pair  $(S_{i+1}, t_0 + T)$ ; this explains the freedom to choose any  $z^*$ , satisfying  $(W_i + z^*) \in V_{T_{i+1}}$ , where  $T_{i+1} = T$ , above.

The following condition implies that each node  $(S_i, t_0 + T_i, W_i)$  is 0-neutral as required in (5):

$$\exists S'_{i+1}, S''_{i+1} \in \mathcal{N}_A(S_i, t_0 + T_i, W_i) \text{ such that } 0 \leq (S'_{i+1} - S_i) \ \& \ (S''_{i+1} - S_i) \leq 0. \quad (31)$$

Analogously to the case of type I models, (31) is not required for the construction of type II models nor is needed in Proposition 6.0.2 below. The property stated in (31) is needed to obtain meaningful price bounds. Notice that if  $(m, q) \in \mathcal{N}_{S,T}^<(T_i)$  then, by (29), there exists  $z^* = z^*(m, q) = \sum_{j=0}^{J-1} |p_j| \delta_0$ , with  $(p_j) \in \mathcal{N}_W(m, q)$ , it then follows that  $z^*(-m, q) \equiv \sum_{j=0}^{J-1} |-p_j| \delta_0$  and  $(-p_j) \in \mathcal{N}_W(-m, q)$ . Therefore, given that  $(-m, q) \in \mathcal{N}_{S,T}^<(T_i)$  as well,  $(S_i + m\delta_0, t_0 + T, W_i + \Delta_i W), (S_i - m\delta_0, t_0 + T, W_i + \Delta_i W) \in \mathcal{N}_A(S_i, T_i, W_i)$  and so the node  $(S_i, T_i, W_i)$  is 0-neutral in case conditions B) or C) execute (this fact

always holds for any type II model). In case A), one could impose the following condition to imply that  $(S_i, T_i, W_i)$  is 0-neutral: there exists  $(m, q), (m', q') \in \mathcal{N}_E$  with  $m m' \leq 0$   $\max(q\Delta, q'\Delta) \leq (T - T_i)$  and there exists  $z$  such that  $N_A(m, q, z) \neq \emptyset$  or  $N_A(m', q', z) \neq \emptyset$ .

### 5.3. Explanations

Once the parameters are assigned to worst case estimates, the constructed models will capture the intended properties of charts; this is proven in Section 6. In this section we provide a preliminary, informal, interpretation of some of the features and conditions used in the construction of the models.

Condition **B** models the situation when there is not enough time for the chart to have a (forthcoming)  $\delta$ -increment. Therefore, when condition **B** executes the trajectory under construction terminates without a  $\delta$ -increment. Condition **A** models the situation when the chart will have a (forthcoming)  $\delta$ -increment. Therefore, execution of condition **A** implies that the trajectories constructed will all have a  $\delta$ -increment. This shows that for any trajectory we have:  $\delta \leq |S_{i+1} - S_i|$  if  $(i + 1) < N(\mathbf{S})$ .

Notice that  $m = 0$  is possible and could happen only in case **B**, moreover, it could be that  $m = 0 = \sum_{j=0}^{J-1} p_j$  with  $J \geq 2$  (recall that  $|p_j| \geq 1$ ). This will account for cases where  $x$  has a number  $J$  of  $\delta_0$ -increments after  $t_i$  but ends up with  $t_{i+1} = T$ ,  $x(T) = x(t_i)$  and so  $m = 0$ . There is still one more case to consider, namely:  $x(T) = x(t_i)$  and so  $m = 0$  but there are no  $\delta_0$ -increments in between. This case we cover by allowing  $J = 0$  (which only has a meaning if  $m = 0$  given that the case  $[m \neq 0 \text{ and } J = 0]$  is impossible).

In other words, the pairs  $(m, q)$  belonging to  $\mathcal{N}_E$  in case A), represent trajectory changes that replicate historical chart's  $\delta$ -increments. The pairs  $(m, q)$  belonging to  $\mathcal{N}_{S,T}^<(T_i)$  in case B), represent trajectory changes that move directly from  $t_0 + T_i$  to  $t_0 + T$  without moving more than  $\delta$  relative to  $S_i$  (but with, possibly, several  $\delta_0$ -increments in between).

The parameter  $s_i$  in (13) represents the number of  $\delta_0$ -increments between two consecutive  $\delta$ -increments. We have replaced  $s_i$  by  $q$  according to (12) and used the variable  $J \leq q$  to account for all possible cases. The constraint  $|\sum_{j=0}^{R-1} p_j| < E$ ,  $R \leq J - 1$  reflects the fact that there is no  $\delta$ -increment between the two nodes under consideration. The case  $J = 1$  represents the case where there is a single  $\delta_0$ -increment which is in effect a  $\delta$ -increment, such a case may be referred as a (spatial) *jump* (as the  $\delta$ -increment can not be anticipated by observing  $\delta_0$ -increments) and so we may have  $m = p_0$  for the case  $J = 1$ .

Once  $\tau$  has been set to its estimated value, we will conclude that if  $T_i < \tau$ , trajectories will move by increments of at least size  $\delta$ . In case that  $T_i \geq \tau$ , it is uncertain if such increments will be happening, we will then also allow for trajectories to “difusse” at  $\delta_0$  granularity. Notice the disjoint dichotomy: to have one more  $\delta$ -increment or to  $\delta_0$ -diffuse directly from  $T_i$  to  $T$  and not to have a  $\delta$ -increment. Such logic for these events is dictated by the operational arrangement and a worst case point of view which are used here to constraint future possibilities.

A main characteristic of type I Models is the fact that the set of admissible nodes is only loosely affected by the current node  $(S_i, t_0 + T_i)$ ,  $T_i < T$ . More specifically, the effect of the current node is to differentiate nodes next to expiration or otherwise. Type II Models impose a stronger dependency on the current node; actually, the role of the  $W_i$  variable is to restrict possible values of  $(m, q)$ . This comment is explained in Section 11.

To make the description of both types of models more symmetric, one could replace the set of pairs

$\mathcal{N}_E$  for set of triples  $(m, q, r)$  where  $r$  is used to encode changes in  $W_i$  in type II models.

## 6. Falsifiable models

This section establishes rigorous results requiring conditions that are empirically observable and implying that the models introduced in previous sections succeed. The meaning of the latter is that with the minimal required initial funds the investor can superhedge a given option by trading along the actual chart values  $x(t_i)$ . The results become empirically relevant when model parameters are set to estimated values or are calibrated. This is done in Sections 8 and 9 which also provide the definite meaning for the models' parameters. We emphasize that the said conditions allow the models to be falsifiable, i.e. one can observe if they do not hold and this can be done dynamically (i.e. as chart samples unfold).

The section requires the availability of the trajectory models (and hence of the parameters used for their construction) and chart values  $x(t)$ ,  $t \in [t_0, t_0 + T]$ . We rely on notation from Section 4.1.

The following definition is needed to enunciate our results.

**Definition 6.1** (Association relationship). We say that  $x$  is *associated* to  $\mathbf{S} \in \mathcal{S}$  if  $\mathbf{X}_i \equiv (x(t_i), t_i, w(x, t_i))$  ( $\mathbf{X}_i \equiv (x(t_i), t_i)$  for type I models) and  $\mathbf{S}_i = (S_i, t_0 + T_i, W_i)$  satisfy  $(x(t_i), t_i) = (S_i, t_0 + T_i)$  for  $0 \leq i \leq N(x, [t_0, t_0 + T]) = N(\mathbf{S})$ . Of course, the coordinate  $W_i$  is dropped for type I models and in that case we can write the association simply as  $\mathbf{X}_i = \mathbf{S}_i$  for  $0 \leq i \leq N(x, [t_0, t_0 + T]) = N(\mathbf{S})$ .

We will say that the model  $\mathcal{M}$  *succeeds* for the chart  $x : [t_0, t_0 + T] \rightarrow \mathbb{R}$  if there is  $\mathbf{S} \in \mathcal{S}$  associated to  $x$ . Otherwise, we will say that the model *fails* for the chart  $x$ . If  $x$  is associated to  $\mathbf{S}$  then (4) holds whenever the coordinates  $S_i$  are replaced by chart values  $x(t_i)$ . That is, given the required initial funds  $\bar{V}(\mathbf{S}_0, Z, \mathcal{M}) + \epsilon$  one can superhedge  $Z$  by trading on chart values  $x(t_i)$ . Notice that the association definition *does not require to match*  $w(x, t_i)$  to  $W_i$  as this variable does not appear explicitly in the superhedging property which is the one we want to uphold. Times  $t_i$  are matched to times  $t_0 + T_i$  as we are dealing with a European option expiring at  $t_0 + T$ .

### 6.1. Validity of Type I models

The notation  $\mathcal{N}_{S,T}^<(T_i)$  used in the proof below was introduced in Section 5.2.1; we also rely on notation from Section 4.1.

**Proposition 6.0.1.** *Consider  $\mathcal{S}$  to be a trajectory set of type I. Assume*

$$\tau \leq \tau(x, [t_0, t_0 + T]), \quad \Gamma(x, [t_0, t_0 + T]) \leq \Gamma. \quad (32)$$

*Furthermore, assume:*

$$(\hat{m}, q) \in \mathcal{N}_E(x, [t_0, t_0 + T]) \text{ then } \exists (m, q) \in \mathcal{N}_E \text{ and } \hat{m} = G m, \quad (33)$$

*and*

$$(x(t_N) - x(t_{N-1})) \in \delta_0 \mathbb{Z}. \quad (34)$$

*Then, there exists  $\mathbf{S} \in \mathcal{S}$  associated to  $x$ .*

*Proof.* Consider first the case when  $\tau(x, [t_0, t_0 + T]) = 0$  then, by (32),  $\tau = 0$  and so condition **C** holds for  $(S_0, t_0 + T_0)$  (recall that  $T_0 = 0$ ). Notice that  $\tau(x, [t_0, t_0 + T]) = 0$  implies  $x(t)$ ,  $t \in [t_0, t_0 + T]$ , has no  $\delta$ -increments; therefore  $N(x) = 1$  and so  $t_1 = t_0 + T$ . (34) implies  $(x(t_0 + T) - x(t_0)) \in \delta_0 \mathbb{Z}$  and  $-F < m \equiv \frac{(x(t_0 + T) - x(t_0))}{\delta_0} < F$ , therefore  $(m, (T - T_0)/\Delta) \in \mathcal{N}_{S,T}^<(T_0)$  which shows that  $\mathbf{S} = \{(S_0 = x(t_0), t_0 + T_0 = t_0), (S_1 = x(t_0 + T), t_0 + T_1 = t_0 + T)\}$  is associated to  $x$  and  $\mathbf{S} \in \mathcal{S}$  because  $(S_1 = x(t_0 + T), t_0 + T_1 = t_0 + T) \in \mathcal{N}_A(S_0, t_0 + T_0)$ , the latter set as in case B).

We may then consider  $\tau(x, [t_0, t_0 + T]) \geq \Delta$  for the remaining of the proof; therefore, there exists  $t_1$  such that  $\delta \leq |x(t_1) - x(t_0)|$ . Moreover  $(\hat{m} \equiv \frac{(x(t_1) - x(t_0))}{\delta_0}, q \equiv \frac{t_1 - t_0}{\Delta}) \in \mathcal{N}_E(x, [t_0, t_0 + T])$ , then (33) implies there exists  $(m, q) \in \mathcal{N}_E$  and  $m = \frac{(x(t_1) - x(t_0))}{\delta_0}$  (recall that  $\delta_0 = G \hat{\delta}_0$ ). Now we continue the proof by induction. For  $(S_i, t_0 + T_i) = (S_0, t_0 + T_0)$ , condition **A** executes due to  $\Gamma \leq (T - T_0) = T$ . Given that  $t_1 - t_0 = q\Delta \leq T - T_0$  we have  $(S_1 = S_0 + m\delta_0, t_0 + T_1 = t_0 + q\Delta) \in \mathcal{N}_A(S_0, t_0 + T_0)$ . This is the first step in the recursive construction and the latter fact is our base case for the following inductive argument.

Assume, inductively, that we have found  $\mathbf{S}' \in \mathcal{S}$  satisfying  $\mathbf{X}_k = \mathbf{S}'_k$  for  $0 \leq k \leq i$  with  $0 < T_i \leq T$ . In the case that  $T_i < T$ , we will show that there exists  $\mathbf{S} \in \mathcal{S}_{(\mathbf{S}', i)}$  and  $\mathbf{X}_{i+1} = \mathbf{S}_{i+1}$  with  $t_{i+1} \leq T$ . One will then conclude that there exists  $\mathbf{S} \in \mathcal{S}$  associated to  $x$ . Therefore, we may assume  $T_i < T$  in what follows.

Consider first the case when the following holds,

$$0[(\Gamma \leq (T - T_i)) \& (\tau \geq \Delta)] \text{ or } (T_i < \tau), \quad (35)$$

we label this the case *a*). Then,  $[(\Gamma(x, [t_0, t_0 + T]) \leq (T - T_i)) \text{ or } (T_i < \tau(x, [t_0, t_0 + T]))]$  holds, which implies there exists  $t_{i+1}$ ,  $t_0 + T \geq t_{i+1} > t_i = t_0 + T_i$  satisfying  $\delta \leq |x(t_{i+1}) - x(t_i)|$ . Therefore,

$$(\hat{m}, q) \equiv \left( \frac{x(t_{i+1}) - x(t_i)}{\hat{\delta}_0}, \frac{t_{i+1} - t_i}{\Delta} \right) \in \mathcal{N}_E(x, [t_0, t_0 + T]),$$

from (33) it follows that there exists  $(m, q) \in \mathcal{N}_E$  so that

$$\hat{m} = G m, \text{ and so } \Delta_i S \equiv m \delta_0 = \frac{\hat{m}}{G} \delta_0 = (x(t_{i+1}) - x(t_i)).$$

We remark that  $q \Delta \leq (T - T_i)$  and so  $q_{\min} \Delta \leq (T - T_i)$  (so case *a*) holds if and only if statement **A** holds). This shows that the construction described in item A) executes and so there exists  $(\Delta_i S, \Delta_i T) = (\Delta_i x, (t_{i+1} - t_i))$  and  $(S_{i+1}, T_{i+1}) \in \mathcal{N}_A(S_i, t_0 + T_i)$ . In view of case a), take  $S_{i+1} \equiv S_i + m\delta_0 = x(t_{i+1}) = x(t_i) + m\delta_0$  and  $t_0 + t_{i+1} \equiv t_0 + T_i + q\Delta = t_i + q\Delta$ . In this way the required  $\mathbf{S}$  is obtained with  $S_k = S'_k$  for  $0 \leq k \leq i$  and  $S_{i+1}$  as it was just defined. Therefore, the inductive hypothesis for  $i + 1$  has been established.

