Electronic Journal of Probability

Electron. J. Probab. **29** (2024), article no. 57, 1–30.
ISSN: 1083-6489 https://doi.org/10.1214/24-EJP1114

An extension of martingale transport and stability in robust finance

Benjamin Jourdain* Gudmund Pammer†

Abstract

While many questions in robust finance can be posed in the martingale optimal transport framework or its weak extension, others like the subreplication price of VIX futures, the robust pricing of American options or the construction of shadow couplings necessitate additional information to be incorporated into the optimization problem beyond that of the underlying asset. In the present paper, we take into account this extra information by introducing an additional parameter to the weak martingale optimal transport problem. We prove the stability of the resulting problem with respect to the risk neutral marginal distributions of the underlying asset, thus extending the results in [9]. A key step is the generalization of the main result in [7] to include the extra parameter into the setting. This result establishes that any martingale coupling can be approximated by a sequence of martingale couplings with specified marginals, provided that the marginals of this sequence converge to those of the original coupling. Finally, we deduce stability of the three previously mentioned motivating examples.

Keywords: martingale optimal transport; adapted Wasserstein distance; robust finance; weak transport; stability; convex order; martingale couplings.
MSC2020 subject classifications: 49Q22; 60G42; 91G80.
Submitted to EJP on April 24, 2023, final version accepted on March 14, 2024.

1 Introduction

In mathematical finance, the evolution of an asset price in a financial market is modeled by an adapted stochastic process (X_t) in a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t))$. To ensure the absence of arbitrage opportunities, risk-neutral measures (also known as equivalent martingale measures) $\mathbb{Q}$ are considered under which the asset price process (X_t) is a martingale, up to assuming zero interest rates. The reason why a transport-type problem arises in robust finance is because the marginals of (X_t) can be derived from market information based on the celebrated observation of Breeden–Litzenberger

*CERMICS, Ecole des Ponts, INRIA, Marne-la-Vallée, France. E-mail: benjamin.jourdain@enpc.fr
†ETH Zürich, Switzerland. E-mail: gudmund.pammer@math.ethz.ch

[11]. According to this observation, the prices of traded vanilla options determine the marginals (μ_t) of (X_t) at their respective maturity times under every risk-neutral measure $\mathbb{Q}$. Instead of considering one specific financial model, a robust approach is then to consider all martingale measures that are compatible with this observation, that is, all filtered probability spaces $(\Omega, \mathcal{F}, \mathbb{Q}, (\mathcal{F}_t))$ and stochastic processes (X_t) such that

$$X \text{ is a } (\mathbb{Q}, (\mathcal{F}_t))\text{-martingale and } X_t \sim \mu_t \text{ at all maturity times } t. \tag{1.1}$$

A reference measure is not needed. Then the robust price bounds for an option with payoff Φ are obtained by solving a transport-type problem [6, 14] where the optimization takes place over the set of all risk-neutral measures that are compatible with the observed prices of vanilla options:

$$\inf / \sup \left\{ \mathbb{E}_{\mathbb{Q}}[\Phi] : (\Omega, \mathcal{F}, \mathbb{Q}, (\mathcal{F}_t), (X_t)) \text{ satisfying } (1.1) \right\}. \tag{1.2}$$

However, as we can only observe the prices of a finite number of derivatives (up to a bid-ask spread), the marginals (μ_t) are merely approximately known. Therefore, it is crucial to establish the stability of the transport-type problem (1.2) with respect to the marginals.

This article is concerned with the one time period setting, that is $t \in \{1, 2\}$. Then, when the payoff Φ is written on the underlying asset X, (1.2) boils down to a martingale optimal transport (MOT) problem

$$\inf_{\pi \in \Pi_M(\mu_1, \mu_2)} / \sup \int \Phi(x, y) \, \pi(dx, dy), \tag{1.3}$$

where $\Pi_M(\mu_1, \mu_2)$ denotes the set of martingale couplings with marginals μ_1 and μ_2, i.e., the set of laws of one-step martingales (X_1, X_2) with $X_t \sim \mu_t$. Continuity of the value of (1.3) w.r.t. the marginal input, which is called stability, has been proved in [4, 28].

Weak martingale optimal transport (WMOT) is a nonlinear generalization of MOT analogous to weak optimal transport, which is a nonlinear generalization of classical optimal transport proposed by Gozlan, Roberto, Samson and Tetali [15], and was considered in [4, 9]. In WMOT one allows for more general payoffs Φ which may depend on the conditional law of X_2 given X_1 in addition to X itself, and the corresponding WMOT problem reads as

$$\inf_{\pi \in \Pi_M(\mu_1, \mu_2)} / \sup \int \Phi\left(x, \pi_x\right) \, \mu_1(dx), \tag{1.4}$$

where π_x comes from the desintegration $\pi(dx, dy) = \mu_1(dx)\pi_x(dy)$. Stability of WMOT has been studied in [9] and was therein used to establish stability of the superreplication price of VIX futures and the stretched Brownian motion.

Even though many problems in robust finance are covered by WMOT, some important examples require that information is included into the optimization problem beyond that of the underlying asset. Accordingly these problems can not be properly treated in the WMOT frameworks. For us, guiding examples of such problems are the subreplication price of VIX futures (see [16]), the robust pricing of American options (see [17]) and the construction of shadow couplings (see [10]). Through augmenting WMOT by an additional parameter, we demonstrate how this extra information can be taken into account, prove stability of the resulting problem, and consequently deduce stability of the three guiding examples. A key step is the generalization of the main result in [7] to our current setting. This result states that any martingale coupling can be approximated by a sequence of martingale couplings with specified marginals, provided that the marginals of this sequence converge to those of the original coupling. As a side product of our approach, we establish the very same result on the level of stochastic processes with general filtrations (c.f. [5]): any one-step martingale on some filtered probability space can be approximated w.r.t. the adapted Wasserstein distance by martingales on (perhaps different) filtered probability spaces, provided that the marginals of this sequence converge to those of the original martingale.

1.1 Notation

Let $(\mathcal{X}, d_\mathcal{X})$ and $(\mathcal{Y}, d_\mathcal{Y})$ be Polish metric spaces and $p \geq 1$ We equip the product $\mathcal{X} \times \mathcal{Y}$ with the product metric $d_{\mathcal{X}\times\mathcal{Y}}((x,y),(\tilde{x},\tilde{y})) := (d_\mathcal{X}(x,\tilde{x})^p + d_\mathcal{Y}(y,\tilde{y})^p)^{1/p}$ which turns $\mathcal{X} \times \mathcal{Y}$ into a Polish metric space. The set of Borel probability measures on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. For $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, we write $\Pi(\mu,\nu)$ for the set of all probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν. We denote by $\mathcal{P}_p(\mathcal{X})$ the subset of $\mathcal{P}(\mathcal{X})$ that finitely integrates $x \mapsto d^p_\mathcal{X}(x,x_0)$ for some (thus any) $x_0 \in \mathcal{X}$ and endow $\mathcal{P}_p(\mathcal{X})$ with the p-Wasserstein distance $\mathcal{W}_p$ so that $(\mathcal{P}_p(\mathcal{X}), \mathcal{W}_p)$ is a Polish metric space where, for $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$,

$$\mathcal{W}_p(\mu,\nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int d_\mathcal{X}(x,y)^p \, \pi(dx,dy) \right)^{1/p}. \tag{1.5}$$

The set of continuous and bounded functions on $\mathcal{X}$ is denoted by $C_b(\mathcal{X})$ and we use the shorthand notation $\mu(f)$ to write the integral of a μ-integrable function $f \colon \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$ w.r.t. a Borel measure μ on $\mathcal{X}$. Given a measurable map $f \colon \mathcal{X} \to \mathcal{Y}$, we denote by $f_\# \mu$ the push-forward measure of μ under f. For Polish spaces $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ and $\pi \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$ and a non-empty subset I of $\{1,2,3\}$, $\mathrm{proj}_I \, \pi$ denotes the image of π by the projection to the coordinates in I, for example, $\mathrm{proj}_1 \, \pi$ is the $\mathcal{X}_1$-marginal of π. Further, we write π_{x_1,x_2} for the disintegration of $\pi(dx_1,dx_2,dx_3) = \mathrm{proj}_{1,2} \, \pi(dx_1,dx_2)\pi_{x_1,x_2}(dx_3)$. Frequently, we use the injection (c.f. [3, Section 2])

$$\begin{aligned} J \colon \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3) &\to \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{P}(\mathcal{X}_3)) \\ \pi &\mapsto ((x_1,x_2,x_3) \mapsto (x_1,x_2,\pi_{x_1,x_2}))_\# \, \pi, \end{aligned}$$

and remark that $J(\pi^k) \to J(\pi)$ in $\mathcal{P}_p(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{P}_p(\mathcal{X}_3))$ implies $\pi^k \to \pi$ in $\mathcal{P}_p(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3)$.

Unless stated otherwise, $\mathbb{R}$ is equipped with the Euclidean distance and Leb denotes the Lebesgue measure on $[0,1]$. Two measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are said to be in convex order and we write $\mu \leq_{cx} \nu$, if

$$\forall \varphi \colon \mathbb{R} \to \mathbb{R} \text{ convex}, \quad \mu(\varphi) \leq \nu(\varphi).$$

We write $\mathrm{mean} \colon \mathcal{P}_1(\mathbb{R}) \to \mathbb{R}$ for $\mathrm{mean}(\rho) = \int y \, \rho(dy)$ and denote by

$$\begin{aligned} \Lambda_M(\mu,\nu) := \Big\{ P \in \mathcal{P}_1(\mathbb{R} \times \mathcal{P}_1(\mathbb{R})) : \int \delta_x \otimes \rho \, P(dx,d\rho) \in \Pi(\mu,\nu), \\ \mathrm{mean}(\rho) = x \, P(dx,d\rho)\text{-a.s.} \Big\}. \end{aligned}$$

1.2 Organization of the paper

Section 2 presents the main results of this paper. First, we introduce in Subsection 2.1 the setup with the additional parameter and state in Theorem 2.1 and Theorem 2.2 the corresponding results related to stability. Furthermore, we present in Subsection 2.2 consequences of these results in the filtered process setting, namely Corollary 2.6. Subsequently, we explain and state stability of the three guiding examples, that are, robust pricing of American options (Subsection 2.3), subreplication of VIX futures (Subsection 2.4), and shadow couplings (Subsection 2.5). Section 3 is concerned with the proofs. In the appendix we collect measure-theoretic auxiliary results.

2 Main results

2.1 An extension of martingale transport

We now introduce a framework that is sufficiently general to deal with the question of stability of our guiding examples. From now on, let $(\mathcal{U}, d_{\mathcal{U}})$ be a Polish metric space that models an extra information parameter $u \in \mathcal{U}$. Given $\bar{\mu} \in \mathcal{P}_1(\mathbb{R} \times \mathcal{U})$ and $\nu \in \mathcal{P}_1(\mathbb{R})$ with $\mathrm{proj}_1 \bar{\mu} \leq_{cx} \nu$, we denote by $\Pi_M(\bar{\mu}, \nu)$ the set of couplings $\pi \in \Pi(\bar{\mu}, \nu)$ such that $\mathrm{mean}(\pi_{x,u}) = x$, $\bar{\mu}(dx, du)$-a.e. Central to establishing the upper (resp. lower) semicontinuity property in our stability results for minimization (resp. maximization) problems is Theorem 3.2, which is a reinforced version of the result below:

Theorem 2.1. *Let $(\bar{\mu}^k, \nu^k)_{k \in \mathbb{N}}$, $\mathrm{proj}_1 \bar{\mu}^k \leq_{cx} \nu^k$, be a convergent sequence in $\mathcal{P}_1(\mathbb{R} \times \mathcal{U}) \times \mathcal{P}_1(\mathbb{R})$ with limit $(\bar{\mu}, \nu)$. Then, every coupling $\pi \in \Pi_M(\bar{\mu}, \nu)$ is the weak limit of a sequence $(\pi^k)_{k \in \mathbb{N}}$ with $\pi^k \in \Pi_M(\bar{\mu}^k, \nu^k)$ and $J(\pi)$ is the weak limit of $(J(\pi^k))_{k \in \mathbb{N}}$.*

In the view of the counter-example by Brückerhoff and Juillet [12], this result does not generalize to higher dimensions, i.e., when $\mathbb{R}$ is replaced by $\mathbb{R}^d$ with $d \geq 2$. This generalization of the main result of [7] to the present framework is also key to establish the stability w.r.t. the marginals of the following variant of WMOT:

$$V_C(\bar{\mu}, \nu) := \inf_{\pi \in \Pi_M(\bar{\mu}, \nu)} \int C(x, u, \pi_{x,u}) \, \bar{\mu}(dx, du). \tag{2.1}$$

As usual, it is necessary to impose regularity on the cost C in order to have a continuous dependence of the optimal value of (2.1) w.r.t. the marginals. Thus, we will suppose the following continuity assumption on the cost function:

Assumption A. We say $C \colon \mathbb{R} \times \mathcal{U} \times \mathcal{P}_p(\mathbb{R}) \to \mathbb{R}$ satisfies Assumption A if C is continuous and there is $K > 0$ such that, for all $(x, u, \rho) \in \mathbb{R} \times \mathcal{U} \times \mathcal{P}_p(\mathbb{R})$ and some $u_0 \in \mathcal{U}$,

$$|C(x, u, \rho)| \leq K \left(1 + |x|^p + d_{\mathcal{U}}^p(u, u_0) + \int |y|^p \, \rho(dy) \right). \tag{2.2}$$

Theorem 2.2. *Let C satisfy Assumption A and $C(x, u, \cdot)$ be convex for all $(x, u) \in \mathbb{R} \times \mathcal{U}$. Then the value function V_C defined in* (2.1) *is attained and continuous on $\{(\bar{\mu}, \nu) : \mathrm{proj}_1 \bar{\mu} \leq_{cx} \nu\} \subseteq \mathcal{P}_p(\mathbb{R} \times \mathcal{U}) \times \mathcal{P}_p(\mathbb{R})$. Furthermore, when $(\bar{\mu}^k, \nu^k)_{k \in \mathbb{N}}$ converges to $(\bar{\mu}, \nu)$ and for $k \in \mathbb{N}$, $\mathrm{proj}_1 \bar{\mu}^k \leq_{cx} \nu^k$ and $\pi^k \in \Pi_M(\bar{\mu}^k, \nu^k)$ is optimal for $V_C(\bar{\mu}^k, \nu^k)$, we have:*

(i) the accumulation points of $(\pi^k)_{k \in \mathbb{N}}$ are optimal for $V_C(\bar{\mu}, \nu)$;

(ii) if additionally $C(x, u, \cdot)$ is strictly convex, then optimizers to (2.1) *are unique. Moreover, $(\pi^k)_{k \in \mathbb{N}}$ and $(J(\pi^k))_{k \in \mathbb{N}}$ weakly converge to the optimizer of $V_C(\bar{\mu}, \nu)$ and its image under J, respectively.*

2.2 Filtered processes

As explained in the introduction, in the robust approach it is natural to consider all martingales that are compatible with market observations. For this reason, we follow the approach in [5], and call in our setting a 5-tuple

$$\mathbf{X} = \left(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \{1,2\}}, X = (X_t)_{t \in \{1,2\}} \right),$$

consisting of a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \{1,2\}})$ and an $(\mathcal{F}_t)$-adapted process X, a filtered process. We say that a filtered process $\mathbf{X}$ is a martingale if X is a $(\mathcal{F}_t)$-martingale under $\mathbb{P}$. When $\mathcal{F}_1$ is larger than the σ-field generated by X_1, the conditional distributions $\mathrm{law}(X_2 | \mathcal{F}_1)$ and $\mathrm{law}(X_2 | X_1)$ may differ and then $\mathrm{law}(X_2 | \mathcal{F}_1)$ is not determined by the law of X. For $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ with $\mu \leq_{cx} \nu$, we write $\mathbf{M}(\mu, \nu)$ for the set of all martingales $\mathbf{X}$ with $X_1 \sim \mu$ and $X_2 \sim \nu$.

Assume that the payoff Φ of the option, which we want to price, depends on the conditional law $\mathrm{law}(X_2|\mathcal{F}_1)$ in addition to the price of the asset X: Φ is a measurable function with domain $\mathbb{R}\times\mathbb{R}\times\mathcal{P}_p(\mathbb{R})$ satisfying the existence of $K>0$ such that $|\Phi(x_1,x_2,\rho)|\le K\left(1+|x_1|^p+|x_2|^p+\int_{\mathbb{R}}|y|^p\,\rho(dy)\right)$ for all $(x_1,x_2,\rho)\in\mathbb{R}\times\mathbb{R}\times\mathcal{P}_p(\mathbb{R})$. The robust price upper bound is given by

$$V_\Phi(\mu,\nu)=\sup_{\mathbf{X}\in\mathbf{M}(\mu,\nu)}\mathbb{E}[\Phi(X,\mathrm{law}(X_2|\mathcal{F}_1))]. \tag{2.3}$$

For $C\colon\mathbb{R}\times\mathcal{P}_p(\mathbb{R})\ni(x,\rho)\mapsto\int\Phi(x,y,\rho)\,\rho(dy)\in\mathbb{R}$, one has

$$\mathbb{E}[\Phi(X,\mathrm{law}(X_2|\mathcal{F}_1))|\mathcal{F}_1]=C(X_1,\mathrm{law}(X_2|\mathcal{F}_1)).$$

We connect the robust price bounds to the optimization problem

$$\hat V_C(\mu,\nu):=\sup_{P\in\Lambda_M(\mu,\nu)}\int C(x,\rho)P(dx,d\rho). \tag{2.4}$$

Indeed, the values of (2.3) and (2.4) coincide. To see this, oberserve that for $\mathbf{X}\in\mathbf{M}(\mu,\nu)$, we have $\mathrm{law}(X_1,\mathrm{law}(X_2|\mathcal{F}_1))\in\Lambda_M(\mu,\nu)$ and therefore $V_\Phi(\mu,\nu)\le\hat V_C(\mu,\nu)$. On the other hand, for $P\in\Lambda_M(\mu,\nu)$, let us endow $\mathbb{R}\times\mathcal{P}_p(\mathbb{R})\times[0,1]$ with $P\otimes\mathrm{Leb}$ and set $(X_1,X_2)\colon\mathbb{R}\times\mathcal{P}_p(\mathbb{R})\times[0,1]\ni(x,\rho,u)\mapsto(x,F_\rho^{-1}(u))\in\mathbb{R}^2$, where F_ρ^{-1} denotes the (left-continuous) quantile function of the probability measure ρ so that $(F_\rho^{-1})_\#\mathrm{Leb}=\rho$ by the inverse transform sampling. Denoting by $\mathcal{F}_1$ and $\mathcal{F}_2$ the respective Borel sigma-field on $\mathbb{R}\times\mathcal{P}_p(\mathbb{R})$ and $\mathbb{R}\times\mathcal{P}_p(\mathbb{R})\times[0,1]$, we have that

$$(\mathbb{R}\times\mathcal{P}_p(\mathbb{R})\times[0,1],P\otimes\mathrm{Leb},\mathcal{F}_2,(\mathcal{F}_t)_{t\in\{1,2\}},(X_t)_{t\in\{1,2\}})\in\mathbf{M}(\mu,\nu).$$

Therefore, we also find $V_\Phi(\mu,\nu)\ge\hat V_C(\mu,\nu)$.

In the current setting, we derive the following analogue to Theorem 2.1.

