Distributionally Robust XVA via Wasserstein Distance

Part 1: Wrong Way Counterparty Credit Risk

Derek Singh, Shuzhong Zhang

Department of Industrial and Systems Engineering, University of Minnesota
singh644@umn.edu, zhangs@umn.edu

Abstract

This paper investigates calculations of robust CVA for OTC derivatives under distributional uncertainty using Wasserstein distance as the ambiguity measure. Wrong way counterparty credit risk can be characterized (and indeed quantified) via the robust CVA formulation. The simpler dual formulation of the robust CVA optimization is derived. Next, some computational experiments are conducted to measure the additional CVA charge due to distributional uncertainty under a variety of portfolio and market configurations. Finally some suggestions for future work, such as robust FVA, are discussed.

Keywords—OTC, CCR, CVA, derivatives, distributional robust optimization, Wasserstein distance, duality

1 Introduction and Overview

1.1 Financial Markets Context and Background

Credit valuation adjustment (CVA) represents the impact on portfolio market value due to counterparty default. It represents the market value of counterparty default risk. Unilateral CVA can be represented mathematically as an integral of discounted expected positive exposure times (incremental) counterparty default probability. The market valuation is a function of counterparty credit risk, the underlying (market) risk factors that drive the portfolio valuation (and hence positive exposure), as well as the correlations between these market risk factors and the counterparty credit risk curves for a given portfolio. CVA is typically measured and reported at the counterparty level.

The “other side” of unilateral CVA is unilateral debit valuation adjustment (DVA). This is the benefit to the firm, of its reduced liability, as measured by discounted expected negative exposure times firm default probability. As above, the market valuation is a function of firm credit risk, underlying market risk factors that drive portfolio valuation, and the correlations. Unilateral DVA can be represented mathematically as an integral of discounted negative exposure times (incremental) firm default probability. DVA is typically measured at the firm level.

Bilateral CVA represents the dual impact on portfolio market value due to counterparty default and firm default. Bilateral CVA can be represented mathematically as the difference between two integrals: (i) discounted expected positive exposure times (incremental) counterparty default probability prior to firm default, (ii) discounted expected negative exposure times (incremental) firm default probability prior to counterparty default. Bilateral CVA is typically measured and reported at the counterparty level, for a given firm.

U.S. regulatory authorities, the Federal Reserve and Office of the Comptroller of the Currency (OCC), periodically assess national banks’ compliance with Market Risk Capital Rule (MRR). Counterparty credit risk (CCR) metrics are key metrics used to evaluate bank risk profiles and balance sheet exposures due to over the counter (OTC) derivatives, securities financing transactions, and other transactions and exposures (Office of the Comptroller of the Currency [2011]). Basel Committee on Banking Supervision has issued supervisory guidance, in the form of its Basel III framework (and supplemental guidance), to quantify capital charges due to CCR. A new element in Basel III was a capital charge due to degradation in CCR for a given portfolio or book of business (Basel Committee on Banking Supervision [2015]). Potential revisions to the Basel framework may include elements to quantify CCR capital charges due to deterioration in market risk exposure.

The Dodd-Frank Wall Street Reform and Consumer Protection Act (July 2010) enacted regulations for the swaps market and authorized creation of centralized exchanges for swaps (and other) derivatives trading. Derivatives that trade on an exchange reference the exchange as the transaction counterparty. Since exchanges clear multiple (typically offsetting) transactions and hedge their risk through other third parties, exchange traded derivatives have minimal CCR risk profile. However, OTC derivatives typically have banks or other financial institutions as counterparties which do have material credit risk profiles. According
to International Swap Dealers Association (ISDA) the OTC derivatives notional outstanding was 544 trillion at year end 2018. Interest rate derivatives notional outstanding was 437 trillion at year end 2018. Current (07/10/19) Bloomberg CDX investment grade and high yield credit spreads are 53 and 323 basis points respectively. Consequently the CCR exposures inherent in the OTC derivatives market represent significant market risk exposures. This motivates the concepts of worst case CVA and wrong way risk (WWR) and the impact of uncertainty in probability distribution on CCR and CVA. It is these considerations that motivate this line of research (Ramzi Ben-Abdallah and Marzouk, 2019), (El Hajjaji and Subbotin, 2015).

An outline of this paper is as follows. Section 1 gives an overview of CVA and WWR as well as a literature review. Section 2 develops the main theoretical results of the paper and provides proof sketches. Section 3 conducts a computational study of WWR for a representative set of (single) derivative instruments, portfolios, and market environments. Section 4 discusses the conclusions and suggestions for future research. All detailed proofs of propositions, corollaries, and theorems are deferred to the Appendix.

1.2 Literature Review

In the past few years some research has been done to investigate and quantify the effect of distributional uncertainty on CVA. Brigo et al. (2013) explicitly incorporate correlation into the stochastic processes driving the market risk and credit default factors. They quantify the effect of dependency structure (and hence wrong way risk) on CVA for a variety of asset classes: interest rate swaps, interest rate swaptions, commodities, equities, and foreign exchange products. Glasserman and Yang (2015) bound the effect of wrong way risk on CVA. Their approach considers a discrete setting for portfolio exposures and counterparty default times and formulates worst case CVA as the solution to a worst case linear program subject to certain constraints (such as fixed marginals for portfolio exposures and default times), where the dependency structure across the risk factors is allowed to vary. As this approach leads to large values for worst case CVA, they introduce a penalty term to modulate or temper the degree of wrong way risk and run some sensitivity analysis to study the effect of the penalty term. Kullback-Leibler (KL) divergence is used to measure the distance between the reference (empirical) and the perturbed distribution. They remark that determining a suitable value for the penalty term would be a topic for further research.

Memartoulie, in his PhD thesis, uses an ordered scenario copula methodology to quantify worst case CVA (Memartoulie, 2017). A particular method of scenario ordering correlates portfolio exposures to company default times (firm, counterparty, or both) and the resulting dependency structure introduces wrong way risk. He chooses to order exposure scenarios by increasing time averaged total exposure and then simulates company default times conditional on the exposure path using pre-specified correlation between the market risk factor(s) and credit risk factor(s). For worst case correlations set to one, he finds results for worst case CVA that are comparable to the method of Glasserman and Yang. In a recent paper, Ben-Abdallah et al. perform a computational study on the effect of wrong way risk on CVA for a portfolio of interest rate swaps, caps, and floors (Ramzi Ben-Abdallah and Marzouk, 2019). They find that the dependency structure between interest rates and default intensity produces material wrong way risk whereas the dependency structure between interest rate volatility and default intensity does not.

Recent results in Lagrangian duality were independently developed by Blanchet and Murthy (2019) and Gao and Kleywegt (2016). These results hold under mild assumptions such as upper semicontinuity in the loss function and lower semicontinuity in the distance metric. Blanchet et al. (2016) applied this duality theory to study a number of classical regression problems in machine learning under distributional uncertainty. In that context, the authors find that distributional uncertainty can be viewed as adding a regularization term analogous to a penalized regression setting. Similarly, Gao et al. (2017) apply the Lagrangian duality theory to problems in statistical learning.

The main innovation in our work is to apply these recent results in Lagrangian duality to worst case CVA using Wasserstein distance as the ambiguity measure. Furthermore, analytical expressions are derived for the solutions to the inner and outer convex optimization problems that comprise worst case CVaR via the Wasserstein approach. A computational study shows the material impact of distributional uncertainty on worst case CVA and illustrates the risk profile.

1.3 Notation and Definitions

1.3.1 Unilateral CVA

Notation and core definitions for unilateral CVA problem setup follow conventions in Glasserman and Yang (2015). Those for the robust CVA problem formulation follow conventions in Blanchet et al. (2018). Unilateral CVA measures expected portfolio loss at time of counterparty default. Let $V^+(\tau_C)$ denote the discounted positive portfolio exposure at time $\tau_C$ and let $R_C \in [0,1)$ denote the recovery rate the firm receives upon counterparty default. The problem setup here assumes a fixed set of observation dates, $0 = t_0 < t_1 < \cdots < t_n = T$. Let $X^+$ denote the vector of recovery adjusted discounted positive exposures and $Y_C$ denote the vector of counterparty default indicators. Let $(x^+_i, y^+_i)$ denote realizations of $(X^+, Y_C)$ along sample paths for $i \in \{1,2,\ldots,N\}$. 

$$x^+_i \in C$$

$$y^+_i \in \{0,1\}$$

$$C = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for } i = 1,2,\ldots,n\}$$

$$\tau_C = \min \{t \geq t_0 : Y_C = 1\}$$

$$R_C = \frac{1}{t_{\tau_C}} \int_{t_0}^{t_{\tau_C}} x(t) \, dt$$

$$V^+(\tau_C) = \frac{1}{t_{\tau_C} - t_0} \int_{t_0}^{t_{\tau_C}} x^+(t) \, dt$$

$\mathbb{R}^n$ denotes the set of real numbers in $n$ dimensions.
The unilateral CVA associated with discounted positive exposure $V^+(\tau_C)$ and counterparty default indicator $\mathbb{1}_{\{\tau_C \leq T\}}$ is

$$CVA^U = \mathbb{E}[(1-R_C)V^+(\tau_C)\mathbb{1}_{\{\tau_C \leq T\}}] = (1-R_C)\int_0^T \mathbb{E}[V^+(t)|\tau_C = t]d\Pi_C(t),$$

where the counterparty default time distribution is given by $\Pi_C(t) = P(\tau_C \leq t)$ \cite{Green2015, Lichters2015, Memartolue2017}. The pair of vectors $(X^+, Y_C) \in (\mathbb{R}^n_+ \times B^1_n)$ is

$$X^+ = ((1-R_C)V^+(\tau_1), \ldots, (1-R_C)V^+(\tau_n)) \quad \text{and} \quad Y_C = (\mathbb{1}_{\{\tau_1 = \tau_1\}}, \ldots, \mathbb{1}_{\{\tau_n = \tau_n\}}).$$

Here $B^1_n$ denotes the set of default time vectors: binary vectors of ones and zeros with $n$ components, and at most one non-zero element. Note that default occurs on at most one observation date within the fixed set of dates in the problem setup. The empirical measure, $P_N$, is

$$P_N(dz) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{(x_i^+, y_i^+)}(dz).$$

Under the empirical measure, $P_N$, unilateral CVA is an expectation of an inner product

$$CVA^U = \mathbb{E}^{P_N}[(X^+, Y_C)].$$

In the context of this work, the uncertainty set for probability measures is

$$\mathcal{U}_{\delta_1}(P_N) = \{P : D_c(P, P_N) \leq \delta_1\}$$

where $D_c$ is the optimal transport cost or Wasserstein discrepancy for cost function $c$ \cite{Blanchet2018}. For convenience the definition for $D_c$ is given as

$$D_c(P, P') = \inf\{\mathbb{E}^{P}[c(A, B)] : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_A = P, \pi_B = P'\}$$

where $\mathcal{P}$ denotes the space of Borel probability measures and $\pi_A$ and $\pi_B$ denote the distributions of $A$ and $B$. Here $A$ denotes $(X^+_1, Y_A) \in (\mathbb{R}^n_+ \times B^1_n)$ and $B$ denotes $(X^+_\delta, Y_B) \in (\mathbb{R}^n_+ \times B^1_n)$ respectively. The analysis in this work uses the cost function $c_{S_1}$ where

$$c_{S_1}(u,v, -y) = S_1(v-y, v-y) + (u-x, u-x).$$

The scale factor $S_1 > 0$ is used to compensate for different domains: $(u,v) \in (\mathbb{R}^n_+ \times B^1_n), (x,y) \in (\mathbb{R}^n_+ \times B^1_n)$.

### 1.3.2 Unilateral DVA

Notation and core definitions for unilateral DVA problem setup mirror those for unilateral CVA problem setup. Unilateral DVA measures expected portfolio benefit at time of firm default. Let $V^-(\tau_F)$ denote the discounted negative portfolio exposure at time $\tau$ and let $R_F \in [0, 1]$ denote the recovery rate the counterparty receives upon firm default. Alternatively, from the counterparty’s perspective, $V^-(\tau_F)$ denotes the counterparty’s discounted positive exposure at time $\tau_F$. The problem setup here assumes a fixed set of observation dates, $0 = t_0 < t_1 < \cdots < t_n = T$. Let $X^-$ denote the vector of recovery adjusted discounted firm negative (counterparty positive) exposures and $Y_F$ denote the vector of firm default indicators. Let $(x^-_i, y^-_j)$ denote realizations of $(X^-, Y_F)$ along sample paths for $i = \{1, 2, \ldots, N\}$.

The unilateral DVA associated with discounted negative exposure $V^-(\tau_F)$ and firm default indicator $\mathbb{1}_{\{\tau_F \leq T\}}$ is

$$DVA^U = -\mathbb{E}[(1-R_F)V^-(\tau_F)\mathbb{1}_{\{\tau_F \leq T\}}] = -(1-R_F)\int_0^T \mathbb{E}[V^-(t)|\tau_F = t]d\Pi_F(t),$$

where firm default time distribution is given by $\Pi_F(t) = P(\tau_F \leq t)$ \cite{Green2015, Lichters2015, Memartolue2017}. The pair of vectors $(X^-, Y_F) \in (\mathbb{R}^n_- \times B^1_n)$ is

$$X^- = ((1-R_F)V^-(\tau_1), \ldots, (1-R_F)V^-(\tau_n)) \quad \text{and} \quad Y_F = (\mathbb{1}_{\{\tau_1 = \tau_1\}}, \ldots, \mathbb{1}_{\{\tau_n = \tau_n\}}).$$

Here $B^1_n$ denotes the set of default time vectors: binary vectors of ones and zeros with $n$ components, and at most one non-zero element. Note that default occurs on at most one observation date within the fixed set of dates in the problem setup. The empirical measure, $Q_N$, is

$$Q_N(dz) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{(x^-_i, y^-_j)}(dz).$$
Under the empirical measure, $Q_N$, unilateral DVA is an expectation of an inner product
\[ \text{DVA}^U = -\mathbb{E}^Q \langle [X^-, Y_F] \rangle. \]

In the context of this work, the uncertainty set for probability measures is
\[ \mathcal{U}_{\delta_2}(Q_N) = \{ Q : D_{\varepsilon}(Q, Q_N) \leq \delta_2 \} \]
where $D_{\varepsilon}$ is the optimal transport cost or Wasserstein discrepancy for cost function $c$ [Blanchet et al., 2018]. For convenience the definition for $D_{\varepsilon}$ is given as
\[ D_{\varepsilon}(Q, Q') = \inf \{ \mathbb{E}^\pi [c(A, B)] : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_A = Q, \pi_B = Q' \} \]
where $\mathcal{P}$ denotes the space of Borel probability measures and $\pi_A$ and $\pi_B$ denote the distributions of $A$ and $B$. Here $A$ denotes $(X_A, Y_A) \in (\mathbb{R}_+ \times B^n_1)$ and $B$ denotes $(X_B, Y_B) \in (\mathbb{R}_+ \times B^n_1)$ respectively. The analysis in this work uses the cost function $c_{S_2}$ where
\[ c_{S_2}((u, v), (x, y)) = S_2(v - y, v - y) + (u - x, u - x). \]
The scale factor $S_2 > 0$ is used to compensate for different domains: $(u, v) \in (\mathbb{R}_+ \times B^n_1), (x, y) \in (\mathbb{R}_+ \times B^n_1)$.

### 1.3.3 Bilateral CVA

Notation and core definitions for bilateral CVA problem setup incorporate those above for unilateral CVA and DVA. Bilateral CVA measures expected portfolio loss (or benefit) due to counterparty and/or firm default. Let $V^+(\tau_C)$ denote the discounted positive portfolio exposure at time $\tau$ and let $R_C \in [0, 1)$ denote the recovery rate the firm receives upon counterparty default. Let $V^-(\tau_F)$ denote the discounted negative portfolio exposure at time $\tau$ and let $R_F \in [0, 1)$ denote the recovery rate the counterparty receives upon firm default. The problem setup here assumes a fixed set of observation dates, $0 = t_0 < t_1 < \ldots < t_n = T$. Let $X^+$ denote the vector of recovery adjusted discounted positive exposures and $Y_C$ denote the vector of counterparty default indicators. Let $(x^+_i, y^+_i)$ denote realizations of $(X^+, Y_C)$ along sample paths for $i = \{1, 2, \ldots, N\}$. Let $X^-$ denote the vector of recovery adjusted discounted firm negative exposures and $Y_F$ denote the vector of firm default indicators. Let $(x^-_i, y^-_i)$ denote realizations of $(X^-, Y_F)$ along sample paths for $i = \{1, 2, \ldots, N\}$.