Similarly, assume statement **B** holds, i.e.:  $[(T - T_i) < q_{\min} \Delta]$  holds, notice this case is potentially non empty as  $T_i > 0$ . Then, by (33),  $[(T - T_i) < q_{\min}(x, [t_0, t_0 + T]) \Delta]$  where  $q_{\min}(x, [t_0, t_0 + T]) \equiv \min\{q : \exists (m, q) \in \mathcal{N}_E(x, [t_0, t_0 + T])\}$ . Therefore, there exists  $t_{i+1} = t_0 + T > t_i = t_0 + T_i$  satisfying  $\delta > |x(t_{i+1}) - x(t_i)|$ , then (34) implies

$$(m, q) \equiv \left( \frac{x(t_{i+1}) - x(t_i)}{\delta_0}, \frac{t_{i+1} - t_i}{\Delta} \right) \in \mathcal{N}_{S,T}^<(T_i).$$

Notice that  $|m| < F$ . This shows that there exists  $(\Delta_i S, \Delta_i T) = (\Delta_i x, (t_{i+1} - t_i))$  and  $(S_{i+1}, t_0 + T_{i+1}) \in \mathcal{N}_A(S_i, t_0 + T_i)$  (this latter set as in case B)).

Finally, consider that statement **C** holds and, we recall, that in this case  $\mathcal{N}_A(S_i, t_0 + T_i)$  is defined as the union of the sets  $\mathcal{N}_A(S_i, t_0 + T_i)$  appearing under the entries *A*) and *B*). From our definition of  $t_{i+1}$  (see Definition 4.2) there are two possibilities:  $\delta \leq |x(t_{i+1}) - x(t_i)|$ ,  $t_i < t_{i+1} \leq t_0 + T$  or  $\delta > |x(t_{i+1}) - x(t_i)|$ ,  $t_i < t_{i+1} = t_0 + T$ . Whichever of these cases holds, the execution of condition **C** in the recursive construction of  $\mathcal{S}$  and the arguments employed above for the cases **A** and **B** imply the existence of  $(S_{i+1}, t_0 + T_{i+1}) \in \mathcal{N}_A(S_i, t_0 + T_i)$  and  $(S_{i+1}, t_0 + T_{i+1}) = (x(t_{i+1}), t_{i+1})$ . So we have established the validity of the inductive hypothesis for  $i + 1$ .  $\square$

Hypothesis (33) implicitly requires  $(x(t_{i+1}) - x(t_i)) \in \delta_0 \mathbb{Z}$  for  $0 \leq i < N(x)$ . We weaken this latter requirement in Section 7. Of course, and more substantially, (33) also requires that the chart  $(\hat{m}, q)$  pairs are matched by corresponding trajectory pairs  $(m, q)$ . The appearance of  $G$  in (33) reflects the fact that chart increments are measured in units of  $\hat{\delta}_0$  while trajectory increments are measured in  $\delta_0$  units.

## 6.2. Validity of Type II models

A similar proposition can also be proved for type II models; towards this end, we will need to rely on notation introduced in Section 4. The proof follows the same logic as the one of Proposition 6.0.1 with the added arguments to handle the coordinate  $W_i$ .

The variation constraint sets  $\{V_\rho\}_{\rho \in (0, T]}$ , introduced in Section 5.2.2, are required to satisfy:

$$\text{if } \rho \leq \rho' \text{ and } w \in V_{\rho'} \setminus V_\rho \text{ then } w \geq y \quad \forall y \in V_\rho, \quad (36)$$

a property that can be easily checked once we assign the sets  $\{V_\rho\}_{\rho \in (0, T]}$  to their estimates in Section 8.6. We also introduce the following notation for  $\rho \leq \rho'$  and  $y \in V_\rho$ :

$$\Delta_{\rho, \rho'} V(y) \equiv \{z \geq 0 : \exists w \in V_{\rho'}, w = z + y\}.$$

Using (36) we obtain

$$V_{\rho'} = \cup_{y \in V_\rho} y + \Delta_{\rho, \rho'} V(y). \quad (37)$$

**Proposition 6.0.2.** *Consider  $\mathcal{S}$  to be a trajectory set of type II. Assume*

$$\tau \leq \tau(x, [t_0, t_0 + T]), \quad \Gamma(x, [t_0, t_0 + T]) \leq \Gamma. \quad (38)$$

*Furthermore assume:*

$$(\hat{m}, q) \in \mathcal{N}_E(x, [t_0, t_0 + T]) \text{ then } \exists (m, q) \in \mathcal{N}_E \text{ and } \hat{m} = G m, \quad (39)$$

$$(x(r_{l+1}) - x(r_l)) \in \delta_0 \mathbb{Z}, \quad 0 \leq l \leq L - 1, \quad (40)$$

*and*

$$\sum_{l_i \leq l, l+1 \leq l_{i+1}} |x(r_{l+1}) - x(r_l)| \in \Delta_{t_i - t_0, t_{i+1} - t_0} V(w(x, t_i)), \quad 0 \leq i \leq N - 1, \quad (41)$$

*also, (36) holds. Then, there exists  $\mathbf{S} \in \mathcal{S}$  associated to  $x$ .*

*Proof.* Consider first the case when  $\tau(x, [t_0, t_0 + T]) = 0$  then, by (38),  $\tau = 0$  and so condition **C** holds for  $(S_0, t_0 + T_0, W_0)$  (recall that  $T_0 = W_0 = 0$ ). Notice that  $\tau(x, [t_0, t_0 + T]) = 0$  implies  $x(t), t \in [t_0, t_0 + T]$ , has no  $\delta$ -increments; therefore  $N(x) = 1$  and so  $t_1 = t_0 + T$ . (40) implies  $(x(t_0 + T) - x(t_0)) \in \delta_0 \mathbb{Z}$  and so  $-F < m \equiv \frac{(x(T) - x(t_0))}{\delta_0} < F$ , therefore  $(m, q \equiv (T - T_0)/\Delta) \in \mathcal{N}_{S,T}^<(T_0)$ . From the definitions and notation in Section 4 we can write (here we rely on (40)):

$$(x(r_{l+j+1}) - x(r_{l+j})) = \hat{\delta}_0 \hat{p}_j = \delta_0 p_j, \text{ and so } \hat{p}_j = G p_j. \quad (42)$$

Moreover

$$\hat{m} = \sum_{j=0}^{s-1} \hat{p}_j = G \sum_{j=0}^{s-1} p_j \equiv Gm, \quad (43)$$

(we have dropped the subscript  $i$  in  $\hat{m}_i$ ,  $l_i$  and  $s_i$  appearing in Section 4, we will do the same in the remaining of the proof). Therefore  $\Delta_0 w \equiv w(x, t_0 + T) - w(x, t_0) = \hat{\delta}_0 \sum_{j=0}^{s-1} |\hat{p}_j| = \delta_0 \sum_{j=0}^{s-1} |p_j|$  (recall that  $w(x, t_0) \equiv 0$ ). By means of the properties listed in (14), and recalling that  $E = F G$ , we can check that the properties in (23) hold and by taking  $J \equiv s$  we verify that  $(p)_{j=0}^{s-1} \in \mathcal{N}_w(m, q)$ . Therefore  $z \equiv \delta_0 \sum_{j=0}^{s-1} |p_j| = \sum_{l_0 \leq l, l+1 \leq l_1} |x(r_{l+1}) - x(r_l)| \in \Delta_{0, t_1 - t_0} V(w(x, t_0))$  where we used (41). It then follows from (37) that  $W_1 \equiv w(x, t_0) + z = W_1 \in V_T$ . We emphasize, again, that the property of association in Definition 6.1 does not involve the coordinate  $W_i$  and, given that  $t_1 = t_0 + T$ , we could then take  $z^* = z$  in (30) (more precisely, we have established that there exists  $z$  satisfying  $(W_0 + z) \in V_T$  such that  $N_A(m, q, z) \neq \emptyset$ ; taking  $z^*$  to be any such  $z$  will do for our argument). Therefore,  $\mathbf{S} = \{(S_0 = x(t_0), t_0 + T_0 = t_0, W_0), (S_1 = x(t_0 + T), t_0 + T_1 = t_0 + T, W_1)\} \in \mathcal{S}$  because  $(S_1 = x(t_0 + T), t_0 + T_1 = t_0 + T, W_1) \in \mathcal{N}_A(S_0, t_0 + T_0, W_0)$ , the latter set as in case *B*). It then follows that  $x$  is associated to  $\mathbf{S}$ .

In what follows we may not provide as many details on arguments involving  $z + W_i$  leading to  $W_{i+1}$ ; they are very similar to the arguments for the case  $\tau(x, [t_0, t_0 + T]) = 0$  above.

We may then assume  $\tau(x, [t_0, t_0 + T]) \geq \Delta$  for the remaining of the proof. Therefore, there exists  $t_1$  such that  $\delta \leq |x(t_1) - x(t_0)|$ . Moreover  $(\hat{m} \equiv \frac{(x(t_1) - x(t_0))}{\delta_0}, q \equiv \frac{t_1 - t_0}{\Delta}) \in \mathcal{N}_E(x, [t_0, t_0 + T])$ , then (39) implies there exists  $(m, q) \in \mathcal{N}_E$  and  $m = \frac{(x(t_1) - x(t_0))}{\delta_0}$  (recall that  $\delta_0 = G \hat{\delta}_0$ ). We continue the proof by induction. For  $(S_i, t_0 + T_i) = (S_0, t_0 + T_0)$ , condition **A** executes due to  $\Gamma \leq (T - T_0) = T$ . Notice that  $x(t_1) = x(t_0) + m\delta_0 = S_1$  and  $t_1 = t_0 + q\Delta$ . The same arguments used above for the case  $\tau(x, [t_0, t_0 + T]) = 0$ , and the fact that  $t_1 - t_0 = q\Delta = T_1 \leq T - T_0$ , show that there exists  $z + W_0 = w(x, t_1)$  where  $z \equiv \delta_0 \sum_{j=0}^{s-1} |p_j| = \sum_{l_0 \leq l, l+1 \leq l_1} |x(r_{l+1}) - x(r_l)| \in \Delta_{0, t_1 - t_0} V(w(x, t_0))$  where we used (41). It then follows from (37) that  $W_1 \equiv w(x, t_0) + z \in V_{T_1}$ . If  $T_1 < T$ , it follows that  $(S_1 = S_0 + m\delta_0, t_0 + T_1 = t_0 + q\Delta, W_1) \in \mathcal{N}_A(S_0, t_0 + T_0, W_0)$  (this latter set as in case *A*). This fact is our base case for the following inductive argument. On the other hand, if  $T_1 = T$ , we could take  $z^* = z$  and argue as in the above case  $\tau(x, [t_0, t_0 + T]) = 0$  and conclude there is  $\mathbf{S}$  associated to  $x$ .

Assume, inductively, that we have found  $\mathbf{S}' \in \mathcal{S}$  satisfying  $\mathbf{X}_k = \mathbf{S}'_k$  for  $0 \leq k \leq i$  with  $0 < T_i < T$ . We will show that there exists  $\mathbf{S} \in \mathcal{S}_{(\mathbf{S}', i)}$  satisfying one of the two alternatives:

- i)  $\mathbf{X}_{i+1} = \mathbf{S}_{i+1}$  with  $t_{i+1} < T$  (i.e. the inductive hypothesis holds at  $i + 1$ ),
  - ii)  $(x(t_{i+1}), t_{i+1}) = (S_{i+1}, T_{i+1})$  and  $T_{i+1} = T$ .
- (44)

One can then conclude that there exists  $\mathbf{S} \in \mathcal{S}$  associated to  $x$ .

Consider first the case when the following holds

$$[(\Gamma \leq (T - T_i)) \& (\tau \geq \Delta)] \text{ or } (T_i < \tau), \quad (45)$$

we label this the case *a*). Then,  $[(\Gamma(x, [t_0, t_0 + T]) \leq (T - T_i)) \text{ or } (T_i < \tau(x, [t_0, t_0 + T]))]$  holds, which implies there exists  $t_{i+1}$ ,  $T \geq t_{i+1} > t_i = T_i$  satisfying  $\delta \leq |x(t_{i+1}) - x(t_i)|$ . Therefore,

$$(\hat{m}, q) \equiv \left( \frac{x(t_{i+1}) - x(t_i)}{\hat{\delta}_0}, \frac{t_{i+1} - t_i}{\Delta} \right) \in \mathcal{N}_E(x, [t_0, t_0 + T]), \quad (46)$$

from (39) it follows that there exists  $(m, q) \in \mathcal{N}_E$  so that

$$\hat{m} = G m, \text{ and so } \Delta_i S \equiv m \delta_0 = \frac{\hat{m}}{G} \delta_0 = (x(t_{i+1}) - x(t_i)). \quad (47)$$

We remark that  $q \Delta \leq (T - T_i)$  and so  $q_{\min} \Delta \leq (T - T_i)$  (so case *a*) holds if and only if statement **A** holds). By case *a*), take  $S_{i+1} \equiv S_i + m \delta_0 = x(t_{i+1}) = x(t_i) + m \delta_0$  and  $T_{i+1} \equiv t_0 + T_i + q \Delta = t_i + q \Delta$ . The same arguments used above for the case  $\tau(x, [t_0, t_0 + T]) = 0$ , and the fact that  $t_{i+1} - t_i = q \Delta = T_{i+1} - T_i \leq T - T_i$ , show that there exists  $z + W_i = w(x, t_{i+1})$  where  $z \equiv \delta_0 \sum_{j=0}^{s-1} |p_j| = \sum_{l_i \leq l, l+1 \leq l_{i+1}} |x(r_{l+1}) - x(r_l)| \in \Delta_{t_i - t_0, t_{i+1} - t_0} V(w(x, t_i))$  where we used (41) (of course, we have  $m = \sum_{j=0}^{s-1} p_j$ ). It then follows from (37) that  $W_{i+1} \equiv w(x, t_i) + z \in V_{T_{i+1}}$ . If  $T_{i+1} < T$  it follows that  $(S_{i+1}, t_0 + T_{i+1}, W_{i+1}) \in \mathcal{N}_A(S_0, t_0 + T_0, W_0)$  (this set as in case *A*). On the other hand, if  $T_{i+1} = T$ , we could take  $z^* = z$  and argue as in the above case  $\tau(x, [t_0, t_0 + T]) = 0$  and conclude, in this case, that there is **S** associated to  $x$ .

Similarly, assume statement **B** holds, i.e.:  $[(T - T_i) < q_{\min} \Delta]$  holds (this is possible as  $T_i > 0$ ). Then  $[(T - T_i) < q_{\min}(x, [t_0, t_0 + T]) \Delta]$  where  $q_{\min}(x, [t_0, t_0 + T]) \equiv \min\{q : \exists (m, q) \in \mathcal{N}_E(x, [t_0, t_0 + T])\}$ . Therefore, there exists  $t_{i+1} = t_0 + T > t_i = t_0 + T_i$  satisfying  $\delta > |x(t_{i+1}) - x(t_i)|$ , then (34) implies

$$(m, q) \equiv \left( \frac{x(t_{i+1}) - x(t_i)}{\delta_0}, \frac{t_{i+1} - t_i}{\Delta} \right) \in \mathcal{N}_{S,T}^<(T_i). \quad (48)$$

Notice that  $|m| < F$ . From the definitions and notation in Section 4 we can write (here we rely on (40)):

$$(x(r_{l+j+1}) - x(r_{l+j})) = \hat{\delta}_0 \hat{p}_j = \delta_0 p_j, \text{ and so } \hat{p}_j = G p_j.$$

Moreover

$$\hat{m}_i = \sum_{j=0}^{s-1} \hat{p}_j = G \sum_{j=0}^{s-1} p_j \equiv G m. \quad (49)$$

Therefore  $\Delta_i w \equiv w(x, t_0 + T) - w(x, t_i) = \hat{\delta}_0 \sum_{j=0}^{s-1} |\hat{p}_j| = \delta_0 \sum_{j=0}^{s-1} |p_j|$ . By means of the properties listed in (14), and recalling that  $E = F G$ , we can check that the properties in (23) hold and by taking  $J \equiv s$  we verify that  $(p)_{j=0}^{J-1} \in \mathcal{N}_w(m, q)$ . Therefore  $z \equiv \delta_0 \sum_{j=0}^{s-1} |p_j| = \sum_{l_i \leq l, l+1 \leq l_{i+1}} |x(r_{l+1}) - x(r_l)| \in \Delta_{t_i - t_0, t_{i+1} - t_0} V(w(x, t_i))$  where we used (41). It then follows from (37) that  $W_{i+1} \equiv w(x, t_i) + z = W_i + z \in V_T$  (where  $w(x, t_i) = W_i$  holds by the inductive hypothesis). We emphasize, again, that the property of association in Definition 6.1 does not involve the coordinate  $W_i$  and, given that  $t_1 = t_0 + T$ , we could then take  $z^* = z$  in (30) (more precisely, we have established that there exists  $z$  satisfying  $(W_i + z) \in V_T$  such that  $\mathcal{N}_A(m, q, z) \neq \emptyset$ ; taking  $z^*$  to be any such  $z$  will do for our argument). Therefore  $(S_{i+1} = x(t_0 + T), t_0 + T_{i+1} = t_0 + T, W_{i+1}) \in \mathcal{N}_A(S_i, t_0 + T_i, W_i)$ , the latter set as in case *B*). It then follows that  $x$  is associated to **S**.