Corollary 2.3. *Let $(\mu^k,\nu^k)_{k\in\mathbb{N}}$, $\mu^k\le_{cx}\nu^k$, be a convergent sequence in $\mathcal{P}_p(\mathbb{R})\times\mathcal{P}_p(\mathbb{R})$ with limit (μ,ν). Then, every $P\in\Lambda_M(\mu,\nu)$ is the $\mathcal{W}_p$-limit of a sequence $(P^k)_{k\in\mathbb{N}}$ with $P^k\in\Lambda_M(\mu^k,\nu^k)$.*

Remark 2.4. The adapted Wasserstein distance between two filtered processes $\mathbf{X}$ and $\mathbf{Y}=\left(\tilde\Omega,\mathcal{G},\mathbb{Q},(\mathcal{G}_t)_{t\in\{1,2\}},Y=(Y_t)_{t\in\{1,2\}}\right)$ is, by [5, Theorem 3.10], given by

$$\mathcal{AW}_p(\mathbf{X},\mathbf{Y})=\mathcal{W}_p(\mathrm{law}(X_1,\mathrm{law}(X_2|\mathcal{F}_1)),\mathrm{law}(Y_1,\mathrm{law}(Y_2|\mathcal{G}_1))).$$

Consequently, the map $\mathbf{X}\mapsto\mathrm{law}(X_1,\mathrm{law}(X_2|\mathcal{F}_1))$ is a surjective isometry from $\mathbf{M}(\mu,\nu)$ onto $\Lambda_M(\mu,\nu)$. Therefore, we may rephrase Corollary 2.3 using $\mathcal{AW}_p$, and obtain under the same assumptions that every process $\mathbf{X}\in\mathbf{M}(\mu,\nu)$ is the $\mathcal{AW}_p$-limit of a sequence of processes $(\mathbf{X}^k)_{k\in\mathbb{N}}$ with $\mathbf{X}^k\in\mathbf{M}(\mu^k,\nu^k)$.

Similar to Theorem 2.2 we get stability of (2.4).

Proposition 2.5. *Let $C\colon\mathbb{R}\times\mathcal{P}_p(\mathbb{R})\to\mathbb{R}$ be continuous and assume that there is a constant $K>0$ such that, for all $(x,\rho)\in\mathbb{R}\times\mathcal{P}_p(\mathbb{R})$,*

$$|C(x,\rho)|\le K\left(1+|x|^p+\int_{\mathbb{R}}|y|^p\rho(dy)\right).$$

Then the value $\hat V_C$ is attained and continuous on $\{(\mu,\nu)\in\mathcal{P}_p(\mathbb{R})\times\mathcal{P}_p(\mathbb{R}):\mu\le_{cx}\nu\}$. Moreover, if $(\mu^k,\nu^k)_{k\in\mathbb{N}}$, $\mu^k\le_{cx}\nu^k$ converges to (μ,ν) in $\mathcal{P}_p(\mathbb{R})\times\mathcal{P}_p(\mathbb{R})$ and $(P^k)_{k\in\mathbb{N}}$, $P^k\in\Lambda_M(\mu^k,\nu^k)$ is a sequence of optimizers of $\hat V_C(\mu^k,\nu^k)$, then its accumulation points are optimizers of $\hat V_C(\mu,\nu)$.

As in Remark 2.4, it is possible to phrase Proposition 2.5 in the language of filtered processes. Since the map $\mathbb{R} \times \mathcal{P}_p(\mathbb{R}) \ni (x, \rho) \mapsto \delta_x \otimes \rho \otimes \delta_\rho \in \mathcal{P}_p(\mathbb{R} \times \mathbb{R} \times \mathcal{P}_p(\mathbb{R}))$ is continuous, adequate continuity and growth assumptions on Φ will imply that $C(x, \rho) := (\delta_x \otimes \rho \otimes \delta_\rho)(\Phi)$ satisfies the assumptions of Proposition 2.5. Hence, we can deduce the following stability result for (2.3).

Corollary 2.6. *Let $\Phi \colon \mathbb{R} \times \mathbb{R} \times \mathcal{P}_p(\mathbb{R}) \to \mathbb{R}$ be continuous and assume that there is a constant $K > 0$ such that, for all $(x_1, x_2, \rho) \in \mathbb{R} \times \mathbb{R} \times \mathcal{P}_p(\mathbb{R})$,*

$$|\Phi(x_1, x_2, \rho)| \le K \left(1 + |x_1|^p + |x_2|^p + \int_{\mathbb{R}} |y|^p \rho(dy)\right).$$

Then the value V_Φ is attained and continuous on $\{(\mu, \nu) \in \mathcal{P}_p(\mathbb{R}) \times \mathcal{P}_p(\mathbb{R}) : \mu \le_{cx} \nu\}$.

2.3 American options

The robust pricing problem of American options as considered by Hobson and Norgilas [17, 18], can be cast in the setting of Subsection 2.2. Given a filtered process $\mathbf{X}$, the filtration $(\mathcal{F}_t)$ models the information that is available to the buyer, who may exercise at only two possible dates, $t \in \{1, 2\}$. For $t \in \{1, 2\}$, let $\Phi_t \colon \mathbb{R}^t \to \mathbb{R}$ be a path-dependent payoff that she receives when exercising at time t. The model-independent price of this American option is given by

$$\mathcal{A}m(\mu, \nu) = \sup_{\mathbf{X} \in \mathbf{M}(\mu, \nu)} \mathrm{price}(\Phi; \mathbf{X}). \tag{2.5}$$

As the buyer can exercise the option at any (stopping) time, the price crucially depends on the information that is available to the buyer and we have that the price of Φ is given by

$$\mathrm{price}(\Phi; \mathbf{X}) := \sup_{\tau\ (\mathcal{F}_t)\text{-stopping time}} \mathbb{E}_{\mathbb{P}} \left[\mathbb{1}_{\{\tau = 1\}} \Phi_1(X_1) + \mathbb{1}_{\{\tau = 2\}} \Phi_2(X) \right]. \tag{2.6}$$

In the case of a Put, that is ($\Phi_1(x) = (K_1 - x)^+$ and $\Phi_2(x, y) = (K_2 - y)^+$), Hobson and Norgilas [17] relate the above suprema to the left-curtain martingale coupling [8] when μ does not weight points. By the Snell-envelope theorem, we have that

$$\mathrm{price}(\Phi; \mathbf{X}) = \mathbb{E}\left[\max\left(\Phi_1(X_1), \mathbb{E}[\Phi_2(X_1, X_2) | \mathcal{F}_1]\right)\right],$$

which allows us to apply here Proposition 2.5 with $C(x, \rho) := \max(\Phi_1(x), \int \Phi_2(x, y)\, \rho(dy))$, and deduce the following stability result:

Corollary 2.7. *Let Φ_1 and Φ_2 be continuous and $\sup_{(x,y) \in \mathbb{R}^2} \left(\frac{\Phi_1(x)}{1 + |x|^p} + \frac{\Phi_2(x,y)}{1 + |x|^p + |y|^p} \right) < \infty$. Then the model independent price $\mathcal{A}m$ is continuous on $\{(\mu, \nu) \in \mathcal{P}_p(\mathbb{R})^2 : \mu \le_{cx} \nu\}$.*

2.4 VIX futures

The VIX is the implied volatility of the 30-day variance swap on the S&P 500. According to Guyon, Menegaux and Nutz [16], the subreplication price at time 0 for the VIX futures contract expiring at T_1 is given by

$$P_{\mathrm{sub}}(\mu, \nu) = \sup\{\mu(\phi) + \nu(\psi)\}, \tag{2.7}$$

where μ and ν denote the risk neutral distributions of the S&P 500 at dates T_1 and T_2 , where T_2 is equal to T_1 plus 30 days, both inferred from the market prices of liquid options. Moreover, the supremum is taken over all $(\phi, \psi) \in L^1(\mu) \times L^1(\nu)$ and measurable maps Δ^S, Δ^L such that, for all $(x, u, y) \in (0, \infty) \times [0, \infty) \times (0, \infty)$,

$$\phi(x) + \psi(y) + \Delta^S(x, v)(y - x) + \Delta^L(x, u)\left(\ell_x(y) - u^2\right) - u \le 0, \tag{2.8}$$

with $\ell_x(y) := \frac{2}{T_2-T_1}\ln(x/y)$. Up to assuming zero interest rates, the S&P 500 is a martingale under the risk neutral measure so that both, μ and ν, have finite first moments and μ is smaller than ν for the convex order. To state the dual problem, we define the set $\Pi_{\mathrm{VIX}}(\mu,\nu)$ of admissible martingale couplings as

$$\{\pi \in \mathcal{P}((0,\infty)\times[0,\infty)\times(0,\infty)) : \mathrm{proj}_1\,\pi = \mu,\ \mathrm{proj}_3\,\pi = \nu, \\ \mathrm{mean}(\pi_{x,u}) = x \text{ and } \pi_{x,u}(\ell_x) = u^2 \text{ for } \mathrm{proj}_{1,2}\,\pi\text{-a.e. } (x,u)\},$$

where $\pi_{x,u}(\ell_x)$ stands for $\int_{(0,\infty)} \ell_x(y)\pi_{x,u}(dy)$. Note that each $\pi \in \Pi_{\mathrm{VIX}}(\mu,\nu)$ satisfies $\pi \in \Pi_M(\mathrm{proj}_{1,2}\,\pi,\nu)$ and we have, by concavity of the logarithm function and Jensen's inequality, for $\mathrm{proj}_{1,2}\,\pi$-a.e. (x,u) that $\pi_{x,u}(\ell_x) \ge 0$. Given probability measures μ,ν on $(0,\infty)$ that are in convex order and finitely integrate $|\ln(x)|+|x|$, the dual problem D_{sub} consists of

$$D_{\mathrm{sub}}(\mu,\nu) = \inf_{\Pi_{\mathrm{VIX}}(\mu,\nu)} \int \sqrt{\pi_{x,u}(\ell_x)}\ \mathrm{proj}_{1,2}\,\pi(dx,du). \tag{2.9}$$

According to [16, Theorem 4.1], the values of $P_{\mathrm{sub}}(\mu,\nu)$ and $D_{\mathrm{sub}}(\mu,\nu)$ coincide. In the present paper, we are going to establish the following stability result with respect to the risk-neutral marginal distributions μ and ν of the S&P 500 at dates T_1 and T_2.

Proposition 2.8. *Let $(\mu^k,\nu^k)_{k\in\mathbb{N}}$, $\mu^k \le_{cx} \nu^k$, be a sequence in $\mathcal{P}((0,\infty))\times\mathcal{P}((0,\infty))$ that converges weakly to $(\mu,\nu)\in\mathcal{P}((0,\infty))\times\mathcal{P}((0,\infty))$. If $\lim_{k\to\infty}\int(|\ln(x)|+|x|)\ \nu^k(dx) = \int(|\ln(x)|+|x|)\ \nu(dx)$, then*

$$\lim_{k\to\infty} D_{\mathrm{sub}}(\mu^k,\nu^k) = D_{\mathrm{sub}}(\mu,\nu).$$

The analogous stability result for the superreplication price of the VIX futures contract is stated in [9, Theorem 1.3] and relies on the reduction of its dual formulation to the value function of a WMOT problem, see [16, Proposition 4.10]. Such a reduction step is, in general, not possible for the dual formulation of the subreplication price and we remark that with the approach in this paper, one can recover [9, Theorem 1.3] without recasting the problem as a WMOT problem.

2.5 Shadow couplings

The shadow couplings introduced by Beiglböck and Juillet in [10] fit into this framework. These couplings admit characterizations in terms of optimality, in terms of geometry of their support sets, and as barrier-type solutions to the Skorokhod embedding problem (c.f. [10, Theorem 1.1]). Let $\mathcal{U}=[0,1]$ and Leb denote the Lebesgue measure on this set. For $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ such that $\mu\le_{cx}\nu$, shadow couplings are solutions to weak martingale transport problems of the form

$$\inf_{\pi\in\Pi_M(\mu,\nu)} \int C_{\bar\mu}(x,\pi_x)\,\mu(dx), \tag{2.10}$$

where $\bar\mu\in\Pi(\mu,\mathrm{Leb})$ and $C_{\bar\mu}(x,\rho) := \inf_{\chi\in\Pi_M(\delta_x\times\bar\mu_x,\rho)}\int(1-u)\sqrt{1+y^2}\,\chi(dx',du,dy)$. The unique solution $\tilde\pi^\star$ to (2.10) is called a shadow coupling with source $\bar\mu$. Attached to each shadow coupling $\tilde\pi^\star$ with source $\bar\mu$ is the (unique) lifted shadow coupling $\pi^\star\in\Pi_M(\bar\mu,\nu)$ that satisfies $\mathrm{proj}_{1,3}\,\pi^\star = \tilde\pi^\star$ and minimizes

$$V_{\mathrm{SC}}(\bar\mu,\nu) := \inf_{\gamma\in\Pi_M(\bar\mu,\nu)} \int_{\mathbb{R}\times[0,1]\times\mathbb{R}} (1-u)\sqrt{1+y^2}\,\gamma(dx,du,dy). \tag{2.11}$$

To emphasize the dependence of $\pi^\star$ on $(\bar\mu,\nu)$, we denote by $SC(\bar\mu,\nu)=\pi^\star$ the unique optimizer of (2.11). According to [10, Theorem 1.1], there is a stochastic basis supporting an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B_t)_{t\geq 0}$ and an $\mathcal{F}_0$-measurable random variable U , uniformly distributed on $[0,1]$, so that $\pi^\star \sim (B_0, U, B_\tau)$ where τ is the hitting time of the process $(B_t, U)_{t\in[0,1]}$ into a left barrier, that is a Borel set $R \subseteq \mathbb{R}\times[0,1]$ such that $(x,u) \in R$, $v \leq u$ implies $(x,v)\in R$. Put differently, there exist two Borel measurable maps $T_1, T_2 \colon \mathbb{R}\times[0,1] \to \mathbb{R}$ satisfying

$$\forall x \in \mathbb{R},\ \forall 0 \leq v \leq u \leq 1,\ T_1(x,u) \leq T_1(x,v) \leq x \leq T_2(x,v) \leq T_2(x,u),$$

such that

$$\pi^\star(dx,du,dy) = \bar\mu(dx,du)\mathrm{Ber}(x, T_1(x,u), T_2(x,u))(dy), \tag{2.12}$$

where, for $x,l,r\in\mathbb{R}$, $\mathrm{Ber}(x,l,r)$ denotes the Bernoulli distribution

$$\mathrm{Ber}(x,l,r) := \begin{cases} \frac{r-x}{r-l}\delta_l + \frac{x-l}{r-l}\delta_r & l<x<r, \\ \delta_x & \text{else.}\end{cases}$$

W.l.o.g. we suppose that $T_1(x,u) = x = T_2(x,u)$ as soon as either $T_1(x,u) = x$ or $T_2(x,u) = x$.

Proposition 2.9. *The optimal value map V_{SC}, the selector SC of optimizers of* (2.11)*, and $J\circ SC$ are continuous on the domain $\{(\bar\mu,\nu) : \mathrm{proj}_1\,\bar\mu \leq_{cx} \nu, \mathrm{proj}_2\,\bar\mu = \mathrm{Leb}\} \subseteq \mathcal{P}_p(\mathbb{R}\times[0,1])\times\mathcal{P}_p(\mathbb{R})$ and with range $\mathbb{R}$, $\mathcal{P}_p(\mathbb{R}\times[0,1]\times\mathbb{R})$, and $\mathcal{P}_p(\mathbb{R}\times[0,1]\times\mathcal{P}_p(\mathbb{R}))$, respectively.*

Furthermore, when $(\bar\mu^k,\nu^k)_{k\in\mathbb{N}}$ with $\mathrm{proj}_1\,\bar\mu^k \leq_{cx} \nu^k$ and $\mathrm{proj}_2\,\bar\mu^k = \mathrm{Leb}$ is a sequence converging to $(\bar\mu,\nu)$ in $\mathcal{P}_p(\mathbb{R}\times\mathcal{U})\times\mathcal{P}_p(\mathbb{R})$ with $\bar\mu^k\to\bar\mu$ in total variation and (T_1^k, T_2^k) (resp. (T_1,T_2)) are the pairs of maps satisfying (2.12) *for $SC(\bar\mu^k,\nu^k)$ (resp. $SC(\bar\mu,\nu)$), then*

$$\begin{aligned} (T_1^k, T_2^k) &\to (T_1,T_2) \quad \textit{in } \bar\mu\textit{-probability on } \{T_1\neq T_2\}, \\ |T_1^k - T_1| \wedge |T_2^k - T_2| &\to 0 \quad \textit{in } \bar\mu\textit{-probability on } \{T_1 = T_2\}. \end{aligned}$$

3 Proofs

3.1 Topological refinements

In order to prove Proposition 2.8, we introduce refinements of the weak topology as detailed below, which we use to establish stronger versions of the results given in the introduction. For the rest of the paper, let $\mathcal{X}$ and $\mathcal{Y}$ be (non-empty) Polish subsets of $\mathbb{R}$ and consider two growth functions $\bar f\colon \mathcal{X}\times\mathcal{U}\to[1,+\infty)$ and $g\colon\mathcal{Y}\to[1,+\infty)$ that are both continuous and

$$\liminf_{\substack{|x|\to\infty \\ x\in\mathcal{X}}} \inf_{u\in\mathcal{U}} \frac{\bar f(x,u)}{|x|} > 0 \quad \text{and} \quad \liminf_{\substack{|y|\to\infty \\ y\in\mathcal{Y}}} \frac{g(y)}{|y|} > 0. \tag{3.1}$$

We define the sets $\mathcal{P}_{\bar f}(\mathcal{X}\times\mathcal{U}) := \{\rho\in\mathcal{P}(\mathcal{X}\times\mathcal{U}) : \rho(\bar f) < \infty\}$ and $\mathcal{P}_g(\mathcal{Y}) := \{\rho\in\mathcal{P}(\mathcal{Y}) : \rho(g)<\infty\}$ and endow them with the initial topology induced by $C_b(\mathcal{X}\times\mathcal{U})\cup\{\bar f\}$ resp. $C_b(\mathcal{Y})\cup\{g\}$, that is,

$$\begin{aligned} \rho^k\to\rho \text{ in } \mathcal{P}_{\bar f}(\mathcal{X}\times\mathcal{U}) &\iff \rho^k\to\rho \text{ weakly and } \rho^k(\bar f)\to\rho(\bar f), \\ \rho^k\to\rho \text{ in } \mathcal{P}_g(\mathcal{Y}) &\iff \rho^k\to\rho \text{ weakly and } \rho^k(g)\to\rho(g). \end{aligned}$$

Similarly, we define $\mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y}) = \{\rho\in\mathcal{P}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y}) : \rho(\bar f\oplus g)<\infty\}$ and $\mathcal{P}_{\bar f\oplus\hat g}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y})) := \{\rho\in\mathcal{P}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y})) : \rho(\bar f\oplus\hat g)<\infty\}$ where $(\bar f\oplus g)(x,u,y) := \bar f(x,u)+g(y)$ and $(\bar f\oplus\hat g)(x,u,\rho) = \bar f(x,u)+\rho(g)$. Again, these spaces are endowed with the topology induced by $C_b(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})\cup\{\bar f\oplus g\}$ resp. $C_b(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y}))\cup\{\bar f\oplus\hat g\}$, that is,

$$\rho^k\to\rho \text{ in } \mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y}) \iff \rho^k\to\rho \text{ weakly and } \rho^k(\bar f\oplus g)\to\rho(\bar f\oplus g),$$
$$\rho^k\to\rho \text{ in } \mathcal{P}_{\bar f\oplus \hat g}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y})) \iff \rho^k\to\rho \text{ weakly and } \rho^k(\bar f\oplus \hat g)\to\rho(\bar f\oplus \hat g).$$

Note that when $\mathcal{X}=\mathbb{R}=\mathcal{Y}$ and $\bar f(x,u)=1+|x|^p+d^p_{\mathcal{U}}(u_0,u)$ for some $u_0\in\mathcal{U}$ and $g(y)=1+|y|^p$, we have $\mathcal{P}_{\bar f}(\mathcal{X}\times\mathcal{U})=\mathcal{P}_p(\mathcal{X}\times\mathcal{Y})$, $\mathcal{P}_g(\mathcal{Y})=\mathcal{P}_p(\mathcal{Y})$, and the topologies on the above introduced spaces coincide with the corresponding p-Wasserstein topologies. Moreover, when $d_{\mathcal{U}}$ is bounded, the growth condition (3.1) provides that these topologies are finer than the corresponding 1-Wasserstein topology. The reader may ignore these refinements of the weak topology by substituting in every statement these refinements with a p-Wasserstein topology.