Due to the linkage, one can write $X = X^+ + X^-$ and decompose sample realizations of $X$ accordingly. Therefore, let $(x_i, y^+_i, y^-_i)$ denote realizations of $(X, Y_C, Y_F)$ along sample paths for $i = \{1, 2, \ldots, N\}$. The relation $x_i = x^+_i + x^-_i$ can be used to decompose $x_i$ into its positive and negative exposures respectively.

The bilateral CVA associated with discounted positive exposure $V^+(\tau_C)$, counterparty default indicator $\mathbb{1}_{\{\tau_C \leq T\} \cap \{\tau_C < \tau_F\}}$, discounted negative exposure $V^-(\tau_F)$, firm default indicator $\mathbb{1}_{\{\tau_F \leq T\} \cap \{\tau_F < \tau_C\}}$, is
\[ \text{CVA}^B = \mathbb{E}[\langle 1 - R_C \rangle V^+(\tau_C) \mathbb{1}_{\{\tau_C \leq T\} \cap \{\tau_C < \tau_F\}}] + \mathbb{E}[\langle 1 - R_F \rangle V^-(\tau_F) \mathbb{1}_{\{\tau_F \leq T\} \cap \{\tau_F < \tau_C\}}]. \]

Equivalently, one can write
\[ \text{CVA}^B = (1 - R_C) \int_0^T \mathbb{E}[V^+(t) | \tau_C = t, \tau_F > t] d\Pi_C(t) + (1 - R_F) \int_0^T \mathbb{E}[V^-(t) | \tau_F = t, \tau_C > t] d\Pi_F(t), \]
where the joint counterparty and firm default time distributions are given by $\Pi_C(t) = P(\tau_C \leq t, \tau_F > \tau_C)$ and $\Pi_F(t) = P(\tau_F \leq t, \tau_C > \tau_F)$ [Green, 2015], [Lichters et al., 2015], [Memartoluc, 2017]. The pair of vectors $(X^+, Y_C) \in (\mathbb{R}_+^n \times B^n_1)$ is
\[ X^+ = ((1 - R_C) V^+(t_1), \ldots, (1 - R_C) V^+(t_n)) \quad \text{and} \quad Y_C = (\mathbb{1}_{\{\tau_C = t_1\} \cap \{\tau_F > \tau_C\}}, \ldots, \mathbb{1}_{\{\tau_C = t_n\} \cap \{\tau_F > \tau_C\}}) \]
and the pair of vectors $(X^-, Y_F) \in (\mathbb{R}_+^n \times B^n_1)$ is
\[ X^- = ((1 - R_F) V^-(t_1), \ldots, (1 - R_F) V^-(t_n)) \quad \text{and} \quad Y_F = (\mathbb{1}_{\{\tau_F = t_1\} \cap \{\tau_C > \tau_F\}}, \ldots, \mathbb{1}_{\{\tau_F = t_n\} \cap \{\tau_C > \tau_F\}}). \]

Here $B^n_1$ denotes the set of default time vectors: binary vectors of ones and zeros with $n$ components, and at most one non-zero element. Note that counterparty or firm default occurs on at most one observation date within the fixed set of dates in the problem setup. The empirical measure, $\Phi_N$, is
\[ \Phi_N(dz) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(x_i, y^+_i, y^-_i)}(dz). \]
Under the empirical measure, $\Phi_N$, bilateral CVA is a sum of expectations of inner products

$$\text{CVA}^B = \mathbb{E}^{\Phi_N}[(X^+, Y_C)] + \mathbb{E}^{\Phi_N}[(X^-, Y_F)].$$

In the context of this work, the uncertainty set for probability measures is

$$\mathcal{P}_{\delta_0}(\Phi_N) = \{P : D_c(\Phi, \Phi_N) \leq \delta_0\}$$

where $D_c$ is the optimal transport cost or Wasserstein discrepancy for cost function $c$ (Blanchet et al. 2018). For convenience the definition for $D_c$ is given as

$$D_c(\Phi, \Phi'_t) = \inf \{ \mathbb{E}^P[c(A, B)] : \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_A = \Phi, \pi_B = \Phi'_t \}$$

where $\mathcal{P}$ denotes the space of Borel probability measures and $\pi_A$ and $\pi_B$ denote the distributions of $A$ and $B$. Here $A$ denotes $(X_A, Y_A, Y'_A) \in (\mathbb{R}^n \times B_n^1 \times B_n^1)$ and $B$ denotes $(X_B, Y_B, Y'_B) \in (\mathbb{R}^n \times B_n^1 \times B_n^1)$ respectively. The analysis in this work uses the cost function $c_{S_3}$ where

$$c_{S_3}(u, v_1, v_2, (x, y_1, v_2)) = S_3(v_1 - y_1, v_1 - y_1) + S_3(v_2 - y_2, v_2 - y_2) + \langle u - x, u - x \rangle.$$  

The scale factor $S_3 > 0$ is used to compensate for different domains: $(u, v_1, v_2) \in (\mathbb{R}^n \times B_n^1 \times B_n^1), (x, y_1, v_2) \in (\mathbb{R}^n \times B_n^1 \times B_n^1)$.

## 2 Theory: Robust CVA and Wrong Way CCR

### 2.1 Unilateral CVA

#### 2.1.1 Inner Optimization Problem

The robust unilateral CVA can be written as

$$\sup_{P \in \mathcal{P}_{\delta_0}(P_N)} \mathbb{E}^P[(X^+, Y_C)]. \tag{P1}$$

Now use recent duality results, noting the inner product $\langle \cdot \rangle$ satisfies the upper semicontinuous condition of the Lagrangian duality theorem, and cost function $c_{S_3}$ satisfies the non-negative lower semicontinuous condition (see Blanchet and Murthy (2019) Assumptions 1 & 2, Gao and Kleywegt (2016)). Hence the dual problem (to sup above) can be written as

$$\inf_{\gamma \geq 0} H(\gamma) := \left[ \gamma \delta_1 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\gamma}(x^+_i, y^-_i) \right] \tag{D1}$$

where

$$\Psi_{\gamma}(x^+_i, y^-_i) = \sup_{u \in \mathbb{R}^n_+, v \in B_n^1} \left[ \langle u, v \rangle - \gamma c_{S_3}(u, (x^+_i, y^-_i)) \right] = \sup_{u \in \mathbb{R}^n_+, v \in B_n^1} \left[ \langle u, v \rangle - \gamma \langle u - x^+_i, u - x^+_i \rangle + S_1(v - y^-_i, v - y^-_i) \right].$$

Now apply change of variables $w_1 = (u - x^+_i)$ and $w_2 = (v - y^-_i)$ to get

$$\Psi_{\gamma}(x^+_i, y^-_i) = \sup_{w_1 \geq -x^+_i, w_2 \in B_n^1} \left[ \langle w_1 + x^+_i, w_2 + y^-_i \rangle - \gamma \langle w_1, w_1 \rangle + S_1(w_2, w_2) \right]$$

where $B_n^1$ denotes the set of ternary vectors of ones, zeros, and minus ones with $n$ components, and at most one +1 and/or one -1 element. Note that $\sup_{w_1} \left[ \cdot \right]$ is attained for $w_1^* \in \mathbb{R}^n_+$ (as will become evident in the proof) hence it suffices to consider this space for $w_1$. It turns out that $\Psi_{\gamma}$ can be expressed as original CVA plus the pointwise max of two convex functions: a hyperbola, and the sum of another hyperbola plus a line of negative slope. $\Psi_{\gamma}$ quantifies the adversarial move in CVA across both time and spatial dimensions while accounting for the associated cost via the $K$ terms.

**Proposition 1.** $\Psi_{\gamma}(x^+_i, y^-_i) = \langle x^+_i, y^-_i \rangle + \frac{||y^-||^2}{4\gamma} + \frac{1}{4\gamma} + (x^+_i - x^-_i - \gamma S_1)K$ or equivalently

$$\Psi_{\gamma}(x^+_i, y^-_i) = \langle x^+_i, y^-_i \rangle + \frac{||y^-||^2}{4\gamma} + \frac{1}{4\gamma} + (x^+_i - x^-_i - \gamma S_1)K$$

where $\tau^*_i = \arg\max_{\tau_i \in \{1, \ldots, n\}, \tau_i \neq \tau_{\tau^*_i}} \{w^+_i\}$ and $\tau_i$ is index $\tau$ such that $y^-_\tau = 1$ else it is 0 if $||y^-|| = 0$. Furthermore, for notational convenience, $\tau_i(i)$ dependency on data point $(i)$ will be dropped. Finally, note $K := (1 + \mathbb{1}_{\tau_i \neq 0})$ and $a \lor b$ denotes max$(a, b)$.

**Proof sketch.** This result follows from jointly maximizing the adversarial exposure $w_1$ and the default time index $w_2$. The structure of $B_n^1$ allows us to decouple this joint maximization and find the critical point to maximize the quadratic in $w_1$ and write down the condition to select the optimal default time index $\tau^*_i$. Finally, consider the two cases $w_2 = 0$ and $w_2 \neq 0$, and take the max to arrive at the solution. The $K$ terms represent the cost associated with the worst case CVA.
2.1.2 Outer Optimization Problem

The goal now is to evaluate

\[ \inf_{\gamma \geq 0} H(\gamma) := \left[ \gamma \delta_1 + \frac{1}{N} \sum_{i=1}^{N} \Psi_\gamma(x_i^+, y_i^+) \right] \]

where

\[ \Psi_\gamma(x_i^+, y_i^+) = \langle x_i^+, y_i^+ \rangle + \left[ \frac{\|x_i^+\|^2}{4\gamma} \right] + \left[ \frac{1 - \|y_i^+\|^2}{4\gamma} + (x_{i_1}^+ - x_{i_2}^+) - \gamma S_1 K \right]^+. \]

The convexity of the objective function \( H(\gamma) \) simplifies the task of solving this optimization problem. The first order optimality condition suffices. As \( \Psi_\gamma \) and hence \( H(\gamma) \) may have non-differentiable kinks at the positive real roots \( \gamma' \) for the last component (hockey stick) function of \( \Psi_\gamma \), we look for \( \gamma' \) such that \( \gamma' < 0 \). This leads to the characterization below for \( \gamma' \) via either left or right derivatives. Inspection of the left and right derivatives for \( H(\gamma) \) reveals that they will cross the origin (as \( \gamma \) sweeps from 0 to \( \infty \)) and hence the sup and inf operators will apply over non-empty sets.

**Proposition 2.** Let \( \gamma' = \sup_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \leq 0 \right\} \), equivalently

let \( \gamma' = \inf_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \geq 0 \right\} \), where

\[
\begin{align*}
K_0 &= \sum_{i=1}^{N} \|x_i^+\|^2, \\
K_1(\gamma) &= \sum_{i=1}^{N} \sum_{0 \leq y_i^+ \leq \gamma} (1 - \|y_i^+\|^2), \\
K_2(\gamma) &= \sum_{i=1}^{N} \left\{ \sum_{0 \leq y_i^+ \leq \gamma} \right\} S_1 K, \\
K_1^- (\gamma) &= \sum_{i=1}^{N} \left\{ \sum_{0 \leq y_i^+ \leq \gamma} \right\} (1 - \|y_i^+\|^2), \quad K_2^- (\gamma) = \sum_{i=1}^{N} \left\{ \sum_{0 \leq y_i^+ \leq \gamma} \right\} S_1 K,
\end{align*}
\]

\( \gamma'_i \) is implicitly defined as the positive (non-negative) root of \( \left[ \frac{1 - \|y_i^+\|^2}{4\gamma} + (x_{i_1}^+ - x_{i_2}^+) - \gamma S_1 K \right] = 0 \).

**Proof sketch.** This result follows from writing down the first order conditions for left and right derivatives for convex function \( H(\gamma) \). Each time \( \gamma \) crosses over \( \gamma'_i \) for some index \( i \) we pick up an additional term in the derivative. The \( K_1, K_2 \) terms quantify this effect. The \( \gamma'_i \) represent the positive roots (non-negative) component (hockey stick) function of \( \Psi_\gamma \).

Putting together the results of these two propositions, we arrive at our first theorem.

**Theorem 1.** The primal problem (P1) has solution

\[ \gamma' \delta_1 + \frac{1}{N} \sum_{i=1}^{N} \Psi_\gamma(x_i, y_i) \]

where \( \gamma' = \sup_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \leq 0 \right\} \) and \( \Psi_\gamma(x_i^+, y_i^+) = \langle x_i^+, y_i^+ \rangle + \left[ \frac{\|x_i^+\|^2}{4\gamma} \right] \vee \left[ \frac{1}{4\gamma} + (x_{i_1}^+ - x_{i_2}^+) - \gamma S_1 K \right]. \)

Expressed in terms of original CVA, this says

\[ \sup_{\gamma' \geq 0} \mathbb{E}^P \left[ \langle X^+, Y^C \rangle \right] = \mathbb{E}^P \left[ \max \left\{ \langle X^+, Y^C \rangle \right\} \right] + \gamma' \delta_1 + \mathbb{E}^P \left[ \left[ \frac{1}{4\gamma} + (X_{i_1}^+ - X_{i_2}^+) - \gamma S_1 K \right] \right] \]

where the additional terms represent a penalty due to uncertainty in probability distribution.

**Proof sketch.** This follows directly from the previous two propositions. \( \Box \)

2.2 Unilateral DVA

2.2.1 Inner Optimization Problem

The robust unilateral DVA can be written as

\[ \inf_{\beta \geq 0} G(\beta) := \beta \delta_2 + \frac{1}{N} \sum_{i=1}^{N} \Psi_\beta(x_i^-, y_i^+) \]

Now use recent duality results, noting the inner product \( \langle \cdot \rangle \) satisfies the upper semicontinuous condition of the Lagrangian duality theorem, and cost function \( c_\beta \) satisfies the non-negative lower semicontinuous condition (see Blachet and Murthy [2019] Assumptions 1 & 2, Gao and Kleywegt [2016]). Hence the dual problem can be written as

\[ \inf_{\beta \geq 0} G(\beta) \]
where
\[
\Psi_\beta(x_i^-, y_i^f) = \sup_{u \in \mathbb{R}^n, v \in B_n^1} [(u, v) - \beta c_S((u, v), (x_i^-, y_i^f))] = \sup_{u \in \mathbb{R}^n, v \in B_n^1} [(u, v) - \beta ((u - x_i^-, u - x^-) + S_2(v - y_i^f, v - y_i^f))].
\]

Now apply change of variables \(w_1 = (u - x_i^-)\) and \(w_2 = (v - y_i^f)\) to get
\[
\Psi_\beta(x_i^-, y_i^f) = \sup_{w_1 \leq -x_i^-, w_2 \in B_2^n} [(w_1 + x_i^-, w_2 + y_i^f) - \beta (w_1, w_1 + S_2(w_2, w_2))]
\]
where sets \(B_n^1\) and \(B_2^n\) are defined as before. Following a similar approach as for UCVA, it turns out that \(\Psi_\beta\) can be expressed as original DVA plus the pointwise max of two convex functions: a line of negative slope, and a piecewise but still convex function. \(\Psi_\beta\) quantifies the adversarial move in DVA across both time and spatial dimensions while accounting for the associated cost via the \(K\) terms.

**Proposition 3.** We have \(\Psi_\beta(x_i^-, y_i^f) = \langle x_i^-, y_i^f \rangle + [-x_i^-, -\beta S_2 K] \vee g_i(\beta)\) where
\[
g_i(\beta) = \begin{cases}
-x_i^-, & \beta \leq \frac{\|y_i^f\|}{2\beta} \\
\frac{\|y_i^f\|^2}{4\beta}, & \beta > \frac{\|y_i^f\|}{2\beta}
\end{cases}
\]
is piecewise (line of negative slope for part 1, hyperbola for part 2) but still convex, and \(\tau_2\) is index \(\tau\) such that \(y^f_{\tau_2} = 1\) else it is 0 if \(\|y_i^f\| = 0\). Finally, note \(K := (\mathbb{I}_{12} + 0)\) and recall \(a \vee b\) denotes \(\max(a, b)\).