Finally, consider that statement **C** holds; we recall that in this case  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$  is defined as the union of the sets  $\mathcal{N}_A(S_i, t_0 + T_i, W_i)$  appearing under the entries **A**) and **B**). From our definition of  $t_{i+1}$  (see Definition 4.2) there are two possibilities:  $\delta \leq |x(t_{i+1}) - x(t_i)|$ ,  $t_i < t_{i+1} \leq t_0 + T$  or  $\delta > |x(t_{i+1}) - x(t_i)|$ ,  $t_i < t_{i+1} = t_0 + T$ . Whichever of these cases holds, the execution of condition **C** in the recursive construction of  $\mathcal{S}$  and the arguments employed above for the cases **A** and **B** imply one of the alternatives in (44).  $\square$

### 6.3. Discussion of rigorous results

The failure or validity of assumptions in Propositions 6.0.1 and 6.0.2 can be determined dynamically; that is, when we arrive to the next node that just unfolded. Hypotheses in the two propositions could fail at any node, so the larger  $N$  the larger the likelihood of the chart  $x$  not being associated to some  $\mathbf{S} \in \mathcal{S}$ . In our setting, the possibility to choose the value of  $\delta$  with  $\delta > \delta_0$  allows to control the maximum value of  $N$ . Restricting  $N$  is also crucial when transaction costs are accounted for.

Section 10 defines worst case parameter estimation and risk taking parameter estimation. The mentioned hypotheses failures could be said to happen *with uncertainty* in the case that all parameters have been estimated in the worst case manner. If, on the other hand, some parameters were estimated in a risk-taking way; the failure may happen with a certain (empirical) frequency.

One special situation, namely the case when there exists  $l$  such that

$$(x(r_{l+1}) - x(r_l)) \notin \{k\delta_0 : k \in \mathbb{Z}\} \quad (50)$$

could lead to violation of the assumptions in Propositions 6.0.1 or 6.0.2. Notice that in practice (50) will occur for our need to take  $\delta_0$  larger in order to achieve reasonable execution times for our implementation (see Appendix 1). So, it will be a case of a parameter to be considered to be risk taking. It turns out that, under some conditions, one can show that weaker versions of the conclusions in Propositions 6.0.1 and 6.0.2 are valid under (50). This is argued in Section 7.

Further discussion on the above issues can be found in Sections 9.2 and 9.3.

## 7. Trajectory approximation error due to $\delta_0 \geq 0.01$

In this section we weaken the assumptions in Proposition 6.0.1 to account for the possibility that  $(x(t_{i+1}) - x(t_i)) \in \delta_0 \mathbb{Z}$  could fail given that we assume  $\delta_0 \geq 0.01$ . As we remarked previously,  $(x(t_{i+1}) - x(t_i)) \in \delta_0 \mathbb{Z}$  is assumed implicitly by (33) and explicitly by (34). In order to afford weaker assumptions, we also need to weaken the conclusion of that proposition while maintaining its practical relevance; this we can achieve by replacing the notion of association between  $x$  and  $\mathbf{S}$  by *weak association* as defined below. We will establish the existence of a value of  $\delta_0 \geq 0.01$  for which one can establish a weak association.

For simplicity, we restrict our analysis to type I models MIDP. A different error analysis is required to accommodate the coordinate  $W_i$  as we will need to account for possible violations of  $V_{t_i-t_0}(x, [t_0, t_0 + T]) \subseteq V_{t_i-t_0}$  in Proposition 6.0.2 due to  $\delta_0 \geq 0.01$ .

We highlight that the superhedging property (4) is crucial to deploy the proposed models; the clear financial relevance is explained in Section 10. It is when  $x$  is associated to  $\mathbf{S}$  that such superhedging property becomes relevant as it holds when we replace the values  $S_i$  for  $x(t_i)$  ( $t_i = T_i$  is also essential).



Therefore, what really matters is that the superhedging property holds when we use the chart samples  $x(t_i)$ , this observation motivates the definitions and developments that follow.

We modify below the association relationship  $x \rightarrow \mathbf{S}$  introduced in Definition 6.1 (as indicated, we presently restrict ourselves to MIDP models). Also, for simplicity,  $Z$  below will be the payoff of a path independent European derivative and so we can consider  $Z : \mathbb{R} \rightarrow \mathbb{R}$ .

**Definition 7.1** (Weak association). Given a chart  $x : [t_0, t_0 + T] \rightarrow \mathbb{R}$ ,  $\beta > 0$  and option  $Z$ , we say that  $\hat{\mathbf{S}} = \hat{\mathbf{S}}(x) \in \mathcal{S}$  is *weakly* associated to  $x$  (for the given  $\beta$  and  $Z$ ) if there exists  $H \in \mathcal{H}$  such that:

$$\bar{V}(S_0, Z, \mathcal{M}) + \beta + \sum_{i=0}^{N(\hat{\mathbf{S}})-1} H_i(\hat{\mathbf{S}}) (x(t_{i+1}) - x(t_i)) \geq Z(x(T)) \quad \text{and} \quad \hat{T}_i = t_i, \quad 0 \leq i \leq N(\hat{\mathbf{S}}), \quad (51)$$

where  $\hat{\mathbf{S}} = \{\hat{\mathbf{S}}_i = (\hat{S}_i, \hat{T}_i)\}$ .

If  $x$  is associated to  $\mathbf{S}$  then, by (4), it will be weakly associated as well for any  $Z$  and  $\beta$ . Actually, the dependencies of  $\hat{\mathbf{S}}$  on  $Z$  and  $\beta$  in Definition 7.1 are incidental, the crucial point is that there exists  $\hat{\mathbf{S}} \in \mathcal{S}$  along which we can invest and that such investments will superhedge  $Z$  when the actual risky asset unfolds along the values  $x(t)$ . Given that  $\beta$  is fixed in the above definition, we refer to  $\hat{\mathbf{S}}$  as an *approximating trajectory*.

### 7.1. Construction of $\hat{\mathbf{S}}(x)$

We describe below a method to set  $\delta_0 \geq 0.01$  and to construct  $\hat{\mathbf{S}}$  weakly associated to  $x$  under weaker assumptions than the ones required in Proposition 6.0.1.

To simplify the notation let  $x_i \equiv x(t_i)$ ,  $0 \leq i \leq N \equiv N(x)$ . We will *assume the following is valid at all nodes*  $(\mathbf{S}, i)$ :

$$\text{if } [|S_i - x_i| \leq \delta_0, t_i = T_i] \text{ then} \quad (52)$$

$$\begin{aligned} [\exists \mathbf{S}^+ = \mathbf{S}^+(x), \mathbf{S}^- = \mathbf{S}^-(x) \in \mathcal{S}_{(\mathbf{S}, i)}, \quad T_{i+1}^+ = T_{i+1}^- = t_{i+1}, \quad x_{i+1} \in [S_{i+1}^-, S_{i+1}^+] \text{ and} \\ 0 < S_{i+1}^+ - S_{i+1}^- \leq \delta_0], \end{aligned}$$

where the interval  $[S_{i+1}^-, S_{i+1}^+]$  is considered here as a subset of  $\mathbb{R}$  (i.e. *not* as a subset of  $\delta_0\mathbb{Z}$ ). Property (52) weakens assumptions required in Proposition 6.0.1. More precisely, in the present case of models MIDP, assumptions (32), (33) and (34) imply (52) that only requires that we can approximate chart values  $x_i$  from above and below with trajectory values  $S_i \in \delta_0\mathbb{Z} + x(t_0)$  instead of assuming that chart changes  $(x_{i+1} - x_i)$  are exactly matched by trajectory coordinates changes  $(S_{i+1} - S_i)$ .

Existence of  $\mathbf{S}^+$  and  $\mathbf{S}^-$  will be available to our construction of  $\hat{\mathbf{S}}$  by the consequent in (52) and the fact that the construction of  $\hat{\mathbf{S}}$  (described in this section) is recursive and the antecedent of (52) will be satisfied in the previous step. For the same reason,  $\hat{T}_i = t_i$  for all  $i$  will be trivial by (52) (and the fact that, by construction, either  $\hat{\mathbf{S}}_{i+1} = \mathbf{S}_{i+1}^+$  or  $\hat{\mathbf{S}}_{i+1} = \mathbf{S}_{i+1}^-$ ) hence, we only explicitly describe the construction of the coordinates  $\hat{S}_i$ .

We will construct an approximating element  $\hat{\mathbf{S}} = \hat{\mathbf{S}}(x)$  in the trajectory set satisfying:

$$\beta \geq Z(x_N) - Z(\hat{S}_N) + \sum_{i=0}^{N-1} H_i(\hat{\mathbf{S}}) (\hat{S}_{i+1} - x_{i+1}) + \sum_{i=0}^{N-1} H_i(\hat{\mathbf{S}}) (x_i - \hat{S}_i), \quad (53)$$

where  $H$  is the superhedging optimal portfolio from (4) for  $Z$  and  $\epsilon = 0$  and  $N = N(\hat{S}) = N_H(\hat{S})$ . Notice that for simplicity we have forced  $\epsilon = 0$  in (4) to avoid introducing a new inessential parameter (this is possible as  $\mathcal{S}$  is finite and this implies the computation of the optimal  $H$  is achieved by an actual minimum in  $\inf_{H \in \mathcal{H}}$ ). Such  $H$  will be fixed for the rest of the discussion. If (53) is fulfilled it guarantees that  $x$  is weakly associated to  $\hat{S}$ ; this is so because (53) implies

$$\bar{V} + \beta + \sum_{i=0}^{N-1} H_i(\hat{S})(x_{i+1} - x_i) \geq \bar{V} + \sum_{i=0}^{N-1} H_i(\hat{S})(\hat{S}_{i+1} - \hat{S}_i) + Z(X_N) - Z(\hat{S}_N) \geq \quad (54)$$

$$Z(\hat{S}_N) + Z(x_N) - Z(\hat{S}_N) = Z(x_N),$$

where the last inequality follows because  $H$  is the optimal portfolio in (4) with  $\epsilon = 0$  and  $\bar{V} \equiv \bar{V}(\mathbf{S}, Z, \mathcal{M})$ .

So, the question becomes: can one choose  $\delta_0 \geq 0.01$  so that (53) holds for some approximating trajectory  $\hat{S}(x)$ ? The answer is in the affirmative and the developments shed light in how some nodes play a relevant role.

## 7.2. Upper bound via sign selection

We proceed recursively, fix  $\hat{S}_0 \equiv S_0 = x_0$ , so  $H_0(\hat{S}_0) = H_0(\mathbf{S}_0)$ . We describe the construction of  $\hat{S}_{i+1}$  under the sole assumption that  $|\hat{S}_i - x_i| \leq \delta_0$ . The construction terminates as soon as  $\hat{T}_i = T$ , it will then be clear that the constructed  $\hat{S}$  belongs to  $\mathcal{S}$  and (53) holds and so  $\hat{S}$  is weakly associated to  $x$ .

Our main goal is to find a suitable upper bound for

$$H_i(\hat{S})(\hat{S}_{i+1} - x_{i+1}) + H_{i+1}(\hat{S})(x_{i+1} - \hat{S}_{i+1}) = (H_{i+1}(\hat{S}) - H_i(\hat{S}))(x_{i+1} - \hat{S}_{i+1}), \quad (55)$$

notice we have joined pairs of terms appearing in (53) and remark that we will have a non paired term, namely  $H_{i+1}(\hat{S})(x_{i+2} - \hat{S}_{i+2})$  for the case of  $i+1 = N-1$ . Therefore, that last term has to be considered as a separate case (something we do after first dealing with the paired terms (55)).

Let us analyze (55) with  $i+1 < N-1$ , and suppose  $S_0, \hat{S}_1(x), \dots, \hat{S}_i(x)$  have been already constructed satisfying the antecedent of (52). We attempt to make (55) non positive by choosing a closest (but there is the issue of choosing the approximation from above or below) value of  $\mathbf{S}'_{i+1}$ ,  $\mathbf{S}' \in \mathcal{S}_{(\hat{S}(x), i)}$ , to the given value of  $x_{i+1}$ . Notice that the value of  $H_{i+1}(\hat{S})$  depends on first choosing the value of  $\hat{S}_{i+1}$ . We have the following possibilities:

- In case there exists  $\mathbf{S}' \in \mathcal{S}_{(\hat{S}(x), i)}$  satisfying  $S'_{i+1} = x_{i+1}$ , we set  $\hat{S}_{i+1} = \hat{S}_{i+1}(x) = S'_{i+1}$ .
- In case that  $\hat{S}^+$  (recall that  $\hat{S}^+ \in \mathcal{S}_{(\hat{S}(x), i)}$ ) satisfies  $[H_{i+1}(\hat{S}^+) - H_i(\hat{S})] \geq 0$  we set  $\hat{S}_{i+1} = \hat{S}_{i+1}(x) = \hat{S}_{i+1}^+$ .
- In case that  $\hat{S}^-$  (recall that  $\hat{S}^- \in \mathcal{S}_{(\hat{S}(x), i)}$ ) satisfies  $[H_{i+1}(\hat{S}^-) - H_i(\hat{S})] \leq 0$  we set  $\hat{S}_{i+1} = \hat{S}_{i+1}(x) = \hat{S}_{i+1}^-$ .

Notice that in all the cases above (55) is non positive and so 0 is, so far, an upperbound for (55).

It remains to define  $\hat{S}_{i+1}(x)$  and bound (55) from above in the following case:

$$\hat{S}_{i+1}^\pm \neq x_{i+1} \ \& \ [H_{i+1}(\hat{S}^+) - H_i(\hat{S})] < 0 \ \& \ [H_{i+1}(\hat{S}^-) - H_i(\hat{S})] > 0. \quad (56)$$

That is,  $H_{i+1}(\mathbf{S}') - H_i(\hat{S})$  goes from being negative to being positive when  $\mathbf{S}' = \hat{S}^+ \rightarrow \mathbf{S}' = \hat{S}^-$ .