Next, we define the injection

$$\begin{aligned} J\colon \mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y}) &\to \mathcal{P}_{\bar f\oplus\hat g}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y})),\\ \pi &\mapsto (x,u,\pi_{x,u})_\#\pi, \end{aligned} \tag{3.2}$$

and observe that $\int C(x,u,\pi_{x,u})\,\mathrm{proj}_{\mathcal{X}\times\mathcal{U}}\,\pi(dx,du)=J(\pi)(C)$ for any $J(\pi)$-integrable $C\colon\mathcal{X}\times\mathcal{U}\times\mathcal{P}(\mathcal{Y})\to\mathbb{R}\cup\{\infty\}$. In our specific setting we treat the $\mathcal{X}$- and $\mathcal{U}$-coordinates similarly as we interpret the $\mathcal{X}$-coordinate as the spatial state (at time 1) and the $\mathcal{U}$-coordinate as the information state (at time 1), whereas we think of the $\mathcal{Y}$-coordinate as the state at time 2. For this reason, we say a sequence $(\pi^k)_{k\in\mathbb{N}}$ in $\mathcal{P}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})$ converges in the adapted weak topology to π if

$$J(\pi^k)\to J(\pi)\quad\text{in } \mathcal{P}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}(\mathcal{Y})). \tag{3.3}$$

The associated adapted p-Wasserstein distance of π^1 and π^2, where $\pi^1,\pi^2\in\mathcal{P}_p(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})$, is given by

$$\mathcal{AW}_p^p(\pi^1,\pi^2):=\inf_{\chi\in\Pi(\bar\mu^1,\bar\mu^2)}\int d^p_{\mathcal{X}}(x_1,x_2)+d^p_{\mathcal{U}}(u_1,u_2)+\mathcal{W}_p^p(\pi^1_{x_1,u_1},\pi^2_{x_2,u_2})\,d\chi(x_1,u_1,x_2,u_2), \tag{3.4}$$

where $\bar\mu^i=\mathrm{proj}_{1,2}\,\pi^i$, and satisfies $\mathcal{AW}_p^p(\pi^1,\pi^2)=\mathcal{W}_p^p(J(\pi^1),J(\pi^2))$ where $\mathcal{W}_p$ is the p-Wasserstein distance on $\mathcal{P}_p(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_p(\mathcal{Y}))$.

3.2 Approximation of extended martingale couplings: proof of Theorem 2.1

Before stating and proving a strengthened version of Theorem 2.1, let us deduce stability of the set of martingale couplings with respect to the marginals. The Hausdorff distance between two closed subsets $\mathcal{A},\mathcal{B}\subseteq\mathcal{P}_p(\mathbb{R}\times\mathcal{U}\times\mathbb{R})$ is denoted by $d_{\mathrm{H}}(\mathcal{A},\mathcal{B}):=\max\left(\sup_{a\in\mathcal{A}}\mathcal{W}_p(a,\mathcal{B}),\sup_{b\in\mathcal{B}}\mathcal{W}_p(b,\mathcal{A})\right)$ where $\mathcal{W}_p(a,\mathcal{B}):=\inf_{b\in\mathcal{B}}\mathcal{W}_p(a,b)$. Note that for $(\bar\mu,\nu)\in\mathcal{P}_p(\mathbb{R}\times\mathcal{U})\times\mathcal{P}_p(\mathbb{R})$, $\Pi_M(\bar\mu,\nu)$ is a compact subset of $\mathcal{P}_p(\mathbb{R}\times\mathcal{U}\times\mathbb{R})$.

Corollary 3.1. *Let $(\bar\mu^k,\nu^k)_{k\in\mathbb{N}}$ with $\mathrm{proj}_1\,\bar\mu^k\le_{cx}\nu^k$ be convergent in $\mathcal{P}_p(\mathbb{R}\times\mathcal{U})\times\mathcal{P}_p(\mathbb{R})$ to $(\bar\mu,\nu)$. Then*

$$\lim_{k\to\infty}d_{\mathrm{H}}\left(\Pi_M(\bar\mu^k,\nu^k),\Pi_M(\bar\mu,\nu)\right)=0.$$

The corresponding statement for couplings without the martingale constraint is straightforward to see as in this case one even has

$$d_{\mathrm{H}}\left(\Pi(\bar\mu,\nu),\Pi(\bar\mu',\nu')\right)\le\left(\mathcal{W}_p^p(\bar\mu,\bar\mu')+\mathcal{W}_p^p(\nu,\nu')\right)^{1/p}\le\mathcal{W}_p(\bar\mu,\bar\mu')+\mathcal{W}_p(\nu,\nu').$$

Proof of Corollary 3.1. Assume that $(\bar\mu^k,\nu^k)_{k\in\mathbb{N}}$ converge in $\mathcal{P}_p(\mathbb{R}\times\mathcal{U})\times\mathcal{P}_p(\mathbb{R})$ to $(\bar\mu,\nu)$. Note that $\bigcup_{k\in\mathbb{N}}\Pi_M(\bar\mu^k,\nu^k)$ is relatively compact as consequence of Prokhorov's theorem.

On the one hand, any subsequence of $(\pi^k)_{k\in\mathbb{N}}$ with $\pi^k \in \Pi_M(\bar\mu^k,\nu^k)$ admits a further subsequence $(\pi^{k_j})_{j\in\mathbb{N}}$ converging to some $\pi \in \Pi_M(\bar\mu,\nu)$ in $\mathcal{P}_p(\mathbb{R}\times\mathcal{U}\times\mathbb{R})$ so that $\lim_{j\to\infty}\mathcal{W}_p\left(\pi^{k_j},\Pi_M(\bar\mu,\nu)\right)=0$. Therefore,

$$\lim_{k\to\infty}\sup_{\pi^k\in\Pi(\bar\mu^k,\nu^k)}\mathcal{W}_p\left(\pi^k,\Pi_M(\bar\mu,\nu)\right)=0.$$

On the other hand, the map $\pi\mapsto\mathcal{W}_p(\pi,\Pi_M(\bar\mu^k,\nu^k))$ is $\mathcal{W}_p$-continuous. Thus, by compactness of the set of martingale couplings there is for every $k\in\mathbb{N}$, $\tilde\pi^k\in\Pi_M(\bar\mu,\nu)$ with $\mathcal{W}_p\left(\tilde\pi^k,\Pi_M(\bar\mu^k,\nu^k)\right)=\sup_{\pi\in\Pi_M(\bar\mu,\nu)}\mathcal{W}_p\left(\pi,\Pi_M(\bar\mu^k,\nu^k)\right)$. Again by compactness, any subsequence of $(\tilde\pi^k)_{k\in\mathbb{N}}$ admits a further subsequence $(\tilde\pi^{k_j})_{j\in\mathbb{N}}$ converging to some $\pi^\star\in\Pi_M(\bar\mu,\nu)$ in $\mathcal{P}_p(\mathbb{R}\times\mathcal{U}\times\mathbb{R})$. By Theorem 2.1 and convergence of the marginals in $\mathcal{P}_p(\mathbb{R}\times\mathcal{U})\times\mathcal{P}_p(\mathbb{R})$, there exist a sequence $(\pi^{k_j})_{j\in\mathbb{N}}$ with $\pi^{k_j}\in\Pi_M(\bar\mu^{k_j},\nu^{k_j})$ such that $\lim_{j\to\infty}\mathcal{W}_p(\pi^{k_j},\pi^\star)=0$. Since

$$\begin{aligned}\sup_{\pi\in\Pi_M(\bar\mu,\nu)}\mathcal{W}_p\left(\pi,\Pi_M(\bar\mu^{k_j},\nu^{k_j})\right)&=\mathcal{W}_p(\tilde\pi^{k_j},\Pi_M(\bar\mu^{k_j},\nu^{k_j}))\\&\le\mathcal{W}_p(\tilde\pi^{k_j},\pi^{k_j})\le\mathcal{W}_p(\tilde\pi^{k_j},\pi^\star)+\mathcal{W}_p(\pi^{k_j},\pi^\star),\end{aligned}$$

one has $\lim_{j\to\infty}\sup_{\pi\in\Pi_M(\bar\mu,\nu)}\mathcal{W}_p\left(\pi,\Pi_M(\bar\mu^{k_j},\nu^{k_j})\right)=0$. Consequently,

$$\lim_{k\to\infty}\sup_{\pi\in\Pi_M(\bar\mu,\nu)}\mathcal{W}_p\left(\pi,\Pi_M(\bar\mu^k,\nu^k)\right)=0. \qquad\square$$

We will prove the following strengthened version of Theorem 2.1 which takes into account general integrability conditions over Polish subsets of $\mathbb{R}$ and is, in fact, an extension of the main result in [7]. For $\mu\in\mathcal{P}(\mathcal{X})$ and $\nu\in\mathcal{P}_g(\mathcal{Y})$, $\mu\le_{cx}\nu$ means that the respective extensions $\mu(\cdot\cap\mathcal{X})$ and $\nu(\cdot\cap\mathcal{Y})$ of μ and ν to the Borel sigma-field on $\mathbb{R}$ satisfy $\mu(\cdot\cap\mathcal{X})\le_{cx}\nu(\cdot\cap\mathcal{Y})$.

Theorem 3.2. *Let $(\bar\mu^k,\nu^k)_{k\in\mathbb{N}}$, $\operatorname{proj}_1\bar\mu^k\le_{cx}\nu^k$, be a convergent sequence in $\mathcal{P}_{\bar f}(\mathcal{X}\times\mathcal{U})\times\mathcal{P}_g(\mathcal{Y})$ with limit $(\bar\mu,\nu)$. Then, every coupling $\pi\in\Pi_M(\bar\mu,\nu)$ is the limit in the adapted weak topology of a sequence $(\pi^k)_{k\in\mathbb{N}}$ with $\pi^k\in\Pi_M(\bar\mu^k,\nu^k)$.*

The proof of Theorem 3.2 relies on the next three auxiliary results, that are Lemma 3.3, Lemma 3.4, and Proposition 3.5.

Here we recall the notion of a pair of measures being *irreducible* and refer to [8, Appendix A] for further details. When $\mu\in\mathcal{M}_1(\mathbb{R})$, we denote by u_μ its potential function, that is the map defined by $u_\mu(y)=\int_{\mathbb{R}}|y-x|\,\mu(dx)$ for $y\in\mathbb{R}$. A pair (μ,ν) of finite positive measures in convex order is called irreducible if $I:=\{x\in\mathbb{R}:u_\mu(x)<u_\nu(x)\}$ is an interval and $\mu(I)=\mu(\mathbb{R})$. For any pair $(\mu,\nu)\in\mathcal{P}_1(\mathbb{R})^2$, $\mu\le_{cx}\nu$, there exists $N\subseteq\mathbb{N}$ and a sequence $(\mu_n,\nu_n)_{n\in N}$ of irreducible pairs of subprobability measures in convex order such that

$$\mu=\eta+\sum_{n\in N}\mu_n\quad\text{and}\quad\nu=\eta+\sum_{n\in N}\nu_n, \tag{3.5}$$

where $(\{u_{\mu_n}<u_{\nu_n}\})_{n\in N}$ are pairwise disjoint and $\eta=\mu|_{\{u_\mu=u_\nu\}}$, c.f. [8, Theorem A.4]. The sequence $(\mu_n,\nu_n)_{n\in N}$ is unique up to rearrangement of the pairs and is called the decomposition of (μ,ν) into irreducible components. The next lemma generalizes [7, Proposition 2.4] to the current setting where we have to keep track of the extra coordinate $u\in\mathcal{U}$.

Lemma 3.3. *Let $(\bar\mu^k,\nu^k)_{k\in\mathbb{N}}$, $\operatorname{proj}_1\bar\mu^k\le_{cx}\nu^k$, be a convergent sequence in $\mathcal{P}_1(\mathbb{R}\times\mathcal{U})\times\mathcal{P}_1(\mathbb{R})$ with limit $(\bar\mu,\nu)$ and write $\operatorname{proj}_1\bar\mu=:\mu$. Let $(\mu_n,\nu_n)_{n\in N}$ where $N\subseteq\mathbb{N}$ be the decomposition of (μ,ν) into irreducible components. Then, for every $k\in\mathbb{N}$, there exists* *a decomposition of* $(\bar\mu^k,\nu^k)$ *into pairs of subprobability measures* $(\bar\mu_n^k,\nu_n^k)_{n\in N}$*,* $(\bar\eta^k,\upsilon^k)$ *such that* $\operatorname{proj}_1\bar\mu_n^k\le_c\nu_n^k$ *for each* $n\in N$*,* $\operatorname{proj}_1\bar\eta^k\le_c\upsilon^k$ *and*

$$\bar\eta^k+\sum_{n\in N}\bar\mu_n^k=\bar\mu^k,\quad \upsilon^k+\sum_{n\in N}\nu_n^k=\nu^k,$$
$$\lim_{k\to\infty}\Big(\sum_{n\in N}\mathcal{W}_1(\bar\mu_n^k,\mu_n\times\bar\mu_x)+\mathcal{W}_1(\nu_n^k,\nu_n)\Big)+\mathcal{W}_1(\bar\eta^k,\eta\times\bar\mu_x)+\mathcal{W}_1(\upsilon^k,\eta)=0.$$

Proof. We set $\bar\eta(dx,du)=\eta(dx)\bar\mu_x(du)$ and $\bar\mu_n(dx,du)=\mu_n(dx)\bar\mu_x(du)$ for $n\in N$. Let us select $\mathcal{W}_1$-optimal couplings $\hat\pi^k\in\Pi(\bar\mu,\bar\mu^k)$ and define

$$\bar\mu_n^k(dx,du):=\int_{\hat x,\hat u}\hat\pi^k_{\hat x,\hat u}(dx,du)\,\bar\mu_n(d\hat x,d\hat u),\quad \bar\eta^k(dx,du):=\int_{\hat x,\hat u}\hat\pi^k_{\hat x,\hat u}(dx,du)\,\bar\eta(d\hat x,d\hat u).$$

We have

$$W_1(\bar\eta^k,\bar\eta)+\sum_{n\in N}\mathcal{W}_1(\bar\mu_n^k,\bar\mu_n)\le\mathcal{W}_1(\bar\mu^k,\bar\mu)\underset{k\to+\infty}{\longrightarrow}0.$$

Pick any $\bar\pi^k\in\bar\Pi_M(\bar\mu^k,\nu^k)$, set

$$\nu_n^k(dy):=\int_{x,u}\bar\pi^k_{x,u}(dy)\,\bar\mu_n^k(dx,du),\quad \upsilon^k(dy):=\int_{x,u}\bar\pi^k_{x,u}(dy)\,\bar\eta^k(dx,du),$$
$$\bar\pi_n^k(dx,du,dy):=\bar\mu_n^k(dx,du)\bar\pi^k_{x,u}(dy)\in\Pi_M(\bar\mu_n^k,\nu_n^k).$$

Let $n\in N$. Since $(\nu^k=\upsilon^k+\sum_{n\in N}\nu_n^k)_{k\in\mathbb{N}}$ converges to ν in $\mathcal{W}_1$, we have tightness of $(\nu_n^k)_{k\in\mathbb{N}}$. As the marginals are tight, we can pass by Prokhorov's theorem to a subsequence and assume that, as $k\to\infty$, $(\bar\pi_n^k)_{k\in\mathbb{N}}$ converges weakly to $\tilde\pi_n\in\Pi_M(\bar\mu_n,\tilde\nu_n)$ where $\tilde\nu_n\in\mathcal{M}_1(\mathbb{R})$. At the same time, as $(\bar\mu^k)_{k\in\mathbb{N}}$ and $(\nu^k)_{k\in\mathbb{N}}$ are $\mathcal{W}_1$-convergent, by passing to a subsequence we can additionally assume convergence of $(\bar\pi^k)_{k\in\mathbb{N}}$ in $\mathcal{W}_1$ to some $\bar\pi\in\Pi_M(\bar\mu,\nu)$. Since $\bar\pi_n^k\le\bar\pi^k$ for each $k\in\mathbb{N}$, passing to weak limits, we have $\tilde\pi_n\le\bar\pi$, which, in view of $\tilde\pi_n(dx,du,\mathbb{R})=\bar\mu_n(dx,du)$, implies $\tilde\pi_n(dx,du,dy)=\bar\mu_n(dx,du)\bar\pi_{x,u}(dy)$ so that $\tilde\nu_n=\nu_n$. By Lemma A.4, we get that $\mathcal{W}_1(\bar\pi_n^k,\tilde\pi_n)$ goes to 0 as $k\to\infty$. Hence, $\mathcal{W}_1(\nu_n^k,\nu_n)$ also goes to 0. The same argument applies to deal with $\mathcal{W}_1(\upsilon^k,\eta)$. □

In order to show Theorem 3.2, it turns out to be beneficial to first demonstrate that a family of couplings with a simpler structure is already dense. We say that a coupling $\pi\in\Pi_M(\bar\mu,\nu)$ is simple if there is $J\in\mathbb{N}$, a measurable partition $(\mathcal{U}_j)_{j=1}^J$ of $\mathcal{U}$ into $\operatorname{proj}_2\bar\mu$-continuity sets and, for $j\in\{1,\dots,J\}$, a martingale kernel $(K_{j,x})_{x\in\mathbb{R}}$ such that

$$\pi(dx,du,dy)=\sum_{j=1}^J\mathbb{1}_{\mathcal{U}_j}(u)\bar\mu(dx,du)\,K_{j,x}(dy).\tag{3.6}$$

Put differently, one may say π is simple if there exist (classical) martingale couplings $\pi^j\in\Pi_M(\mu,\nu_j)$, $j\in\{1,\dots,J\}$, and a measurable partition $(\mathcal{U}_j)_{j=1}^J$ of $\mathcal{U}$ in $\operatorname{proj}_2\bar\mu$-continuity sets such that

$$\pi(dx,du,dy)=\sum_{j=1}^J\pi^j(dx,dy)\bar\mu_x(du\cap\mathcal{U}_j).$$

The next lemma establishes that these simple couplings are already dense in $\Pi_M(\bar\mu,\nu)$.

Lemma 3.4. *Let* $\bar\mu\in\mathcal{P}_1(\mathbb{R}\times\mathcal{U})$ *and* $\nu\in\mathcal{P}_1(\mathbb{R})$*. Then, the set of couplings satisfying* (3.6) *is dense in* $\Pi_M(\bar\mu,\nu)$ *w.r.t. the adapted weak topology.*

Proof. We denote by $\lambda = \mathrm{proj}_2\, \bar\mu \in \mathcal{P}_1(\mathcal{U})$. Let $u_0 \in \mathcal{U}$ and $\varepsilon > 0$. We claim that there is a finite partition $(\mathcal{U}_j)_{j=1}^J$, $J \in \mathbb{N}$, of $\mathcal{U}$ into λ-continuity sets such that

$$\sup\{d_{\mathcal{U}}(u,\hat u)\colon u,\hat u \in \mathcal{U}_j\} \le \frac{\epsilon}{2} \text{ for } j \in \{1,\dots,J-1\}, \text{ and } \int_{\mathcal{U}_J} d_{\mathcal{U}}(u,u_0)\,\lambda(du) \le \frac{\epsilon}{4}. \quad (3.7)$$

To this end, note that since the map $u \mapsto d_{\mathcal{U}}(u,u_0)$ is integrable w.r.t. λ, we can choose $M_\epsilon \in [0,+\infty)$ with

$$\int_{\{u\in\mathcal{U}\colon\, d_{\mathcal{U}}(u,u_0)>M_\varepsilon\}} d_{\mathcal{U}}(u,u_0)\,\lambda(du) \le \frac{\varepsilon}{8}.$$

By inner regularity of λ there exists a compact subset $\mathcal{K}$ of $A := \{u \in \mathcal{U}\colon d_{\mathcal{U}}(u,u_0) \le M_\epsilon\}$ such that $\lambda(A\setminus\mathcal{K}) \le \frac{\varepsilon}{8M_\varepsilon}$. Therefore, we have

$$\int_{A\setminus\mathcal{K}} d_{\mathcal{U}}(u,u_0)\,\lambda(du) \le \lambda(A\setminus\mathcal{K})M_\epsilon \le \frac{\epsilon}{8}.$$

Next, we choose for each $u \in \mathcal{K}$ a radius $r_u \in (0,\frac{\epsilon}{4}]$ such that the boundary of the ball $B_{r_u}(u) := \{\hat u \in \mathcal{U}\colon d_{\mathcal{U}}(u,\hat u) < r_u\}$ has zero measure under λ. The family $(B_{r_u}(u))_{u\in\mathcal{K}}$ is an open cover of the compact set $\mathcal{K}$, which permits us to extract from this family a finite subcover of $\mathcal{K}$ denoted by $(A_j)_{j=1}^I$, $I \in \mathbb{N}$. Let $J := I+1$, $\mathcal{U}_J := \bigcap_{j=1}^I A_j^c \subset \mathcal{K}^c$, and set recursively, $\mathcal{U}_j := A_j \cap \left\{\bigcup_{i=1}^{j-1} A_i\right\}^c$ for $j \in \{1,\dots,J-1\}$. By this procedure we have constructed a partition $(\mathcal{U}_j)_{j=1}^J$ of $\mathcal{U}$ into measurable sets. Moreover, as for each $i \in \{1,\dots,J\}$ the boundary of $\mathcal{U}_i$ is contained in the union of the boundaries of the balls $(A_j)_{j=1}^J$, it must have zero λ-measure. Finally, for each $j \in \{1,\dots,J-1\}$ we get

$$\sup\{d_{\mathcal{U}}(u,\hat u)\colon u,\hat u \in \mathcal{U}_j\} \le \sup\{d_{\mathcal{U}}(u,\hat u)\colon u,\hat u \in A_j\} \le \frac{\epsilon}{2},$$

and compute

$$\int_{\mathcal{U}_J} d_{\mathcal{U}}(u,u_0)\,\lambda(du) \le \int_{A^c} d_{\mathcal{U}}(u,u_0)\,\lambda(du) + \int_{A\setminus\mathcal{K}} d_{\mathcal{U}}(u,u_0)\,\lambda(du) \le \frac{\epsilon}{8} + \frac{\epsilon}{8} = \frac{\epsilon}{4}.$$

We have shown the claim (3.7).