Equivalently, \(\Psi_\beta(x_i^-, y_i^f) = \langle x_i^-, y_i^f \rangle + [-x_{\tau_2}^-, -\beta S_2 K] + \tilde{g}_i(\beta)\) where
\[
\tilde{g}_i(\beta) = \begin{cases}
[\beta S_2 K - \frac{2\beta}{\|y_i^f\|}x_i^f_{\tau_2}], & \beta \leq \frac{\|y_i^f\|}{2\beta} \\
\beta S_2 K - \frac{\|y_i^f\|^2}{4\beta} + x_{\tau_2}^-, & \beta > \frac{\|y_i^f\|}{2\beta}
\end{cases}
\]
section

**Proof sketch.** This result follows from jointly maximizing the adversarial exposure \(w_1\) and the default time index \(w_2\). The structure of \(B_{2}^n\) allows us to decouple this joint maximization and find the critical point to maximize the quadratic in \(w_1\) and write down the condition to select the optimal default time index \(\tau_2^+\). Finally, consider the two cases \(w_2 = 0\) and \(w_2 \neq 0\), and take the max to arrive at the solution. For \(w_2 = 0\), the constraint \(w_1 \leq -x_i^-\) leads to the piecewise structure for \(\tilde{g}_i(\beta)\). For \(w_2 \neq 0\), it turns out that \(\tau_2^+ = 0\) which leads to the simpler linear term in the solution. The \(K\) terms represent the cost associated with the worst case DVA.

### 2.2.2 Outer Optimization Problem

The goal now is to evaluate
\[
-\inf_{\beta \geq 0} G(\beta) := \beta \delta_2 + \frac{1}{N} \sum_{i=1}^{N} \Psi_\beta(x_i^-, y_i^f)
\]
where
\[
\Psi_\beta(x_i^-, y_i^f) = \langle x_i^-, y_i^f \rangle + [-x_{\tau_2}^-, -\beta S_2 K] + \tilde{g}_i(\beta).
\]

The convexity of the objective function \(G(\beta)\) simplifies the task of solving this optimization problem. The first order optimality condition suffices. As \(\Psi_\beta\) and hence \(G(\beta)\) may have non-differentiable kinks at the positive real roots \(\beta^*_i\) (where they exist) for the last component (hockey stick) function of \(\Psi_\beta\), we look for \(\beta^*\) such that \(0 \in \partial G(\beta^*)\). One additional complication for DVA is that the roots \(\beta^*_i\) for \(\tilde{g}_i(\beta)\) may be imaginary due to the plus sign in the +\(\beta S_2 K\) term. Nonetheless, we continue with the characterization below for \(\beta^*\) via either left or right derivatives. However, note the indicator functions reference values \(0 \leq \tilde{g}_i(\beta)\) as opposed to inequalities \(0 \leq \gamma \leq \gamma^*\). Inspection of the left and right derivatives for \(G(\beta)\) reveals that they may not always cross the origin (as \(\beta\) sweeps from 0 to \(\infty\)), and hence the sup operator may apply over an empty set. In such a degenerate case, the corner point solution is \(G(0) = 0\) where \(\beta^* = 0\).

**Proposition 4.** Let \(\beta^* = \sup_{\beta \geq 0} \{\beta : \delta_2 - \frac{1}{N} S_2 K_0 + \frac{1}{N} [K_1(\beta) + K_2(\beta)] \leq 0\}\), equivalently
let \(\beta^* = \inf_{\beta \geq 0} \{\beta : \delta_2 - \frac{1}{N} S_2 K_0 + \frac{1}{N} [K_1(\beta) + K_2(\beta)] \geq 0\}\), where \(\beta^* = \frac{\|y_i^f\|}{2\beta}, K_0 = \frac{1}{N} \sum_{i=1}^{N} \|y_i^f\|\).
\[ K_1(\beta) = \sum_{i=1}^{N} \mathbb{1}_{\{x_{1i}^2 \leq \beta_i^2\}} \mathbb{1}_{\{0 \leq \bar{e}_i(\beta)\}}[S_2 K - (x_{1i}^2)^2], \]
\[ K_2(\beta) = \sum_{i=1}^{N} \mathbb{1}_{\{x_{1i}^2 > \beta_i^2\}} \mathbb{1}_{\{0 \leq \bar{e}_i(\beta)\}}[S_2 K - (x_{1i}^2)^2], \]
\[ K_1'(\beta) = \sum_{i=1}^{N} \mathbb{1}_{\{x_{1i}^2 \leq \beta_i^2\}} \mathbb{1}_{\{0 < \bar{e}_i(\beta)\}}[S_2 K - (x_{1i}^2)^2], \]
\[ K_2'(\beta) = \sum_{i=1}^{N} \mathbb{1}_{\{x_{1i}^2 > \beta_i^2\}} \mathbb{1}_{\{0 < \bar{e}_i(\beta)\}}[S_2 K - (x_{1i}^2)^2]. \]

In the degenerate case, where \( \sup_{\beta \geq 0} \) is taken over an empty set, select \( \beta^* = 0 \) where \( G(0) = 0 \).

**Proof sketch.** This result follows from writing down the first order conditions for left and right derivatives for convex objective function \( G(\beta) \). Each time \( g_i(\beta) \) crosses over zero for some index \( i \) we pick up an additional term in the derivative. The \( K_1, K_2 \) terms quantify this effect. Recall in the degenerate case, the corner point solution is \( G(0) = 0 \) where \( \beta^* = 0 \).

Putting together the results of these two propositions, we arrive at our next theorem.

**Theorem 2.** The primal problem \(\text{(P2)}\) has solution

\[
- \sup_{Q \in \mathcal{Q}_2(\mathcal{Q}_N)} \mathbb{E}^Q[(X^-, Y_F)] = -[\beta^* \delta_2 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\beta^*}(x_i, y_i)]
\]

where \( \beta^* = \sup_{\beta \geq 0} \{ \beta : \delta_2 - \frac{1}{2} |S_2 K_0| + \frac{1}{2} [K_1(\beta) + K_2(\beta)] \leq 0 \} \), and \( \Psi_{\beta^*}(x^*_i, y^*_i) = \langle x^*_i, y^*_i \rangle + g(\beta) \vee [-x^*_i - \beta S_2 K] \).

Expressed in terms of original DVA, this says

\[
- \sup_{Q \in \mathcal{Q}_2(\mathcal{Q}_N)} \mathbb{E}^Q[(X^-, Y_F)] = -\mathbb{E}^Q[(X^-, Y_F)] - \beta^* \delta_2 - \mathbb{E}^Q[X^- - \beta^* S_2 K] \vee g(\beta^*)
\]

where

\[
g(\beta) = \begin{cases} 
- X^-_i - \beta (X^-_i)^2, & -X^-_i \leq \|Y_F\| / 2\beta \\
\|Y_F\|^2 / 4\beta^2, & -X^-_i > \|Y_F\| / 2\beta
\end{cases}
\]

and the additional terms represent a penalty due to uncertainty in probability distribution.

**Proof sketch.** This follows directly from the previous two propositions.

### 2.3 Bilateral CVA

#### 2.3.1 Inner Optimization Problem

The robust bilateral CVA can be written as

\[
\sup_{\Phi \in \mathcal{P}_{\delta}(\Phi_N)} \mathbb{E}^{\Phi}[(X^+, Y_C) + (X^-, Y_F)].
\]

(P3)

Similar to before, use recent duality results, noting that the inner product \( \langle \cdot, \cdot \rangle \) satisfies the upper semicontinuous condition of the Lagrangian duality theorem, and cost function \( c_S \) satisfies the non-negative lower semicontinuous condition (see [Blanchet and Murthy] (2019) Assumptions 1 & 2, [Gao and Kleywegt] (2016)). Hence the dual problem can be written as

\[
\inf_{\alpha \geq 0} F(\alpha) := \left[ \alpha \delta_3 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\alpha}(x_i, y^C_i, y^F_i) \right]
\]

(D3)

where

\[
\Psi_{\alpha}(x_i, y^C_i, y^F_i) = \sup_{u \in \mathbb{R}^n, v_i \in B^1, v_2 \in B^1} \left[ \langle u^+, 1_{v_1 < x_i, v_2} \rangle v_1 + \langle u^-, 1_{v_1 < x_i, v_2} \rangle v_2 - \alpha c_S((u, v_1, v_2), (x_i, y^C_i, y^F_i)) \right]
\]

\[
= \sup_{u \in \mathbb{R}^n, v_i \in B^1, v_2 \in B^1} \left[ \langle u^+, 1_{v_1 < x_i, v_2} \rangle v_1 + \langle u^-, 1_{v_1 < x_i, v_2} \rangle v_2 - \alpha (\langle u - x_i, u - x_i \rangle + S_3(v_1 - y^C_i, v_1 - y^F_i) + S_3(v_2 - y^C_i, v_2 - y^F_i)) \right]
\]
Note that default times \((v_1, v_2)\) are compared via the indicator function \(\mathbb{1}_{\{v_1 \leq v_2\}}\) by comparing indices (into the fixed dates array \(0 < t_1 < \cdots < t_n = T\)) of the respective default times. So if \(v_1\) has a one element in index \(i\) and either \(\|v_2\| = 0\) or \(v_2\) has a one element in index \(j\) and \(i < j\) then \(\mathbb{1}_{\{v_1 < v_2\}} = 1\) else if \(i > j\) or \(\|v_1\| = 0\) then \(\mathbb{1}_{\{v_1 < v_2\}} = 0\). The probability that \(i = j\) for any \(i, j \in \{1, \ldots, n\}\) is zero in continuous time, hence this case is not considered here. Also \(\|v_1\| = 1\) implies default time \(v_1 \leq t_n = T\), the maturity date of the CVA calculation. Similar analysis applies to \(v_2\).

Now apply change of variables \(w_1 = (u - x_i), w_2 = (v_1 - y_i^f),\) and \(w_3 = (v_2 - y_i^f)\) to get

\[
\Psi_\infty(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in B_y, w_3 \in B_x} \left[\left((w_1 + x_i)^+, \mathbb{1}_{\{w_2 + y_i^f < w_3 + y_i^f\}} w_2 + y_i^f\right) + ((w_1 + x_i)^-, \mathbb{1}_{\{w_2 + y_i^f < w_3 + y_i^f\}} w_2 + y_i^f)\right]

\]

\[
- \alpha((w_1, w_1) + S_3(w_2, w_2) + S_3(w_3, w_3)).
\]

Following a similar approach as for UCVA and UDVA, it turns out that \(\Psi_\infty\) can be expressed as the pointwise max of four functions of more complex forms. The four functions represent the four logical cases for \(w_2\) and \(w_3\) each being zero or non-zero. Furthermore, we need to consider the sub-cases where the counterparty defaults before the firm, as in \(\Psi_0\). Again, \(\Psi_\infty\) quantifies the adversarial moves in CVA and DVA across both time and spatial dimensions while accounting for the associated cost via the \(K\) terms.

**Remark 1.** Please note this result involves some lengthy and tedious derivations and requires some time to go through. However, there are some patterns across the various cases and sub-cases which does simplify the analysis to some extent.

### Table 1: Lookup table of optimization sub-problems

| Optimization | Objective Function | Solution |
|--------------|--------------------|----------|
| \(\sup_{w_1 \in \mathbb{R}^n}\langle w_1, y_i\rangle - \alpha \langle w_1, w_1 \rangle\) | \(\|y_i\|^2_{2\alpha}\) | \(\|x_i\|^2_{2\alpha}\) |
| \(\sup_{w_1 \leq x_i} \langle w_1, y_i\rangle - \alpha \langle w_1, w_1 \rangle\) | \([x_i, \mathbb{1}_{\{\|y_i\|_{2\alpha}\}} - \alpha [x_i, \|y_i\|_{2\alpha}]^2\] | \([\frac{1}{4\alpha} + \langle x_i, y_i \rangle]^+\) |
| \(\sup_{w_1 \in \mathbb{R}^n} ((w_1 + x_i)^+, y_i\rangle - \alpha \langle w_1, w_1 \rangle\) | \(\{[\langle x_1, y_1 + x_i, y_i \rangle]\langle x_1, y_1 \rangle - 1\} \langle \alpha (\langle x_1, y_1 \rangle)^2\) | \(\{[\frac{1}{4\alpha} + \langle x_1, y_1 \rangle + (x_1 - x_i)\} \langle \alpha (x_1)^2\) |
| \(\sup_{w_1 \in \mathbb{R}^n} ((w_1 + x_i)^-, y_i\rangle - \alpha \langle w_1, w_1 \rangle\) | \(\{[\langle x_1, y_1 + x_i, y_i \rangle]\langle x_1, y_1 \rangle - 1\} \langle \alpha (\langle x_1, y_1 \rangle)^2\) | \(\langle x_1 - x_i\rangle \langle \alpha (x_1)^2\) |

**Proposition 5.** We have \(\Psi_\infty(x_i, y_i^f, y_i^f) = \bigvee_{k=1}^{4} \Psi_k(x_i, y_i^f, y_i^f)\) where

\[
\Psi_1(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{c < y_i\} - \alpha \psi_1(x_i, y_i^f, y_i^f)} + \mathbb{1}_{\{y_i < c\} - \alpha \psi_1(x_i, y_i^f, y_i^f)},
\]

\[
\Psi_2(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{c < y_i\} - \alpha \psi_2(x_i, y_i^f, y_i^f)} + \mathbb{1}_{\{y_i < c\} - \alpha \psi_2(x_i, y_i^f, y_i^f)},
\]

\[
\Psi_3(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{c < y_i\} - \alpha \psi_3(x_i, y_i^f, y_i^f)} + \mathbb{1}_{\{y_i < c\} - \alpha \psi_3(x_i, y_i^f, y_i^f)},
\]

\[
\Psi_4(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{c < y_i\} - \alpha \psi_4(x_i, y_i^f, y_i^f)} + \mathbb{1}_{\{y_i < c\} - \alpha \psi_4(x_i, y_i^f, y_i^f)},
\]

and (suppressing arguments for brevity):

\[
\psi_1 = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle\right]^+, \quad \psi_2 = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle + \langle x_1 - x_i \rangle\right]^+ - \alpha S_3 K_{2a}, \quad \psi_3 = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle + \langle x_1 - x_i \rangle\right]^+ - \alpha S_3 K_{3a},
\]

\[
\psi_4 = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle\right]^+ - \alpha S_3 K_{4a}, \quad \psi_{2a} = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle\right]^+ - \alpha S_3 K_{2b}, \quad \psi_{3a} = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle\right]^+ - \alpha S_3 K_{3b}, \quad \psi_{4a} = \left[\frac{1}{4\alpha} + \langle x_i, y_i \rangle\right]^+ - \alpha S_3 K_{4b}.
\]
Note parameter $\tau_1^*$ and constant $K$ are defined within the proof by cases (see Supplementary Material), and are omitted here for brevity. Recall $\tau_2$ is index $\tau$ such that $y_i^{(c,f)} = 1$ else it is 0 if $\|y_i^{(c,f)}\| = 0$. The selection in $\{c,f\}$ is determined by context.

Proof sketch. This result follows from jointly maximizing the adversarial exposure $w_1$ and the default time indices $w_2, w_3$. The structure of $B^+_{n,3}$ allows us to decouple this joint maximization and find the critical point to maximize the quadratic in $w_1$ and write down the condition to select the optimal default time index $\tau_1^*$ for either the counterparty (in sub-case a) or the firm (in sub-case b), as determined by first to default. Finally, take the max over the four logical cases for $w_2$ and $w_3$ to arrive at the solution. The $K$ terms represent the cost associated with the worst case BCVA.

\[ \square \]

### 2.3.2 Outer Optimization Problem

The goal now is to evaluate

$$\inf_{\alpha \geq 0} F(\alpha) := \left[ \alpha \delta_3 + \frac{1}{N} \sum_{i=1}^{N} \Psi_\alpha(x_i, y_i^c, y_i^f) \right]$$

where the $\Psi_\alpha$ functions are given as the solutions to Proposition 2.5. Although Lagrangian duality implies the convexity of this objective function, due to its complexity, computational methods and solvers are used to evaluate this expression. Nonetheless, the solution can be expressed as below.

**Theorem 3.** The primal problem $P3$ has solution $\left[ \alpha^* \delta_3 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\alpha^*}(x_i, y_i^c, y_i^f) \right]$ where $\alpha^* = \arg\min_{\alpha \geq 0} \left[ \alpha \delta_3 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\alpha}(x_i, y_i^c, y_i^f) \right]$ and $\Psi_{\alpha^*}(x_i, y_i^c, y_i^f) = \sqrt{\sum_{k=1}^{d} \Psi_{\alpha^*}(x_i, y_i^c, y_i^f)}$.