If (56) holds we define  $\hat{S}_{i+1} = \hat{S}_{i+1}(x) = \hat{S}_{i+1}^+$  whenever  $0 < \hat{S}_{i+1}^+ - x_{i+1} \leq \delta_0/2$  and set  $\hat{S}_{i+1} = \hat{S}_{i+1}(x) = \hat{S}_{i+1}^-$  whenever  $0 < x_{i+1} - \hat{S}_{i+1}^- < \delta_0/2$ . This completes the definition of  $\hat{S}_i$  for  $0 \leq i \leq N(x) - 1$ .

Notice that in all cases  $|\hat{S}_i - x_i| \leq \delta_0$  holds for  $0 \leq i \leq N(x) - 1$  which gives access to the consequent in (52) through the recursion/induction.

We gain some insight by analyzing (56) in some more detail, towards this goal we introduce some notation that will be used shortly. Given a node  $(\mathbf{S}, i)$ , any existing pair  $\mathbf{S}^\uparrow$  and  $\mathbf{S}^\downarrow$  satisfying  $0 < S_{i+1}^\uparrow - S_{i+1}^\downarrow \leq \delta_0$ ,  $\mathbf{S}^\uparrow, \mathbf{S}^\downarrow \in \mathcal{S}_{(\mathbf{S}, i)}$ ,  $T_{i+1}^\uparrow = T_{i+1}^\downarrow$  and

$$[H_{i+1}(\mathbf{S}^\uparrow) - H_i(\mathbf{S})] < 0 \quad \& \quad [H_{i+1}(\mathbf{S}^\downarrow) - H_i(\mathbf{S})] > 0 \quad (57)$$

will be called a *portfolio turning pair* for such node. Notice that if the pair  $\mathbf{S}^\uparrow = \mathbf{S}^\uparrow(\mathbf{S})$  and  $\mathbf{S}^\downarrow = \mathbf{S}^\downarrow(\mathbf{S})$  exists, then it is unique and that they are analogous to  $\mathbf{S}^+(x)$  and  $\mathbf{S}^-(x)$  introduced earlier.

As an informal remark, notice that a turning pair for a node is expected to happen in isolated circumstances. For example, a plausible continuity of  $H_{i+1}$  gives  $H_{i+1}(S_0, \dots, S_i, S_{i+1}^\uparrow) \approx H_{i+1}(S_0, \dots, S_i, S_{i+1}^\downarrow)$  given that  $0 < S_{i+1}^\uparrow - S_{i+1}^\downarrow \leq \delta_0$ . That is, by taking  $\delta_0$  small enough any case satisfying (56) disappears by continuity. In the actual discrete case this is clearly always true by taking  $\delta_0 = 0.01$  so the portfolio investment is always continuous (independently of the option) with respect to the natural (spatial) proximity of trajectories (by the discrete nature of real trajectories).

For  $i$  fixed, let:

$$E_i^\downarrow \equiv \{\mathbf{S} \in \mathcal{S} : \exists \mathbf{S}^\uparrow, \mathbf{S}^\downarrow \in \mathcal{S}_{(\mathbf{S}, i)}, 0 < S_{i+1}^\uparrow - S_{i+1}^\downarrow \leq \delta_0, \quad (58)$$

$$[H_{i+1}(\mathbf{S}^\uparrow) - H_i(\mathbf{S})] < 0 \quad \& \quad [H_{i+1}(\mathbf{S}^\downarrow) - H_i(\mathbf{S})] > 0 \quad \& \quad S_{i+1} = S_{i+1}^\downarrow\} \quad (59)$$

and,

$$E_i^\uparrow \equiv \{\mathbf{S} \in \mathcal{S} : \exists \mathbf{S}^\uparrow, \mathbf{S}^\downarrow \in \mathcal{S}_{(\mathbf{S}, i)}, 0 < S_{i+1}^\uparrow - S_{i+1}^\downarrow \leq \delta_0, \quad (60)$$

$$[H_{i+1}(\mathbf{S}^\uparrow) - H_i(\mathbf{S})] < 0 \quad \& \quad [H_{i+1}(\mathbf{S}^\downarrow) - H_i(\mathbf{S})] > 0 \quad \& \quad S_{i+1} = S_{i+1}^\uparrow\}. \quad (61)$$

For  $(\mathbf{S}, i)$  given we introduce:

$$w_i(\delta_0, \mathbf{S}) \equiv \mathbf{1}_{E_i^\uparrow}(\mathbf{S}) |H_{i+1}(\mathbf{S}^\uparrow) - H_i(\mathbf{S})| + \mathbf{1}_{E_i^\downarrow}(\mathbf{S}) |H_{i+1}(\mathbf{S}^\downarrow) - H_i(\mathbf{S})|. \quad (62)$$

Consider (55), we have already obtained an upperbound by 0 when (56) *does not hold*. For the upperbound inequality below we then assume the validity of (56), therefore

$$[H_{i+1}(\hat{\mathbf{S}}) - H_i(\hat{\mathbf{S}})](x_{i+1} - \hat{S}_{i+1}) = \mathbf{1}_{\{\hat{S}_{i+1} = \hat{S}_{i+1}^+\}}(\hat{\mathbf{S}}) [H_{i+1}(\hat{\mathbf{S}}) - H_i(\hat{\mathbf{S}})](x_{i+1} - \hat{S}_{i+1}) + \quad (63)$$

$$\mathbf{1}_{\{\hat{S}_{i+1} = \hat{S}_{i+1}^-\}}(\hat{\mathbf{S}}) [H_{i+1}(\hat{\mathbf{S}}) - H_i(\hat{\mathbf{S}})](x_{i+1} - \hat{S}_{i+1}) = \quad (64)$$

$$\mathbf{1}_{\{\hat{S}_{i+1} = \hat{S}_{i+1}^+\}}(\hat{\mathbf{S}}) [H_{i+1}(\hat{\mathbf{S}}^+) - H_i(\hat{\mathbf{S}})](x_{i+1} - \hat{S}_{i+1}) + \quad (65)$$

$$\mathbf{1}_{\{\hat{S}_{i+1} = \hat{S}_{i+1}^-\}}(\hat{\mathbf{S}}) [H_{i+1}(\hat{\mathbf{S}}^-) - H_i(\hat{\mathbf{S}})](x_{i+1} - \hat{S}_{i+1}) \leq \quad (66)$$

$$\mathbf{1}_{E_i^\uparrow}(\hat{\mathbf{S}}) |H_{i+1}(\hat{\mathbf{S}}^\uparrow) - H_i(\hat{\mathbf{S}})| \frac{\delta_0}{2} + \mathbf{1}_{E_i^\downarrow}(\hat{\mathbf{S}}) |H_{i+1}(\hat{\mathbf{S}}^\downarrow) - H_i(\hat{\mathbf{S}})| \frac{\delta_0}{2}.$$

Adding up terms we obtain the following bound:

$$\sum_{i=0}^{N-2} ((H_{i+1}(\hat{\mathbf{S}}) - H_i(\hat{\mathbf{S}}))(x_{i+1} - \hat{S}_{i+1}) \leq w(\delta_0) \equiv \frac{\delta_0}{2} \sup_{\mathbf{S} \in \mathcal{S}} \left( \sum_{i=0}^{N-2} w_i(\delta_0, \mathbf{S}) \right). \quad (67)$$

The definition of  $w(\delta_0)$  only involves  $\mathcal{M}$  (and  $Z$  indirectly), that is, its evaluation only requires access to  $\mathcal{S}$  and the optimal portfolio  $H$  (for option  $Z$ ). We comment on the meaning and practicality of (67) in the next section.

### 7.3. Constraints on $\delta_0$ for weak association

Here we put together previous estimates; to sum up, we want to construct a trajectory  $\hat{\mathbf{S}}(x)$  which is weakly associated to  $x$ . This allows to extend the superhedging property, available in  $\mathcal{M}$  when using  $\bar{V}(S_0, Z, \mathcal{M})$ , to the case of using the actual trajectory samples  $x_i$ . The construction of  $\hat{\mathbf{S}}(x)$  described in the previous section works as long as  $\delta_0$  is small enough, in fact we show below that we obtain constraints on  $\delta_0$  which depend only on  $\mathcal{M}, \beta > 0$  and  $Z$ .

To make use of the obtained upperbound, we write (53) as follows

$$\beta \geq a(x) + b(x), \quad (68)$$

where

$$a(x) = Z(x_N) - Z(\hat{S}_N) + H_{N-1}(\hat{\mathbf{S}}) (x_N - \hat{S}_N) \quad (69)$$

and

$$b(x) = \sum_{i=0}^{N-2} (H_{i+1}(\hat{\mathbf{S}}) - H_i(\hat{\mathbf{S}})) (x_{i+1} - \hat{S}_{i+1}). \quad (70)$$

Bounding  $a(x)$  is not problematic as we can always make  $H_{N-1}(\hat{\mathbf{S}}) (x_N - \hat{S}_N) \leq 0$  by choosing  $\hat{S}_N$  above or below  $x_N$  depending on the sign of  $H_{N-1}(\hat{\mathbf{S}})$  (which does not depend on  $\hat{S}_N$ ). Then, we can bound  $Z(x_N) - Z(\hat{S}_N)$  by the modulus of continuity of  $Z$  as  $|x_N - \hat{S}_N| \leq \delta_0$ . All this we can do thanks to the fact that the antecedent in (52) holds for  $i = N(x) - 1$ . In the case of a European call  $H_i \geq 0$ , so we can choose  $\hat{S}_N \geq x_N$  hence  $Z(x_N) - Z(\hat{S}_N) \leq 0$  and so  $a(x) \leq 0$ . Therefore, for a European call option  $\beta \geq b(x) \geq a(x) + b(x)$  and so, for this reason we concentrate only on  $b(x)$  below.

We have seen that  $b(x) \leq w(\delta_0)$  (see (67)), therefore we look for a value of  $\delta_0$  satisfying

$$w(\delta_0) \leq \beta. \quad (71)$$

As we have argued, taking  $\delta_0 = 0.01$  makes  $w(\delta_0) = 0$  and allows the construction of  $\hat{\mathbf{S}}$  under the hypothesis (52). This situation is consistent with Proposition 6.0.1.

### 7.4. Meaning of Constraint $w(\delta_0) \leq \beta$

Admittedly the computation of  $w(\delta_0)$  is impractical but the analysis leading to this quantity is useful. Actually, to avoid any dependency on  $\hat{\mathbf{S}}$ , and so on  $x$ , the sets  $E_i^\uparrow$  and  $E_i^\downarrow$  were defined larger than could be necessary if we allow our models to have some extra flexibility. Assume we allow  $\delta_0$  to depend conditionally on a given node  $(\mathbf{S}, i)$ , in that scenario, we can modify its value iteratively as we construct  $\hat{\mathbf{S}}_0, \hat{\mathbf{S}}_1, \dots$ . This added flexibility allows us to replace the condition  $\mathbf{S} \in \mathcal{S}$  in the definition of the sets  $E_i^\uparrow$  and  $E_i^\downarrow$  by  $\mathbf{S} \in \mathcal{S}_{(\mathbf{S}, i)}$ . With such modification  $w_i(\delta_0, \mathbf{S}) = 0$  if  $\mathbf{S} \notin \mathcal{S}_{(\mathbf{S}, i)}$ ; for example, one could identify nodes  $(\mathbf{S}, i)$  where  $w_i(\delta_0, \mathbf{S}) \neq 0$  and refine  $\delta_0$  at those nodes only. Crucially, this refinement can be done as the samples  $x_i$  unfold.

Notice that we have defined weak association in terms of superhedging but we can think in terms of profits as follows: let  $\text{Profits} \equiv P(t_0, K) - \bar{V}(S_0, Z, \mathcal{M}) - \beta$  where  $P(t_0, K)$  is the market selling price

of  $Z$  at  $t_0$  and  $\beta > 0$ . One can then set up parameters in  $\mathcal{M}$  to risk-taking values so that  $\text{Profits} > 0$ . Then (71) allows to select a value of  $\delta_0$  such that  $\text{Profits}$  do not disappear due to errors in trajectory matching (of course with risk taking parameter there is a risk for (52) to fail). This is so as (71) implies (53) which in turn implies (54).

## 8. Measurements on charts for estimation

The present and the next sections specify all the inputs required to construct each of the two types of models described in Section 5.2. Actually, by making different instantiations of some of the models' inputs we will be able to describe several specific models.

Towards these tasks we assume the availability of chart values  $x^*(t)$  with  $t \in \{k\Delta : k \in \mathbb{Z}\}$ . We will also use the notation  $\mathcal{T} \subseteq \{k\Delta : k \in \mathbb{Z}\}$  to denote a historical *interval* of times for which we have access to values  $x^*(t)$ , used for estimation. We will specify further characteristics of  $\mathcal{T}$  whenever relevant. We also assume a given  $T = M_T\Delta$ . Notice that treating  $\mathcal{T}$  as an interval will incorporate overnight effects that could be best modelled separately (something we neglect to do).

All models rely on the following parameters:  $\tau, \Gamma$ , the set  $\mathcal{N}_E$  and the collection of sets  $\{V_\rho\}_{0 \leq \rho \leq T}$  (this later collection is only needed for type II models). In some cases, for simplicity, we refer to this family of parameters and sets generically as *input parameters*, they need to be estimated. The remaining parameters are of a different type, they are:  $\delta, \delta_0, \Delta, T$ . The parameters  $\delta$  and  $\Delta$  are investor dependent and they should be calibrated, guidelines to calibrate  $\delta$  are available in Section 9. The parameter  $\delta_0$  is a parameter of practical concern discussed in several places in the paper and in particular in Section 7. Finally  $T$  depends on the intended financial application.

Our approach relies on *worst case estimation* and *risk taking estimation* (see definitions in Section 9). Instead of relying on general notions from the start, we cover first some parameters in a case by case basis. We expect, in this way, to illustrate the straightforward methodology. Hence, for the time being, we will use the word *estimate* in quite general terms to refer to a computation that uses historical data.

Given historical data  $x^*(t)$ ,  $t \in \mathcal{T}$ , and a value  $\delta_0$  we obtain sampling times  $\{r_l\}$  that depend on that specific value of  $\delta_0$ . These sample times are evaluated by fixing a window of  $M_T$  data points (the number of data points in a time window of length  $T$ ) and then shifting the window by one time unit and starting again. Each set of sampling times  $\{r_l\}$ , one for each such window, will have a variable cardinality/length of  $L(x^*, [t, t + T]) + 1$ . Notice that the notation  $\{r_l\}$  does not indicate the dependency on the window (which, in these cases, is left implicit for simplicity).

The following sections rely on notation introduced in Section 4.1.

### 8.1. Rounding

Estimates are evaluated, as much as possible, by manipulating the (exact) data values  $x(t) \in \hat{\delta}_0 \mathbb{Z}$ ; whenever rounding to the grid  $\delta_0 \mathbb{Z}$  is required it will take place at the last step of the computations. Rounding is required in two instances, in one case for the estimation of  $\mathcal{N}_E$  and in the other case for the estimation of  $\{V_\rho\}$  both quantities are model dependent and so associated to  $\delta_0$ . The quantities  $\mathcal{N}_E(x, [t_0, t_0 + T])$  (from Definition 4.3) and  $\sum_{l_i \leq l, l+1 \leq l_{i+1}} |x(r_{l+1}) - x(r_l)|$  (from (41)) are chart dependent and do not require any rounding. We note in passing that, for the latter quantity, we could have introduced the chart variation notation  $V_{t_{i+1}-t_0}(x, [t_0, t_0 + T]) - V_{t_i-t_0}(x, [t_0, t_0 + T])$  but we have not done

so.