Let $\bar\pi \in \Pi_M(\bar\mu,\nu)$. Consider the disintegration kernel $\bar\mu_x$ such that $\bar\mu(dx,du) = \mu(dx)\bar\mu_x(du)$, and, for $j \in \{1,\dots,J\}$, denote by $\bar\mu_j(dx,du)$ the restrictions $\mathbb{1}_{\mathbb{R}\times\mathcal{U}_j}(x,u)\bar\mu(dx,du)$. Since $\bar\mu_j(dx,du)$-a.e. $\bar\mu_x(\mathcal{U}_j) > 0$ we can define

$$\bar\pi^J(dx,du,dy) := \sum_{j=1}^J \bar\mu_j(dx,du)K_{j,x}(dy) \text{ where } K_{j,x}(dy) := \int_{\mathcal{U}_j} \bar\pi_{x,u}(dy)\,\frac{\bar\mu_x(du)}{\bar\mu_x(\mathcal{U}_j)},$$

and remark that $\bar\pi^J \in \Pi_M(\bar\mu,\nu)$. The last step is to estimate the $\mathcal{AW}_1$-distance between $\bar\pi$ and $\bar\pi^J$.

Let Id denote the identity function. Using that $(\mathrm{Id},\mathrm{Id})_\#\bar\mu$ is an admissible coupling in the minimization problem that constitutes the $\mathcal{AW}_1$-distance between $\bar\pi$ and $\bar\pi^J$, and Jensen's inequality (see for example [9, Proposition A.9]), we obtain the estimate

$$\begin{aligned}
\mathcal{AW}_1(\bar\pi,\bar\pi^J) &\le \sum_{j=1}^J \int_{\mathbb{R}\times\mathcal{U}_j} \mathcal{W}_1(\bar\pi_{x,u},K_{j,x})\,\bar\mu(dx,du)\\
&\le \sum_{j=1}^J \int_{\mathbb{R}\times\mathcal{U}_j}\int_{\mathcal{U}_j} \mathcal{W}_1(\bar\pi_{x,u},\bar\pi_{x,\hat u})\,\frac{\bar\mu_x(d\hat u)}{\bar\mu_x(\mathcal{U}_j)}\,\bar\mu(dx,du)\\
&= \int \mathcal{W}_1(\bar\pi_{x,u},\bar\pi_{\hat u,\hat x})\,\gamma(dx,du,d\hat x,d\hat u)
\end{aligned}$$

where $\gamma := \sum_{j=1}^J \gamma_j$ with the subprobability couplings γ_j defined for $j \in \{1, \dots, J\}$ by

$$\gamma_j := ((x, u, \hat{u}) \mapsto (x, u, x, \hat{u}))_\# \left(\frac{\bar{\mu}_x(d\hat{u})}{\bar{\mu}_x(\mathcal{U}_j)} \bar{\mu}_j(dx, du) \right) \in \Pi(\bar{\mu}_j, \bar{\mu}_j).$$

We have $\gamma \in \Pi(\bar{\mu}, \bar{\mu})$ and, by (3.7),

$$\begin{aligned}
\int d_{\mathcal{U}}(u, \hat{u}) + |x - \hat{x}| \, \gamma(dx, du, d\hat{x}, d\hat{u}) &= \sum_{j=1}^J \int d_{\mathcal{U}}(u, \hat{u}) \, \gamma_j(dx, du, d\hat{x}, d\hat{u}) \\
&\leq \frac{\epsilon}{2} \sum_{j=1}^{J-1} \gamma_j(\mathbb{R} \times \mathcal{U} \times \mathbb{R} \times \mathcal{U}) + \int d_{\mathcal{U}}(u, u_0) + d_{\mathcal{U}}(u_0, \hat{u}) \, \gamma_J(dx, du, d\hat{x}, d\hat{u}) \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon. \qquad (3.8)
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, Lemma A.1 gives the conclusion. □

Finally, we also require the following approximation result.

Proposition 3.5. *Let $(\mu^k, \nu^k)_{k \in \mathbb{N}}$, $\mu^k \leq_{cx} \nu^k$, be a sequence in $\mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R})$ with limit (μ, ν) being irreducible. For $1 \leq j \leq J \in \mathbb{N}$, let $(\mu_j^k)_{k \in \mathbb{N}}$ be a convergent sequence in $\mathcal{M}_1(\mathbb{R})$ with limit μ_j and $\sum_{j=1}^N \mu_j^k = \mu^k$. Let $(\nu_j)_{j=1}^J$, $\mu_j \leq_{cx} \nu_j$, be a family in $\mathcal{M}_1^*(\mathbb{R})$ such that $\nu = \sum_{j=1}^J \nu_j$. Then, for $1 \leq j \leq J$, there exist a convergent sequence $(\nu_j^k)_{k \in \mathbb{N}}$ in $\mathcal{M}_1(\mathbb{R})$ with limit ν_j such that*

$$\mu_j^k \leq_c \nu_j^k \quad \text{and} \quad \sum_{j=1}^J \nu_j^k = \nu^k. \qquad (3.9)$$

The proof of Proposition 3.5 is rather technical and therefore postponed to Subsection 3.7. On closer inspection of the statement, this technicality is not completely surprising: in the setting of Proposition 3.5, let $(\mu_j)_{j=1}^J$ and $(\mu_j^k)_{j=1}^J$ be families of measures with $\mu_j(\{x_j\}) = \mu_j(\mathbb{R})$ and $\mu_j^k(\{x_j^k\}) = \mu_j^k(\mathbb{R})$ for some $x_j, x_j^k \in \mathbb{R}$ so that the points $(x_j)_{j=1}^J$ are distinct. For $\pi \in \Pi_M(\mu, \nu)$, we define $\nu_j := \pi_{x_j}$. Invoking Proposition 3.5 we obtain $(\nu_j^k)_{j=1}^J$ and set $\pi^k := \sum_{j=1}^J \mu_j^k \otimes \nu_j^k$. Since μ_j^k is concentrated on a single point and $\mu_j^k \leq_{cx} \nu_j^k$, π^k defines a martingale coupling in $\Pi_M(\mu^k, \nu^k)$ and, as $\nu_j^k \to \nu_j$ and $\mu_j^k \to \mu_j$ in $\mathcal{M}_1(\mathbb{R})$, $(\pi^k)_{k \in \mathbb{N}}$ converges in $\mathcal{AW}_1$ to π. Hence, we recover in this particular setting the main result of [7], which states that, as long as $\mu^k \leq_{cx} \nu^k$, $\mu^k \to \mu$, $\nu^k \to \nu$ in $\mathcal{W}_1$, any martingale coupling in $\Pi_M(\mu, \nu)$ can be approximated in $\mathcal{AW}_1$ by a sequence $(\pi^k)_{k \in \mathbb{N}}$ with $\pi^k \in \Pi_M(\mu^k, \nu^k)$.

Proof of Theorem 3.2. By following the reasoning outlined in [7, Lemma 5.2], incorporating the additional coordinate and replacing [7, Proposition 2.5] by Lemma 3.3, one can confirm that it suffices to establish the conclusion when $(\bar{\mu}, \nu)$ is such that $(\mathrm{proj}_1 \bar{\mu}, \nu)$ is irreducible. As the argument runs almost verbatim to the proof of [7, Lemma 5.2], we omit the details and assume from now on that $(\mathrm{proj}_1 \bar{\mu}, \nu)$ is irreducible.

Let us suppose that $d_{\mathcal{U}}$ denotes some bounded complete metric compatible with the topology on $\mathcal{U}$ and check that we may suppose w.l.o.g. that $\mathcal{X} = \mathbb{R} = \mathcal{Y}$, $\bar{f}(x, u) = 1 + |x| + d_{\mathcal{U}}(u, u_0)$ for some $u_0 \in \mathcal{U}$ and $g(y) = 1 + |y|$. The convergence of $(\bar{\mu}^k)_k$ (resp. $(\nu^k)_k$) to $\bar{\mu}$ in $\mathcal{P}_{\bar{f}}(\mathcal{X} \times \mathcal{U})$ (resp. ν in $\mathcal{P}_g(\mathcal{Y})$) implies that $(\bar{\mu}^k(\cdot \cap \mathcal{X} \times \mathcal{U}))_k$ (resp. $(\nu^k(\cdot \cap \mathcal{Y}))_k$) converges to $\bar{\mu}(\cdot \cap \mathcal{X} \times \mathcal{U})$ in $\mathcal{P}_1(\mathbb{R} \times \mathcal{U})$ (resp. $\nu(\cdot \cap \mathcal{Y})$ in $\mathcal{P}_1(\mathbb{R})$). We set $\tilde{\pi}(\cdot) = \pi(\cdot \cap \mathcal{X} \times \mathcal{U} \times \mathcal{Y}) \in \Pi_M(\bar{\mu}(\cdot \cap \mathcal{X} \times \mathcal{U}), \nu(\cdot \cap \mathcal{Y}))$. Let $(\tilde{\pi}^k)_k$ be a sequence such that $\tilde{\pi}^k \in \Pi(\bar{\mu}^k(\cdot \cap \mathcal{X} \times \mathcal{U}), \nu^k(\cdot \cap \mathcal{Y}))$ and $(\tilde{J}(\tilde{\pi}^k) = (x, u, \tilde{\pi}_{x,u}^k)_\# \tilde{\pi}^k)_k$ converges to

$\tilde{J}(\tilde{\pi}) = (x, u, \tilde{\pi}_{x,u})_{\#}\tilde{\pi}$ in $\mathcal{P}_1(\mathbb{R} \times \mathcal{U} \times \mathcal{P}_1(\mathbb{R}))$. Since $\mathcal{X}$, $\mathcal{U}$ and $\mathcal{Y}$ are Polish, the Borel sigma-fields satisfy

$$\begin{aligned}\mathcal{B}(\mathcal{X} \times \mathcal{U} \times \mathcal{Y}) &= \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{Y}) \\ &= \sigma\left(\{\{A \cap \mathcal{X}\} \times B \times \{C \cap \mathcal{Y}\} :\ A, C \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}(\mathcal{U})\}\right) \\ &= \{D \cap \mathcal{X} \times \mathcal{U} \times \mathcal{Y} : D \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathbb{R})\} \\ &= \{D \cap \mathcal{X} \times \mathcal{U} \times \mathcal{Y} : D \in \mathcal{B}(\mathbb{R} \times \mathcal{U} \times \mathbb{R})\}.\end{aligned}$$

By Alexandrov's theorem, $\mathcal{X}$ and $\mathcal{Y}$ are countable intersections of open subsets of $\mathbb{R}$. Hence $\mathcal{X}, \mathcal{Y} \in \mathcal{B}(\mathbb{R})$ and $\mathcal{X} \times \mathcal{U} \times \mathcal{Y} \in \mathcal{B}(\mathbb{R} \times \mathcal{U} \times \mathbb{R})$ so that $\mathcal{B}(\mathcal{X}) \subset \mathcal{B}(\mathbb{R})$, $\mathcal{B}(\mathcal{X}) \subset \mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\mathcal{X} \times \mathcal{U} \times \mathcal{Y}) \subset \mathcal{B}(\mathbb{R} \times \mathcal{U} \times \mathbb{R})$. Let $\pi^k \in \Pi(\bar{\mu}^k, \nu^k)$ be defined by $\pi^k = \tilde{\pi}^k|_{\mathcal{B}(\mathcal{X} \times \mathcal{U} \times \mathcal{Y})}$. By [9, Lemma A.7], the sequence $(J(\pi^k))_k$ is relatively compact in $\mathcal{P}_{\bar{f} \oplus \hat{g}}(\mathcal{X} \times \mathcal{U} \times \mathcal{P}_g(\mathcal{Y}))$. Let $(J(\pi^{k_j}))_j$ denote some subsequence converging to Q. Since the injection $i : \mathcal{X} \times \mathcal{U} \times \mathcal{P}(\mathcal{Y}) \ni (x, u, \rho) \mapsto (x, u, \rho(\cdot \cap \mathcal{Y})) \in \mathbb{R} \times \mathcal{U} \times \mathcal{P}(\mathbb{R})$ is continuous, $i_{\#}J(\pi^k) = \tilde{J}(\tilde{\pi}^k)$ and $i_{\#}J(\pi) = \tilde{J}(\tilde{\pi})$, we have for any continuous and bounded function φ on $\mathbb{R} \times \mathcal{U} \times \mathcal{P}(\mathbb{R})$,

$$Q(\varphi \circ i) = \lim_{j \to \infty} J(\pi^{k_j})(\varphi \circ i) = \lim_{j \to \infty} \tilde{J}(\tilde{\pi}^{k_j})(\varphi) = \tilde{J}(\tilde{\pi})(\varphi) = J(\pi)(\varphi \circ i).$$

The equality between the left-most and right-most terms remains valid when φ is measurable and bounded. Let $A \in \mathcal{B}(\mathcal{X})$, $B \in \mathcal{B}(\mathcal{U})$ and $C \in \mathcal{B}(\mathcal{P}(\mathcal{Y}))$. Since $\mathcal{P}(\mathbb{R}) \ni \rho \mapsto \rho(\mathcal{Y})$ is measurable, $\mathcal{P}_{\mathcal{Y}}(\mathbb{R}) = \{\eta \in \mathcal{P}(\mathbb{R}) : \eta(\mathcal{Y}) = 1\}$ is a Borel subset of $\mathcal{P}(\mathbb{R})$. Since $\mathcal{P}(\mathcal{Y}) \ni \rho \mapsto \rho(\cdot \cap \mathcal{Y}) \in \mathcal{P}_{\mathcal{Y}}(\mathbb{R})$ is an homeomorphism with inverse $r : \mathcal{P}_{\mathcal{Y}}(\mathbb{R}) \ni \rho \mapsto \rho|_{\mathcal{B}(\mathcal{Y})} \in \mathcal{P}(\mathcal{Y})$ when $\mathcal{P}_{\mathcal{Y}}(\mathbb{R})$ is endowed with the topology induced by $\mathcal{P}(\mathbb{R})$, $r^{-1}(C) \in \mathcal{B}(\mathcal{P}_{\mathcal{Y}}(\mathbb{R})) \subset \mathcal{B}(\mathcal{P}(\mathbb{R}))$. For the choice $\varphi = \mathbb{1}_{A \times B \times r^{-1}(C)}$, we deduce that $Q(A \times B \times C) = J(\pi)(A \times B \times C)$. Since $\mathcal{B}(\mathcal{X} \times \mathcal{U} \times \mathcal{P}(\mathcal{Y})) = \mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{P}(\mathcal{Y}))$, this implies that $Q = J(\pi)$. Therefore the sequence $(J(\pi^k))_k$ converges to $J(\pi)$ in $\mathcal{P}_{\bar{f} \oplus \hat{g}}(\mathcal{X} \times \mathcal{U} \times \mathcal{P}_g(\mathcal{Y}))$.

Therefore, we assume from now on that $\mathcal{X} = \mathbb{R} = \mathcal{Y}$, $\bar{f}(x, u) = 1 + |x| + d_{\mathcal{U}}(u, u_0)$ and $g(y) = 1 + |y|$.

Moreover, by using Lemma 3.4 we may assume that π admits the representation (3.6). Let $(\mathcal{U}_j)_{j=1}^J$ be the associated finite measurable partition of $\mathcal{U}$. Without loss of generality, e.g. by replacing one element of the partition $\mathcal{U}_k$ such that $\bar{\mu}(\mathbb{R} \times \mathcal{U}_k) > 0$ with the union of $\mathcal{U}_k$ with all elements $\mathcal{U}_j$ that satisfy $\bar{\mu}(\mathbb{R} \times \mathcal{U}_j) = 0$ and removing the latter, we can assume that $\min_{1 \le j \le J} \bar{\mu}(\mathbb{R} \times \mathcal{U}_j) > 0$. For $j \in \{1, \dots, J\}$ and $k \in \mathbb{N}$, we define

$$\bar{\mu}_j := \mathbb{1}_{\mathbb{R} \times \mathcal{U}_j} \bar{\mu}, \quad \bar{\mu}_j^k := \mathbb{1}_{\mathbb{R} \times \mathcal{U}_j} \bar{\mu}^k, \quad \mu_j := \operatorname{proj}_1 \bar{\mu}_j \quad \text{and} \quad \mu_j^k := \operatorname{proj}_1 \bar{\mu}_j^k.$$

As $(\mathcal{U}_j)_{j=1}^J$ is comprised of continuity sets for the first marginal of $\bar{\mu}$, the weak convergence of $(\bar{\mu}^k)_{k \in \mathbb{N}}$ to $\bar{\mu}$ implies that $(\bar{\mu}_j^k)_{k \in \mathbb{N}}$ converges weakly to $\bar{\mu}_j$ and, due to the continuity of the first coordinate mapping, $(\mu_j^k)_{k \in \mathbb{N}}$ converges weakly to μ_j for each $j \in \{1, \cdots, J\}$. All the requirements of Proposition 3.5 are satisfied, allowing us to identify, for each $j \in \{1, \dots, J\}$, a sequence of subprobability measures $(\nu_j^k)_{k \in \mathbb{N}}$ such that

$$\nu_j^k \underset{k \to +\infty}{\longrightarrow} \nu_j, \quad \mu_j^k \le_c \nu_j^k \quad \text{and} \quad \sum_{j=1}^J \nu_j^k = \nu^k.$$

From now on we will assume that k is large enough so that $\min_{1 \le j \le J} \mu_j^k(\mathbb{R}) > 0$. Weak convergence of the original sequences yields, for each $j \in \{1, \cdots, J\}$, that the normalized sequence $(\bar{\mu}_j^k / \mu_j^k(\mathbb{R}))_{k \in \mathbb{N}}$ (resp. $(\nu_j^k / \mu_j^k(\mathbb{R}))_{k \in \mathbb{N}}$) converges weakly to $\bar{\mu}_j / \mu_j(\mathbb{R})$ (resp. $\nu_j / \mu_j(\mathbb{R})$) as $k \to \infty$. As $(\bar{\mu}^k)_{k \in \mathbb{N}}$ and $(\nu^k)_{k \in \mathbb{N}}$ are $\mathcal{W}_1$-convergent sequences, it then follows easily from Lemmas A.3 and A.4 that the normalized sequences converge in $\mathcal{W}_1$. Thus, we can apply [7, Theorem 2.6] and obtain an $\mathcal{AW}_1$-convergent sequence $(\gamma_j^k)_{k \in \mathbb{N}}$

of martingale couplings with limit γ_j where

$$\gamma_j := \frac{\mu_j \otimes K_{j,x}}{\mu_j(\mathbb{R})} \in \Pi_M\left(\frac{\mu_j}{\mu_j(\mathbb{R})}, \frac{\nu_j}{\mu_j(\mathbb{R})}\right) \quad \text{and} \quad \gamma_j^k \in \Pi_M\left(\frac{\mu_j^k}{\mu_j^k(\mathbb{R})}, \frac{\nu_j^k}{\mu_j^k(\mathbb{R})}\right),$$

where we write $\eta \otimes K(dx,dy) := \eta(dx)K_x(dy)$ for the gluing of a measure $\eta \in \mathcal{P}(\mathcal{X})$ with a measurable kernel $K\colon \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. Further, write

$$\bar{\gamma}_j^k := \frac{\bar{\mu}_j^k \otimes \gamma_{j,x}^k}{\mu_j^k(\mathbb{R})} \quad \text{and} \quad \bar{\gamma}_j := \frac{\mathbb{1}_{\mathbb{R}\times\mathcal{U}_j\times\mathbb{R}}\pi}{\mu_j(\mathbb{R})} = \frac{\bar{\mu}_j \otimes K_{j,x}}{\mu_j(\mathbb{R})}.$$