Expressed in terms of original BCVA, this says

$$\sup_{\Phi \in \mathcal{W}_3(\Phi_N)} E^{\Phi}[\{X^+, Y^C\} + (X^-, Y^F)] = E^{\Phi_N}[\{X^+, Y^C\} + (X^-, Y^F)] + \alpha^* \delta_3 + E^{\Phi_N}[\Psi_{\alpha^*}(X, Y^C, Y^F) - (X^+, Y^C) + (X^-, Y^F)]^+$$

where the additional terms represent a penalty due to uncertainty in probability distribution.

Proof sketch. This follows directly from the previous proposition. $\square$

Empirical results for our worst case CVA studies are provided in Section 3. From the authors’ perspective the computational study was illuminating to understand the magnitude and shape of worst case CVA profiles as a function of uncertainty. Some recent work in probability theory was done to map Wasserstein radii into statistical significance levels which turns out to be quite useful. See the discussion on this topic in Section 3.1.

### 3 Computational Study: Robust CVA and Wrong Way CCR

This computational study uses the Matlab Financial Instruments Toolbox and extends WWR portfolio analysis (Brigo et al., 2013, section 5.3) to consider uncertainty in probability distribution. Other key concepts that will be leveraged in this section are the measure concentration results and association of Wasserstein radius $\delta$ with confidence level $1 - \beta$ for some $\beta \in (0, 1)$. As of July 10, 2019, 5y par interest rate swaps are 1.88% (see www.interestrateswapstoday.com). The full table is shown below.

| Swap Tenor | 1y  | 2y  | 3y  | 5y  | 7y  | 10y | 30y |
|------------|-----|-----|-----|-----|-----|-----|-----|
| Swap Rate  | 2.13% | 1.95% | 1.89% | 1.88% | 1.94% | 2.05% | 2.27% |

Furthermore, Bloomberg shows U.S. CDX investment grade and high yield 5y credit default swap spreads as below.

The computational studies in this section will investigate (and quantify) worst case unilateral CVA, DVA, and bilateral CVA for different market environments and portfolios of interest rate swaps. The current swaps curve (shown above) will be used in conjunction with monte carlo simulation of a one factor Hull-White model for interest rates. The counterparty credit curve selection will vary between investment grade and high yield (as shown above). The different portfolio setups will be described in the following sections. All calculations are done in Matlab as an extension of the example provided in the financial instruments toolbox (Matlab, 2019).
3.1 Wasserstein Radius and Significance Levels

A natural question to ask when computing worst case CVA is how to interpret the size of the Wasserstein radius $\delta$. Substantial research has been done towards answering this question. Some key results are mentioned here (Carlsson et al., 2018, Section 3). A rough procedure for selecting $\delta$ involves sampling two independent data sets $D_1$ and $D_2$, and setting $\delta = \alpha c^*$ where $\alpha \in [1/2, 1]$ and $c^*$ denotes the cost of the minimum bipartite matching between $D_1$ and $D_2$ (Carlsson et al., 2018, Canas and Rosasco, 2012). Under an additional assumption of bound $A$ for a light tailed distribution $\Phi$, Esfahani and Kuhn (2018) provide concentration inequalities that characterize the radius of the smallest Wasserstein ball $B_\delta(\Phi_N)$ (centered at the empirical probability measure $\Phi_N$) that contains $\Phi$ (the true probability measure) with significance (confidence) level $\beta \in (0, 1)$. The following result is due to Fournier and Guillin (2015) and the constants $c_1, c_2$ below can be calculated explicitly by following the proof:

\[
P[D_\varepsilon(\Phi, \Phi_N) \geq \delta] \leq \begin{cases} \exp(c_1 \varepsilon^2 N^{\max\{n, 2\}}) & \text{if } \delta \leq 1, \\ c_1 \exp(-c_2 N\delta^{\alpha}) & \text{if } \delta > 1 \end{cases}
\]

$\forall N \geq 1, n \neq 2$, and $\delta > 0$ where $c_1 > 0, c_2 > 0$ depend only on $\alpha, A$, and $n$. Esfahani and Kuhn (2018) discuss how equating the RHS above to $\beta$ and solving for $\delta$ gives

\[
\delta_N(\beta) = \begin{cases} \varepsilon \beta^{-1/\max\{n, 2\}} & \text{if } N \geq \frac{\log(c_1 \beta^{-1})}{c_2}, \\ \varepsilon \beta^{-1/\alpha} & \text{if } N < \frac{\log(c_1 \beta^{-1})}{c_2} \end{cases}
\]

however these bounds are overly conservative, and result in a radius $\delta^*$ much larger than necessary.

As an alternative approach, we follow a method that does not require additional assumptions on tail bounds and provides a more explicit mapping between $\delta$ and $\beta$ (Carlsson et al., 2018, Section 3). Theorem 6.15 of Villani (2008) gives a bound on Wasserstein distance between two pdfs $\Phi, \Phi'$ as

\[
D(\Phi, \Phi') \leq \int_{\mathcal{X}} ||x_0 - x|| \cdot |\Phi(x) - \Phi'(x)|\, dA.
\]

Theorem 1(i) of Villani (2008) relates the RHS above to the relative entropy $H(\Phi|\Phi')$ and Bolley et al. (2007) show that for any distribution $\Phi$ with empirical distribution $\Phi_N$,

\[
\limsup_{N \to \infty} \frac{1}{N} \log P[D(\Phi, \Phi_N) \geq \delta] \leq -\alpha(\delta)
\]

where

\[
\alpha(\delta) = \inf_{\Phi^* : D(\Phi, \Phi^*) \geq \delta} H(\Phi, \Phi^*).
\]

Carlsson et al. (2018) show

\[
H(\Phi_1|\Phi_2) \geq \frac{8r - 2\sqrt{16r^2 + 16r\delta} + 24r + 12\delta + 9 + 4\delta + 6}{3 + 4r}
\]

where

\[
r = \max_{x_0 \in \mathcal{X}, x \in \mathcal{X}} ||x - x_0||
\]

denotes the radius of domain $\mathcal{X}$ whereby

\[
\log \int_{\mathcal{X}} \exp^{2 ||x - x_0||} \Phi_N(x)\, dA \leq \log \exp^{2r} = 2r.
\]

Finally, Carlsson et al. (2018) get the result

\[
P[D(\Phi, \Phi_N) \geq \delta] \leq \exp\left(-N \frac{8r - 2\sqrt{16r^2 + 16r\delta} + 24r + 12\delta + 9 + 4\delta + 6}{3 + 4r}\right)
\]

Table 3: CDS Spreads

| CDX Index | IG | HY |
|-----------|----|----|
| CDS Spread | 53 | 323 |
which is the approach used in this study. Therefore, for a desired significance (confidence) level \( \beta \in (0, 1) \), find \( \delta_\beta \) such that

\[
1 - \beta = \exp \left( -N \frac{8r - 2\sqrt{16r^2 + 16r\delta_\beta + 24r + 12\delta_\beta + 9 + 4\delta_\beta + 6}}{3 + 4r} \right).
\]

However, in our problem setting, the domain \( \mathcal{R} \) of possible realizations of discounted exposures and default time pairings \((X, \tau_C, \tau_F)\) is difficult to bound. The approach taken here is to use the empirical domain \( \mathcal{R}_N \) determined by the empirical measure \( \Phi_N \) as a proxy for domain \( \mathcal{R} \). For sample size \( N \) large enough, \( \mathcal{R}_N \sim \mathcal{R} \) and it follows that \( \limsup_{N \to \infty} \mathcal{R}_N \to \mathcal{R} \). Therefore we write

\[
r_N = \max_{x_0 \in \mathcal{R}_N, x \in \mathcal{R}_N} \|x - x_0\|
\]

and use the approximate relation

\[
1 - \beta \sim \exp \left( -N \frac{8r_N - 2\sqrt{16r_N^2 + 16r_N\delta_\beta + 24r_N + 12\delta_\beta + 9 + 4\delta_\beta + 6}}{3 + 4r_N} \right).
\]

### 3.2 Unilateral CVA

#### 3.2.1 Portfolio of Interest Rate Swaps, Investment Grade Counterparty

The portfolio here consists of a dozen interest rate swaps, with a mix of receiving fixed and paying fixed swaps, at different coupons, maturities, and notionals. The fixed coupons range between 2% and 2.5%, the maturities range between 4y and 12y, the notional range between 400k USD and 1mm USD. The investment grade counterparty credit spread is set to 50 basis points. The table of confidence levels \( \beta \) and their corresponding Wasserstein radii \( \delta_\beta \) follows.

| Table 4: UCVA Investment Grade Wasserstein Radii |
|-----------------------------------------------|
| Confidence Level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 | 0.999 |
| W Radius delta   | 1.0  | 1.1  | 1.2  | 1.4  | 1.7  | 2.1   |

The scale factor \( S_1 \) is set (by default) to 1 and the portfolio exposures are scaled to be in units of hundreds of thousands of dollars. Again, the intent of scaling is to provide appropriate penalty to the adversarial change in joint distribution of portfolio exposures and default times that promotes worst case CVA and wrong way risk. Further work may conduct a sensitivity analysis regarding the pairings of \( S_1 \) and units of portfolio exposures to investigate suitable (unsuitable) ranges that preserve (distort) the shape of the robust CVA profile. Matlab plots characterizing the CVA exposure profile and trajectory of worst case CVA as a function of Wasserstein radius are shown. Again, we think about worst case CVA (which incorporates probability of counterparty default) as compared to the PFE (Potential Future Exposure) which shows tail percentiles of portfolio exposure (not multiplied by probability of counterparty default).

The baseline CVA for this portfolio is small (about 1.6k USD) and represents the dot product of the discounted positive portfolio exposure profile times counterparty default probability. The worst case CVA curve is shown below. Note the worst case CVA is approximately 68% the size of Max PFE for Wasserstein radius \( \delta \) about 1.4 which maps to a significance level around 95%. So the takeaway here is worst case CVA is still a significant percentage of PFE for swap portfolios with low counterparty default curves (investment grade).

#### 3.2.2 Portfolio of Interest Rate Swaps, High Yield Counterparty

The portfolio here consists of a dozen interest rate swaps, with a mix of receiving fixed and paying fixed swaps, at different coupons, maturities, and notional. The fixed coupons range between 2% and 2.5%, the maturities range between 4y and 12y, the notional range between 400k USD and 1mm USD. The high yield counterparty credit spread is set to 320 basis points. The table of confidence levels \( \beta \) and their corresponding Wasserstein radii \( \delta_\beta \) follows. The scale factor \( S_1 \) is set (by default) to 1 and the portfolio exposures are scaled to be in units of hundreds of thousands of dollars.
Figure 1: Swaps Portfolio Positive Exposure Profiles

Figure 2: Swaps Portfolio Worst Case UCVA Profile

Figure 3: Swaps Portfolio Positive Exposure Profiles
Table 5: UCVA High Yield Wasserstein Radii

| Confidence Level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 | 0.999 |
|------------------|------|------|------|------|------|-------|
| W Radius delta   | 1.0  | 1.1  | 1.2  | 1.4  | 1.7  | 2.1   |

Figure 4: Swaps Portfolio Worst Case UCVA Profile

The baseline CVA for this portfolio is around 8.5k USD and represents the dot product of the discounted positive portfolio exposure profile times counterparty default probability. The worst case CVA curve is shown. Note the worst case CVA is still approximately 64% the size of Max PFE for Wasserstein radius $\delta$ about 1.4 which maps to a significance level around 95%. So the takeaway here is worst case CVA is a significant percentage of PFE for swap portfolios with moderately high counterparty default curves (high yield). For this particular experiment, not much deviation in worst case CVA was observed, for investment grade vs. high yield counterparty credit. Keep in mind, the correlations between default times and portfolio exposures are what primarily drive wrong way risk, more so than differences in the marginal distributions (e.g., investment grade vs. high yield counterparty credit). Although, as a second order effect, the high yield default curves incur a somewhat higher penalty (under worst case CVA scenarios) in moving the occurrence of default times to maximize wrong way risk than investment grade credits (for which less default events occur). This contributes to a slightly lower worst case CVA when compared to the investment grade counterparty credit.

3.3 Unilateral DVA

3.3.1 Portfolio of Interest Rate Swaps, Investment Grade Firm

The portfolio here consists of a dozen interest rate swaps, with a mix of receiving fixed and paying fixed swaps, at different coupons, maturities, and notional amounts. The fixed coupons range between 2% and 2.5%, the maturities range between 4y and 12y, the notional amounts range between 4mm USD and 10mm USD. The investment grade firm credit spread is set to 50 basis points. The table of confidence levels $\beta$ and their corresponding Wasserstein radii $\delta$ follows.

| Confidence Level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 | 0.999 |
|------------------|------|------|------|------|------|-------|
| W Radius delta   | 0.9  | 1.0  | 1.1  | 1.2  | 1.5  | 1.9   |

The scale factor $S_2$ is set (by default) to 1 and the portfolio exposures are scaled to be in units of millions of dollars. Same comments as above, regarding scaling, apply. Matlab plots characterizing the DVA exposure profile and trajectory of worst case DVA as a function of Wasserstein radius are shown.
The baseline DVA for this portfolio is -9.5k USD and represents the dot product of the discounted negative portfolio exposure profile times firm default probability. The worst case DVA plot is shown. Again, the plot illustrates that worst case DVA quickly attains its lower bound (in magnitude) of zero (no liability benefit to the firm for DVA).

### 3.3.2 Portfolio of Interest Rate Swaps, High Yield Firm

The reference portfolio here is the same one used in the previous subsection, albeit with notional from 4mm to 10mm USD. The high yield firm credit spread is set to 320 basis points. The table of confidence levels $\beta$ and their corresponding Wasserstein radii $\delta$ is shown. The scale factor $S_2$ is set (by default) to 1 and the portfolio exposures are scaled to be in units of millions of dollars.

#### Table 7: UDVA High Yield Wasserstein Radii

| Confidence Level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 | 0.999 |
|------------------|------|------|------|------|------|-------|
| W Radius delta   | 0.9  | 1.0  | 1.1  | 1.2  | 1.5  | 1.9   |

The baseline DVA for this portfolio is -45k USD and represents the dot product of the discounted negative portfolio exposure profile times firm default probability. The worst case DVA plot is shown. Once again, the plot illustrates that worst case DVA quickly attains its lower bound (in magnitude) of zero (no liability benefit to the firm for DVA).
3.4 Bilateral CVA

3.4.1 Portfolio of Interest Rate Swaps, Investment Grade Counterparty and Firm

The corresponding UCVA portfolio is used for comparison. The portfolio consists of a dozen interest rate swaps, with a mix of receiving fixed and paying fixed swaps, at different coupons, maturities, and notional sizes. The fixed coupons range between 2% and 2.5%, the maturities range between 4y and 12y, the notional sizes range between 400k USD and 1mm USD. The investment grade counterparty credit spread is set to 50 basis points. The firm credit spread is set to 25 basis points. The table of confidence levels $\beta$ and their corresponding Wasserstein radii $\delta$ follows.

| Confidence Level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 | 0.999 |
|------------------|------|------|------|------|------|-------|
| W Radius delta   | 1.6  | 1.8  | 2.0  | 2.2  | 2.8  | 3.5   |

The scale factor $S_3$ is set (by default) to 1 and the portfolio exposures are scaled to be in units of hundreds of thousands of dollars. Same comments as above, for UCVA and UDVA, regarding scaling, apply. Matlab plots characterizing the BCVA positive and negative exposure profiles and trajectory of worst case BCVA as a function of Wasserstein radius are shown.

The baseline BCVA for this portfolio is small (around 1k USD) and represents the dot product of the discounted positive portfolio exposure profile times counterparty default probability plus dot product of the discounted negative portfolio exposure times firm default probability. The worst case BCVA curve is shown below. Note the worst case CVA is approximately 82% the size of Max PFE (Potential Future Exposure) for Wasserstein radius $\delta$ about 2.2 which maps to a significance level around
95%. So the takeaway here is worst case BCVA is still a significant percentage of PFE for swap portfolios with low counterparty default curves (investment grade). It is also interesting to compare worst case BCVA vs. worst case UCVA for a given delta. For example, for $\delta$ of 1.4, which represents the 95% significance level for UCVA, we see BCVA of around 1.4 which is below UCVA of 1.6. This agrees with intuition that BCVA should be less than UCVA due to liability benefit from DVA.
3.4.2 Portfolio of Interest Rate Swaps, High Yield Counterparty

The corresponding UCV A portfolio is used for comparison. The portfolio consists of a dozen interest rate swaps, with a mix of receiving fixed and paying fixed swaps, at different coupons, maturities, and notional. The fixed coupons range between 2% and 2.5%, the maturities range between 4y and 12y, the notional range between 400k USD and 1mm USD. The high yield counterparty credit spread is set to 320 basis points. The firm credit spread is set to 25 basis points. The table of confidence levels $\beta$ and their corresponding Wasserstein radii $\delta$ follows.

| Confidence Level | 0.80 | 0.85 | 0.90 | 0.95 | 0.99 | 0.999 |
|------------------|------|------|------|------|------|-------|
| W Radius delta   | 1.7  | 1.8  | 2.0  | 2.3  | 2.9  | 3.5   |

The scale factor $S_3$ is set (by default) to 1 and the portfolio exposures are scaled to be in units of hundreds of thousands of dollars. Same comments as above, for UCVA and UDVA, regarding scaling, apply. Matlab plots characterizing the BCVA positive and negative exposure profiles and trajectory of worst case BCVA as a function of Wasserstein radius are shown.