To emphasize: the sampling times  $\{r_i\}$  and transaction times  $\{t_i\}$  are evaluated using the exact values  $x(t)$ . Similarly, the values of  $\tau(x, [t_0, t_0 + T])$  and  $\Gamma(x, [t_0, t_0 + T])$  in (17) and (18) respectively, are in no need of any rounding.

Given  $y(t) \in \hat{\delta}_0 \mathbb{Z}$  let  $\|y(t)\| \in \delta_0 \mathbb{Z}$  denote the closest value to  $y(t)$  (i.e. closest in  $\delta_0 \mathbb{N}$ ). In case of ambiguity, namely when both conditions  $\lfloor y(t) \rfloor - y(t) = \lceil y(t) \rceil - y(t) = \delta_0/2$  hold (where  $\lfloor \cdot \rfloor$  and  $\lceil \cdot \rceil$  are floor and ceiling functions rounding in  $\delta_0 \mathbb{N}$ ), any possible disambiguation will be fine.

With some abuse of notation, when compared to Definition 4.3, we introduce  $\mathcal{N}_E(\|x\|, [t_0, t_0 + T])$  which is the first instance when we are required to use rounded values:

$$\mathcal{N}_E(\|x\|, [t_0, t_0 + T]) \equiv \left\{ \left( \frac{\|x(t_{i+1}) - x(t_i)\|}{\delta_0}, \frac{t_{i+1} - t_i}{\Delta} \right) : 0 \leq i \leq P - 1 \right\}, \quad (72)$$

where the  $\{t_i\}$  are as in Definition (4.2) and  $P \equiv P(x, [t_0, t_0 + T]) \equiv N \equiv N(x, [t_0, t_0 + T])$  (the latter as introduced in Definition 4.2) if  $\delta \leq |x(t_N) - x(t_{N-1})|$  or  $P \equiv P(x, [t_0, t_0 + T]) \equiv N - 1$  otherwise.

The second instance when we rely on rounded values is in (79) below.

## 8.2. Estimate for $\mathcal{N}_E$

Define,

$$\mathcal{N}_E^* \equiv \cup_{[t, t+T] \subseteq \mathcal{T}} \mathcal{N}_E(\|x^*\|, [t, t + T]), \quad (73)$$

where  $\mathcal{N}_E(\|x^*\|, [t, t + T])$  was introduced in (72). Define  $q_* \equiv \min\{q : \exists(m, q) \in \mathcal{N}_E^*\}$ ,  $q^* \equiv \max\{q : \exists(m, q) \in \mathcal{N}_E^*\}$ ,  $m_* \equiv \min\{m : \exists(m, q) \in \mathcal{N}_E^*\}$  and  $m^* \equiv \max\{m : \exists(m, q) \in \mathcal{N}_E^*\}$ .

The set  $\mathcal{N}_E$  will be assigned to two different estimates, doing so for each of the model types gives four different models. This multiplicity illustrates the obvious need to explore the stability/robustness of these alternatives. This is something we do not do in this paper and it may be best approached with machine learning techniques. Notice also that one can detect if the model fails in a dynamic way, this is a very important feature as it could be used to learn or adapt the investor behaviour. We do not pursue such a possibility in the paper.

1. **Direct Product (DP) Choice:**  $\mathcal{N}_E = \{m \in \mathbb{Z} : m_* \leq m \leq m^*\} \times \{q \in \mathbb{N} : q_* \leq q \leq q^*\}$  (and so  $\mathcal{N}_E^* \subseteq \mathcal{N}_E$ ).
2. **Observable (OBS) Choice:**  $\mathcal{N}_E = \mathcal{N}_E^*$ .

Either of the two choices for  $\mathcal{N}_E$  is referred to as as a *worst case* estimate while the sets that remain after removing any arbitrary subset from the worst case estimates are referred to as risk taking estimates.

## 8.3. Estimate for $\Gamma$ [Implying: $q_{min} \Delta \leq \Gamma$ ]

Define

$$\Gamma^* \equiv \max_{[t, t+T] \subseteq \mathcal{T}} \Gamma(x^*, [t, t + T]), \quad (74)$$

where  $\Gamma(x^*, [t, t + T])$  is given by (18). That is,  $\Gamma^*$  is the smallest time duration such that for *any*  $t$  satisfying  $[t, t + \Gamma^*] \subseteq \mathcal{T}$  there exists  $s \in (t, t + \Gamma^*]$  such that  $\delta \leq |x^*(s) - x^*(t)|$ . Unless  $\delta$  is taken

sufficiently small, we will find  $\Gamma^* = T$  which will make the parameter  $\Gamma^*$  irrelevant. This later remark should be clear from the definitions of the models.

We set

$$\Gamma \equiv \Gamma^*, \quad (75)$$

such choice is referred as *worst case*. One can also consider  $\Gamma < \Gamma^*$ , this will entail some risk-taking as defined and analyzed in Sections 6 and 10. Notice that the choice  $\Gamma \geq \Gamma^*$  is possible and is more conservative than the worst case choice.

Given the above definitions and recalling that  $q_{min} \equiv \min\{q : (m, q) \in \mathcal{N}_E\}$ , it then follows that  $q_{min} \Delta$  is the smallest time length, in the historical data set, for  $x^*$  to have a  $\delta$ -increment. Therefore, the definition of  $\Gamma^*$  implies

$$q_{min} \Delta \leq \Gamma^* = \Gamma. \quad (76)$$

This inequality is an assumption required for the two types of models and is thus established from (75).

#### 8.4. Estimate for $\tau$

Define

$$\tau_* \equiv \min_{[t, t+T] \subseteq \mathcal{T}} \tau(x^*, [t, t+T]) \quad (77)$$

where  $\tau(x^*, [t, t+T])$  was introduced in (17). We then set

$$\tau \equiv \tau_*. \quad (78)$$

One can also consider  $\tau > \tau_*$  or  $\tau < \tau_*$ , the first choice will entail some risk-taking while the second choice being more conservative (both relative to worst case). Finally, notice that  $\tau_* = 0$  implies  $\Gamma^* = T$ .

#### 8.5. Models: MIDP, MIOBS

Given the above choices of  $\mathcal{N}_E$ ,  $\Gamma$  and  $\tau$ , we define two models of type I which we refer to by MIDP and MIOBS. The former obviously refers to the DP choice for  $\mathcal{N}_E$  and the latter to the OBS choice for  $\mathcal{N}_E$ .

#### 8.6. Estimates for $\{V_\rho\}$ , $\rho \in [0, T]$ [Models: MIIDPVAR, MIIOBSVAR]

Here we introduce two models of type II. As we did with type I models we set  $\mathcal{N}_E$  exactly as in items (1) and (2) in Section 8.2. These choices we label MIIDP and MIIOBS, it remains to specify the estimate for  $\{V_\rho\}$ .

For  $0 \leq \rho \leq T$  define

$$V_\rho(\|x^*\|, [t, t+T]) \equiv \left\| \sum_{l=0}^{r_{l+1} \leq t+\rho} |x(r_{l+1}) - x(r_l)| \right\| \quad (79)$$

where, as we have done many times before, the sampling times are as in Definition 4.1 and the notation  $\|y\|$  means the rounding of  $y$  to the grid  $\delta_0 \mathbb{Z}$  (already introduced in Section 8.1). Moreover, define

$$V_\rho^* \equiv \cup_{[t, t+T] \subseteq \mathcal{T}} V_\rho(\|x^*\|, [t, t+T]). \quad (80)$$

1. We introduce model MIIDPVAR (Model II Direct Product with Variation) by setting:  $\mathcal{N}_E = \{m \in \delta_0 \mathbb{Z} : m_* \leq m \leq m^*\} \times [q_*, q^*]$ .  $\{V_\rho\} \equiv \{V_\rho^*\}, \rho \in [0, T]$
2. We introduce model MIIOBSVAR (Model II Observed with Variation) by setting:  $\mathcal{N}_E = \mathcal{N}_E^*$ .  $\{V_\rho\} \equiv \{V_\rho^*\}, \rho \in [0, T]$ .

That is, we have set  $\mathcal{N}_E$  as we have done for MIDP and MIOBS. Examples of the sets  $\{V_\rho\}$  are displayed in Figures 18 and 19.

## 9. Data used, calibration of $\delta$ and output for estimates

This section completes the specification of the models introduced by mentioning some criteria to calibrate  $\delta$  and providing some output for the estimated parameters. We also take the opportunity to introduce some general definitions of *worst case estimation*, such discussion should provide a perspective to our approach.

The simple estimation methodology used in the paper has been already introduced in Section 8 and we expand on it in the present section. Unless indicated otherwise, whenever we refer to probabilities in relation to parameter estimation or a risk-reward analysis, we mean empirical frequencies. In particular, probabilities are not necessarily given by a probability measure.

The word *calibration* refers to a process for assigning values to model parameters that can be chosen by the investor, in our case,  $\delta$ . All other parameters have been set to worst case estimates in Section 8.

At this point in the paper, it seems appropriate to be more specific regarding the data we used and how it affects estimation. We will deploy the models in a setting with  $\Delta = 1$  hour or  $\Delta = 1/2$  hour and  $T$  of the order of several days. These values were imposed by the data we had access to, a negative side effect is long execution time for pricing. To illustrate how to alleviate this situation, in the case of type II models, we have increased the value of  $\delta_0$  as this choice allows for reasonable fast execution times for our implementation (we have also sub-sampled the sets  $\mathcal{N}_E$ ). Notice that  $\delta_0$  has two usages: *i*) it represents the discretization used in the model's grid (and it is in this role where  $\delta_0$  needs to be curtailed to reduce computational time) *ii*) it defines the sampling times, changing  $\delta_0$  has then an impact on estimation (in particular for type II models which uses the variation coordinate). Details are provided at appropriate places in the paper.

The requirement (29) for type II models is enforced in our implementation along the lines described at the end of Section 5.2.2. We also enforce  $S_i \geq 0$  by setting  $S_j = 0, i \leq j$  where  $i$  is the smallest index such that  $S_{i-1} + m\delta_0 < 0$ .

### 9.1. Data employed

The two sources of data used for calibration and estimation are:

- An approximately 8 month long stretch of hourly data ticks for Facebook Inc (FB) from 2016-12-06 to 2017-07-03. At the end of this period, call option ask prices were captured across a variety of strikes for comparison with an expiration of 15 calendar days into the future.
- The same time interval of the Standard and Poor's Depository Receipts S & P 500 Trust ETF (SPY) of 30 minutes ticks. Again, at the end of this period, call option ask prices were captured across a variety of strikes for comparison with an expiration of 15 calendar days into the future.



Let  $|M|$  be the total number of data points and  $M_T$  the number of data points corresponding to an interval  $[t_0, t_0 + T]$ . Notice we label past data points by  $M, M + 1, \dots, -1$  i.e.  $M < 0$ .

We neglect overnight and weekend effects in the sense that we do not model any of its characteristics but their peculiarities will impact our estimates. Given that in our data there are 7 hourly samples per business day for the FB stock and there are 11 business days in the 15 calendar days we have  $T = 77 \Delta$ , with  $\Delta = 1$  hr (so  $M_T = 77$ ). Similarly for the 1/2 hourly data of SPY we have  $T = 154 \Delta$  with  $\Delta = 1/2$  hr (so  $M_T = 154$ ).

We use the notation  $\mathcal{T} = \{k\Delta : k = M, M + 1, \dots, -1\}$  and whenever we indicate *historically estimated from chart data* below we mean that we made use of data  $x^*(t)$  with  $t \in \mathcal{T}$ .

## 9.2. Comments on worst case estimation

We hope the previous section's introduction by example of what we mean with worst case estimation has helped to show that our approach is simple minded and should not require, for our purposes, a theoretical justification. The present section and later sections do introduce some definitions on the matter but they are intended to provide due precision and with no intention of complicating the framework.

Worst case estimation as well as calibration (to a lesser extent) depend on a set of historical data. We use 8 months of data and hence, it can be argued, our estimates are not really worst case as it does not contemplate all possible historically available data. Our reasoning and conclusions, for the considered specific models, only depend on the property that the estimated sets  $\mathcal{N}_E$  have finite cardinality (or, more generally, are bounded). So, this should be considered the underlying assumption: these sets of reachable nodes are bounded even in the case of unlimited availability of historical data.

One expects that the worst case estimation as defined will be robust in matching future unfolding trajectories. Failure to match the unfolding chart to a modeling trajectory will only happen if the combined time and stock movements of the unfolding chart are more extreme than any in the past data. As we mentioned previously, a main issue we need to confront is that the presence of extreme price movements (and they do appear in our data set) will translate into non informative bounds for option pricing. Section 11 explains the specific key feature in type II models that blocks the effect of extreme parameter values implying unrealistic price bounds.

Informally, parameters associated to an operational setting and used to construct models, should have stable worst case ranges. That is, their ranges should not contain non informative values. For example, our investor can select  $\delta$  and, up to some extent  $T$  as well, so that  $\tau_* > 0$  and  $\Gamma_* < T$ . The extreme values  $\tau_* = 0$  and  $\Gamma_* = T$  make these parameters non informative for model construction. The case when a worst case parameter estimate (e.g.  $N^*$ ) grows unbounded as historical data is aggregated is a clear indication that some characteristics of the model need to be re-considered. To emphasize: selecting the model variables keeping in mind that their worst case ranges should substantially constrain the trajectory set is central to the proposed methodology.

## 9.3. Criteria to select $\delta$ and $\delta_0$

Recall that  $\delta$  is the minimum chart change before an investor considers rebalancing. The implications of possible choices of  $\delta$  are worth analyzing.

Choosing  $\delta$  too large will cause the conditional set to collapse as there will be a value of  $\delta$  large

enough so that  $\mathcal{N}_E$  will contain a single element (in particular, this will create an arbitrage). Increasing  $\delta$  further means that  $\mathcal{N}_E$  will be empty. Reducing  $\delta$  from this extreme value could produce a binomial tree which will give a complete model. In this case the upper and lower bounds will coincide but this common value will not be necessarily equal to the market price. Of course, even in our worst case estimation setting, the issue arises of the effect on trajectory matching if we use a large value of  $\delta$  leading to a small set  $\mathcal{N}_E$ .

For the above reasons we discuss some particular criteria as well as general guidelines to calibrate a value of  $\delta$ . A general criteria, which we actually use to create our output examples, is to choose a value of  $\delta$  which gives values of model parameters that constrain the size of the trajectory set. Towards this end we plot how the parameters  $\tau, \Gamma, m_{max}, m_{min}$  change with  $\delta$ . Moreover, we already mentioned that the choice of  $\delta$  controls the maximum value of  $N(\mathbf{S})$ . In case we estimate some parameters to risk taking values, there will be a historical frequency of model failure. In this case, controlling the maximum value of  $N(\mathbf{S})$  allows to estimate the likelihood of model failure along a trajectory  $\mathbf{S}$  (which is upperbounded by  $N(\mathbf{S})$  times the mentioned frequency of failure). In general terms, one can select a value of  $\delta$  where observable and operational parameters are in between realistic worst case bounds. For example the minimum number of transactions  $N_*$  and the maximum number of transactions  $N^*$  represent quantities that can be achieved by our investor. Notice that the usual calibration practice will be to try to match the option prices to market prices; we could also use this criteria as a calibration target. Namely we could select  $\delta$  so that our bounds are tight and contain the observable market value in between. A main benefit of the market models in the paper is the fact that one can use complementary criteria to calibrate given that most features of the models are observable and have a direct operational meaning.