To prove that $(\bar{\gamma}_j^k)_{k\in\mathbb{N}}$ converges in $\mathcal{AW}_1$ to $\bar{\gamma}_j$, we choose

$$\chi_j^k \in \Pi\left(\frac{\bar{\mu}_j^k}{\mu_j^k(\mathbb{R})}, \frac{\bar{\mu}_j}{\mu_j(\mathbb{R})}\right) \quad \text{resp.} \quad \hat{\chi}_j^k \in \Pi\left(\frac{\mu_j^k}{\mu_j^k(\mathbb{R})}, \frac{\mu_j}{\mu_j(\mathbb{R})}\right)$$

that are $\mathcal{W}_1$-optimal between their marginals resp. optimal for $\mathcal{AW}_1(\gamma_j^k, \gamma_j)$. For ease of notation, we moreover define $\check{\chi}_j^k$ as $\check{\chi}_j^k(dx,du,d\hat{x},d\hat{u},dz) := \chi_j^k(dx,du,d\hat{x},d\hat{u})\hat{\chi}_{j,x}^k(dz)$ and compute

$$\begin{aligned}
\mathcal{AW}_1(\bar{\gamma}_j^k, \bar{\gamma}_j) &\le \mathcal{W}_1\left(\frac{\bar{\mu}_j^k}{\mu_j^k(\mathbb{R})}, \frac{\bar{\mu}_j}{\mu_j(\mathbb{R})}\right) + \int \mathcal{W}_1(\gamma_{j,x}^k, K_{j,\hat{x}})\, \chi_j^k(dx,du,d\hat{x},d\hat{u}) \\
&\le \mathcal{W}_1\left(\frac{\bar{\mu}_j^k}{\mu_j^k(\mathbb{R})}, \frac{\bar{\mu}_j}{\mu_j(\mathbb{R})}\right) + \int \left(\mathcal{W}_1(\gamma_{j,x}^k, K_{j,z}) + \mathcal{W}_1(K_{j,z}, K_{j,\hat{x}})\right)\, \check{\chi}_j^k(dx,du,d\hat{x},d\hat{u},dz) \\
&\le \mathcal{W}_1\left(\frac{\bar{\mu}_j^k}{\mu_j^k(\mathbb{R})}, \frac{\bar{\mu}_j}{\mu_j(\mathbb{R})}\right) + \mathcal{AW}_1(\gamma_j^k, \gamma_j) + \int \mathcal{W}_1(K_{j,z}, K_{j,\hat{x}})\, \check{\chi}_j^k(dx,du,d\hat{x},d\hat{u},dz). \quad (3.10)
\end{aligned}$$

By Lemma A.1 and since

$$\int |z-\hat{x}|\check{\chi}_j^k(dx,du,d\hat{x},d\hat{u},dz) \le \mathcal{W}_1\left(\frac{\bar{\mu}_j^k}{\mu_j^k(\mathbb{R})}, \frac{\bar{\mu}_j}{\mu_j(\mathbb{R})}\right) + \mathcal{AW}_1(\gamma_j^k, \gamma_j) \xrightarrow[k\to\infty]{} 0,$$

the right-hand side of (3.10) goes to 0 as $k \to \infty$.

In the next step, we revert the normalization by setting

$$\pi^k := \sum_{j=1}^J \mu_j^k(\mathbb{R})\bar{\gamma}_j^k \in \bar{\Pi}_M(\bar{\mu}^k, \nu^k).$$

Let $\varepsilon \in (0, \min_{1\le j\le J}\mu_j(\mathbb{R}))$ be arbitrary and assume that k is sufficiently large so that $\max_{1\le j\le J}|\mu_j^k(\mathbb{R}) - \mu_j(\mathbb{R})| \le \varepsilon$. We split each of $\bar{\pi}^k$ and $\bar{\pi}$ into two parts:

$$\pi^k = \sum_{j=1}^J (\mu_j(\mathbb{R}) - \varepsilon)\,\bar{\gamma}_j^k + \sum_{j=1}^J \left(\mu_j^k(\mathbb{R}) - \mu_j(\mathbb{R}) + \varepsilon\right)\bar{\gamma}_j^k \text{ and } \pi = \sum_{j=1}^J (\mu_j(\mathbb{R}) - \varepsilon)\,\bar{\gamma}_j + \varepsilon\sum_{j=1}^J \bar{\gamma}_j.$$

Because $(\bar{\mu}_j)_{j=1}^J$ are pairwise singular, we can apply [7, Lemma 3.7] to deduce that

$$\lim_{k\to\infty} \mathcal{AW}_1\left(\sum_{j=1}^J (\mu_j(\mathbb{R}) - \varepsilon)\,\bar{\gamma}_j^k, \sum_{j=1}^J (\mu_j(\mathbb{R}) - \varepsilon)\,\bar{\gamma}_j\right) = 0.$$

With the help of [7, Lemma 3.6] and [7, Lemma 3.1 (a)(c)], we conclude that

$$\limsup_{k\to\infty} \mathcal{AW}_1\left(\pi^k, \pi\right) \le C\left(I^1_{J_\epsilon}(\bar{\mu}) + I^1_{J_\epsilon}(\nu)\right), \quad (3.11)$$

where $C > 0$ is a constant that does not depend on (k, ϵ) and, for a Polish space $\mathcal{X}$, $\delta > 0$, $x_0 \in \mathcal{X}$, and $\eta \in \mathcal{M}_1(\mathcal{X})$, $I^1_\delta(\eta)$ is given by

$$I^1_\delta(\eta) := \sup_{\tau \in \mathcal{M}_1(\mathcal{X}), \tau \le \eta, \tau(\mathcal{X}) \le \delta} \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0)\, \tau(dx).$$

As $\epsilon > 0$ was arbitrary and we have by [7, Lemma 3.1 (b)] that $\lim_{\epsilon \searrow 0} \left(I^1_{J\epsilon}(\bar{\mu}) + I^1_{J\epsilon}(\nu) \right) = 0$, we can infer from (3.11) that $(\bar{\pi}^k)_{k \in \mathbb{N}}$ converges to $\bar{\pi}$ in $\mathcal{AW}_1$. □

3.3 Proofs of Corollary 2.3 and Proposition 2.5

We are first going to prove the following stronger variants of Corollary 2.3 and Proposition 2.5 before deducing Proposition 2.8. Let $f \colon \mathcal{X} \to [1, +\infty)$ be a continuous growth function such that

$$\liminf_{\substack{|x| \to \infty \\ x \in \mathcal{X}}} \frac{f(x)}{|x|} > 0.$$

The topological space $\mathcal{P}_f(\mathcal{X})$ is defined like $\mathcal{P}_g(\mathcal{Y})$ with $\mathcal{X}$ and f replacing $\mathcal{Y}$ and g. The topological space $\mathcal{P}_{f \oplus \hat{g}}(\mathcal{X} \times \mathcal{P}_g(\mathcal{Y}))$ is defined analogously to $\mathcal{P}_{\bar{f} \oplus \hat{g}}(\mathcal{X} \times \mathcal{U} \times \mathcal{P}_g(\mathcal{Y}))$ but without the u coordinate.

Corollary 3.6. *Let $(\mu^k, \nu^k)_{k \in \mathbb{N}}$, $\mu^k \le_{cx} \nu^k$, be a convergent sequence in $\mathcal{P}_f(\mathcal{X}) \times \mathcal{P}_g(\mathcal{Y})$ with limit (μ, ν). Then, every $P \in \Lambda_M(\mu, \nu)$ is the limit in $\mathcal{P}_{f \oplus \hat{g}}(\mathcal{X} \times \mathcal{P}_g(\mathcal{Y}))$ of a sequence $(P^k)_{k \in \mathbb{N}}$ with $P^k \in \Lambda_M(\mu^k, \nu^k)$.*

Proof. Let $P = \mu(dx) P_x(d\rho) \in \Lambda_M(\mu, \nu)$. By Lemma 3.22 [23], there exists a measurable mapping $\mathbb{R} \times [0,1] \ni (x, u) \mapsto \pi_{x,u} \in \mathcal{P}_1(\mathbb{R})$ such that ${\pi_{x,u}}_{\#}\mathrm{Leb}(du) = P_x$. Then for $\bar{\mu} = \mu \otimes \mathrm{Leb}$, $\pi(dx, du, dy) = \bar{\mu}(dx, du) \pi_{x,u}(dy) \in \Pi_M(\bar{\mu}, \nu)$. The sequence $(\bar{\mu}^k = \mu^k \otimes \mathrm{Leb})_{k \in \mathbb{N}}$ converges to $\bar{\mu}$ in $\mathcal{P}_{\bar{f}}(\mathcal{X} \times [0,1])$ where $\bar{f}(x, u) = f(x)$. By Theorem 3.2, π is the limit in the adapted weak topology of a sequence $(\pi^k)_{k \in \mathbb{N}}$ with $\pi^k \in \Pi_M(\bar{\mu}^k, \nu^k)$. Therefore, we have $P^k = J(\pi^k) \in \Lambda_M(\mu^k, \nu^k)$ and get that $(P^k)_{k \in \mathbb{N}}$ converges to P in $\mathcal{P}_{f \oplus \hat{g}}(\mathcal{X} \times \mathcal{P}_g(\mathcal{Y}))$. □

Proposition 3.7. *Assume that $C \colon \mathcal{X} \times \mathcal{P}_g(\mathcal{Y}) \to \mathbb{R}$ is continuous and that there is a constant $K > 0$ such that, for all $(x, \rho) \in \mathcal{X} \times \mathcal{P}_g(\mathcal{Y})$,*

$$|C(x, \rho)| \le K \left(1 + f(x) + \rho(g)\right).$$

Then, the value $\hat{V}_C$ defined in (2.4) is attained and continuous on $\{(\mu, \nu) \in \mathcal{P}_f(\mathcal{X}) \times \mathcal{P}_g(\mathcal{Y}) : \mu \le_{cx} \nu\}$. Moreover, if $(\mu^k, \nu^k)_{k \in \mathbb{N}}$ with $\mu^k \le_{cx} \nu^k$ converges to (μ, ν) in $\mathcal{P}_f(\mathcal{X}) \times \mathcal{P}_g(\mathcal{Y})$ and $P^k \in \Lambda_M(\mu^k, \nu^k)$ is a sequence of optimizers of $\hat{V}_C(\mu^k, \nu^k)$, then its accumulation points are optimizers of $\hat{V}_C(\mu, \nu)$.

Proof. Note that the mapping $\mathcal{P}_{f \oplus \hat{g}}(\mathcal{X} \times \mathcal{P}_g(\mathcal{Y})) \ni P \mapsto P(C)$ is continuous. Using [9, Lemma A.7] for the relative compactness, we easily check that for $(\mu, \nu) \in \mathcal{P}_f(\mathcal{X}) \times \mathcal{P}_g(\mathcal{Y})$ with $\mu \le_{cx} \nu$, $\Lambda_M(\mu, \nu)$ is compact and non-empty. Therefore $\hat{V}_C$ is attained.

Let $(\mu^k, \nu^k)_{k \in \mathbb{N}}$ with $\mu^k \le_{cx} \nu^k$ be convergent in $\mathcal{P}_f(\mathcal{X}) \times \mathcal{P}_g(\mathcal{Y})$ with limit (μ, ν). Let $P^\star \in \Lambda_M(\mu, \nu)$ be optimal for $\hat{V}_C(\mu, \nu)$. By Corollary 3.6, there exists a sequence $(P^k)_{k \in \mathbb{N}}$ with $P^k \in \Lambda_M(\mu^k, \nu^k)$ that converges to $P^\star$ in $\mathcal{P}_{f \oplus \hat{g}}(\mathcal{X} \times \mathcal{P}_g(\mathcal{Y}))$. Therefore

$$\hat{V}_C(\mu, \nu) = P^\star(C) = \lim_{k \to \infty} P^k(C) \le \liminf_{k \to \infty} \hat{V}_C(\mu^k, \nu^k). \tag{3.12}$$

Now choosing $P^{k,\star} \in \Lambda_M(\mu^k, \nu^k)$ such that $\hat{V}_C(\mu^k, \nu^k) = P^{k,\star}(C)$, we may extract a subsequence $(P^{k_j,\star})_{j \in \mathbb{N}}$ converging to some P^∞ in $\mathcal{P}_{f \oplus \hat{g}}(\mathcal{X} \times \mathcal{P}_g(\mathcal{Y}))$ and such that $\lim_{j \to \infty} P^{k_j,\star}(C) = \limsup_{k \to \infty} \hat{V}_C(\mu^k, \nu^k)$. Then $P^\infty \in \Lambda_M(\mu, \nu)$ and

$$\hat{V}_C(\mu, \nu) \ge P^\infty(C) = \lim_{j \to \infty} P^{k_j}(C) = \limsup_{k \to \infty} \hat{V}_C(\mu^k, \nu^k).$$

With (3.12), we deduce that $\hat{V}_C(\mu,\nu) = \lim_{k\to\infty} \hat{V}_C(\mu^k,\nu^k)$ so that $\hat{V}_C$ is continuous on $\{(\eta,\theta) : \eta \leq_{cx} \theta\} \subseteq \mathcal{P}_f(\mathcal{X}) \times \mathcal{P}_g(\mathcal{Y})$. Moreover $\hat{V}_C(\mu,\nu) = P^\infty(C)$ and this equality remains true for any accumulation point P^∞ of $(P^{k,\star})_{k\in\mathbb{N}}$. □

3.4 Proof of Proposition 2.8 about the stability of the VIX futures subreplication price

Proof of Proposition 2.8. Set $\mathcal{X} = (0,+\infty)$, $f : \mathcal{X} \ni x \mapsto |\ln(x)| + |x|$ and $C_{\mathrm{VIX}} : \mathcal{X} \times \mathcal{P}_f(\mathcal{X}) \ni (x,\rho) \mapsto -\sqrt{\rho(\ell_x) \vee 0}$ where we recall that $\mathcal{X} \ni y \mapsto \ell_x(y) = \frac{2}{T_2 - T_1} \ln(x/y)$ for $x \in \mathcal{X}$. Let $\mu,\nu \in \mathcal{P}_f(\mathcal{X})$. For $\pi \in \Pi_{\mathrm{VIX}}(\mu,\nu)$, we have that $(x, \pi_{x,u})_\# \pi = P \in \Lambda_M(\mu,\nu)$ with

$$\int \sqrt{\pi_{x,u}(\ell_x)}\ \mathrm{proj}_{1,2}\,\pi(dx,du) = -P(C_{\mathrm{VIX}}). \tag{3.13}$$

Therefore $D_{\mathrm{sub}}(\mu,\nu) \geq -\hat{V}_{C_{\mathrm{VIX}}}(\mu,\nu)$.

On the other hand, if $P \in \Lambda_M(\mu,\nu)$ then

$$\pi := \Big((x,\rho,y) \mapsto (x, \sqrt{\rho(\ell_x)}, y)\Big)_\# P \otimes \rho \in \Pi_{\mathrm{VIX}}(\mu,\nu),$$

since we have, by Jensen's inequality combined with $\int \rho(f) P(dx,d\rho) = \nu(f) < \infty$, $\rho(\ell_x) \in [0,+\infty)$ $P(dx,d\rho)$-a.e. and, for $\varphi\colon \mathcal{X} \times \mathbb{R}_+ \to \mathbb{R}$ measurable and bounded,

$$\begin{aligned}\int_{\mathcal{X}\times\mathbb{R}_+\times\mathcal{X}} &\varphi(x,u)(u^2 - \ell_x(y))\pi(dx,du,dy)\\ &= \int_{\mathcal{P}_{f\oplus\hat{f}}(\mathcal{X}\times\mathcal{P}_f(\mathcal{X}))} \varphi(x,\sqrt{\rho(\ell_x)}) \int_{\mathcal{X}} (\rho(\ell_x) - \ell_x(y))\,\rho(dy)\,P(dx,d\rho) = 0.\end{aligned}$$

Therefore (3.13) again holds. We conclude that $D_{\mathrm{sub}}(\mu,\nu) = -\hat{V}_{C_{\mathrm{VIX}}}(\mu,\nu)$ and deduce from Proposition 3.7 applied with $\mathcal{Y} = \mathcal{X} = (0,\infty)$ and $g = f$ the continuity of D_{sub}. □

3.5 Stability of extended weak martingale optimal transport problems: proof of Theorem 2.2

This section is dedicated to the proof of a stronger variant of Theorem 2.2, that is Theorem 3.8 below.

Assumption B. We say that a cost function $C\colon \mathcal{X} \times \mathcal{U} \times \mathcal{P}_g(\mathcal{Y}) \to \mathbb{R}$ satisfies Assumption B if C is continuous and there is a constant $K > 0$ such that, for all $(x,u,\rho) \in \mathcal{X} \times \mathcal{U} \times \mathcal{P}_g(\mathcal{Y})$,

$$|C(x,u,\rho)| \leq K\left(1 + \bar{f}(x,u) + \rho(g)\right). \tag{3.14}$$

Theorem 3.8. *Let C satisfy Assumption B and $C(x,u,\cdot)$ be convex. Then the value function V_C defined in* (2.1) *is attained and continuous on $\{(\bar{\mu},\nu) : \mathrm{proj}_1\,\bar{\mu} \leq_{cx} \nu\} \subseteq \mathcal{P}_{\bar{f}}(\mathcal{X} \times \mathcal{U}) \times \mathcal{P}_g(\mathcal{Y})$.*

Furthermore, when $(\bar{\mu}^k,\nu^k)_{k\in\mathbb{N}}$ converges to $(\bar{\mu},\nu)$ in $\mathcal{P}_{\bar{f}}(\mathcal{X} \times \mathcal{U}) \times \mathcal{P}_g(\mathcal{Y})$ and for $k \in \mathbb{N}$, $\mathrm{proj}_1\,\bar{\mu}^k \leq_{cx} \nu^k$ and $\pi^k \in \Pi_M(\bar{\mu}^k,\nu^k)$ is optimal for $V_C(\bar{\mu}^k,\nu^k)$, we have:

- *(i) the accumulation points of $(\pi^k)_{k\in\mathbb{N}}$ are optimal for $V_C(\bar{\mu},\nu)$;*
- *(ii) if $C(x,u,\cdot)$ is strictly convex, then optimizers of* (2.1) *are unique and $(\pi^k)_{k\in\mathbb{N}}$ converges to the optimizer of $V_C(\bar{\mu},\nu)$ in the adapted weak topology.*

Proof. By [9, Proposition A.12 (b)], the map

$$\mathcal{P}_{\bar{f}\oplus g}(\mathcal{X} \times \mathcal{U} \times \mathcal{Y}) \ni \pi \mapsto \int C(x,u,\pi_{x,u})\ \mathrm{proj}_{1,2}\,\pi(dx,du)$$

is lower semicontinuous. Since $\Pi_M(\bar\mu,\nu)$ is a compact subset of $\mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})$, we deduce that the value function is attained.