The baseline BCVA for this portfolio is around 7.8k USD and represents the dot product of the discounted positive portfolio exposure profile times counterparty default probability plus dot product of the discounted negative portfolio exposure times firm default probability. The worst case BCVA curve is shown below. Note the worst case CVA is approximately 80% the size of Max PFE (Potential Future Exposure) for Wasserstein radius $\delta$ about 2.3 which maps to a significance level around 95%. So the takeaway here is worst case BCVA is still a significant percentage of PFE for swap portfolios with high yield counterparty
default curves. It is also interesting to compare worst case BCV A vs. worst case UCVA for a given $\delta$. For example, for $\delta$ of 1.4, which represents the 95% significance level for UCVA, we see BCV A of around 1.35 which is below UCVA of 1.45. This agrees with intuition that BCV A should be less than UCVA due to liability benefit from DVA.

4 Conclusions and Further Work

This work has developed theoretical results and investigated calculations of robust CVA and wrong way risk for OTC derivatives under distributional uncertainty using Wasserstein distance as an ambiguity measure. The financial market overview and foundational notation and wrong way risk (robust CVA) primal problem definitions were introduced in Section 1. Using recent duality results [Blanchet and Murthy, 2019], the simpler dual formulation and its analytic solutions for UCVA, UDVA, and BCVA were derived in Section 2. After that, in Section 3, some computational experiments were conducted to measure the additional CVA charge (and/or DVA impairment) due to distributional uncertainty for a variety of portfolio and market configurations for UCVA, UDVA, and BCVA. Using some probability results on bounding Wasserstein distance between distributions [Carlsson et al., 2018], a mapping between Wasserstein radii $\delta$ and significance levels $\beta$ was devised to study the trajectories of wrong way risk as a function of radius $\delta$. UCVA increased to a significant percentage of PFE. UDVA quickly reached its lower bound of zero liability benefit. BCVA was below UCVA (as expected) but still showed an upward (apparently concave) trajectory as radius $\delta$ increased. Finally, we conclude with some commentary on directions for further research.

One direction for future research, as has been previously discussed, is a thorough study (including sensitivity analysis) regarding the pairings of scale factors $(S_1, S_2, S_3)$ and units of portfolio exposures to investigate suitable (unsuitable) ranges that preserve (distort) the shape of the robust UCVA, UDVA, BCVA profiles (as a function of Wasserstein radii, and hence distributional uncertainty) respectively. As a reminder, the intent of scaling is to provide appropriate penalty to the adversarial change in joint distribution of portfolio exposures and default times that promotes worst case UCVA, UDVA, BCVA and wrong way risk. Another direction for future research would be to develop (and apply) similar theoretical machinery as used for robust CVA and wrong way risk in this work towards robust FVA (Funding Valuation Adjustment) and wrong way risk in that context. Intuitively, wrong way risk arises in that context when the market cost of funding the derivatives position increases at the same time as the funding exposure increases.
References

Bartl, D., Drapeau, S., and Tangpi, L. (2017). Computational aspects of robust optimized certainty equivalents and option pricing.

Basel Committee on Banking Supervision (2015). Review of the credit valuation adjustment risk framework.

Blanchet, J., Chen, L., and Zhou, X. Y. (2018). Distributionally robust mean-variance portfolio selection with wasserstein distances.

Blanchet, J., Kang, Y., and Murthy, K. (2016). Robust wasserstein profile inference and applications to machine learning. arXiv preprint arXiv:1610.05627.

Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. Mathematics of Operations Research, 44(2):565–600.

Bolley, F., Guillin, A., and Villani, C. (2007). Quantitative concentration inequalities for empirical measures on non-compact spaces. Probability Theory and Related Fields, 137(3-4):541–593.

Brigo, D., Morini, M., and Pallavicini, A. (2013). Counterparty credit risk, collateral and funding: with pricing cases for all asset classes, volume 478. John Wiley & Sons.

Canas, G. and Rosasco, L. (2012). Learning probability measures with respect to optimal transport metrics. In Advances in Neural Information Processing Systems, pages 2492–2500.

Carlsson, J. G., Behroozi, M., and Mihic, K. (2018). Wasserstein distance and the distributionally robust tsp. Operations Research, 66(6):1603–1624.

El Hajjaji, O. and Subbotin, A. (2015). Cva with wrong way risk: Sensitivities, volatility and hedging. International Journal of Theoretical and Applied Finance, 18(03):1550017.

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming, 171(1-2):115–166.

Fournier, N. and Guillin, A. (2015). On the rate of convergence in wasserstein distance of the empirical measure. Probability Theory and Related Fields, 162(3-4):707–738.

Gao, R., Chen, X., and Kleywegt, A. J. (2017). Wasserstein distributional robustness and regularization in statistical learning. arXiv preprint arXiv:1712.06050.

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with wasserstein distance. arXiv preprint arXiv:1604.02199.

Glasserman, P. and Yang, L. (2015). Bounding wrong-way risk in measuring counterparty risk. Office of Financial Research Working Paper, (15-16):15–76.

Green, A. (2015). XVA: Credit, Funding and Capital Valuation Adjustments. John Wiley & Sons.

Lichters, R., Stamm, R., and Gallagher, D. (2015). Modern derivatives pricing and credit exposure analysis: theory and practice of CSA and XVA pricing, exposure simulation and backtesting. Springer.

Matlab (2019). Matlab, counterparty credit risk and cva. https://www.mathworks.com/help/fininst/counterparty-credit-risk-and-cva.htm Accessed: 2019-07-30.

Memartoluie, A. (2017). Computational methods in finance related to distributions with known marginals.

Office of the Comptroller of the Currency (2011). Interagency supervisory guidance on counterparty credit risk management. Bulletin 2011-30a.

Ramzi Ben-Abdallah, M. B. and Marzouk, O. (2019). Wrong-way risk of interest rate instruments. Journal of Credit Risk, pages 21–44.

Villani, C. (2008). Optimal transport: old and new, volume 338. Springer Science & Business Media.
A Supplement for Theory: Robust CVA and Wrong Way CCR (Section 2)

Proposition 1 \( \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \left[ \frac{\| y_1^+ \|^2}{4\gamma} \right] \vee \left[ \frac{1}{2\gamma} + (x_1^+ - x_{i_1}) - \gamma S_1 K \right] \) or equivalently

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \left[ \frac{\| y_1^+ \|^2}{4\gamma} \right] + \left[ \frac{1}{2\gamma} + (x_1^+ - x_{i_1}) - \gamma S_1 K \right] \]

where \( \tau_1^* = \arg \max_{\tau_1 \in \{1, \ldots, n\}, \tau_1 \neq \tau_2} \\langle x_{\tau_1}^+, y_1^+ \rangle \) and \( \tau_2 \) is index \( \tau \) such that \( y_1^+ \tau_2 = 1 \) else it is 0 if \( \| y_1^+ \| = 0 \). Furthermore, for notational convenience, \( \tau_j(i) \) dependency on data point \( i \) will be dropped. Finally, note \( K := (1 + \mathbb{1}_{\{\tau_2 \neq 0\}}) \) and \( a \vee b \) denotes \( \max(a, b) \).

Proof. The particular structure of \( B_1^a \) and \( B_2^a \) will be exploited to evaluate the sup above. The analysis proceeds by considering different cases for optimal values \( (w_1^a, w_2^a) \).

Case 1. Suppose \( w_2^a = 0 \).

Then \( \tau_1^* = \tau_2 \) and

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \sup_{w_1 \in \mathbb{R}_+^n} \left[ \langle w_1, y_1^+ \rangle - \gamma \langle w_1, w_1 \rangle \right]. \]

Applying the Cauchy-Schwarz Inequality gives

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \sup_{\| w_1 \|} \| y_1^+ \| - \gamma \| w_1 \|^2. \]

Evaluating the critical point \( \| w_1^a \| = \frac{\| y_1^+ \|}{2\gamma} \in \mathbb{R}_+^n \) for the quadratic gives

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \frac{\| y_1^+ \|^2}{4\gamma}. \]

Case 2. Now consider \( w_2^a \neq 0 \).

Then \( w_2^a \) has +1 in position \( \tau_1^*(i) \) and -1 in position \( \tau_2(i) \), where \( \tau_j(i) \) is 0 means the value ±1 does not occur. Furthermore, \( \tau_1^* \neq \tau_2 \) otherwise \( w_2^a = 0 \). Observe

\[ \langle w_1, w_2 + y_1^+ \rangle = w_{1\tau_1}. \]

The structure of finite set \( B_2^a \) implies

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \sup_{w_1 \in \mathbb{R}_+^n, \tau_1 \in \{1, \ldots, n\}, \tau_1 \neq \tau_2} \left[ w_{1\tau_1} + (x_{\tau_1}^+ - x_{\tau_2}^+) - \gamma \langle w_1, w_1 \rangle + S_1 K \right]. \]

Note that the specification \( \tau_1 = 0, \tau_2 \in \{1, \ldots, n\} \) is never optimal since

\[ z = \left[ -x_{\tau_2}^+ - \gamma \langle (w_1, w_1) + S_1 (\mathbb{1}_{\{\tau_2 \neq 0\}}) \rangle \right] \leq 0, \]

which implies \( \Psi_\gamma \) would attain a higher value for the case \( w_2^a = 0 \).

Again, using that \( B_2^a \) is a finite set, one can write

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \max_{\tau_1 \in \{1, \ldots, n\}, \tau_1 \neq \tau_2} \sup_{w_1 \in \mathbb{R}_+^n} \left[ w_{1\tau_1} + (x_{\tau_1}^+ - x_{\tau_2}^+) - \gamma \langle w_1, w_1 \rangle + S_1 K \right]. \]

Observing that the only positive component of \( w_1 \in \mathbb{R}_+^n \) inside the sup is \( \tau_1 \) gives

\[ \sup_{w_1 \in \mathbb{R}_+^n} \left[ w_{1\tau_1} - \gamma \langle w_1, w_1 \rangle \right] = \sup_{w_{1\tau_1} \in \mathbb{R}_+^n} \left[ w_{1\tau_1} - \gamma (w_{1\tau_1}^2) \right]. \]

Evaluating at the critical point \( w_{1\tau_1} = \frac{1}{2\gamma} \in \mathbb{R}_+ \) for the above quadratic gives

\[ \sup_{w_{1\tau_1} \in \mathbb{R}_+^n} \left[ w_{1\tau_1} - \gamma (w_{1\tau_1}^2) \right] = \frac{1}{4\gamma}. \]

Therefore one can write

\[ \Psi_\gamma(x_1^+, y_1^+) = \langle x_1^+, y_1^+ \rangle + \max_{\tau_1 \in \{1, \ldots, n\}, \tau_1 \neq \tau_2} \left[ \frac{1}{4\gamma} + (x_{\tau_1}^+ - x_{\tau_2}^+) - \gamma S_1 K \right]. \]
Furthermore, \( \tau_1^* \) is determined as
\[
\tau_1^* = \arg \max_{\tau_1 \in \{1, \ldots, n\}, \tau_1 \neq 0} |x_{r_1}^+|.
\]

Substituting back into expression for \( \Psi_\gamma \) gives
\[
\Psi_\gamma(x_i^+, y_i^+) = (x_i^+, y_i^+) + \left[ \frac{1}{4\gamma} + (x_i^+ - x_i^+ - \gamma S_1 K) \right].
\]

Finally, taking the max values for \( \Psi_\gamma \) over cases \( w_1^2 = 0; w_2^2 \neq 0 \) gives
\[
\Psi_\gamma(x_i^+, y_i^+) = (x_i^+, y_i^+) + \left[ \frac{1}{4\gamma} + (x_i^+ - x_i^+ - \gamma S_1 K) \right].
\]

\[\square\]

**Proposition 2** Let \( \gamma^* = \sup_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \leq 0 \right\} \), equivalently

let \( \gamma^* = \inf_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \geq 0 \right\} \), where

\[
K_0 = \sum_{i=1}^{N} \|y_i^\|^2, K_1(\gamma) = \sum_{i=1}^{N} \frac{1}{\gamma} (1 - \|y_i^\|^2), K_2(\gamma) = \sum_{i=1}^{N} \frac{1}{\gamma} S_i K, \quad K_1(\gamma) = \sum_{i=1}^{N} \frac{1}{\gamma} S_i K, \quad \text{and}
\]

\( \gamma^* \) is implicitly defined as the positive (non-negative) root of \( \left[ \frac{1}{4\gamma^2} + (x_i^+ - x_i^+) - \gamma S_1 K \right] = 0 \).

**Proof.** First note the convexity of \( H(\gamma) = \gamma \delta_1 + \frac{1}{N} \sum_{i=1}^{N} \Psi_\gamma \) as a function of \( \gamma \) hence the first order optimality condition will suffice to determine \( \gamma^* \). Each \( \Psi_\gamma \) is convex in \( \gamma \) since it is the pointwise max of two convex functions. The first of these is a hyperbola and the second of these is a sum of a hyperbola plus a negative linear term (both convex). Finally, \( H \) is the sum of a linear term and a sum of convex functions \( \Psi_\gamma \). \( H \) may not differentiable at \( \gamma \) hence we look for \( \gamma^* \) such that \( 0 \in \partial H(\gamma^*) \).

Proceed to rewrite \( \Psi_\gamma \) as
\[
\Psi_\gamma(x_i^+, y_i^+) = (x_i^+, y_i^+) + \left[ \frac{1}{4\gamma} + \frac{1 - \|y_i^\|^2}{4\gamma^2} + \frac{x_i^+ - x_i^+ - \gamma S_1 K}{4\gamma^2} \right].
\]

Define \( J_1(\gamma) := \{ i \in \{1, \ldots, N\} : \frac{1 - \|y_i^\|^2}{4\gamma^2} + (x_i^+ - x_i^+) - \gamma S_1 K \geq 0 \}, \quad J_1(\gamma) := \{ i \in \{1, \ldots, N\} : \frac{1 - \|y_i^\|^2}{4\gamma^2} + (x_i^+ - x_i^+) - \gamma S_1 K \geq 0 \} \).

Now the first order condition says
\[
\delta - \frac{1}{N} \sum_{i=1}^{N} \frac{\|y_i^\|^2}{4\gamma^2} + \frac{1}{\gamma} \sum_{i \in J_1} \left[ \frac{1 - \|y_i^\|^2}{4\gamma^2} + S_i K \right] \leq 0 \leq \delta - \frac{1}{N} \sum_{i=1}^{N} \frac{\|y_i^\|^2}{4\gamma^2} + \frac{1}{\gamma} \sum_{i \in J_1} \left[ \frac{1 - \|y_i^\|^2}{4\gamma^2} + S_i K \right].
\]

The LHS can be expressed as
\[
\delta - \frac{1}{N} \sum_{i=1}^{N} \frac{\|y_i^\|^2}{4\gamma^2} + \frac{1}{\gamma} \sum_{i \in J_1} \left[ \frac{1 - \|y_i^\|^2}{4\gamma^2} + S_i K \right] \leq 0 \leq \delta - \frac{1}{N} \sum_{i=1}^{N} \frac{\|y_i^\|^2}{4\gamma^2} + \frac{1}{\gamma} \sum_{i \in J_1} \left[ \frac{1 - \|y_i^\|^2}{4\gamma^2} + S_i K \right]
\]

where \( \gamma^* \) is implicitly defined via the set \( J_1 \) as
\[
\left[ \frac{1 - \|y_i^\|^2}{4\gamma^*} + (x_i^+ - x_i^+) - \gamma S_1 K \right] = 0.
\]

For the above quadratic, it is the case that \( \gamma^* \) is real valued. On LHS, Substitution gives
\[
\delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \leq 0.
\]

Now the LHS is an increasing function in \( \gamma \). Hence one can write
\[
\gamma^* = \sup_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \leq 0 \right\}.
\]
On RHS, substitution gives
\[ \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \geq 0. \]

As RHS is also an increasing function in \( \gamma \), equivalently, one can write
\[ \gamma^* = \inf_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \geq 0 \right\}. \]

\( \square \)

**Theorem** 1. The primal problem \( P \) has solution \( \gamma^* \delta_1 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\tau}(x_i, y_i) \) where \( \gamma^* = \sup_{\gamma \geq 0} \left\{ \gamma : \delta_1 - \frac{1}{N} \left[ \frac{(K_0 + K_1(\gamma))}{4\gamma^2} + K_2(\gamma) \right] \leq 0 \right\} \) and \( \Psi_{\tau}(x_i^+, y_i^+) = (x_i^+, y_i^+) + \frac{\|y_i^\|}{2\gamma} \vee \left[ \frac{1}{4\gamma^2} + (x_i^+ - x_i^-) - \gamma S_i \right]. \)

Expressed in terms of original CVA, this says
\[ \sup_{P \in \mathcal{G}_1(P_N)} \mathbb{E}^P[(X^+, Y^C)] = \mathbb{E}^P[(X^+, Y^C)] + \gamma^* \delta_1 + \mathbb{E}^P\left[ \left[ \frac{1}{4\gamma^2} + (X_i^+ - X_i^-) - \gamma^* S_i \right] \right] \]

where the additional terms represent a penalty due to uncertainty in probability distribution.