The set  $\mathcal{N}_E$  should become stable before we exhaust all our historical data. In the paper we deal with two dimensions,  $(m, q)$ , and one can monitor the convex hull of these pairs as a measure of stability. The convex hull could also be used as an alternative to the direct product construction in order to obtain a solid set  $\mathcal{N}_E$  (as opposed to discrete points).

Given the information we store in the coordinates of the two types of models, and in order to control the maximum value of  $N(\mathbf{S})$ , we remark that  $N(\mathbf{S}) q_{min} \Delta \leq M_T \Delta = T$ . Therefore, we could select  $\delta$  so that  $\frac{M_T}{q_{min}}$  implies at most a few rebalances per day on average (if the value of  $q_{min}$  allows us to do this).

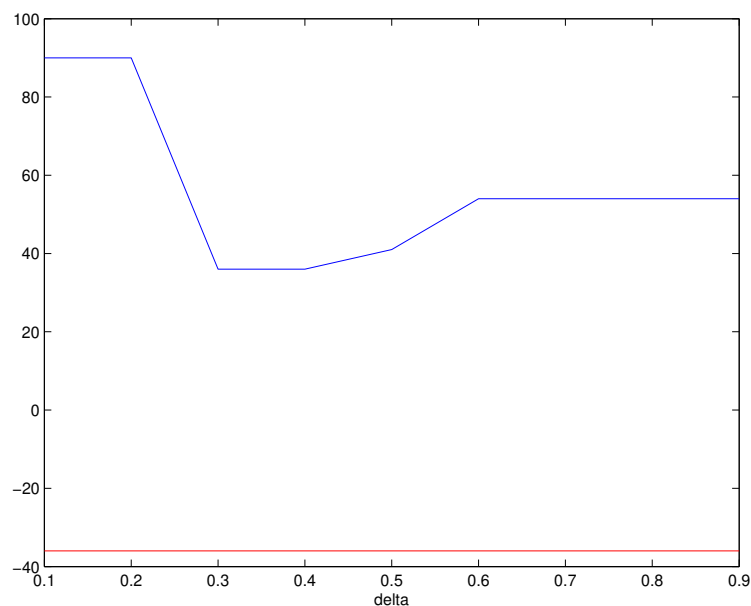
Finally, for type I models and FB data, we have provided output for the choice  $\delta = 0.5, \delta_0 = 0.1$ . This value of  $\delta$  is consistent with the above explanations as it follows from Figures 5 and 6. The analogous output for SPY data are Figures 9 and 10. Type II models require a larger data structure implying longer computational times. To alleviate this situation we took  $\delta_0 = 0.5$  which allows to take  $\delta = 1$  (for the FB data set) and  $\delta_0 = 0.4, \delta = 0.8$  (for the SPY data set) given smaller sets  $\mathcal{N}_E$ . Furthermore, we also sub-sampled the sets  $\mathcal{N}_E$  (as explained later in the paper). For consistency of comparison, Figure 15 provides pricing bounds for the model MIDP for the SPY data set with the values  $\delta = 0.8, \delta_0 = 0.4$  (this output was produced with the same sub-sampling as the one used for type II models).

Based on experimentation, we expect that none of these choices, will significantly alter the price bounds we report (relative to taking  $\delta_0 = \hat{\delta}_0 = 0.01$ ). Moreover, using sub-sampled sets  $\mathcal{N}_E$  that include the extreme values (as opposed to the full sets) should not alter the price bounds reported as the latter are only affected by the extreme stock value changes (given that the payoffs are convex).

#### 9.4. Output for $\mathcal{N}_E$ , $\tau$ , $\Gamma$ , $N_*$ and $N^*$

As indicated, the output uses  $\Delta = 1$  hr for FB data and  $\Delta = 1/2$  hr for SPY data. Figures 2, 3, 5 and 6 illustrate some structure induced in the set of pairs  $(m, q) \in \mathcal{N}_E$  (as well as in  $\tau, \Gamma, N_*$  and  $N^*$ ) which takes place by changing  $\delta$  for the FB data. The analogous output for SPY data are Figures 9 and 10 (further output can be found in Appendix 2). In all these cases  $\delta_0 = 0.1$ . Figures 4 and 7 illustrate the sets  $\mathcal{N}_E$ , with  $\delta = 0.5, \delta_0 = 0.1$  for the FB and SPY data respectively.

As already mentioned, for type II models we have taken  $\delta = 1, \delta_0 = 0.5$  for the FB case and  $\delta = 0.8, \delta_0 = 0.4$  for the SPY case. The analogous outputs, to the ones mentioned in the previous paragraph, can be found in Appendix 2; here we only provide Figure 8 that displays  $\mathcal{N}_E$  taking  $\delta = 0.8, \delta_0 = 0.4$  for the SPY data set.

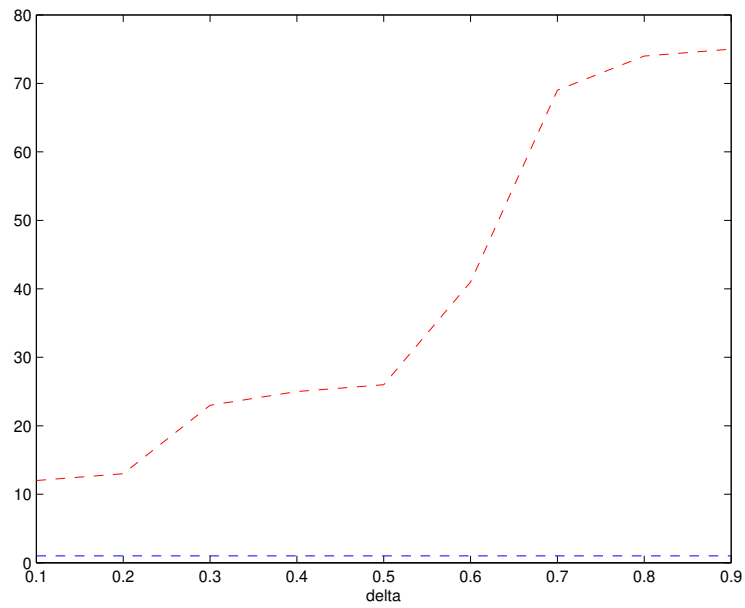


**Figure 2.** maximum and minimum of  $m_i$  with respect to  $\delta$ . FB data,  $\delta_0 = 0.1, \hat{\delta}_0 = 0.01, \Delta = 1$  hr.

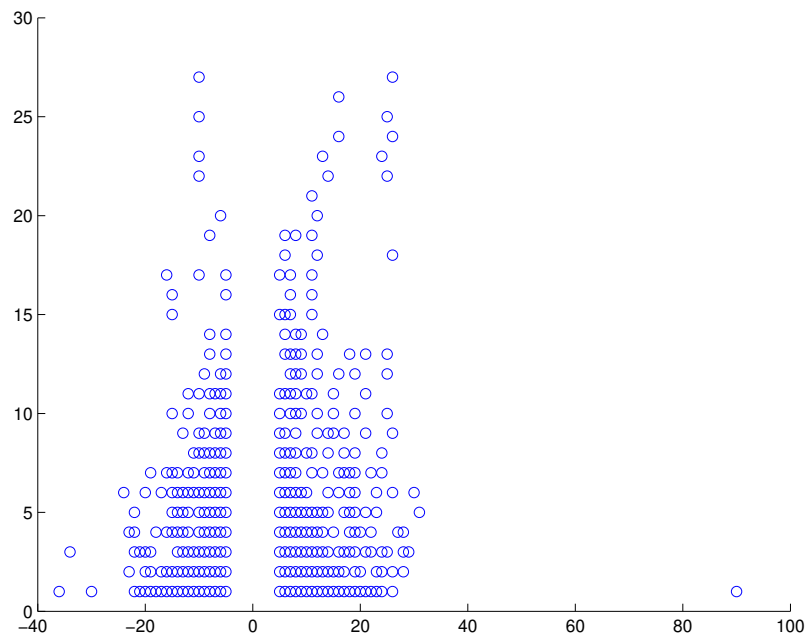
#### 9.5. Price bounds output

Here we comment and explain on the different models' prices obtained as function of the option's strikes. For comparison, the references Britten-Jones and Neuberger (1996) as well as Rebonato (2012), provide output for related models and the same minmax methodology. Figures 11, 12, 13 and 14 provide lower and upper bounds (in red) as functions of the strike for the models: MIOBS, MIDP, MIIDP and MIIOBS respectively for the FB data. In Appendix 2 we have provided more related output, namely output for MIOBS, MIDP but now for  $\delta = 1, \delta_0 = 0.5$  and  $\mathcal{N}_E$  sub-sampled as we do for type II models (in this way one can directly compare FB output with the output for type II models provided here).

Figures 15, 16 and 17 provide lower and upper bounds (in red) as functions of the strike for the models: MIDP, MIIDP and MIIOBS respectively for the SPY data.



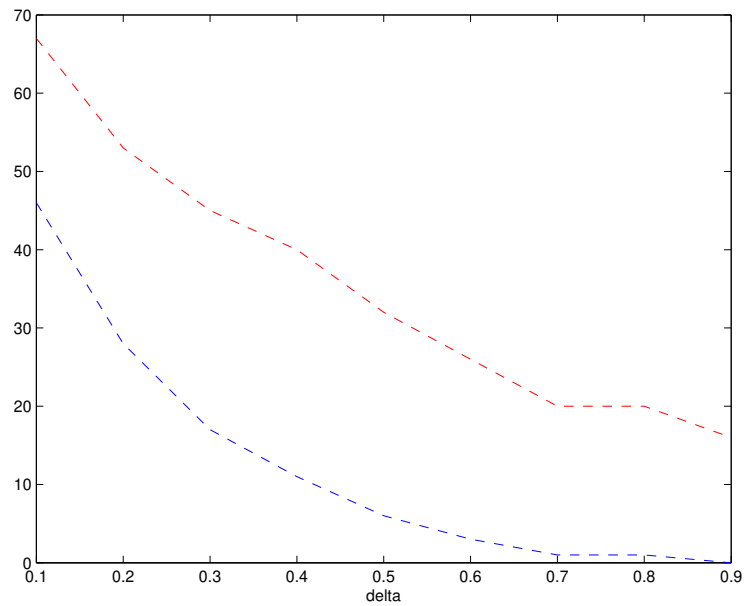
**Figure 3.** maximum and minimum of  $q_i$  with respect to  $\delta$ . FB data.  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$ .



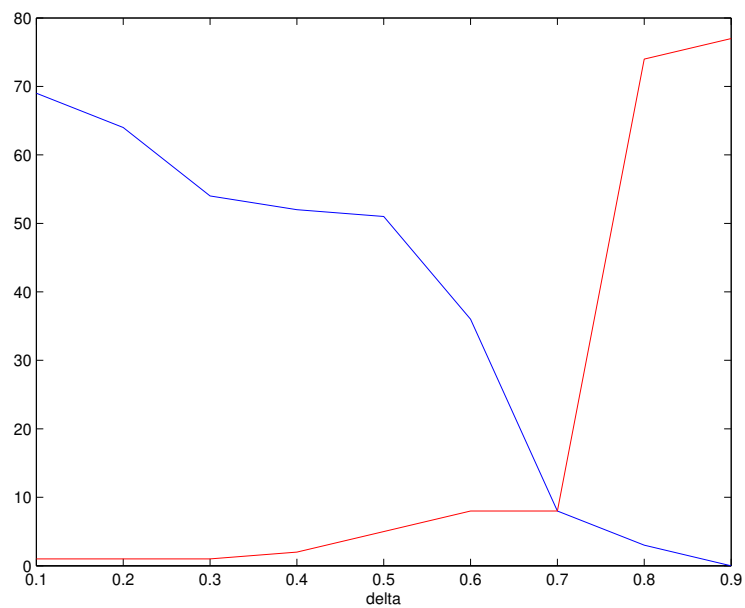
**Figure 4.** Set  $\mathcal{N}_E$  of observable pairs  $(m, q)$ . FB data.  $\delta = 0.5$ ,  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$ .

In all the above listed figures, the option market prices are displayed in blue as a function of the strikes.

The values of the relevant parameters are displayed in the captions, notice that for type II models

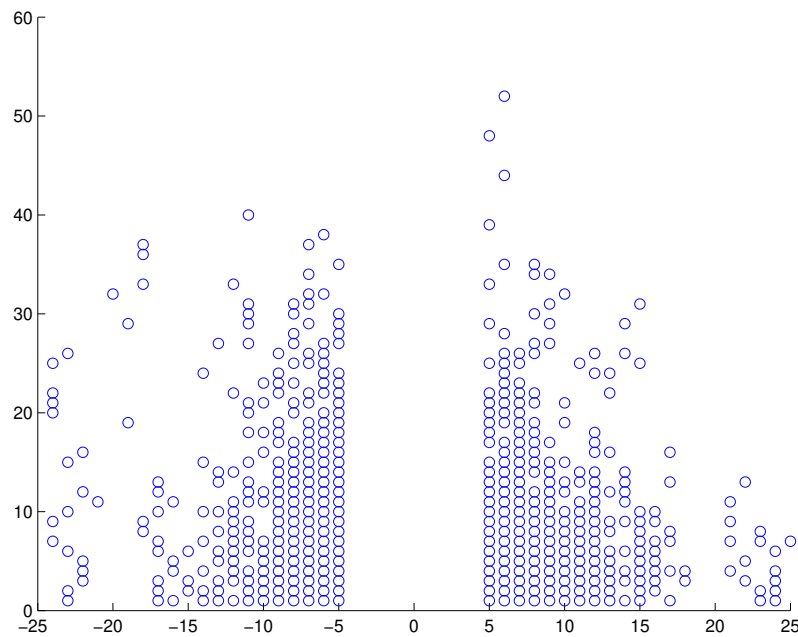


**Figure 5.**  $N_*$  and  $N^*$ , i.e. minimum and maximum number of  $\delta$ -increments over historical data with respect to  $\delta$ . FB data.  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$ .