Let $((\bar\mu^k,\nu^k))_{k\in\mathbb{N}}$ in $\{(\bar\eta,\theta):\mathrm{proj}_1\,\bar\eta\le_{cx}\theta\}\subseteq\mathcal{P}_{\bar f}(\mathcal{X}\times\mathcal{U})\times\mathcal{P}_g(\mathcal{Y})$ be convergent with limit $(\bar\mu,\nu)$. Let $\pi^\star\in\Pi^M(\bar\mu,\nu)$ be such that $V_C(\bar\mu,\nu)=J(\pi^\star)(C)$. By Theorem 3.2, there exists a sequence $(\pi^k)_{k\in\mathbb{N}}$ with $\pi^k\in\Pi_M(\bar\mu^k,\nu^k)$ such that $J(\pi^k)\to J(\pi^\star)$ in $\mathcal{P}_{\bar f\oplus\hat g}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y}))$. Since the mapping $\mathcal{P}_{\bar f\oplus\hat g}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y}))\ni P\mapsto P(C)$ is continuous, we deduce that

$$\limsup_{k\to\infty}V_C(\bar\mu^k,\nu^k)\le\lim_{k\to\infty}J(\pi^k)(C)=J(\pi^\star)(C)=V_C(\bar\mu,\nu). \tag{3.15}$$

Choosing now $\pi^{k,\star}\in\Pi_M(\bar\mu^k,\nu^k)$ such that $V_C(\bar\mu^k,\nu^k)=\int C(x,u,\pi^{k,\star}_{x,u})\bar\mu^k(dx,du)$, we may extract a subsequence $(\pi^{k_j,\star})_{j\in\mathbb{N}}$ such that $(\pi^{k_j,\star})_{j\in\mathbb{N}}$ converges to π^∞ in $\mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})$ and $\lim_{j\to\infty}\int C(x,u,\pi^{k_j,\star}_{x,u})\bar\mu^{k_j}(dx,du)=\liminf_{k\to\infty}V_C(\mu^k,\nu^k)$. The limit π^∞ belongs to $\Pi_M(\bar\mu,\nu)$ and, by lower semicontinuity of

$$\mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})\ni\pi\mapsto\int C(x,u,\pi_{x,u})\ \mathrm{proj}_{1,2}\,\pi(dx,du),$$

we deduce that

$$\begin{aligned}V_C(\bar\mu,\nu)&\le\int C(x,u,\pi^\infty_{x,u})\,\bar\mu(dx,du)\\&\le\lim_{j\to\infty}\int C(x,u,\pi^{k_j,\star}_{x,u})\,\bar\mu^{k_j}(dx,du)=\liminf_{k\to\infty}V_C(\mu^k,\nu^k).\end{aligned}$$

With (3.15), we deduce that $\lim_{k\to\infty}V_C(\bar\mu^k,\nu^k)=V_C(\mu,\nu)$ so that V_C is continuous on $\{(\bar\eta,\theta):\mathrm{proj}_1\,\bar\eta\le_{cx}\theta\}\subseteq\mathcal{P}_{\bar f}(\mathcal{X}\times\mathcal{U})\times\mathcal{P}_g(\mathcal{Y})$. Moreover, $V_C(\bar\mu,\nu)=\int C(x,u,\pi^\infty_{x,u})\bar\mu(dx,du)$ and this equality remains true for any accumulation point π^∞ of $(\pi^{k,\star})_{k\in\mathbb{N}}$ in $\mathcal{P}_{\bar f\oplus g}(\mathcal{X}\times\mathcal{U}\times\mathcal{Y})$.

Let us now assume that $C(x,u,\cdot)$ is strictly convex and prove (ii). When $\pi\in\Pi_M(\bar\mu,\nu)$ is another optimizer of $V_C(\bar\mu,\nu)$, then $\frac{\pi+\pi^\star}{2}\in\Pi_M(\bar\mu,\nu)$ and $\int C(x,u,(\frac{\pi+\pi^\star}{2})_{x,u})\bar\mu(dx,du)\ge V_C(\bar\mu,\nu)$ so that

$$\int\bigg(C\big(x,u,\frac{\pi_{x,u}+\pi^\star_{x,u}}{2}\big)-\frac{1}{2}(C(x,u,\pi_{x,u})+C(x,u,\pi^\star_{x,u}))\bigg)\,\bar\mu(dx,du)\ge0.$$

With the strict convexity, this implies that $\bar\mu(dx,du)$-a.e., $\pi_{x,u}=\pi^\star_{x,u}$ and $\pi=\pi^\star$.

Let us finally assume that the sequence $(\pi^k)_{k\in\mathbb{N}}$ of optimizers of $V_C(\bar\mu^k,\nu^k)$ admits a subsequence which does not have $\pi^\star$ as an accumulation point w.r.t. the adapted weak topology. By [9, Lemma A.7], this particular subsequence admits a subsequence $(\pi^{k_j})_{j\in\mathbb{N}}$ such that $(J(\pi^{k_j}))_{j\in\mathbb{N}}$ converges in $\mathcal{P}_{\bar f\oplus\hat g}(\mathcal{X}\times\mathcal{U}\times\mathcal{P}_g(\mathcal{Y}))$ to P. We define $\tilde\pi\in\Pi_M(\bar\mu,\nu)$ by $\tilde\pi=\bar\mu\times\tilde\pi_{x,u}$ with $\tilde\pi_{x,u}=\int\rho(dy)\,P_{x,u}(d\rho)$. As $C(x,u,\cdot)$ is convex and continuous, we have by Jensen's inequality

$$\begin{aligned}\int C(x,u,\tilde\pi_{x,u})\,\bar\mu(dx,du)&\le\int C(x,u,\rho)\,P(dx,du,d\rho)\\&=\lim_{k\to\infty}\int C(x,u,\pi^k_{x,u})\,\bar\mu^k(dx,du)=V_C(\bar\mu,\nu).\end{aligned}$$

In particular, $\tilde\pi$ is an optimizer of $V_C(\bar\mu,\nu)$ and, by strict convexity of $C(x,u,\cdot)$, we have $J(\tilde\pi)=P$ and uniqueness of optimizers. Thus, $\tilde\pi=\pi^\star$, and we also get $J(\pi^\star)=P$. Hence, $(\pi^{k_j})_{j\in\mathbb{N}}$ converges in the adapted weak topology to $\pi^\star$, which is a contradiction and completes the proof. ☐

3.6 Stability of the shadow couplings: proof of Proposition 2.9

Let us first state a consequence of Proposition 2.9 concerning the shadow couplings.

In view of Sklar's theorem, it is natural to parametrize the dependence structure between μ and the Lebesgue measure on $[0,1]$ in the lift $\bar{\mu} \in \Pi(\mu, \mathrm{Leb})$ of μ by copulas i.e. probability measures on $[0,1] \times [0,1]$ with both marginals equal to the Lebesgue measure. We call shadow coupling between μ and ν with copula χ the shadow coupling between μ and ν with source equal to the image $\bar{\mu}_\chi$ of χ by $[0,1] \times [0,1] \ni (v,u) \mapsto (F_\mu^{-1}(v), u) \in \mathbb{R} \times [0,1]$, where F_μ^{-1} denotes the quantile function of μ , i.e., the left-continuous pseudo-inverse of its cumulative distribution function.

Corollary 3.9. *The shadow coupling with copula χ is continuous on the domain $\{(\mu,\nu) : \mu \leq_{cx} \nu\} \subseteq \mathcal{P}_p(\mathbb{R}) \times \mathcal{P}_p(\mathbb{R})$ and with range $(\mathcal{P}_p(\mathbb{R} \times \mathbb{R}), \mathcal{W}_p)$ and even continuous in $\mathcal{AW}_p$ at each couple (μ,ν) such that μ does not weight points.*

The proof of this corollary is postponed after the one of Proposition 2.9. For the Hoeffding-Fréchet copula, $\chi(dv,du) = \mathrm{Leb}(du)\delta_u(dv)$, we recover the stability w.r.t. the marginals μ and ν of the left-curtain coupling proved by Juillet in [22]. For the independence copula $\chi(dv,du) = \mathrm{Leb}(dv) \otimes \mathrm{Leb}(du)$, we deduce the continuity of the sunset coupling.

The proof that the selector SC of the lifted shadow coupling is continuous when the codomain $\mathcal{P}_p(\mathbb{R} \times [0,1] \times \mathbb{R})$ is endowed with the adapted Wasserstein distance $\mathcal{AW}_p$ relies on the fact that, by (2.12), the selector SC takes values in the following extremal set of extended martingale couplings

$$\Pi^{\mathrm{ext}}_{M,p} := \{\pi \in \mathcal{P}_p(\mathbb{R} \times \mathcal{U} \times \mathbb{R}) : \#\,\mathrm{Supp}(\pi_{x,u}) \in \{1,2\} \text{ and } \mathrm{mean}(\pi_{x,u}) = x \ \pi\text{-a.s.}\}.$$

The set $\Pi^{\mathrm{ext}}_{M,p}$ is extremal in the following sense: when $\pi \in \Pi^{\mathrm{ext}}_{M,p}$ and $P \in \mathcal{P}_p(\mathbb{R} \times \mathcal{U} \times \mathcal{P}_p(\mathbb{R}))$ with $I(P) = \pi$, where $I(P)$ is the unique measure that satisfies

$$\int f(x,u,y)\, I(P)(dx,du,dy) = \int \int f(x,u,y)\, \rho(dy)\, P(dx,du,d\rho),$$

for all $f \in C_b(\mathbb{R} \times \mathcal{U} \times \mathbb{R})$ and $\mathrm{mean}(\rho) = x$ P-a.s., then we already have $P = J(\pi)$. Proceeding from this observation, the next lemma shows that on $\Pi^{\mathrm{ext}}_{M,p}$ the p-Wasserstein topology coincides with the p-adapted Wasserstein topology, which we in turn use to prove Proposition 2.9.

Lemma 3.10. *The identity map Id on $\mathcal{P}_p(\mathbb{R} \times \mathcal{U} \times \mathbb{R})$ is $(\mathcal{W}_p, \mathcal{AW}_p)$-continuous at any $P \in \Pi^{\mathrm{ext}}_{M,p}$. In particular, the metric spaces $(\Pi^{\mathrm{ext}}_{M,p}, \mathcal{W}_p)$ and $(\Pi^{\mathrm{ext}}_{M,p}, \mathcal{AW}_p)$ are topologically equivalent.*

Proof. We follow a similar line of reasoning as used in [26, Lemma 1.9]. As $\mathcal{W}_p \leq \mathcal{AW}_p$, it suffices to show that, given a sequence $(\pi^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_p(\mathbb{R} \times \mathcal{U} \times \mathbb{R})$ with $\mathrm{mean}(\pi^k_{x,u}) = x$ π^k-a.s. and $\pi \in \Pi^{\mathrm{ext}}_{M,p}$,

$$\lim_{k\to\infty} \mathcal{W}_p(\pi^k, \pi) = 0 \implies \lim_{k\to\infty} \mathcal{AW}_p(\pi^k, \pi) = 0.$$

So, let $(\pi^k)_{k\in\mathbb{N}}$ and π be as above and assume that $\pi^k \to \pi$ in $\mathcal{W}_p$. Observe that $J|_{\Pi^{\mathrm{ext}}_{M,p}}$ is bijective onto

$$J(\Pi^{\mathrm{ext}}_{M,p}) = \{P \in \mathcal{P}_p(\mathbb{R} \times \mathcal{U} \times \mathcal{P}_p(\mathbb{R})) : I(P) \in \Pi^{\mathrm{ext}}_{M,p} \text{ and } \mathrm{mean}(\rho) = x \ P(dx,du,d\rho)\text{-a.e.}\}, \tag{3.16}$$

with inverse I. Using [3, Lemma 2.3], we find that the sequence $(J(\pi^k))_{k\in\mathbb{N}}$ is $\mathcal{W}_p$-relatively compact in $\mathcal{P}_p(\mathbb{R} \times \mathcal{U} \times \mathcal{P}_p(\mathbb{R}))$. Therefore, there is a subsequence $(\pi^{k_j})_{j\in\mathbb{N}}$ such that $J(\pi^{k_j}) \to P$. Since $\pi^{k_j} \to I(P) = \pi \in \Pi^{\mathrm{ext}}_M$ and $\mathrm{mean}(\rho) = x$ $P(dx,du,d\rho)$-a.e., we get by (3.16) that $P \in J(\Pi^{\text{ext}}_{M,p})$ which yields by bijectivity of $J|_{\Pi^{\text{ext}}_{M,p}}$ that $P = J(\pi)$. Hence, $J(\pi^{k_j}) \to J(\pi)$ in $\mathcal{W}_p$ which means that $\pi^{k_j} \to \pi$ in $\mathcal{AW}_p$.

Since any subsequence of $(\pi^k)_{k\in\mathbb{N}}$ admits, by above reasoning, an $\mathcal{AW}_p$-convergent subsequence with limit π, we conclude that $\pi^k \to \pi$ in $\mathcal{AW}_p$. □

The proof of Proposition 2.9 also relies on the following lemma, the proof of which is postponed to the end of the current section.

Lemma 3.11. *Let $x, y, z \in \mathbb{R}$ with $y < x < z$, and $((y^k, z^k))_{k\in\mathbb{N}}$ be a $(-\infty, x] \times [x, +\infty)$-valued sequence such that for each k, either $y^k < x < z^k$ or $y^k = x = z^k$. Then we have*

(i) $\mathcal{W}_1(\mathrm{Ber}(x, y^k, z^k), \mathrm{Ber}(x, y, z)) \to 0 \iff |y^k - y| + |z^k - z| \to 0$,

(ii) $\mathcal{W}_1(\mathrm{Ber}(x, y^k, z^k), \delta_x) \to 0 \iff (z^k - x) \wedge (x - y^k) \to 0$.

Proof of Proposition 2.9. As optimizers of V_{SC} are unique, we immediately obtain from Theorem 3.8, applied with $C(x, u, \rho) = \int_{\mathbb{R}} (1-u)\sqrt{1+y^2}\rho(dy)$, continuity of

$$V_{\mathrm{SC}} \colon \{(\bar\mu, \nu) \in \mathcal{P}_p(\mathbb{R} \times [0,1]) \times \mathcal{P}_p(\mathbb{R}) : \mathrm{proj}_1 \bar\mu \le_{cx} \nu, \mathrm{proj}_2 \bar\mu = \mathrm{Leb}\} \to \mathbb{R}, \tag{3.17}$$

$$SC \colon \{(\bar\mu, \nu) \in \mathcal{P}_p(\mathbb{R} \times [0,1]) \times \mathcal{P}_p(\mathbb{R}) : \mathrm{proj}_1 \bar\mu \le_{cx} \nu, \mathrm{proj}_2 \bar\mu = \mathrm{Leb}\} \to \mathcal{P}_p(\mathbb{R} \times [0,1] \times \mathbb{R}), \tag{3.18}$$

when the domain is endowed with the product of the corresponding Wasserstein p-topologies. Since SC is a continuous function taking values in Π^{ext}_M, Lemma 3.10 ensures that it is still continuous when the codomain is endowed with the stronger $\mathcal{AW}_p$-distance. Therefore we have that

$$SC(\bar\mu^k, \nu^k) \to SC(\bar\mu, \nu) \quad \text{in } \mathcal{AW}_1,$$

which is equivalent to $\mathcal{W}_1$-convergence of

$$(\mathrm{Id}, \mathrm{Ber}(X, T_1^k, T_2^k))_\# \bar\mu^k = J(SC(\bar\mu^k, \nu^k)) \to J(SC(\bar\mu, \nu)) = (\mathrm{Id}, \mathrm{Ber}(X, T_1, T_2))_\# \bar\mu,$$

where $X \colon \mathbb{R} \times [0,1] \ni (x, u) \mapsto x \in \mathbb{R}$. Applying Lemma A.2 in the setting

$$\mathcal{V} = \mathbb{R} \times [0,1], \quad \mathcal{Z} = \mathcal{P}_1(\mathbb{R}), \quad \theta^k = \bar\mu^k,$$
$$\theta = \bar\mu, \quad \varphi^k = \mathrm{Ber}(X, T_1^k, T_2^k) \quad \text{and} \quad \varphi = \mathrm{Ber}(X, T_1, T_2),$$

yields $\mathrm{Ber}(X, T_1^k, T_2^k) \to \mathrm{Ber}(X, T_1, T_2)$ in $\bar\mu$-probability. There exists a subsequence such that this convergence holds $\bar\mu$-a.s. Hence, we can invoke Lemma 3.11 and derive the assertion in the second statement of the proposition for this particular subsequence. By the above reasoning any subsequence admits a subsubsequence which fulfills the conclusion of the second statement of the proposition, which readily implies the statement. □

Proof of Corollary 3.9. For the continuity in $\mathcal{W}_p$, it is enough to combine Proposition 2.9 with

$$\forall \mu, \mu' \in \mathcal{P}_p(\mathbb{R}),\ \mathcal{W}_p^p(\bar\mu_\chi, \bar\mu'_\chi) \le \int_{[0,1]\times[0,1]} |F_\mu^{-1}(v) - F_{\mu'}^{-1}(v)|^p \chi(dv, du) = \mathcal{W}_p^p(\mu, \mu'),$$
$$\forall \pi, \pi' \in \mathcal{P}_p(\mathbb{R} \times [0,1] \times \mathbb{R}),\ \mathcal{W}_p(\mathrm{proj}_{1,3}\pi, \mathrm{proj}_{1,3}\pi') \le \mathcal{W}_p(\pi, \pi') \le \mathcal{AW}_p(\pi, \pi').$$

To prove the reinforced continuity in $\mathcal{AW}_p$, we consider a sequence $((\mu^k, \nu^k)_k)$ in $\mathcal{P}_p(\mathbb{R}) \times \mathcal{P}_p(\mathbb{R})$ with $\mu^k \le_{cx} \nu^k$ converging to (μ, ν) where μ does not weight points. As seen in the proof of Proposition 2.9, the function SC is still continuous when the codomain is endowed with the $\mathcal{AW}_p$-distance, so that $\mathcal{AW}_p(SC(\bar\mu^k_\chi, \nu^k), SC(\bar\mu_\chi, \nu)) \to 0$.