**Proof.** This follows by direct substitution of \( \gamma^* \) as characterized in Proposition 2.2 into Proposition 2.1 and then the dual problem \( D_1 \).

**Proposition 3** We have \( \Psi_{\beta}(x_i^+, y_i^+) = (x_i^+, y_i^+) + \left[ -x_i^- \beta S_i K \right] \vee \bar{g}_i(\beta) \) where
\[ g_i(\beta) = \begin{cases} -x_i^- \beta (x_i^-)^2, & -x_i^- \leq \frac{y_i^\|}{2\beta} \\
\frac{y_i^\|}{2\beta}, & -x_i^- > \frac{y_i^\|}{2\beta} \end{cases} \]
is piecewise (line of negative slope for part 1, hyperbola for part 2) but still convex, and \( \tau_2 \) is index \( \tau \) such that \( y_{i_{\tau_2}} = 1 \) else it is 0 if \( \|y_i^\| = 0 \). Finally, note \( K := (1_{\tau_{\tau_2} = 0}) \) and recall \( a \vee b \) denotes \( \max(a, b) \).

Equivalently, \( \Psi_{\beta}(x_i^-, y_i^+) = (x_i^-, y_i^+) + \left[ -x_i^- \beta S_i K \right] \vee \bar{g}_i^+(\beta) \) where
\[ \bar{g}_i(\beta) = \begin{cases} \beta S_i K - \beta (x_i^-)^2, & -x_i^- \leq \frac{y_i^\|}{2\beta} \\
\frac{y_i^\|}{2\beta} + x_i^- \beta S_i K, & -x_i^- > \frac{y_i^\|}{2\beta}. \end{cases} \]

**Proof.** The particular structure of \( B_i^1 \) and \( B_i^2 \) will be exploited to evaluate the sup above. The analysis proceeds by considering different cases for optimal values \( (w_i^1, w_i^2) \).

**Case 1** Suppose \( w_i^2 = 0 \). Then \( \tau_{i_{\tau_2}} = \tau_2 \) and
\[ \Psi_{\beta}(x_i^-, y_i^+) = (x_i^-, y_i^+) + \sup_{w_1 \leq -x_i^-} [(w_1, y_i^+) - \beta (w_1, w_1)]. \]

First look at the unconstrained problem,
\[ \Psi_{\beta}(x_i^-, y_i^+) = (x_i^-, y_i^+) + \sup_{w_1} [(w_1, y_i^+) - \beta (w_1, w_1)]. \]

Applying the Cauchy-Schwarz Inequality gives
\[ \Psi_{\beta}(x_i^-, y_i^+) = (x_i^-, y_i^+) + \sup_{w_1} [(w_1, y_i^+) - \beta (w_1, w_1)]. \]

Evaluating the critical point \( \|w_i^1\| = \frac{\|y_i^\|}{2\beta} \in \mathbb{R}_+^+ \), for the quadratic gives
\[ \Psi_{\beta}(x_i^-, y_i^+) = (x_i^-, y_i^*) + \frac{\|y_i^\|^2}{4\beta}. \]
Now let us return to the constrained problem, $\Psi_\beta$. Recall that $\tau_2$ is index $\tau$ such that $y^{f}_{\tau} = 1$ else it is 0 if $\|y^f\| = 0$. Observe that the only positive component of $w_1$ inside the sup is $\tau_2$. Proceed to write

$$
\Psi_\beta(x^*_i, y^f_i) = \langle x^*_i, y^f_i \rangle + \sup_{w_{1 \tau_2} \leq -x_{1 \tau_2}} (\langle w_1, y^f_i \rangle - \beta \langle w_1, w_1 \rangle).
$$

Deduce that

$$
w_1^{\tau_2} = \left[-x_{1 \tau_2} \wedge \frac{\|y^f_i\|}{2\beta}\right].
$$

Therefore

$$
\Psi_\beta(x^*_i, y^f_i) = \langle x^*_i, y^f_i \rangle + \left[-x_{1 \tau_2} \wedge \frac{\|y^f_i\|}{2\beta}\right] - \beta \left[-x_{1 \tau_2} \wedge \frac{\|y^f_i\|}{2\beta}\right]^2.
$$

Next, let us do some simplification for

$$
g_i(\beta) = \left[-x_{1 \tau_2} \wedge \frac{\|y^f_i\|}{2\beta}\right] - \beta \left[-x_{1 \tau_2} \wedge \frac{\|y^f_i\|}{2\beta}\right]^2.
$$

Considering the two cases, it follows that:

$$
g_i(\beta) = \begin{cases} 
-x_{1 \tau_2} - \beta (x_{1 \tau_2})^2, & -x_{1 \tau_2} \leq \frac{\|y^f_i\|}{2\beta} \\
\frac{\|y^f_i\|^2}{2\beta}, & -x_{1 \tau_2} > \frac{\|y^f_i\|}{2\beta}.
\end{cases}
$$

Note that $g_i(\beta)$ is a convex function!

In the degenerate case, $-x_{1 \tau_2} = 0$, then $g_i(\beta) = 0 \forall \beta \geq 0$. In the degenerate case, $\|y^f_i\| = 0$, then $g_i(\beta) = 0 \forall \beta \geq 0$. Otherwise, $g_i(\beta)$ is piecewise (line of negative slope for part 1, hyperbola for part 2) but still convex.

$$
g_i'(\beta) = \begin{cases} 
-(x_{1 \tau_2})^2, & -x_{1 \tau_2} \leq \frac{\|y^f_i\|}{2\beta} \\
\frac{\|y^f_i\|^2}{2\beta}, & -x_{1 \tau_2} > \frac{\|y^f_i\|}{2\beta}.
\end{cases}
$$

Remarkably, these slopes are equal when $-x_{1 \tau_2} = \frac{\|y^f_i\|}{2\beta}$ hence the convexity of $g_i(\beta)$ holds. Proceed to rewrite $\Psi_\beta$ as

$$
\Psi_\beta(x^*_i, y^f_i) = \langle x^*_i, y^f_i \rangle + g_i(\beta).
$$

**Case 2** Now consider $w^* \neq 0$.

Then $w^*_i$ has +1 in position $\tau_1^*$ and -1 in position $\tau_2$, where $\tau_i = 0$ means the value ±1 does not occur.

Furthermore, $\tau^*_1 \neq \tau_2$ otherwise $w^*_2 = 0$. Observe

$$
\langle w_1, w_2 + y^f_i \rangle = w_1 \tau_1.
$$

The structure of finite set $B^2_\beta$ implies

$$
\Psi_\beta(x^*_i, y^f_i) = \langle x^*_i, y^f_i \rangle + \sup_{w_1 \leq -x_{1 \tau_1}} \left[w_1 \tau_1 + (x_{1 \tau_1} - x_{1 \tau_2}) - \beta (\langle w_1, w_1 \rangle + S_2 K)\right].
$$

Note that the specification $\tau_1 \neq 0$ is optimal if

$$
z = \sup_{w_1 \leq -x_{1 \tau_1}} \left[w_1 \tau_1 + x_{1 \tau_1} - \beta (w_{1 \tau_1}^2 + S_2 I_{\{\tau_1 \neq 0\}})\right] > 0.
$$

Since this never occurs, the conclusion is: $\tau^*_1 = 0$. Substituting back into expression for $\Psi_\beta$ gives

$$
\Psi_\beta(x^*_i, y^f_i) = \langle x^*_i, y^f_i \rangle - \left[x_{1 \tau_2} + \beta S_2 I_{\{\tau_2 \neq 0\}}\right].
$$

**Max**

Finally, taking the max values for $\Psi_\beta$ over cases $w^*_2 = 0$; $w^*_2 \neq 0$ gives

$$
\Psi_\beta(x^*_i, y^f_i) = \langle x^*_i, y^f_i \rangle + g_i(\beta) \vee \left[-x_{1 \tau_2} - \beta S_2 K\right]
$$
where \( K := (\mathbb{1}_{\tau_2 \neq 0}) \). Inspection suggests that \( \Psi_\beta \) can be simplified further. Proceed to rewrite \( \Psi_\beta \) as

\[
\Psi_\beta(x_{l}, y_{l}') = \langle x_{l}, y_{l}' \rangle + \left[ -x_{l,2} - \beta S_{2}K \right] + \delta_{l}^+(\beta)
\]

where

\[
\delta_{l}(\beta) = \begin{cases} 
\beta S_{2}K - \beta (x_{l,2})^2, & -x_{l,2} \leq \frac{\|y_{l}'\|}{2\beta}, \\
\frac{\|y_{l}'\|^2}{4\beta} + x_{l,2} + \beta S_{2}K, & -x_{l,2} > \frac{\|y_{l}'\|}{2\beta}.
\end{cases}
\]

**Proposition 4** Let \( \beta^* = \sup_{\beta \geq 0} \{ \beta : S_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N}\left[ K_{1}(\beta) + K_{2}(\beta) \right] \leq 0 \} \), equivalently

let \( \beta^* = \inf_{\beta \geq 0} \{ \beta : S_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N}\left[ K_{1}(\beta) + K_{2}(\beta) \right] \geq 0 \} \), where \( \beta = \frac{\|y_{l}'\|}{2\beta}, K_{0} = \sum_{i=1}^{N} \|y_{l}'\| \).

\[
K_{1}(\beta) = \sum_{i=1}^{N} \mathbb{1}_{\{x_{l,2} \leq \beta\}} \mathbb{1}_{\{0 \leq \delta_{l}(\beta)\}} [S_{2}K - (x_{l,2})^2],
\]

\[
K_{2}(\beta) = \sum_{i=1}^{N} \mathbb{1}_{\{x_{l,2} > \beta\}} \mathbb{1}_{\{0 \leq \delta_{l}(\beta)\}} [S_{2}K - \frac{\|y_{l}'\|^2}{4\beta^2}].
\]

**Proof.** Define \( \beta' = \frac{\|y_{l}'\|}{2\beta} \). Recall that \( g_{l}(\beta) \) is a convex function \( \forall i \in \{1, \ldots, N\} \) (see Proof for Case 1, Proposition 2.3). Hence one can deduce that \( G(\beta) = \beta S_{2} - \frac{1}{N} \sum_{i=1}^{N} \Psi_{\beta}(x_{l}, y_{l}') \) is a convex function of \( \beta \). This follows easily since \( \Psi_{\beta} \) is a constant plus pointwise max of two convex functions, \( g_{l}(\beta) \) and a linear term. And \( G(\beta) \) is a linear term plus a sum of convex functions \( \Psi_{\beta} \). Hence the first order optimality condition will suffice to determine \( \beta^* \). \( G \) may not be differentiable at \( \beta \) hence we look for \( \beta^* \) such that \( 0 \in \partial G(\beta^*) \). The first order condition, LHS, says

\[
\delta_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N} \sum_{i=1}^{N} \left[ \mathbb{1}_{\{x_{l,2} \leq \beta\}} \mathbb{1}_{\{0 \leq \delta_{l}(\beta)\}} [S_{2}K - (x_{l,2})^2] + \mathbb{1}_{\{x_{l,2} > \beta\}} \mathbb{1}_{\{0 \leq \delta_{l}(\beta)\}} [S_{2}K - \frac{\|y_{l}'\|^2}{4\beta^2}] \right] \leq 0.
\]

Note the convexity of \( G(\beta) \) is enough to ensure that LHS is an increasing (non-decreasing) function of \( \beta \). In the degenerate case, where \( \sup_{\beta \geq 0} \) is taken over an empty set, select \( \beta^* = 0 \) since \( G(0) = 0 \) is the minimum value for \( G \). Otherwise, write

\[
\beta^* = \sup_{\beta \geq 0} \left\{ \beta : \delta_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N} \left[ K_{1}(\beta) + K_{2}(\beta) \right] \leq 0 \right\}.
\]

Substituting for \( K_{1}(\beta) \) and \( K_{2}(\beta) \) this simplifies to

\[
\beta^* = \sup_{\beta \geq 0} \left\{ \beta : \delta_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N} \left[ K_{1}(\beta) + K_{2}(\beta) \right] \leq 0 \right\}.
\]

The first order condition, RHS, says

\[
\delta_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N} \sum_{i=1}^{N} \left[ \mathbb{1}_{\{x_{l,2} \leq \beta\}} \mathbb{1}_{\{0 \leq \delta_{l}(\beta)\}} [S_{2}K - (x_{l,2})^2] + \mathbb{1}_{\{x_{l,2} > \beta\}} \mathbb{1}_{\{0 \leq \delta_{l}(\beta)\}} [S_{2}K - \frac{\|y_{l}'\|^2}{4\beta^2}] \right] \geq 0.
\]

Note the convexity of \( G(\beta) \) is enough to ensure that RHS is an increasing (non-decreasing) function of \( \beta \). Inspection of \( G(\beta) \) and the derivative condition below reveal that \( \inf_{\beta > 0} \) is not taken over an empty set. In particular, for \( \beta \) sufficiently large, \( 0 < \delta_{l}(\beta) \) for index \( i \) such that \( K > 0 \), the \( S_{2}K_{0} \) and \( S_{2}K \) terms cancel and RHS \( \geq 0 \) there. If no such index \( i \) exist, then RHS equals \( \delta_{2} > 0 \). Therefore, proceed to write

\[
\beta^* = \inf_{\beta \geq 0} \left\{ \beta : \delta_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N} \left[ K_{1}^{-1}(\beta) + K_{2}^{-1}(\beta) \right] \geq 0 \right\}.
\]

Substituting for \( K_{1}^{-1}(\beta) \) and \( K_{2}^{-1}(\beta) \) this simplifies to

\[
\beta^* = \inf_{\beta \geq 0} \left\{ \beta : \delta_{2} - \frac{1}{N}[S_{2}K_{0}] + \frac{1}{N} \left[ K_{1}^{-1}(\beta) + K_{2}^{-1}(\beta) \right] \geq 0 \right\}.
\]
Theorem 2. The primal problem \( P2 \) has solution

\[
- \sup_{Q \in \mathcal{Q}_2(a)} \mathbb{E}^Q[(X^-, Y_F)] = -[\beta^* \delta_2 + \frac{1}{N} \sum_{i=1}^{N} \Psi_{\beta^*}(x_i, y_i)]
\]

where \( \beta^* = \sup_{\beta \geq 0} \{ \beta : \delta_2 - \frac{1}{N} [S_2K_0 + \frac{1}{N} K_1(\beta) + K_2(\beta)] \leq 0 \} \), and \( \Psi_{\beta}(x_i, y_i') = \langle x_i, y_i' \rangle + g_i(\beta) \vee \left[ -x_i - \beta S_2K \right] \).

Expressed in terms of original DVA, this says

\[
- \sup_{Q \in \mathcal{Q}_2(a)} \mathbb{E}^Q[(X^-, Y_F)] = -\mathbb{E}^{Q_n}[(X^-, Y_F)] - \beta^* \delta_2 - \mathbb{E}^{Q_n} \left[ -X_{S_2} - \beta S_2K \right] \vee g(\beta^*)
\]

where

\[
g(\beta) = \begin{cases} 
-X_{S_2} - \beta (X_{S_2})^2, & X_{S_2} \leq \frac{\|Y_F\|}{2\beta}, \\
X_{S_2} > \frac{\|Y_F\|}{2\beta}, & \|Y_F\| > 0
\end{cases}
\]

and the additional terms represent a penalty due to uncertainty in probability distribution.