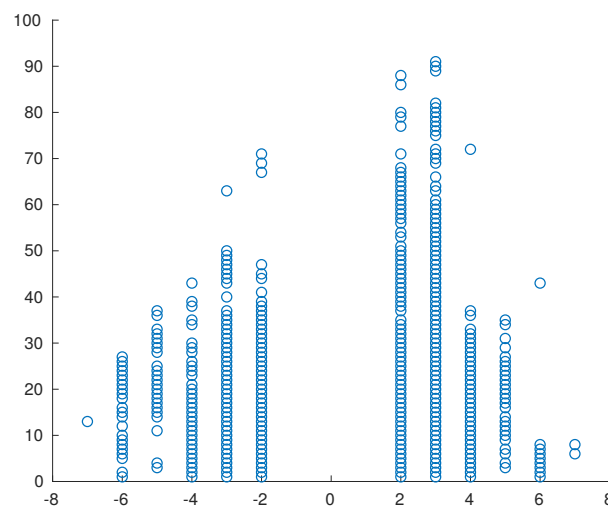


**Figure 6.**  $\tau$  (blue, decreases) and  $\Gamma$  (red, increases) with respect to  $\delta$ . FB data.  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$ .

the values of  $\delta_0, \delta$  are larger than in the case of type I models. As explained before, this was done to speed up the execution times from the order of several days to a few hours (details are provided in Appendix 1). In the case of type II models we have also sub-sampled  $\mathcal{N}_E$  as follows. For the



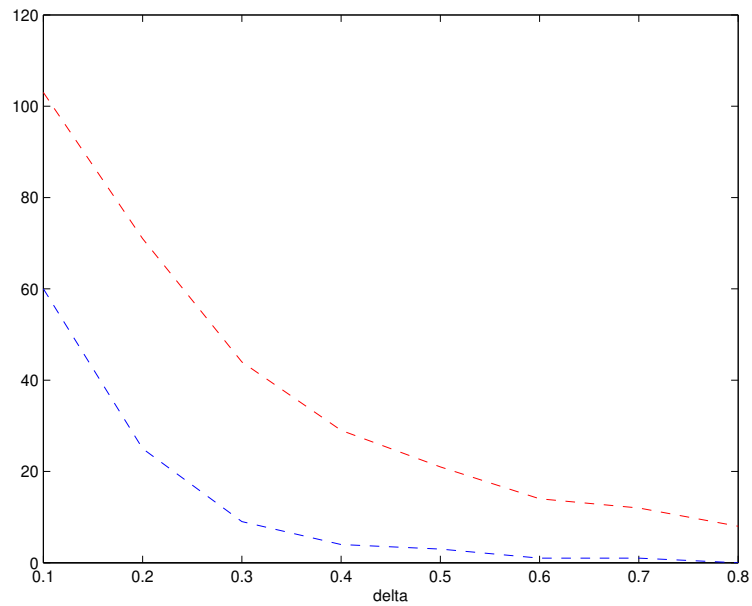
**Figure 7.** Set  $\mathcal{N}_E$  of observable pairs  $(m, q)$ . SPY data.  $\delta = 0.5$ ,  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$



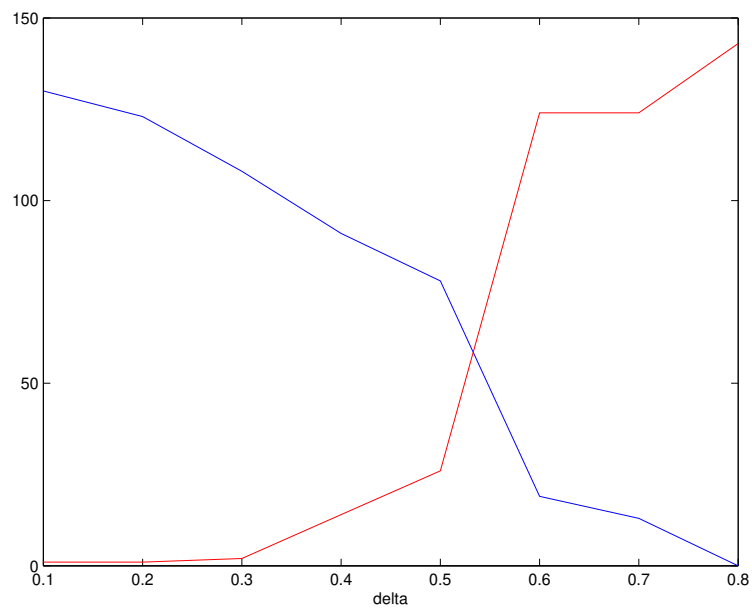
**Figure 8.** Set  $\mathcal{N}_E$  of observable pairs  $(m, q)$ . SPY data.  $\delta = 0.8$ ,  $\delta_0 = 0.4$ ,  $\hat{\delta}_0 = 0.01$

DP model we skipped every other value of  $q$  and kept the extreme values of  $m$ . For the OBS model, for every  $q$  value we kept only the extreme values of  $m$ . Another shortcut we took when generating values for  $\Delta_i W = \sum_{j=0}^{s-1} |p_j| \delta_0$  with the constraints (23) in effect. An explanation of a simple algorithm approximating this computation is described in Appendix 1, the parameter  $K$  (not to be confused with the option's strike) in that algorithm was taken to be  $K = 50$ .

The model MIDP is quite conservative given the construction of  $\mathcal{N}_E$  as a direct product; the model's

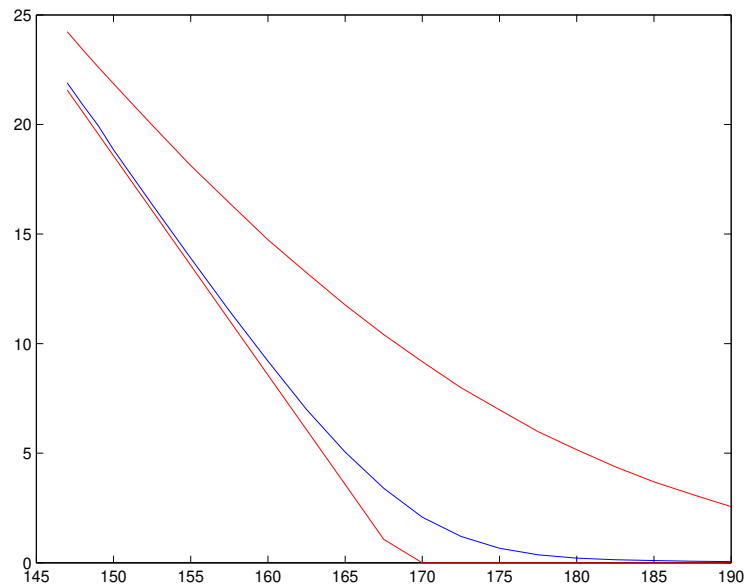


**Figure 9.**  $N_*$  and  $N^*$ , i.e. minimum and maximum number of  $\delta$ -increments over historical data with respect to  $\delta$ . SPY data,  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$ .

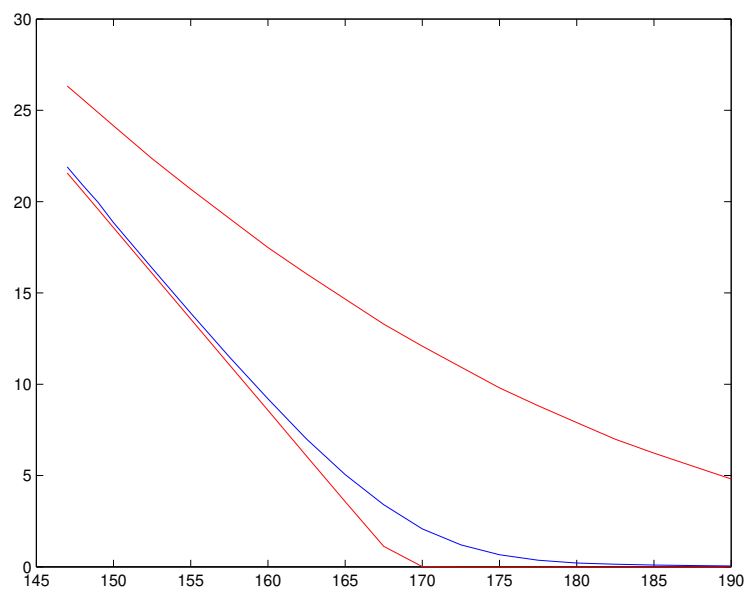


**Figure 10.**  $\tau_*$  (blue, decreases) and  $\Gamma$  (red, increases) with respect to  $\delta$ . SPY data,  $\delta_0 = 0.1$ ,  $\hat{\delta}_0 = 0.01$ .

bounds reflect this fact. In contrast, the effect of using variation for constraining trajectories is illustrated in MIIDP which substantially tightens up the bounds (see Section 11 for the explanation). Consistently, even tighter bounds are given by the model MII OBS as the historically observed pairs



**Figure 11.** FB data model type I, MIOBS. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 0.5, \delta_0 = 0.1$

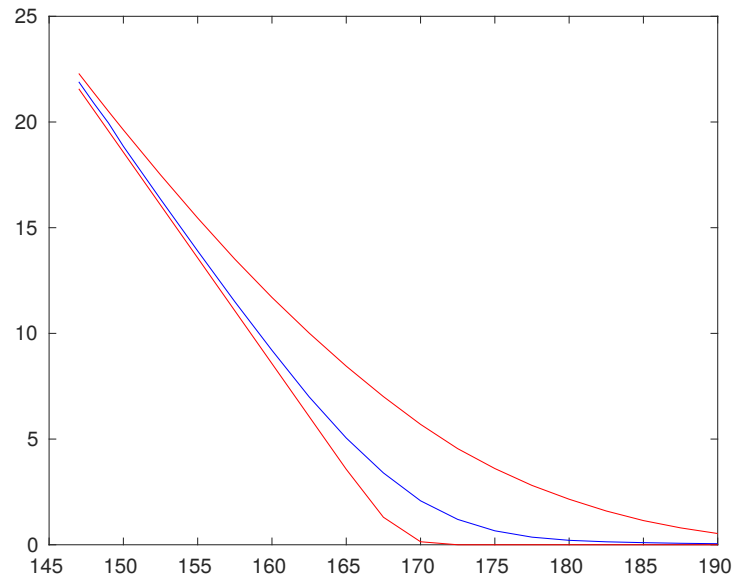


**Figure 12.** FB data model type I, MIDP. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 0.5, \delta_0 = 0.1$

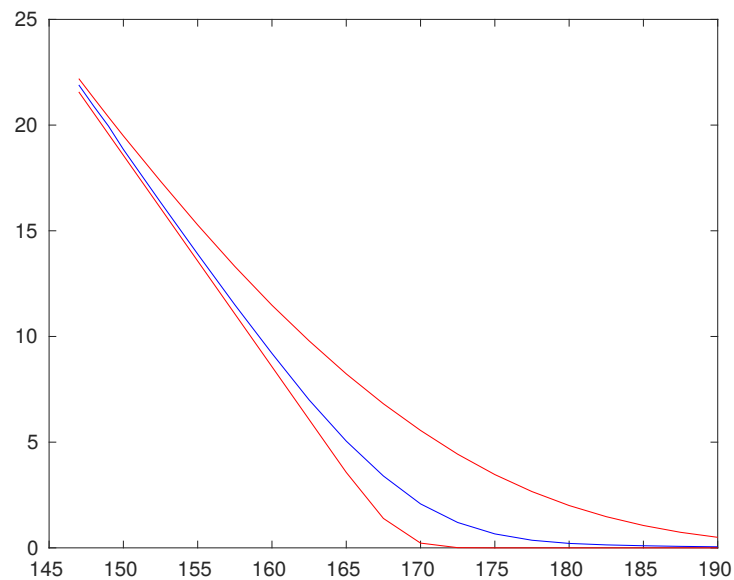
$(m, q)$  are a subset of  $\mathcal{N}_E$  when the latter is given by a direct product. Figures 18 and 19 show the variation values used for the output for type II models.

Some details on execution times are provided in Appendix 1, for example the model MIOBS took



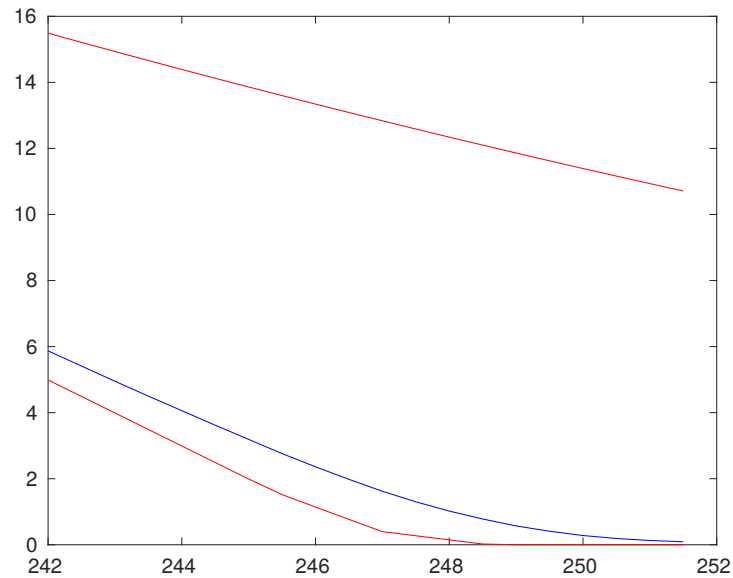


**Figure 13.** FB data model type II, MIIDP. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 1$ ,  $\delta_0 = 0.5$

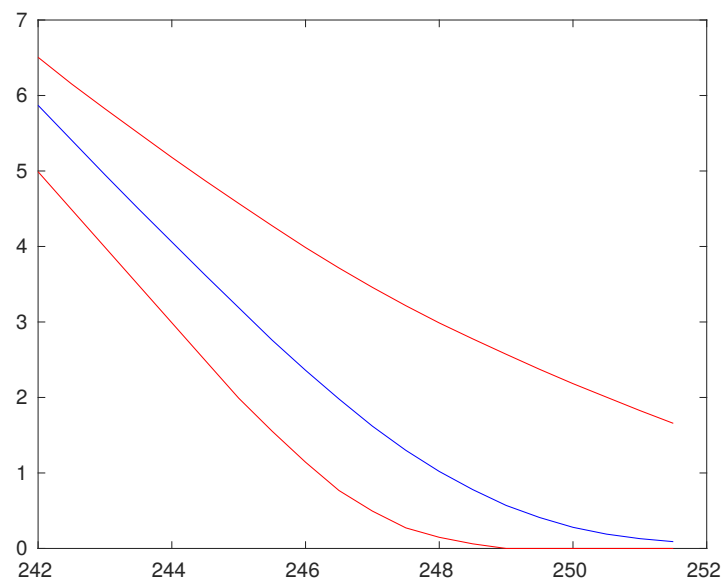


**Figure 14.** FB data model type II, MII OBS. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 1$ ,  $\delta_0 = 0.5$

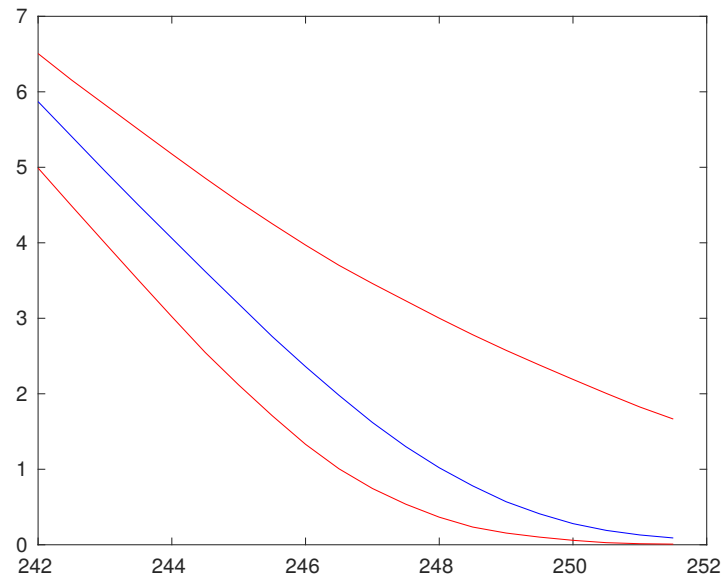
77 minutes in 4 CPU's at 2.60GHz and 7.5 GiB of RAM plus 4 GB of Swap memory (in a Linux system). Model MII OBS took 315 minutes (these times are for the SPY data set). While it may seem that compared to classical finite differences and Monte Carlo methods these times are prohibitively long, we would like to point out the following: computing maximal hedging bounds (i.e. the pricing