For notational simplicity, we now denote SC^k and SC respectively in place of $SC(\bar{\mu}^k_\chi, \nu^k)$ and $SC(\bar{\mu}_\chi, \nu)$. Let $\eta^k \in \Pi(\bar{\mu}^k_\chi, \bar{\mu}_\chi)$ be optimal for $\mathcal{AW}_p(SC^k, SC)$. We have

$$\int_{[0,1]\times[0,1]} \mathcal{W}_p^p(SC^k_{F^{-1}_{\mu^k}(v),u}, SC_{F^{-1}_\mu(v),u})\chi(dv,du) \leq 2^{p-1}\mathcal{AW}_p^p(SC^k, SC) \\ + 2^{p-1}\int_{[0,1]\times[0,1]\times\mathbb{R}\times[0,1]} \mathcal{W}_p^p(SC_{x,w}, SC_{F^{-1}_\mu(v),u})\chi(dv,du)\eta^k_{F^{-1}_{\mu^k}(v),u}(dx,dw)$$

The second term on the right-hand side goes to 0 according to Lemma A.1 since $\bar{\mu}_\chi$ is the image of χ by $[0,1]\times[0,1] \ni (v,u) \mapsto (F^{-1}_\mu(v), u) \in \mathbb{R}\times[0,1]$ and, using $|x - F^{-1}_\mu(v)|^p \leq 2^{p-1}(|x - F^{-1}_{\mu^k}(v)|^p + |F^{-1}_{\mu^k}(v) - F^{-1}_\mu(v)|^p)$, we have

$$\int_{[0,1]^2\times\mathbb{R}\times[0,1]} |x - F^{-1}_\mu(v)|^p + |w-u|^p\chi(dv,du)\eta^k_{F^{-1}_{\mu^k}(v),u}(dx,dw) \\ \leq 2^{p-1}\left(\mathcal{AW}_p^p(SC^k, SC) + \mathcal{W}_p^p(\mu^k,\mu)\right) \to 0.$$

Hence,

$$\int_{[0,1]\times[0,1]} \mathcal{W}_p^p\left(SC^k_{F^{-1}_{\mu^k}(v),u}, SC_{F^{-1}_\mu(v),u}\right)\chi(dv,du) \to 0.$$

Let π^k (resp. π) denote the shadow coupling with copula χ between μ^k and ν^k (resp. μ and ν) and for $(x,w)\in\mathbb{R}\times[0,1]$, $\vartheta^k(v,w) = F_{\mu^k}(F^{-1}_{\mu^k}(v)-) + w\mu^k(\{F^{-1}_{\mu^k}(v)\})$. The image of the Lebesgue measure on $[0,1]\times[0,1]$ by ϑ^k is the Lebesgue measure on $[0,1]$ and for each $v\in(0,1)$, $F^{-1}_{\mu^k}(\vartheta^k(v,w)) = F^{-1}_{\mu^k}(v),\ dw$ a.e.. Hence dv a.e.,

$$\pi^k_{F^{-1}_{\mu^k}(v)} = \int_{[0,1]\times[0,1]} SC^k_{F^{-1}_{\mu^k}(\vartheta^k(v,w)),u}\chi_{\vartheta^k(v,w)}(du)dw.$$

Since μ does not weight points, F^{-1}_μ is one-to-one and $\pi_{F^{-1}_\mu(v)} = \int_{[0,1]} SC_{F^{-1}_\mu(v),u}\chi_v(du)$, dv a.e.. By the triangle inequality, we have

$$\mathcal{W}_p^p\left(\pi^k_{F^{-1}_{\mu^k}(v)}, \pi_{F^{-1}_\mu(v)}\right) \\ \leq 2^{p-1}\int_{[0,1]\times[0,1]} \mathcal{W}_p^p\left(SC^k_{F^{-1}_{\mu^k}(\vartheta^k(v,w)),u}, SC_{F^{-1}_\mu(\vartheta^k(v,w)),u}\right)\chi_{\vartheta^k(v,w)}(du)\,dw \\ + 2^{p-1}\int_{[0,1]} \mathcal{W}_p^p\left(\int_{[0,1]} SC_{F^{-1}_\mu(\vartheta^k(v,w)),u}\,\chi_{\vartheta^k(v,w)}(du), \int_{[0,1]} SC_{F^{-1}_\mu(v),u}\,\chi_v(du)\right) dw.$$

Using again that the image of the Lebesgue measure on $[0,1]\times[0,1]$ by ϑ^k is the Lebesgue measure on $[0,1]$, we deduce that

$$\mathcal{AW}_p^p(\pi,\pi^k) \leq \int_{[0,1]} |F^{-1}_\mu(v) - F^{-1}_{\mu^k}(v)|^p + \mathcal{W}_p^p(\pi_{F^{-1}_\mu(v)}, \pi^k_{F^{-1}_{\mu^k}(v)})\,dv \\ \leq \mathcal{W}_p^p(\mu,\mu^k) + 2^{p-1}\int_{[0,1]\times[0,1]} \mathcal{W}_p^p\left(SC^k_{F^{-1}_{\mu^k}(v),u}, SC_{F^{-1}_\mu(v),u}\right)\chi(dv,du) + \\ 2^{p-1}\int_{[0,1]\times[0,1]} \mathcal{W}_p^p\left(\int_{[0,1]} SC_{F^{-1}_\mu(\vartheta^k(v,w)),u}\,\chi_{\vartheta^k(v,w)}(du), \int_{[0,1]} SC_{F^{-1}_\mu(v),u}\,\chi_v(du)\right) dw\,dv.$$

The sum of the first two terms on the right-hand side goes to 0 as $n\to\infty$. Since, by the proof of [19, Proposition 4.2] (see the equation just above (4.12) where $\theta(F^{-1}_\mu(v), w) = v$ since F_μ is continuous), $dvdw$ -a.e., $\vartheta^k(v,w)\to v$, we have $\int_{[0,1]\times[0,1]} |\vartheta^k(v,w) - v|^p\,dv\,dw \to 0$ by Lebesgue's theorem, so that the third term on the right-hand side also goes to 0 by Lemma A.1. □

Remark 3.12. Like in the proof of [19, Proposition 4.2], we could check that $\mathcal{AW}_p(\pi^k, \pi)$ still goes to 0 as $n \to \infty$ when

$$\forall x \in \mathbb{R},\ \mu^k(\{x\}) > 0 \Rightarrow \exists (x^k)_k \in \mathbb{R}^{\mathbb{N}},\ F_{\mu^k}(x^k) \wedge F_\mu(x) - F_{\mu^k}(x^k-) \vee F_\mu(x-) \to \mu(\{x\}).$$

Proof of Lemma 3.11. To show (i) and (ii) we may assume w.l.o.g. that $y^k < x < z^k$, since $y^k = x = z^k$ implies $\mathrm{Ber}(x, y^k, z^k) = \delta_x \neq \mathrm{Ber}(x, y, z)$. Then we compute

$$\begin{aligned} \mathcal{W}_1\left(\mathrm{Ber}(x, y^k, z^k), \mathrm{Ber}(x, y, z)\right) &= \left(\tfrac{z-x}{z-y} \wedge \tfrac{z^k-x}{z^k-y^k}\right)|y - y^k| + \left(\tfrac{z-x}{z-y} - \tfrac{z^k-x}{z^k-y^k}\right)^+ (z^k - y) \\ &+ \left(\tfrac{z^k-x}{z^k-y^k} - \tfrac{z-x}{z-y}\right)^+ (z - y^k) + \left(\tfrac{x-y}{z-y} \wedge \tfrac{x-y^k}{z^k-y^k}\right)|z - z^k|. \end{aligned} \tag{3.19}$$

When $|y^k - y| + |z^k - z| \to 0$ then each summand in (3.19) also vanishes, showing the reverse implication of (i).

Conversely, when $\mathcal{W}_1(\mathrm{Ber}(x, y^k, z^k), \mathrm{Ber}(x, y, z)) \to 0$ then all four summands on the right-hand side of (3.19) have to go to 0 individually as $k \to \infty$. Thus, since $z^k - y \geq x - y > 0$ and $z - y^k \geq z - x > 0$, we find, due to the second and third terms in (3.19), that $\frac{z^k-x}{z^k-y^k} \to \frac{z-x}{z-y}$. Then using that the first and fourth terms also have to converge to 0, we get $|y^k - y| + |z^k - z| \to 0$, which completes the proof of (i).

On the other hand, we have

$$\frac{\mathcal{W}_1\left(\mathrm{Ber}(x, y^k, z^k), \delta_x\right)}{2} = \frac{(z^k - x)(x - y^k)}{z^k - y^k} = \frac{(z^k - x) \vee (x - y^k)}{z^k - y^k}\big((z^k - x) \wedge (x - y^k)\big).$$

Since $z^k - y^k \geq (z^k - x) \vee (x - y^k) \geq \frac{1}{2}(z^k - y^k)$, we deduce that

$$1 \leq \frac{\mathcal{W}_1\left(\mathrm{Ber}(x, y^k, z^k), \delta_x\right)}{(z^k - x) \wedge (x - y^k)} \leq 2,$$

which yields (ii) and completes the proof. □

3.7 Proof of Proposition 3.5

The proof of Proposition 3.5 relies on Lemma 3.13 below and Wasserstein projections in convex order. By [1, Proposition 4.2] there is a map $\mathcal{J} : \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R}) \to \mathcal{P}_1(\mathbb{R})$ satisfying

$$\mu \leq_c \mathcal{J}(\mu, \nu) \text{ and } \mathcal{W}_1(\mathcal{J}(\mu, \nu), \nu) = \inf_{\mu \leq_c \eta} \mathcal{W}_1(\eta, \nu), \tag{3.20}$$

which is called a Wasserstein projection in convex order. According to [20, Theorem 1.1], $\mathcal{J}$ is Lipschitz continuous: for $\mu, \nu, \mu', \nu' \in \mathcal{P}_p(\mathbb{R})$ with $p \geq 1$, we have

$$\mathcal{W}_p(\mathcal{J}(\mu, \nu), \mathcal{J}(\mu', \nu')) \leq \ \mathcal{W}_p(\mu, \mu') + 2\mathcal{W}_p(\nu, \nu'). \tag{3.21}$$

For $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ which share the same mean, we also denote by $\mu \wedge_c \nu$ the minimum of μ and ν in the convex order (see [24]). Its potential function is given by $u_{\mu \wedge_c \nu} = \mathrm{co}(u_\mu \wedge u_\nu)$ where co denotes the convex hull operator.

Lemma 3.13. *Let $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$, $a < b$, and $\rho \in \mathcal{P}_1(\mathbb{R})$ be concentrated on $[a, b]$ with mean $x \in \mathbb{R}$. Let $(a^m)_{m \in \mathbb{N}}, (b^m)_{m \in \mathbb{N}}$, $a < a^m < x < b^m < b$, be monotone sequences with $a^m \to a$ and $b^m \to b$. Then*

$$\mathcal{W}_1\left(\rho \wedge_c \left(\frac{b^m - x}{b^m - a^m}\delta_{a^m} + \frac{x - a^m}{b^m - a^m}\delta_{b^m}\right), \rho\right) \to 0.$$

Proof of Lemma 3.13. Let for each $m \in \mathbb{N}$, $\eta^m = \frac{b^m - x}{b^m - a^m}\delta_{a^m} + \frac{x - a^m}{b^m - a^m}\delta_{b^m}$ and $\rho^m = \rho \wedge_c \eta^m$. Let us check that $\lim_{m\to\infty} \sup_{y\in(-\infty,a^m]}(u_\rho(y) - (x-y)) = 0$. If $a = -\infty$, this is a consequence of $\lim_{y\to-\infty}(u_\rho(y) - (x-y)) = 0$. If $a > -\infty$, $u_\rho(y) = x - y$ for $y \le a$ and since u_ρ is 1-Lipschitz,

$$\forall y \in [a, a^m],\ u_\rho(y) \le u_\rho(a) + (y-a) = x - a + y - a \le x - y + 2(a^m - a).$$

Since $u_{\eta^m}(y) = x - y$ for $y \in (-\infty, a^m]$ and using a symmetric reasoning to deal with the supremum over $[b^m, +\infty)$ we deduce that $\lim_{m\to\infty} \sup_{y\in(-\infty,a^m]\cup[b^m,+\infty)}(u_\rho(y) - u_{\eta^m}(y)) = 0$. By convexity of u_ρ and since u_{η^m} is affine on $[a^m, b^m]$, $\sup_{y\in\mathbb{R}}(u_\rho(y) - u_{\eta^m}(y)) = \sup_{y\in(-\infty,a^m]\cup[b^m,+\infty)}(u_\rho(y) - u_{\eta^m}(y))$ so that

$$\lim_{m\to\infty} \sup_{y\in\mathbb{R}}(u_\rho(y) - u_{\eta^m}(y)) = 0.$$

Since the convex function $u_\rho - \sup_{y\in\mathbb{R}}(u_\rho(y) - u_{\eta^m}(y))$ is not greater than $u_\rho \wedge u_{\eta^m}$,

$$u_\rho - \sup_{y\in\mathbb{R}}(u_\rho(y) - u_{\eta^m}(y)) \le \operatorname{co}(u_\rho \wedge u_{\eta^m}) = u_{\rho^m} \le u_\rho.$$

Hence u_{ρ^m} converges uniformly to u_ρ as $m \to \infty$, which implies that $\mathcal{W}_1(\rho^m, \rho) \underset{m\to+\infty}{\longrightarrow} 0$. $\square$

Proof of Proposition 3.5. For every $j = 1, \dots, J$, we have $\mu_j^k \to \mu_j$ in $\mathcal{M}_1(\mathbb{R})$ with $\mu_j(\mathbb{R}) > 0$, so that $\mu_j^k(\mathbb{R}) \to \mu_j(\mathbb{R}) > 0$ and, for k large enough, $\min_{1\le j\le J} \mu_j^k(\mathbb{R}) > 0$. We are going to check the existence of $M < \infty$ such that for each $\varepsilon \in (0,1)$, we can find sequences $(\nu_j^{k,\varepsilon})_{k\in\mathbb{N}}$ in $\mathcal{M}_1(\mathbb{R})$ that satisfy

$$\mu_j^k \le_c \nu_j^{k,\varepsilon}, \quad \sum_{j=1}^J \nu_j^{k,\varepsilon} = \nu^k \quad \text{and} \quad \limsup_{k\to+\infty} \sum_{j=1}^J \mu_j(\mathbb{R}) \mathcal{W}_1\left(\frac{\nu_j^{k,\varepsilon}}{\mu_j^k(\mathbb{R})}, \frac{\nu_j}{\mu_j(\mathbb{R})}\right) \le M\varepsilon. \tag{3.22}$$

As ε is arbitrary, the conclusion follows easily from (3.22).

Before jumping into the various steps of proving (3.22), we fix the following notation: Let $a \in \{-\infty\} \cup \mathbb{R}$ and $b \in \mathbb{R} \cup \{+\infty\}$ be the endpoints of the irreducible component (a, b) of (μ, ν). Further, let

$$\pi^j = \frac{\mu_j}{\mu_j(\mathbb{R})} \otimes \pi_x^j \in \Pi_M\left(\frac{\mu_j}{\mu_j(\mathbb{R})}, \frac{\nu_j}{\nu_j(\mathbb{R})}\right).$$

Up to modifying $x \mapsto \pi_x^j$ on a μ-null set, we suppose w.l.o.g. that for all $x \in (a,b)$, π_x^j is concentrated on $[a,b]$ and $\operatorname{mean}(\pi_x^j) = x$. Finally, for $m \in \mathbb{N}$, pick $a^m, b^m \in (a,b)$, $a^m < b^m$, with $a^m \searrow a$, and $b^m \nearrow b$, so that $\mu_j([a^m, b^m]) > 0$ and $\mu_j(\{a^m, b^m\}) = 0$ for each $j = 1, \dots, J$.

Step 1: We claim that when m is sufficiently large, there exists $\tilde\nu_j \in \mathcal{M}_1(\mathbb{R})$ with

$$\mathcal{W}_1(\tilde\nu_j, \nu_j) < \epsilon, \quad \tilde\nu_j \le_c \nu_j, \quad \mu_j|_{[a^m,b^m]} \le_c \tilde\nu_j|_{[a^m,b^m]} \text{ and } \tilde\nu_j|_{\mathbb{R}\setminus[a^m,b^m]} = \mu_j|_{\mathbb{R}\setminus[a^m,b^m]}. \tag{3.23}$$

To show (3.23) we define q_x^m as the unique probability measure supported on $\{a^m, b^m\}$ with $\operatorname{mean}(q_x^m) = x$ when $x \in [a^m, b^m]$, and δ_x otherwise, i.e.,

$$q_x^m := \begin{cases} \frac{b^m - x}{b^m - a^m}\delta_{a^m} + \frac{x - a^m}{b^m - a^m}\delta_{b^m} & \text{if } x \in (a^m, b^m), \\ \delta_x & \text{else.} \end{cases}$$

Set $\pi^{j,m}(dx, dy) := \mu_j(dx)\,(\pi_x^j \wedge_c q_x^m)(dy)$. The measure $\pi^{j,m}$ is a martingale coupling between μ_j and its second marginal, which we denote by $\nu_{j,m}$ and thus $\nu_{j,m} \le_c \nu_j$. Thanks to Lemma 3.13 we have for every $x \in (a,b)$ that $\mathcal{W}_1(\pi^j_x, \pi^j_x \wedge_c q^m_x) \to 0$. Furthermore, by the triangle inequality and convexity of the absolute value we have

$$\mathcal{W}_1(\pi^j_x, \pi^j_x \wedge_c q^m_x) \le \mathcal{W}_1(\pi^j_x, \delta_0) + \mathcal{W}_1(\delta_0, \pi^j_x \wedge_c q^m_x) \le 2\mathcal{W}_1(\pi^j_x, \delta_0),$$

where the right-hand side is μ_j-integrable. Hence, we get by dominated convergence

$$\mathcal{AW}_1(\pi^j, \pi^{j,m}) \le \int_{\mathbb{R}} \mathcal{W}_1(\pi^j_x, \pi^j_x \wedge_c q^m_x)\, \mu_j(dx) \underset{m\to+\infty}{\longrightarrow} 0. \tag{3.24}$$

Letting m be sufficiently large, (3.24) yields that $\tilde{\nu}_j := \nu_{j,m}$ satisfies $\mathcal{W}_1(\tilde{\nu}_j, \nu_j) < \varepsilon$. Since, for $x \in [a^m, b^m]$, $(\pi^j_x \wedge_c q^m_x)([a^m, b^m]) = 1$ and for $x \in \mathbb{R}\backslash[a^m, b^m]$, $\pi^j_x \wedge_c q^m_x = \delta_x$, $\tilde{\nu}_j|_{[a^m,b^m]}$ is the second marginal of $\mu_j|_{[a^m,b^m]} \otimes (\pi^j_x \wedge_c q^m_x)$ and (3.23) holds. Observe that $\tilde{\nu}_j([a^m, b^m]) = \mu_j([a^m, b^m])$ implies that

$$\operatorname{mean}\left(\frac{\mu_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}\right) = \operatorname{mean}\left(\frac{\nu_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}\right) =: \tilde{x}^m_j.$$

Step 2: Next we construct, for $j \in \{1, \dots, J\}$, sequences $(\tilde{\nu}^k_j)_{k\in\mathbb{N}}$ in $\mathcal{M}_1(\mathbb{R})$ such that

$$\mu^k_j \le_c \tilde{\nu}^k_j, \quad \sum_{j=1}^J \tilde{\nu}^k_j \le_c \nu^k \quad \text{and } \tilde{\nu}^k_j \underset{k\to\infty}{\to} (1-\epsilon)\tilde{\nu}_j + \epsilon\mu_j \text{ in } \mathcal{M}_1(\mathbb{R}). \tag{3.25}$$

Since $\mu_j(\{a^m, b^m\}) = 0$, we have for every $h \in C_b(\mathbb{R})$ that the discontinuities of $h\mathbb{1}_{[a^m,b^m]}$ are a μ_j-null set, whence we get by Portmanteau's theorem

$$\int \mathbb{1}_{[a^m,b^m]}(x)h(x)\,\mu^k_j(dx) \to \int \mathbb{1}_{[a^m,b^m]}(x)h(x)\,\mu_j(dx).$$

With Lemma A.4, we deduce that $\mu^k_j|_{[a^m,b^m]}$ converges to $\mu_j|_{[a^m,b^m]}$ in $\mathcal{M}_1(\mathbb{R})$ as $k \to \infty$. When k is sufficiently large, we have $\mu^k_j([a^m, b^m]) > 0$ and define

$$x^{k,m}_j := \operatorname{mean}\left(\frac{\mu^k_j|_{[a^m,b^m]}}{\mu^k_j([a^m,b^m])}\right) \quad \text{and} \quad \hat{\nu}^k_j := \mathcal{J}\left(\frac{\mu^k_j|_{[a^m,b^m]}}{\mu^k_j([a^m,b^m])}, \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}\right) \wedge_c q^m_{x^{k,m}_j}.$$

To simplify notation, we use the above definitions also when $\mu^k_j([a^m, b^m]) = 0$ under the convention that the undefined term $\frac{\mu^k_j|_{[a^m,b^m]}}{\mu^k_j([a^m,b^m])}$ is replaced by $\frac{\mu_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}$. By Lemma A.3 and $\mathcal{W}_1$-Lipschitz continuity of $\mathcal{J}$, c.f. (3.21), we have

$$\mathcal{J}\left(\frac{\mu^k_j|_{[a^m,b^m]}}{\mu^k_j([a^m,b^m])}, \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}\right) \underset{k\to+\infty}{\longrightarrow} \mathcal{J}\left(\frac{\mu_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}, \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])}\right) = \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])},$$

in $\mathcal{W}_1$, where the last equality follows from the fact that $\mu|_{[a^m,b^m]} \le_c \tilde{\nu}_j|_{[a^m,b^m]}$. Again by Lemma A.3, we obtain that $x^{k,m}_j \to \tilde{x}^m_j$ and therefore $q^m_{x^{k,m}_j} \to q^m_{\tilde{x}^m_j}$ in $\mathcal{W}_1$. Thus, [7, Lemma 4.1], i.e., continuity of $\wedge_c$, provides that

$$\hat{\nu}^k_j \underset{k\to+\infty}{\longrightarrow} \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])} \wedge_c q^m_{\tilde{x}^m_j} \quad \text{in } \mathcal{W}_1.$$

Since $\tilde{\nu}_j|_{[a^m,b^m]}$ is concentrated on $[a^m, b^m]$ with mass $\mu_j([a^m, b^m])$ and mean $\tilde{x}^m_j$, we have $\tilde{\nu}_j|_{[a^m,b^m]} \le_c \mu_j([a^m, b^m])q^m_{\tilde{x}^m_j}$. Hence,

$$\hat{\nu}^k_j \underset{k\to+\infty}{\longrightarrow} \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu_j([a^m,b^m])} \quad \text{in } \mathcal{W}_1. \tag{3.26}$$