Proof. This follows by direct substitution of \( \beta^* \) as characterized in Proposition 2.4 into Proposition 2.3 and then the dual problem \( D2 \).

Proposition 5. We have \( \Psi_{\alpha}(x_i, y_i, y_i') = \sqrt{\alpha} \Psi_{\alpha}^0(x_i, y_i, y_i') \) where \( \Psi_{\alpha}^0(x_i, y_i, y_i') = \sum_{i=1}^{N} \Psi_{\alpha}^0(x_i, y_i, y_i') \)

\[
\Psi_{\alpha}(x_i, y_i, y_i') = \sum_{i=1}^{N} \Psi_{\alpha}^0(x_i, y_i, y_i') + \frac{1}{\sqrt{\alpha}} \Psi_{\alpha}^1(x_i, y_i, y_i'),
\]

where

\[
\Psi_{\alpha}^0(x_i, y_i, y_i') = \frac{1}{\sqrt{\alpha}} \Psi_{\alpha}^0(x_i, y_i, y_i'), \Psi_{\alpha}^1(x_i, y_i, y_i') = \frac{1}{\sqrt{\alpha}} \Psi_{\alpha}^1(x_i, y_i, y_i'), \Psi_{\alpha}^2(x_i, y_i, y_i') = \frac{1}{\sqrt{\alpha}} \Psi_{\alpha}^2(x_i, y_i, y_i'), \Psi_{\alpha}^3(x_i, y_i, y_i') = \frac{1}{\sqrt{\alpha}} \Psi_{\alpha}^3(x_i, y_i, y_i'), \Psi_{\alpha}^4(x_i, y_i, y_i') = \frac{1}{\sqrt{\alpha}} \Psi_{\alpha}^4(x_i, y_i, y_i'),
\]

and (suppressing arguments for brevity):

\[
\Psi_{\alpha}^{1a} = \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') \right]^{+}, \Psi_{\alpha}^{1b} = \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') \right]^{-} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') \right]^{+}, \Psi_{\alpha}^{2a} = \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') + (x_i - x_i') \right]^{+} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') \right]^{-} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') + (x_i - x_i') \right]^{-}, \Psi_{\alpha}^{3a} = \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') + (x_i - x_i') \right]^{+} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') \right]^{-} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') + (x_i - x_i') \right]^{-}, \Psi_{\alpha}^{4a} = \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') + (x_i - x_i') \right]^{+} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') \right]^{-} - \left[ \frac{1}{\sqrt{\alpha}} + (x_i, y_i') + (x_i - x_i') \right]^{-}.
\]

Note parameter \( t_i^* \) and constant \( K \) are defined within the proof by cases (see Supplementary Material), and are omitted here for brevity. Recall \( \tau_2 \) is index \( x \) such that \( \langle x, \cdot \rangle^+ = 0 \) else it is \( 0 \) if \( \|x\| > 0 \). The selection in \( \{c, f\} \) is determined by context.

Proof. The particular structure of \( B_{\alpha}^1 \) and \( B_{\alpha}^2 \) will be exploited to evaluate the sup above. The analysis proceeds by considering different cases for optimal values \((w_1^*, w_2^*, w_3^*)\).

Case 1. Suppose \( w_2^* = 0, w_3^* = 0 \). Then

\[
\Psi_{\alpha}(x_i, y_i, y_i') = \sup_{w_1 \in \mathbb{R}^\alpha} \left[ \langle (w_1 + x_i)^+, 1 \rangle_{\|y_i'\| < y_i'} \right] + \left[ \langle (w_1 + x_i)^-, 1 \rangle_{\|y_i'\| < y_i'} \right] - \alpha \langle w_1, w_1 \rangle.
\]

a) Suppose \( 1_{\|y_i'\| < y_i'} = 1 \). Then

\[
\Psi_{\alpha}(x_i, y_i, y_i') = \sup_{w_1 \in \mathbb{R}^\alpha} \left[ \langle (w_1 + x_i)^+, 1 \rangle_{\|y_i'\| < y_i'} \right] - \alpha \langle w_1, w_1 \rangle.
\]

Therefore \( \|y_i'\| = 1 \). Let \( \tau_2 \) denote default time for \( y_i' \). Simplify further to get

\[
\Psi_{\alpha}(x_i, y_i, y_i') = \sup_{w_1 \tau_2 \in \mathbb{R}} \left[ (w_1 \tau_2 + x_i \tau_2)^+ - \alpha (w_1 \tau_2)^2 \right].
\]

Now follow the approach in Bartl et al. (2017) to write down the first order optimality condition:

\[
1_{\{\alpha \geq 0\}}(w_1 \tau_2 + x_i \tau_2) - 2 \alpha w_1 \tau_2 \leq 0 \leq 1_{\{\alpha \geq 0\}}(w_1 \tau_2 + x_i \tau_2) - 2 \alpha w_1 \tau_2.
\]
i) Suppose \((w_1^{*\tau_2} + x_{2\tau_2}) < 0\). Then \(w_1^{*\tau_2} = 0\). So \(x_{2\tau_2} < 0 \implies w_2^{*\tau_2} = 0\).

ii) Suppose \((w_1^{*\tau_2} + x_{2\tau_2}) > 0\). Then \(w_1^{*\tau_2} = \frac{1}{2\alpha}\). So \(x_{2\tau_2} > -\frac{1}{2\alpha} \implies w_2^{*\tau_2} = \frac{1}{2\alpha}\).

iii) Note \((w_1^{*\tau_2} + x_{2\tau_2}) = 0\) is not possible (does not satisfy first order optimality condition).

Considering the intervals for \(x_{2}\) above, there are three cases as below.

i) \(x_{2\tau_2} \geq 0 \implies w_1^{*\tau_2} = \frac{1}{2\alpha} \implies \Psi = \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+\).

ii) \(x_{2\tau_2} \leq -\frac{1}{2\alpha} \implies w_2^{*\tau_2} = 0 \implies \Psi = \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+\).

iii) \((-\frac{1}{2\alpha} < x_{2\tau_2} < 0) \implies \Psi = \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+\).

In summary, considering all cases above, conclude that
\[
\Psi_{\alpha}(x_i, y_i^f, y_i^f) = \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+.
\]

This can also be expressed as
\[
\Psi_{\alpha}(x_i, y_i^f, y_i^f) = \left[\frac{1}{4\alpha} + (x_i, y_i^f)\right]^+.
\]

b) Suppose \(\mathbb{1}_{\{y_i^f < y_i^f\}} = 1\). Then
\[
\Psi_{\alpha}(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}^n} \left[\langle w_1 + x_i, y_i^f \rangle - \alpha(w_1, w_1)\right].
\]

Therefore \(\|y_i^f\| = 1\). Let \(\tau_2\) denote default time for \(y_i^f\). Simplify further to get
\[
\Psi_{\alpha}(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}^n} \left[\langle w_1 + x_i, y_i^f \rangle - \alpha(w_1, w_1)\right].
\]

Now follow the approach in [Bartl et al., 2017] to write down the first order optimality condition:
\[
\mathbb{1}_{\{x_i \geq 0\}}(w_1^{\tau_2} + x_{2\tau_2}) - 2\alpha w_1^{\tau_2} \leq 0 \leq \mathbb{1}_{\{x_i \leq 0\}}(w_1^{\tau_2} + x_{2\tau_2}) - 2\alpha w_1^{\tau_2}.
\]

i) Suppose \((w_1^{*\tau_2} + x_{2\tau_2}) > 0\). Then \(w_1^{*\tau_2} = 0\). So \(x_{2\tau_2} > 0 \implies w_2^{*\tau_2} = 0\).

ii) Suppose \((w_1^{*\tau_2} + x_{2\tau_2}) < 0\). Then \(w_1^{*\tau_2} = \frac{1}{2\alpha}\). So \(x_{2\tau_2} < -\frac{1}{2\alpha} \implies w_1^{*\tau_2} = \frac{1}{2\alpha}\).

iii) Note \((w_1^{*\tau_2} + x_{2\tau_2}) = 0\) is not possible (does not satisfy first order optimality condition).

Considering the intervals for \(x_{2}\) above, there are three cases as below.

i) \(x_{2\tau_2} > 0 \implies w_1^{*\tau_2} = 0 \implies \Psi = 0\).

ii) \(x_{2\tau_2} < -\frac{1}{2\alpha} \implies w_2^{*\tau_2} = \frac{1}{2\alpha} \implies \Psi = \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+\).

iii) \([-\frac{1}{2\alpha} \leq x_{2\tau_2} \leq 0) \implies w_1^{*\tau_2} = [x_{2\tau_2}]\).

Note the slope \((1 - 2\alpha w_1^{\tau_2})\) is positive for \(0 \leq w_1^{\tau_2} < \frac{1}{2\alpha}\), and equals zero at \(w_1^{\tau_2} = \frac{1}{2\alpha}\).

However, \((w_1^{\tau_2} + x_{2\tau_2})^+\) attains its max value of zero for \(w_1^{\tau_2} = \left|x_{2\tau_2}\right|\) so stop there.

In summary, considering all cases above, conclude that
\[
Psi_{\alpha}(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{x_{2\tau_2} < -\frac{1}{2\alpha}\}}(x_{2\tau_2}) \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+ - \mathbb{1}_{\{0 \leq x_{2\tau_2} \leq 0\}}(x_{2\tau_2})^2\right].
\]

This can also be expressed as
\[
Psi_{\alpha}(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{x_{2\tau_2} < -\frac{1}{2\alpha}\}}(x_{2\tau_2}) \left[\frac{1}{4\alpha} + x_{2\tau_2}\right]^+ - \mathbb{1}_{\{0 \leq x_{2\tau_2} \leq 0\}}(x_{2\tau_2})^2\right].
\]

c) Suppose \(\mathbb{1}_{\{y_i^f = \|y_i^f\| = 0\}} = 1\).

In this trivial case, \(\Psi = 0\). Note there is no third subcase for Cases 2-4 below since that would imply \(w_2^2 = 0, w_3^3 = 0\).
Finally, to sum up Case 1, considering parts a) and b), let us write:

\[ \Psi^l_\alpha(x_i, y_i', y'_i) = \mathbb{1}_{(y'_i < y'_p)} \Psi^l_{\alpha}(x_i, y_i', y'_i) + \mathbb{1}_{(y'_i < y'_p)} \Psi^l_{\alpha}(x_i, y_i', y'_i). \]

**Case 2** Suppose \( w^2_i \neq 0, w^2_i = 0. \)

Then \( w^2_i \) has +1 in position \( \tau^*_i \) and -1 in position \( \tau_2 \), where \( \tau^*_i = 0 \) means the value \( \pm 1 \) does not occur. Furthermore, \( \tau^*_i \neq \tau_2 \) otherwise \( w^2_i = 0. \)

\[ \Psi_\alpha(x_i, y_i', y'_i) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in B^2_n} \left[ ((w_1 + x_i)^+, w_2 + y'_i) + ((w_1 + x_i)^-, w_1, w_2) - \alpha(\langle w_1, w_1 \rangle + S_3(w_2, w_2)) \right]. \]

a) Suppose \( \mathbb{1}_{(w_2 + y'_i < y'_p)} = 1. \) Then

\[ \Psi_\alpha(x_i, y_i', y'_i) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in B^2_n} \left[ ((w_1 + x_i)^+, w_2 + y'_i) - \alpha(\langle w_1, w_1 \rangle + S_3(w_2, w_2)) \right]. \]

Recall \( (w_1 + x_i, (w_2 + y'_i)) = (w_1 \tau_1 + x_ \tau_1). \) Also recall \( \tau_1 \) and \( \tau_2 \) are associated with \( y'_i \). Let \( \tau_{2,f} \) denote default time (index) for \( y'_i \). The default time constraint implies \( \tau_1 < \tau_{2,f}. \) Therefore \( \tau_1 > 0. \) The structure of finite set \( B^2_n \) implies

\[ \Psi_\alpha(x_i, y_i', y'_i) = \sup_{w_1 \in \mathbb{R}^n, 0 < \tau_1 < \tau_{2,f}, \tau_1 \neq \tau_2} \left[ ((w_1 + x_i)^+, w_1, w_2) - \alpha(\langle w_1, w_1 \rangle + S_3(w_2, w_2)) \right]. \]

Observe the only positive component for \( w_1 \in \mathbb{R}^n \) in sup above is \( \tau_1. \)

\[ \sup_{w_1 \in \mathbb{R}^n} \left[ ((w_1 + x_\tau_1)^+, w_1, w_1) - \alpha(\langle w_1, w_1 \rangle + S_3(w_2, w_2)) \right]. \]

Evaluating at the critical point \( w^*_{1\tau_1} = \frac{1}{2\alpha} \in \mathbb{R} \) for the above quadratic gives

\[ \sup_{w_1 \in \mathbb{R}^n} \left[ ((w_1 + x_\tau_1)^+, w_1, w_1) - \alpha(\langle w_1, w_1 \rangle + S_3(w_2, w_2)) \right] = \left[ \frac{1}{4\alpha} + x_\tau_1 \right]^+. \]

Therefore one can write

\[ \Psi_\alpha(x_i, y_i', y'_i) = \max_{0 < \tau_1 < \tau_{2,f}} \left[ \left[ \frac{1}{4\alpha} + x_\tau_1 \right]^+ - \alpha S_3 K^2a \right] \]

where \( K^2 a := (1 + \mathbb{1}_{(\tau_{2,f} = 0)}). \) Furthermore, \( \tau^*_i \) is determined as

\[ \tau^*_i = \arg\max_{0 < \tau_1 < \tau_{2,f}, \tau_1 \neq \tau_2} \left[ x_\tau_1 \right]. \]

Substituting back into expression for \( \Psi_\alpha \) gives

\[ \Psi_\alpha^{2a}(x_i, y_i', y'_i) = \left[ \left[ \frac{1}{4\alpha} + x_\tau_1 \right]^+ - \alpha S_3 K^2a \right]. \]

This can also be expressed as

\[ \Psi_\alpha^{2a}(x_i, y_i', y'_i) = \left[ \left[ \frac{1}{4\alpha} + (x_i, y_i') + (x_{1\tau_1}^* - x_\tau_1) \right]^+ - \alpha S_3 K^2a \right]. \]

b) Suppose \( \mathbb{1}_{(y'_i < w_2 + y'_p)} = 1. \) Then

\[ \Psi_\alpha(x_i, y_i', y'_i) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in B^2_n} \left[ ((w_1 + x_i)^-, y'_i) - \alpha(\langle w_1, w_1 \rangle + S_3(w_2, w_2)) \right]. \]

Recall \( \tau_1 \) and \( \tau_2 \) are associated with \( y'_i \). Let \( \tau_{2,f} \) denote the default time (index) for \( y'_i \). The default time constraint implies \( \tau_{2,f} < \tau_1. \) Therefore \( \tau_{2,f} > 0 \) and \( \| y'_i \| = 1. \) Note the only non-zero component of \( ||y'_i|| \) is \( \tau_{2,f}. \) Hence set \( w^*_{1f} = 0 \forall \tau \neq \tau_{2,f}. \) Simplifying further

\[ \Psi_\alpha(x_i, y_i', y'_i) = \sup_{w_{1\tau_{2,f}} \in \mathbb{R}, w_2 \in B^2_n} \left[ (w_{1\tau_{2,f}} + x_{i\tau_{2,f}})^- - \alpha((w_{1\tau_{2,f}})^2 + S_3 K^2b) \right]. \]
where $K^{2b} := (\mathbb{1}_{\{\tau_1 \neq 0\}} + \mathbb{1}_{\{\tau_2 = 0\}}) = 1$. For $K^{2b}$, if $\tau_2 = 0$, then $\tau_1 \neq 0$ since $w_3^* \neq 0$. Otherwise set $\tau_1 = 0$ if $\tau_2 \neq 0$ to maximize $\sup_{w_3}$ above. Following the calculations in Case 1b) above, conclude that

$$\Psi_{2a}(x_i, y_i^f, y_i^f) = \left[ \mathbb{1}_{\{(\nu_{\tau_0} < \frac{1}{\alpha}) \cup (\nu_{\tau_0} > 0)\}} \left\{ \frac{1}{4\alpha} + x_{i\tau} \right\} - \mathbb{1}_{\{\frac{1}{\alpha} \leq \nu_{\tau_2} \leq 0\}} \left\{ \alpha(\langle x_{i\tau} \rangle^2) \right\} \right].$$