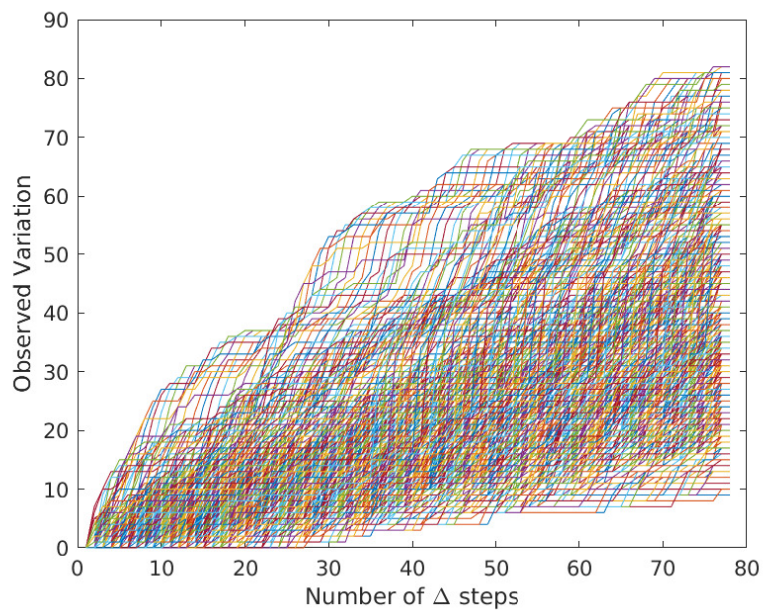
**Figure 15.** SPY data model type I, MIDP. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 0.8$ ,  $\delta_0 = 0.4$ . In contrast to the output for FB, and in order to have one consistent comparison, this type I model output was generated in the same way as the corresponding type II models (in particular subsampling  $\mathcal{N}_E$ ). Appendix 2 has the output for the full model.



**Figure 16.** SPY data model type II, MIIDP. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 0.8$ ,  $\delta_0 = 0.4$

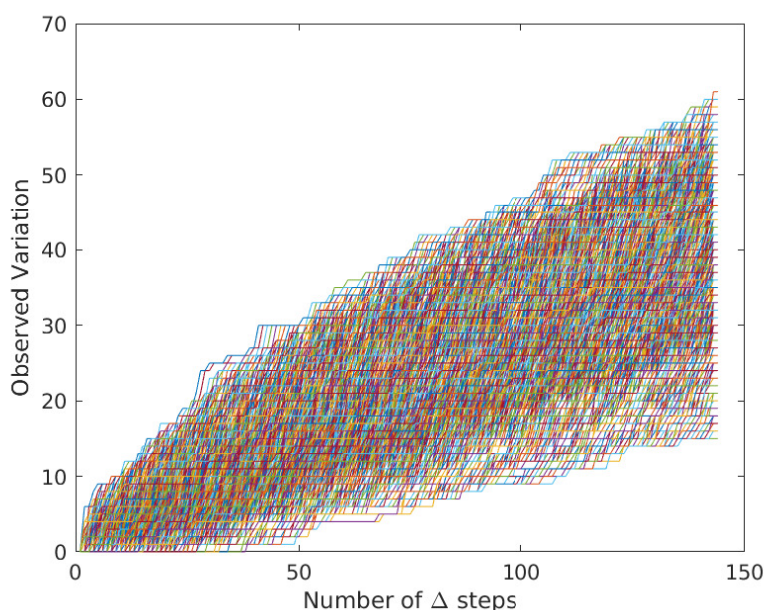


**Figure 17.** SPY data model type II, MIIOBS. Price bounds and market price (in blue) as functions of European call strikes.  $\delta = 0.8$ ,  $\delta_0 = 0.4$



**Figure 18.** Historical variation constraint sets for FB data. Obtained for  $\delta = 1$ ,  $\delta_0 = 0.5$ . The  $x$ -axis represent the time steps.

component of our algorithm) is of comparable complexity to the classical pricing algorithm as it can be viewed as a linear dual problem to computing the maximal price in an incomplete discrete time framework. The later problem being itself a convex hull problem with the risk neutral probabilities the hull weights. On the other hand, and as a way of contrast with stochastic models, we have not made any



**Figure 19.** Historical variation constraint sets for SPY data. Obtained for  $\delta = 0.8$ ,  $\delta_0 = 0.4$ . The  $x$ -axis represent the time steps.

assumptions on the dynamics of the underlying. Correspondingly, we have no assumptions to draw on that will help us find bounds in our models. In addition, our models make use of either every available observed movement or make worse case assumptions and so there is an extra cost to build the graph data structure required to encode the trajectory set. More general stochastic models (e.g. general Levy processes) are often similarly computationally expensive but offering less generality than our setting. Conversely, binomial and trinomial trees often used in industry to calibrate American option prices would obviously have much shorter execution times than the models presented here.

Some details on execution times are provided in Appendix 1.

## 10. Risk-reward under worst case

Let us consider the case of an investor selling a European call option  $Z$  (with an analogous analysis for the case of buying an option) at time  $t_0$  with strike  $K$  and expiring at  $t_0 + T$ . She will set up a portfolio that will be rebalanced as described in Section 2 and Appendix 1 with a chosen parameter  $\delta$ . Let  $P(t_0, K)$  be the actual market price at which one option contract can be sold and  $\mathcal{M} \equiv \mathcal{S} \times \mathcal{H}$  a trajectorial model. If  $P(t_0, K) > \bar{V}(S_0, Z, \mathcal{M})$  and the unfolding chart  $x(t)$ ,  $t \in [t_0, t_0 + T]$ , is associated to some  $\mathbf{S} \in \mathcal{S}$ , then the investor will profit  $P(t_0, K) - \bar{V}(S_0, Z, \mathcal{M})$  per contract. This profit is guaranteed by (4) under the assumption that *worst case estimates are not violated*. Under such assumption, if  $\mathcal{S}$  is one of the two trajectory sets types described in the paper and its parameters are estimated to worst case, then  $x \rightarrow \mathbf{S}$  holds. This is so by either the conclusion of Proposition 6.0.1 or Proposition 6.0.2 which hold given that the assumption that worst case estimates are not violated implies the hypothesis of the said propositions. The conclusion is that  $P(t_0, K) > \bar{V}(S_0, Z, \mathcal{M})$  implies a profit with certainty.

The other logical possibility, in case that  $P(t_0, K) \leq \bar{V}(S_0, Z, \mathcal{M})$ , is to set the parameters in  $\mathcal{S}$  to risk

taking estimates (all such possibilities were introduced in Section 8). This entails setting parameter values so that  $P(t_0, K) > \bar{V}(S_0, Z, \mathcal{M}')$  holds (here  $\mathcal{M}'$  is the model with the said parameters set to risk taking values). It should be clear that the risk being taken is relative to the worst case and that empirical frequencies of failure can be evaluated.

Notice that nothing prevents the models to fail just because the future may not be included in our worst case past estimations. Clearly, distinct models will behave differently regarding this possibility.

The above discussion on profits also applies when we replace the notion of association between  $x$  and  $\mathbf{S}$  by weak association introduced in Section 7.

We have not explored the profit and loss function that results by starting the superhedging portfolio with an initial value  $v = P(t_0, K) \leq \bar{V}(S_0, Z, \mathcal{M})$ . Our main remark is that our framework permits to take investment decisions that have an objective meaning in terms of risk-reward.

## 11. Type II models: obstruction for local worst case estimates to propagate to unrealistic bounds

This short section explains why, in the case of type II models, extreme parameter values do not translate into unrealistic bounds. It is the effect of the constraints  $W_i \in V_{T_i}$  that limits the appearance of extreme events. This is easily seen as follows, consider  $(m_{max}, q) \in \mathcal{N}_E$  and  $m_i = m_{max} = \sum_{j=0}^{J-1} p_j$ . Then  $|m_{max}|\delta_0 \leq \sum_{j=0}^{J-1} |p_j|\delta_0 = \Delta_i W$ , hence a trajectory  $\mathbf{S}$  can not have  $\Delta_i S = m_{max}\delta_0$  at too many coordinates  $i$ . Otherwise,  $W_{T_k} \in V_{T_k}$  will not hold at some  $k$  (more precisely, this condition will fail as soon as the element  $\mathbf{S}$  under consideration does not reflect the historical worst case constraint being enforced by  $V_{T_k}$ ). A similar analysis applies to  $m_{min}$  (under the assumption that  $|m_{min}|$  is an extreme value). To summarize: an empirical constraint in variation (i.e.  $W_k \in V_{T_k}$ ) limits the number of nodes  $(\mathbf{S}, i)$  at which the investor has to account for the worst case possible stock movements (i.e. large values of  $|m_i|$ ). This has the effect of reducing the required funds  $\bar{V}_i(\mathbf{S}, Z, \mathcal{M})$ , this fact can actually be proven rigorously for convex payoffs  $Z$ . The resulting effect is a smaller price interval  $[\underline{V}_0, \bar{V}_0]$ . The net effect is to limit the appearance of a large number of extreme values of  $m$  along any trajectory, notice that those large moves could appear anywhere in a trajectory and we do not assume any particular distribution. It will be of much interest to find a local condition (i.e. only involving consecutive times  $T_i, T_{i+1}$ ) that achieves a similar result.

One could replace our use of *variation* by *quadratic variation*.

## 12. Discussion

The methodology presented in the paper shows, by example, the possibility to construct models with a financial operational meaning. The elements used to construct the models result from portfolio rebalances and updates of data summaries (i.e. observation and updates for the variables of interest). There is then an association, via operations, between financial events and the model components. In other words, we assume a class of agents that act and react to financial events in a general and well defined manner. This setup allows the agents to obtain prices relative to their operations; the risk and reward characteristic of these prices are then linked to observable events. The price bounds so obtained have an objective meaning for these agents as models' features correspond to specific set of conditions faced by the agents. This characteristic makes the models of great value to investors. Clearly, it is

easy to argue about the existence of this type of models without going to the trouble, i.e. analysis and implementation, that we have taken in our paper. The key comment is that the proposed models, when calibrated to worst case conditions (i.e. a robust calibration), can provide tight bounds as shown in the paper. These tight bounds can translate into attractive profit and loss characteristics. The latter type of analysis is a time consuming exercise not developed in the paper and better approached in higher frequency setting with several stocks and with more practical goals in mind. We remark that it is possible to replace the option by another stock and extend the analysis to a multidimensional setting making the whole methodology easier to deploy for testing. This exercise has been performed in Crisci (2019) and will be reported elsewhere. Nonetheless, we should emphasize that, in our opinion, the essential ingredient to obtain useful operational models is the selection of key variables affecting the agent's portfolio rebalances. Profit and loss analysis leading to investment decisions should be reliable and rewarding in relation to the relevance of the variables used in the model's construction.

### Acknowledgments

S. Ferrando acknowledges partial financial support from NSERC for the completion of this project.

### Conflict of interest

All authors declare no conflicts of interest in this paper.

### References

- Brigo D, Mercurio F (2006) Interest rate models-theory and practice: with smile, inflation and credit.
- Britten-Jones M, Neuberger A (1996) Arbitrage pricing with incomplete markets. *App Math Financ* 3: 347–363.
- Campbell JY, Lo AW, MacKinlay CA, et al. (1997) *The Econometrics of financial markets*, Princeton, NJ: Princeton University Press.
- Carassus L, Gobet E, Temam E (2006) A class of financial products and models where super-replication prices are explicit. *Proceedings of the Ritsumeikan International Symposium on Stochastic Processes and Applications to Mathematical Finance*, 67–84.
- Carassus L, Vargiolu T (2018) Super-replication price: it can be ok. *Esaim Proc Surv* 64: 54–64.
- Crisci D (2019) Trajectory based market models for two stocks. MSc. Thesis, Mathematics, Ryerson University.
- Cutland NJ, Roux A (2012) Derivative pricing in discrete time.
- Degano IL, Ferrando SE, González AL (2018) Trajectory based models. evaluation of minmax pricing bounds. *Dyn Cont Discrete Impulsive Ser B Appl & Algorithms* 25: 97–128.
- Eberlein E, Jacod J (1997) On the range of option prices. *Financ Stoch* 1: 131–140.

- Feng L, Li B, Podobnik B, et al. (2012) Linking agent-based models and stochastic models of financial markets. *Proceedings of the National Academy of Sciences of the United States of America* 109: 347–363.
- Ferrando S, et al. (2019a) Appendix 1: Algorithm, trajectorial asset models in matlab. *Supplementary material to: “Trajectorial Asset Models with Operational Assumptions”*.
- Ferrando S, et al. (2019b) Appendix 2: Matlab software bundle implementing trajectorial models. file: `Operational_root.zip`. *Supplementary material to: “Trajectorial Asset Models with Operational Assumptions”*.
- Ferrando SE, González AL, Degano IL, et al. (2019) Trajectorial market models: arbitrage and pricing intervals. *Revista de la Unión Matemática Argentina* 60: 149–185.
- Föllmer H, Schied A (2013) *Stochastic finance: an introduction in discrete time*. third edition, 3.
- Gilboa I (2009) *Theory of decision under uncertainty*.
- Kahalé N (2017) Super-replication of financial derivatives via convex programming. *Manage Sci* 63: 2323–2339.
- Merton R (1973) Theory of rational option pricing. *Bell J Econ Manage Sci* 4: 141–183.
- Pfleiderer P (2014) Chameleons: the misuse of theoretical models in finance and economics. *Revista de Econ Institucional* 16: 23–60.
- Rebonato R (2012) *Volatility and correlation: The perfect hedger and the fox*, second edition.
- Schoutens W, Simons E, Tistaert J (2006) A perfect calibration! now what? *Best Wilmott* 2: 281–304.
- Sniedovich M (2016) From statistical decision theory to robust optimization: A maximin perspective on robust decision-making, *Robustness Analysis in Decision Aiding, Optimization and Analytics*. Editors, M. Doumpos, C. Zopounidis and E. Grigoroudis., International Series in Operations Research & Management: 59–85.

## Supplementary Material

### A. Appendix 1

This appendix links to a separate file (Ferrando and et. al. (2019a)) providing details of a Matlab software implementation for the models introduced in the paper.

### B. Appendix 2

This appendix links to a separate compressed zip file (Ferrando and et. al. (2019b)). The zip file is named `Operational_root` and contains a software bundle written in Matlab. The software implements the methods and algorithms described in the paper and can be used to reproduce the output displayed in the main paper. The software bundle also contains a `README.pdf` file, a copy of

---

Appendix 1, and some other files as described in README.pdf. The data necessary to reproduce the outcomes displayed in the paper is also included in the bundle. There is also the file OutputOperationalModels.pdf, this is the document entitled "Output for Trajectorial Asset Models with Operational Assumptions" and contains extra output beyond the one contained in the main paper.



AIMS Press

©2019 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)