This can also be expressed as

$$\Psi_{2b}(x_i, y_i^f, y_i^f) = \left( \mathbb{1}_{\{\nu_{\tau_2} < \frac{1}{\alpha} \cup (\nu_{\tau_2} > 0)\}} \left\{ \frac{1}{4\alpha} + \langle x_i, y_i^f \rangle \right\} - \mathbb{1}_{\{\frac{1}{\alpha} \leq \nu_{\tau_2} \leq 0\}} \left\{ \alpha(\langle x_i, y_i^f \rangle^2) \right\} \right).$$

Finally, to sum up Case 2, considering parts a) and b), let us write:

$$\Psi_2(x_i, y_i^f, y_i^f) = \mathbb{1}_{\{w_2 + y_i^f < y_i^f\}} \Psi_{2a}(x_i, y_i^f, y_i^f) + \mathbb{1}_{\{y_i^f < w_2 + y_i^f\}} \Psi_{2b}(x_i, y_i^f, y_i^f).$$

**Case 3** Suppose $w_2^* = 0$, $w_3^* \neq 0$. Then $w_3^*$ is +1 in position $\tau_1^*$ and -1 in position $\tau_2$, where $\tau_j = 0$ means the value $\pm 1$ does not occur. Furthermore, $\tau_1^* \neq \tau_2$ otherwise $w_3^* = 0$.

$$\Psi_2(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, w_3 \in \mathbb{B}_a^*} \left[ \langle w_1 + x_i, y_i^f \rangle + \langle w_3 + y_i^f, y_i^f \rangle - \alpha(\langle w_1, w_1 \rangle + S_3(w_3, w_3)) \right].$$

a) Suppose $\mathbb{1}_{\{y_i^f < w_3 + y_i^f\}} = 1$. Then

$$\Psi_2(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, w_3 \in \mathbb{B}_a^*} \left[ \langle w_1 + x_i, y_i^f \rangle - \alpha(\langle w_1, w_1 \rangle + S_3(w_3, w_3)) \right].$$

Recall $\langle w_1 + x_i, y_i^f \rangle = (w_1 + x_{i\tau_2} + x_{i\tau_1}).$ Also recall $\tau_1$ and $\tau_2$ are associated with $y_i^f$. Let $\tau_{2,c}$ denote the default time (index) for $y_i^f$. The default time constraint implies $\tau_{2,c} < \tau_1$. Therefore $\tau_{2,c} > 0$ and $\|y_i^f\| = 1$. Note the only positive component of $\|y_i^f\|$ is $\tau_{2,c}$. Hence set $w_{1c}^* = 0 \forall \tau \neq \tau_{2,c}$. Simplicity further to get

$$\Psi_2(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}, w_3 \in \mathbb{B}_a^*} \left[ \langle w_1 + x_{i\tau_{2,c}}, y_i^f \rangle + \alpha(\langle w_{1c}, w_1 \rangle^2 + S_3(w_3, w_3)) \right]$$

where $K^{3a} := (\mathbb{1}_{\{\tau_1 \neq 0\}} + \mathbb{1}_{\{\tau_2 \neq 0\}}) = 1$, following logic in Case 2b) above. Evaluating at the critical point $w_{1c}^* = \frac{1}{4\alpha} \in \mathbb{R}$ gives

$$\sup_{w_{1c} \in \mathbb{R}} \left[ \langle w_{1c} + x_{i\tau_{2,c}}, y_i^f \rangle + \alpha(\langle w_{1c}, w_1 \rangle^2 + S_3(w_3, w_3)) \right] = \left[ \frac{1}{4\alpha} + x_{i\tau_{2,c}} \right]^+. $$

Therefore one can write

$$\Psi_2(x_i, y_i^f, y_i^f) = \left[ \frac{1}{4\alpha} + x_{i\tau_{2,c}} \right] + \alpha S_3^{3a}. $$

This can also be expressed as

$$\Psi_2(x_i, y_i^f, y_i^f) = \left[ \frac{1}{4\alpha} + \langle x_i, y_i^f \rangle \right] - \alpha S_3^{3a}. $$

b) Suppose $\mathbb{1}_{\{w_3 + y_i^f < y_i^f\}} = 1$. Then

$$\Psi_2(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, w_3 \in \mathbb{B}_a^*} \left[ \langle w_1 + x_i, y_i^f \rangle - \alpha(\langle w_1, w_1 \rangle + S_3(w_3, w_3)) \right].$$

Recall $\langle w_1 + x_i, (w_3 + y_i^f) \rangle = (w_1 + x_{i\tau_1} + x_{i\tau_2}).$ Also recall $\tau_1$ and $\tau_2$ are associated with $y_i^f$. Let $\tau_{2,c}$ denote default time (index) for $y_i^f$. The default time constraint implies $\tau_1 < \tau_{2,c}$. Therefore $\tau_1 > 0$ and

$$\Psi_2(x_i, y_i^f, y_i^f) = \sup_{w_1 \in \mathbb{R}, 0 < \tau_1 < \tau_{2,c}, \tau_1 \neq \tau_2} \left[ \langle w_{1\tau_1} + x_{i\tau_1}, y_i^f \rangle - \alpha(\langle w_{1\tau_1}, w_1 \rangle^2 + S_3(w_3, w_3)) \right]$$

where $K^{3b} := (1 + \mathbb{1}_{\{\tau_2 \neq 0\}}).$ Following the calculations in Case 2b) above, conclude that

$$\Psi_2(x_i, y_i^f, y_i^f) = \left[ \mathbb{1}_{\{\nu_{\tau_1} < \frac{1}{\alpha} \cup (\nu_{\tau_1} > 0)\}} \left\{ \frac{1}{4\alpha} + x_{i\tau_1} \right\} - \mathbb{1}_{\{\frac{1}{\alpha} \leq \nu_{\tau_1} \leq 0\}} \left\{ \alpha(\langle x_{i\tau_1} \rangle^2) \right\} \right].$$
Furthermore, $\tau^*_i$ is determined as

$$\tau^*_i = \arg\max_{0 < \tau \leq \tau^*_i, \tau \neq \tau_2} [x_{r_1}].$$

Therefore one can write

$$\Psi^{3b}_\alpha(x_i, y_i^c, y_i^f) = \left[ \mathbb{1}_{\{x_{r_1} < \frac{1}{4\alpha}\}} \frac{1}{4\alpha} + x_{r_1} \right] - \mathbb{1}_{\{\frac{1}{4\alpha} \leq x_{r_1} \leq 0\}} [\alpha(x_{r_1})^2 + \alpha S_3 K^{3b}].$$

This can also be expressed as

$$\Psi^{3b}_\alpha(x_i, y_i^c, y_i^f) = \left[ \mathbb{1}_{\{x_{r_1} < \frac{1}{4\alpha}\}} \frac{1}{4\alpha} + (x_i y_i^c + (x_{r_1} - x_2)) \right] - \mathbb{1}_{\{\frac{1}{4\alpha} \leq x_{r_1} \leq 0\}} [\alpha(x_{r_1})^2 + \alpha S_3 K^{3b}].$$

Finally, to sum up Case 3, considering parts a) and b), let us write:

$$\Psi^{3a}_\alpha(x_i, y_i^c, y_i^f) = \left[ \left( y_i^f + \mathbb{1}_{\{w_3 + y_i^f \leq t\}} \right) \Psi^{3a}_\alpha(x_i, y_i^c, y_i^f) + \mathbb{1}_{\{w_3 + y_i^f \leq t\}} \Psi^{3b}_\alpha(x_i, y_i^c, y_i^f) \right].$$

**Case 4**

Suppose $w_2^* \neq 0, w_3^* \neq 0$.

Then $w_2^*$ has $+1$ in position $\tau^*_i$, and $-1$ in position $\tau_2$, where $\tau_{j,c} = 0$ means the value $\pm 1$ does not occur.

Furthermore, $\tau^*_i \neq \tau_2$, otherwise $w_3^* = 0$.

And $w_3^*$ has $+1$ in position $\tau^*_f$, and $-1$ in position $\tau_2$, where $\tau_{j,f} = 0$ means the value $\pm 1$ does not occur.

Furthermore, $\tau^*_f \neq \tau_2$, otherwise $w_3^* = 0$.

$$\Psi(x_i, y_i^c, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in \mathbb{B}^2_w, w_3 \in \mathbb{B}^2_w} \left[ (w_1 + 1)^+, \mathbb{1}_{\{w_2 + y_i^c \leq w_3 + y_i^f\}} w_2 + y_i^f \right] + \mathbb{1}_{\{w_2 + y_i^c \leq w_3 + y_i^f\}} w_3 + y_i^f \right] - \Psi(x_i, w_1) + S_3 (w_2, w_3) \right].$$

a) Suppose $\mathbb{1}_{\{w_2 + y_i^c \leq w_3 + y_i^f\}} = 1$. Then

$$\Psi(x_i, y_i^c, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in \mathbb{B}^2_w, w_3 \in \mathbb{B}^2_w} \left[ (w_1 + 1)^+, w_2 + y_i^f \right] - \Psi(x_i, w_1) + S_3 (w_2, w_3) \right].$$

Recall $\langle (w_1 + 1), (w_2 + y_i^f) \rangle = (w_1 + 1, x_{r_1})$. The default time constraint implies $\tau_{1,c} < \tau_{1,f}$. Therefore $\tau_{1,c} > 0$.

The structure of finite set $\mathbb{B}^2_w$ implies

$$\Psi(x_i, y_i^c, y_i^f) = \sup_{w_1 \in \mathbb{R}^n, 0 < \tau_{1,c} < \tau_{1,f}, \tau_{1,c} \neq \tau_2} \left[ (w_1 + 1) + x_{r_1})^+ \Psi(x_i, w_1) + S_3 (w_2, w_3) \right].$$

Observe the only positive component for $w_1 \in \mathbb{R}^n$ in sup above is $\tau_{1,c}$.

$$\sup_{w_1 \in \mathbb{R}^n} \left[ (w_1 + 1)^+ + \Psi(x_i, w_1) \right] = \sup_{w_1 \in \mathbb{R}^n} \left[ (w_1 + 1)^+ \Psi(x_i, w_1) \right].$$

Evaluating at the critical point $w_1^{*} = \frac{1}{4\alpha} \in \mathbb{R}$ for the above quadratic gives

$$\sup_{w_1 \in \mathbb{R}^n} \left[ (w_1 + 1)^+ \Psi(x_i, w_1) \right] = \left[ \frac{1}{4\alpha} + x_{r_1} \right].$$

Therefore one can write

$$\Psi(x_i, y_i^c, y_i^f) = \max_{0 < \tau_{1,c} < \tau_{1,f}, \tau_{1,c} \neq \tau_2} \left[ \frac{1}{4\alpha} + x_{r_1} \right] - \alpha S_3 K^{4a}.$$

where $K^{4a} := (\mathbb{1}_{\{\tau_{1,c} = 0\}} + \mathbb{1}_{\{\tau_{2,c} = 0\}} + \mathbb{1}_{\{\tau_{1,f} = 0\}} + \mathbb{1}_{\{\tau_{2,f} = 0\}}) = (2 + \mathbb{1}_{\{\tau_2 = 0\}})$ following logic as in Case 3a) above.

Furthermore, $\tau^*_i$ is determined as

$$\tau^*_i = \arg\max_{0 < \tau \leq \tau^*_i, \tau \neq \tau_2} [x_{r_1}].$$

Substituting back into expression for $\Psi(x_i, y_i^c, y_i^f)$ gives

$$\Psi^{4a}_\alpha(x_i, y_i^c, y_i^f) = \left[ \frac{1}{4\alpha} + x_{r_1} \right] - \alpha S_3 K^{4a}. $$

30
Let $\tau_2 = \tau_{2, c}$. Then this can also be expressed as
\[
\Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l) = \left[ \frac{1}{4\alpha} + (x_i, y_i^c) + (x_i, y_i^l) \right] - \alpha S_3 K^{4b},
\]

b) Suppose $\mathbb{I}_{(w_3 + y_i^c < w_2 + y_i^l)} = 1$. Then
\[
\Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l) = \sup_{w_1 \in \mathbb{R}^n, w_2 \in \mathbb{B}_n^c, w_3 \in \mathbb{B}_n^c} \left[ (w_1 + x_i)^{-1}, w_3 + y_i^l \right] - \alpha \langle w_1 + s_3, w_2 + s_3, w_3 \rangle.
\]

Recall $\langle (w_1 + x_i), (w_3 + y_i^l) \rangle = (w_1, x_i, y_i^l)$. The default time constraint implies $\tau_{1, f} < \tau_{1, c}$. Therefore $\tau_{1, f} > 0$. The structure of finite set $\mathbb{B}_n^c$ implies
\[
\Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l) = \sup_{w_1 \in \mathbb{R}^n, 0 < \tau_{1, f} < \tau_{c, f}, \tau_{1, f} \neq \tau_{2, f}} \left[ (w_1, x_i, y_i^l) - \alpha (w_1 + s_3) \right].
\]

where $K := \sum_{k=1}^{N} \mathbb{I}_{\{\tau_{1, k} \neq 0\}} + \mathbb{I}_{\{\tau_{2, k} \neq 0\}} + \mathbb{I}_{\{\tau_{1, k} = 0\}} + \mathbb{I}_{\{\tau_{2, k} = 0\}} = (2 + \mathbb{I}_{\{\tau_{2, k} \neq 0\}})$ following logic as in Case 4a) above. Following the calculations in Case 3b) above, conclude that
\[
\Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l) = \sum_{k=1}^{N} \mathbb{I}_{\{\tau_{1, k} \neq 0\}} \left[ (w_1 + x_i, y_i^c) - \alpha (w_1 + s_3) \right].
\]

Furthermore, $\tau_i^*$ is determined as
\[
\tau_i^* = \arg\max_{0 < \tau_{1, f} < \tau_{c, f}, \tau_{1, f} \neq \tau_{2, f}} [x_i, y_i^c].
\]

Therefore one can write
\[
\Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l) = \sum_{k=1}^{N} \mathbb{I}_{\{\tau_{1, k} \neq 0\}} \left[ (w_1 + x_i, y_i^c) - \alpha (w_1 + s_3) \right].
\]

Let $\tau_2 = \tau_{2, f}$. Then this can also be expressed as
\[
\Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l) = \sum_{k=1}^{N} \mathbb{I}_{\{\tau_{1, k} \neq 0\}} \left[ (w_1 + x_i, y_i^c) - \alpha (w_1 + s_3) \right].
\]

Finally, to sum up Case 4, considering parts a) and b), let us write:
\[
\Psi_{\alpha}^{4d}(x_i, y_i^c, y_i^l) = \mathbb{I}_{(w_2 + y_i^c < w_3 + y_i^l)} \Psi_{\alpha}^{4a}(x_i, y_i^c, y_i^l) + \mathbb{I}_{(w_3 + y_i^c < w_2 + y_i^l)} \Psi_{\alpha}^{4b}(x_i, y_i^c, y_i^l).
\]

**Theorem 3.** The primal problem $P_3$ has solution $[\alpha^* \delta_1 + \frac{1}{N} \sum_{k=1}^{N} \Psi_{\alpha^*}(x_i, y_i^c, y_i^l)]$ where $\alpha^* = \arg\min_{\alpha \geq 0} [\alpha \delta_1 + \frac{1}{N} \sum_{k=1}^{N} \Psi_{\alpha^*}(x_i, y_i^c, y_i^l)]$ and $\Psi_{\alpha^*}(x_i, y_i^c, y_i^l) = \sum_{k=1}^{N} \Psi_{\alpha^*}(x_i, y_i^c, y_i^l)$.

Expressed in terms of original BCVA, this says
\[
\sup_{\Phi \in \mathcal{H}_d(\Phi_N)} \mathbb{E}^\Phi[\{X^+, Y^C\} + \{X^-, Y^F\}] = \mathbb{E}^\Psi[\{X^+, Y^C\} + \{X^-, Y^F\}] + \alpha^* \delta_3 + \mathbb{E}^\Psi[\Psi_{\alpha^*}(X, Y^C, Y^F) - \{X^+, Y^C\} + \{X^-, Y^F\}]^+
\]

where the additional terms represent a penalty due to uncertainty in probability distribution.

**Proof.** This follows by direct substitution of $\alpha^*$ as characterized above into the dual problem $D_3$. 

\[\square\]