We set

$$\tilde{\nu}_j^k := (1-\epsilon)\left(\mu_j^k|_{\mathbb{R}\setminus[a^m,b^m]} + \mu_j^k([a^m,b^m])\hat{\nu}_j^k\right) + \epsilon\mu_j^k. \tag{3.27}$$

By definition of $\hat{\nu}_j^k$ and since $\frac{\mu_j^k|_{[a^m,b^m]}}{\mu_j^k([a^m,b^m])} \leq_c q^m_{x_j^{k,m}}$, we have $\frac{\mu_j^k|_{[a^m,b^m]}}{\mu_j^k([a^m,b^m])} \leq_c \hat{\nu}_j^k$, which yields

$$\mu_j^k = (1-\epsilon)\left(\mu_j^k|_{\mathbb{R}\setminus[a^m,b^m]} + \mu_j^k|_{[a^m,b^m]}\right) + \epsilon\mu_j^k \leq_c \tilde{\nu}_j^k. \tag{3.28}$$

In order to complete step 2, it remains to show that, when k is sufficiently large, then

$$\sum_{j=1}^J \tilde{\nu}_j^k \leq_c \nu^k. \tag{3.29}$$

As $\mu_j^k \to \mu_j$ and $\mu_j^k|_{[a^m,b^m]} \to \mu_j|_{[a^m,b^m]}$ in $\mathcal{M}_1(\mathbb{R})$ and by (3.26) and (3.23), we also obtain that $\tilde{\nu}_j^k \to (1-\epsilon)\tilde{\nu}_j + \epsilon\mu_j$ in $\mathcal{M}_1(\mathbb{R})$. In turn, this implies locally uniform convergence of the sequence of potential functions $(u_{\tilde{\nu}_j^k})_{k\in\mathbb{N}}$ to $u_{(1-\epsilon)\tilde{\nu}_j+\epsilon\mu_j}$. At the same time, as $\nu^k \to \nu$ in $\mathcal{P}_1(\mathbb{R})$, we have uniform convergence of u_{ν^k} to u_ν. Thus, we find, for each $\delta > 0$, an index $k(\delta) \in \mathbb{N}$ such that for all $k \geq k(\delta)$ and $j \in \{1, \cdots, J\}$ we have

$$u_{\tilde{\nu}_j^k}|_{[a^m,b^m]} \leq (1-\epsilon)u_{\tilde{\nu}_j}|_{[a^m,b^m]} + \epsilon u_{\mu_j}|_{[a^m,b^m]} + \frac{\delta}{J} \quad \text{and} \quad u_\nu \leq u_{\nu^k} + \delta. \tag{3.30}$$

As (μ,ν) is irreducible with component $(a,b) \supseteq [a^m,b^m]$, we can fix a $\delta > 0$ such that

$$\epsilon u_\mu|_{[a^m,b^m]} \leq \epsilon u_\nu|_{[a^m,b^m]} - 2\delta. \tag{3.31}$$

Let $k \geq k(\delta)$, and compute, for $y \in [a^m,b^m]$,

$$\begin{aligned}
\sum_{j=1}^J u_{\tilde{\nu}_j^k}(y) &\leq (1-\epsilon)\sum_{j=1}^J u_{\tilde{\nu}_j}(y) + \epsilon\sum_{j=1}^J u_{\mu_j}(y) + \delta \\
&\leq (1-\epsilon)u_\nu(y) + \epsilon u_\nu(y) - \delta \\
&= u_\nu(y) - \delta \leq u_{\nu^k}(y),
\end{aligned}$$

where the first and last inequalities follow from (3.30), the second from $\sum_{j=1}^J \mu_j = \mu$, $\sum_{j=1}^J \tilde{\nu}_j \leq_c \sum_{j=1}^J \nu_j = \nu$ and (3.31). Next, let $y \in \mathbb{R}\setminus[a^m,b^m]$. Since $\mu_j^k([a^m,b^m])\hat{\nu}_j^k$ and $\mu_j^k|_{[a^m,b^m]}$ are both concentrated on $[a^m,b^m]$ with the same mass and barycentre, we obtain that their potential functions take the same value at y. We have

$$\begin{aligned}
\sum_{j=1}^J u_{\tilde{\nu}_j^k}(y) &= (1-\epsilon)\sum_{j=1}^J \left(u_{\mu_j^k|_{\mathbb{R}\setminus[a^m,b^m]}}(y) + \mu_j^k([a^m,b^m])u_{\hat{\nu}_j^k}(y)\right) + \epsilon\sum_{j=1}^J u_{\mu_j^k}(y) \\
&= (1-\epsilon)\sum_{j=1}^J \left(u_{\mu_j^k|_{\mathbb{R}\setminus[a^m,b^m]}}(y) + u_{\mu_j^k|_{[a^m,b^m]}}(y)\right) + \epsilon u_{\mu^k}(y) = u_{\mu^k}(y) \leq u_{\nu^k}(y).
\end{aligned}$$

Summarizing, we have $\sum_{j=1}^J u_{\tilde{\nu}_j^k} \leq u_{\nu^k}$ for $k \geq k(\delta)$, which yields (3.29).

Step 3: The final step consists in modifying $(\tilde{\nu}_j^k)_{j=1}^J$ to $(\nu_j^{k,\epsilon})_{j=1}^J$ that fulfills (3.22). Denote by $\chi^k \in \Pi_M(\sum_{j=1}^J \tilde{\nu}_j^k, \nu^k)$ the inverse transform martingale coupling (see [21]), which by [21, Theorem 2.11] satisfies

$$\int_{\mathbb{R}\times\mathbb{R}} |y-x|\,\chi^k(dx,dy) \leq 2\mathcal{W}_1\left(\sum_{j=1}^J \tilde{\nu}_j^k, \nu^k\right). \tag{3.32}$$

We define $\nu_j^{k,\varepsilon}$ as the second marginal of $\tilde{\nu}_j^k \otimes \chi_x^k$, that is

$$\nu_j^{k,\varepsilon}(dy) := \int_{\mathbb{R}} \chi_x^k(dy)\, \tilde{\nu}_j^k(dx).$$

Using (3.28) we have

$$\mu_j^k \leq_c \tilde{\nu}_j^k \leq_c \nu_j^{k,\varepsilon} \quad \text{and} \quad \sum_{j=1}^J \nu_j^{k,\varepsilon} = \nu^k. \tag{3.33}$$

To prove the remaining claim in (3.22), we estimate

$$\mu_j(\mathbb{R})\mathcal{W}_1\left(\frac{\nu_j^{k,\varepsilon}}{\mu_j^k(\mathbb{R})}, \frac{\nu_j}{\mu_j(\mathbb{R})}\right) \leq \frac{\mu_j(\mathbb{R})}{\mu_j^k(\mathbb{R})}\mathcal{W}_1\left(\nu_j^{k,\varepsilon}, \tilde{\nu}_j^k\right) + \mathcal{W}_1\left(\frac{\mu_j(\mathbb{R})}{\mu_j^k(\mathbb{R})}\tilde{\nu}_j^k, \tilde{\nu}_j\right) + \mathcal{W}_1\left(\tilde{\nu}_j, \nu_j\right). \tag{3.34}$$

Because of (3.25), $\sum_{j=1}^J \tilde{\nu}_j^k$ converges to $(1-\varepsilon)\sum_{j=1}^J \tilde{\nu}_j + \varepsilon \sum_{j=1}^J \mu_j$ in $\mathcal{W}_1$. We compute

$$\begin{aligned}
\limsup_{k\to\infty} \sum_{j=1}^J \frac{\mu_j(\mathbb{R})}{\mu_j^k(\mathbb{R})}\mathcal{W}_1\left(\nu_j^{k,\varepsilon}, \tilde{\nu}_j^k\right) &= \limsup_{k\to\infty} \sum_{j=1}^J \mathcal{W}_1\left(\tilde{\nu}_j^k, \nu_j^{k,\varepsilon}\right) \\
&\leq \limsup_{k\to+\infty} \int_{\mathbb{R}\times\mathbb{R}} |y-x|\, \chi^k(dx,dy) \\
&\leq 2 \limsup_{k\to+\infty} \mathcal{W}_1\left(\sum_{j=1}^J \tilde{\nu}_j^k, \nu^k\right) \\
&\leq 2\mathcal{W}_1\left((1-\varepsilon)\sum_{j=1}^J \tilde{\nu}_j + \varepsilon \sum_{j=1}^J \mu_j, \nu\right) \\
&\leq 2(1-\varepsilon)\sum_{j=1}^J \mathcal{W}_1(\tilde{\nu}_j, \nu_j) + 2\varepsilon \mathcal{W}_1(\mu, \nu) \\
&\leq 2\left(J + \mathcal{W}_1(\mu,\nu)\right)\varepsilon,
\end{aligned}$$

where the first equality is due to $\mu_j^k(\mathbb{R}) \to \mu_j(\mathbb{R})$, the first inequality holds because $\sum_{j=1}^J \tilde{\nu}_j^k \otimes \chi_x^k = \chi^k$, the second due to (3.32), the second last by the convexity of the 1-Wasserstein distance and $\nu = \sum_{j=1}^J \nu_j$, and the last by (3.23). As $\varepsilon < 1$, we obtain by (3.25) and convexity of the 1-Wasserstein distance that

$$\limsup_{k\to\infty} \mathcal{W}_1\left(\frac{\mu_j(\mathbb{R})}{\mu_j^k(\mathbb{R})}\tilde{\nu}_j^k, \tilde{\nu}_j\right) \leq \varepsilon \mathcal{W}_1(\mu_j, \tilde{\nu}_j) \leq \varepsilon\big(\mathcal{W}_1(\mu_j,\nu_j) + \varepsilon\big) \leq \varepsilon\big(\mathcal{W}_1(\mu_j,\nu_j) + 1\big).$$

Plugging these estimates into (3.34) yields

$$\limsup_{k\to\infty} \sum_{j=1}^J \mu_j(\mathbb{R})\mathcal{W}_1\left(\frac{\nu_j^{k,\varepsilon}}{\mu_j^k(\mathbb{R})}, \frac{\nu_j}{\mu_j(\mathbb{R})}\right) \leq \left(4J + 2\mathcal{W}_1(\mu,\nu) + \sum_{j=1}^J \mathcal{W}_1(\mu_j,\nu_j)\right)\varepsilon$$

so that (3.22) holds with $M = 4J + 2\mathcal{W}_1(\mu,\nu) + \sum_{j=1}^J \mathcal{W}_1(\mu_j,\nu_j)$. □

A Measure-theoretic auxiliary results

Throughout the paper we make repeat use of the following reformulation of [13, Lemma 2.7].

Lemma A.1. *Let $(\mathcal{V}, d_\mathcal{V})$ and $(\mathcal{Z}, d_\mathcal{Z})$ be Polish metric spaces, $\mu \in \mathcal{P}_p(\mathcal{V})$ and $\varphi\colon \mathcal{V} \to \mathcal{Z}$ be a measurable function such that $\varphi_\# \mu \in \mathcal{P}_p(\mathcal{Z})$. Then*

$$\Pi(\mu,\mu) \ni \pi \mapsto \int_{\mathcal{V}\times\mathcal{V}} d^p_\mathcal{Z}(\varphi(v), \varphi(\hat v))\, \pi(dv, d\hat v)$$

vanishes when $\int_{\mathcal{V}\times\mathcal{V}} d^p_\mathcal{V}(v,\hat v)\, \pi(dv, d\hat v)$ goes to 0.

Proof. It was shown in [13, Lemma 2.7] that

$$\sup\left\{ \int_{\mathcal{V}\times\mathcal{V}} d^p_\mathcal{Z}(\varphi(v), \varphi(\hat v))\, \pi(dv, d\hat v) : \pi \in \Pi(\mu,\mu) \text{ with } \int_{\mathcal{V}\times\mathcal{V}} d^p_\mathcal{V}(v,\hat v)\, \pi(dv, d\hat v) \le \delta \right\}$$

goes to 0 as $\delta \searrow 0$, which proves our claim. □

To comment on our use of Lemma A.1, we typically apply it in the specific setting where φ is a disintegration kernel of some fixed coupling, cf. for example the end of the proof of Lemma 3.4. For this reason, the lemma proves very useful to check convergence in the adapted Wasserstein topology. For more details on the adapted weak topologies and the adapted Wasserstein distance, we refer the interested reader to [2, 5, 26].

The next result relates weak convergence of couplings concentrated on the graph of measurable functions, to convergence in probability of said functions. This lemma is a generalization of the classical result [25, Lemma 1].

Lemma A.2. *Let $\mathcal{V}, \mathcal{Z}$ be Polish spaces, $(\theta^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}(\mathcal{V})$ that converges in total variation to θ, and let $\varphi^k \colon \mathcal{V} \to \mathcal{Z}$, $k \in \mathbb{N}$, and $\varphi\colon \mathcal{V} \to \mathcal{Z}$ be measurable functions. Then*

$$(\mathrm{Id}, \varphi^k)_\# \theta^k \to (\mathrm{Id}, \varphi)_\# \theta \text{ in } \mathcal{P}(\mathcal{V}\times\mathcal{Z}) \implies \varphi^k \to \varphi \text{ in } \theta\text{-probability}.$$

Proof. As $\theta^k \to \theta$ in total variation, we have that the total variation distance between $(\mathrm{Id}, \varphi^k)_\# \theta^k$ and $(\mathrm{Id}, \varphi^k)_\# \theta$ vanishes as $k \to \infty$. Thus, since the sequence $((\mathrm{Id}, \varphi^k)_\# \theta^k)_{k\in\mathbb{N}}$ converges to $(\mathrm{Id}, \varphi)_\# \theta =: \eta$ in $\mathcal{P}(\mathcal{V}\times\mathcal{Z})$, the same holds for the sequence $(\eta^k)_{k\in\mathbb{N}}$ where $\eta^k := (\mathrm{Id}, \varphi^k)_\# \theta$. W.l.o.g. we assume that the metrics $d_\mathcal{X}$ and $d_\mathcal{Y}$ are both bounded, so that $\eta^k \to \eta$ in $\mathcal{W}_1$ and can pick couplings $\chi^k \in \Pi(\theta,\theta)$ such that

$$\mathcal{W}_1(\eta^k, \eta) = \int_{\mathcal{V}\times\mathcal{V}} d_\mathcal{V}(v,\hat v) + d_\mathcal{Z}(\varphi^k(v), \varphi(\hat v))\, \chi^k(dv, d\hat v).$$

By the triangle inequality we have

$$\begin{aligned} \int d_\mathcal{Z}(\varphi^k(v), \varphi(v))\, \theta(dv) &= \int d_\mathcal{Z}(\varphi^k(v), \varphi(v))\, \chi^k(dv, d\hat v) \\ &\le \int d_\mathcal{Z}(\varphi^k(v), \varphi(\hat v)) + d_\mathcal{V}(\varphi(\hat v), \varphi(v))\, \chi^k(dv, d\hat v) \\ &= \mathcal{W}_1(\eta^k, \eta) + \int d_\mathcal{Z}(\varphi(v), \varphi(\hat v))\, \chi^k(dv, d\hat v). \end{aligned} \tag{A.1}$$

The first summand in (A.1) vanishes as $k \to \infty$ since $\eta^k \to \eta$ in $\mathcal{W}_1$, whereas the second summand vanishes as consequence of Lemma A.1 since $\int d_\mathcal{V}(v,\hat v)\, \chi^k(dv, d\hat v) \to 0$. □

A.1 Convergence of subprobability measures

Occasionally it is advantageous to work with subprobability measures. Therefore, we denote by $\mathcal{M}_p(\mathcal{X})$ the set of finite non-negative Borel measures on $\mathcal{X}$ that have finite p-th moments and by $\mathcal{M}^*_p(\mathcal{X})$ the subset of measures with positive mass. We say that a sequence $(\rho^k)_{k\in\mathbb{N}}$ converges in $\mathcal{M}_p(\mathcal{X})$ to ρ if one of the following equivalent conditions holds:

(a) $(\rho^k)_{k\in\mathbb{N}}$ converges weakly to ρ and, for some $x_0 \in \mathcal{X}$, $\lim_{k\to\infty} \int d_{\mathcal{X}}^p(x,x_0)\,\rho^k(dx) = \int d_{\mathcal{X}}^p(x,x_0)\,\rho(dx)$;

(b) for every continuous function $\varphi\colon \mathcal{X}\to\mathbb{R}$ such that, for some $x_0\in\mathcal{X}$ and all $x\in\mathcal{X}$, $|\varphi(x)| \le 1 + d_{\mathcal{X}}^p(x,x_0)$, $\lim_{k\to\infty} \rho^k(\varphi) = \rho(\varphi)$.

Further, when $\rho,\tilde{\rho} \in \mathcal{M}_p^*(\mathcal{X})$ have equal mass, we can consider their p-Wasserstein distance given by

$$\mathcal{W}_p(\rho,\tilde{\rho}) := \rho(\mathcal{X})^{\frac{1}{p}}\,\mathcal{W}_p\left(\frac{\rho}{\rho(\mathcal{X})}, \frac{\tilde{\rho}}{\tilde{\rho}(\mathcal{X})}\right),$$

and similarly define the p-adapted Wasserstein distance $\mathcal{AW}_p$ between measures $\pi,\tilde{\pi} \in \mathcal{M}_p^*(\mathcal{X}\times U\times\mathcal{Y})$ that have equal mass.

Lemma A.3. *Let $p \ge 1$, $\rho \in \mathcal{M}_p^*(\mathcal{X})$, and $(\rho^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{M}_p^*(\mathcal{X})$ with $\lim_{k\to\infty}\rho^k(\mathcal{X}) = \rho(\mathcal{X})$. Then the following are equivalent:*

(i) $(\rho^k)_{k\in\mathbb{N}}$ converges in $\mathcal{M}_p(\mathcal{X})$ to ρ;

(ii) the normalized sequence $\left(\frac{\rho^k}{\rho^k(\mathcal{X})}\right)_{k\in\mathbb{N}}$ converges to $\frac{\rho}{\rho(\mathcal{X})}$ in $\mathcal{P}_p(\mathcal{X})$.

Proof. Since $\lim_{k\to\infty}\rho^k(\mathcal{X}) = \rho(\mathcal{X})$, we have in either case that $(\rho^k)_{k\in\mathbb{N}}$ and the normalized sequence $(\rho^k/\rho^k(\mathcal{X}))_{k\in\mathbb{N}}$ are weakly convergent with limit ρ and $\rho/\rho(\mathcal{X})$, respectively. For some $x_0\in\mathcal{X}$, we then have

$$\lim_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,\rho^k(dx) = \int d_{\mathcal{X}}(x,x_0)^p\,\rho(dx)$$

if and only if

$$\lim_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,\frac{\rho^k(dx)}{\rho^k(\mathcal{X})} = \int d_{\mathcal{X}}(x,x_0)^p\,\frac{\rho(dx)}{\rho(\mathcal{X})}.$$

Thus, the equivalence of (i) and (ii) follows from [27, Definition 6.8]. □

Lemma A.4. *Let $p\ge 1$ and $\mathcal{X}$ be a Polish space. Let $(\rho^k)_{k\in\mathbb{N}}$ be a convergent sequence in $\mathcal{M}_p(\mathcal{X})$ and $(q^k)_{k\in\mathbb{N}}$ be a weakly convergent sequence with $q^k \le \rho^k$ for every $k\in\mathbb{N}$. Then, $(q^k)_{k\in\mathbb{N}}$ converges in $\mathcal{M}_p(\mathcal{X})$.*

Proof. Write ρ and q for the weak limits of $(\rho^k)_{k\in\mathbb{N}}$ and $(q^k)_{k\in\mathbb{N}}$ respectively. Consider the sequence $\tilde{q}^k := \rho^k - q^k \in \mathcal{M}_p(\mathcal{X})$, $k\in\mathbb{N}$, which is also weakly convergent with limit $\tilde{q} := \rho - q$. By Portmanteau's theorem we have

$$\int d_{\mathcal{X}}(x,x_0)^p\,q(dx) \le \liminf_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,q^k(dx),$$

$$\int d_{\mathcal{X}}(x,x_0)^p\,\tilde{q}(dx) \le \liminf_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,\tilde{q}^k(dx).$$

Hence,

$$\begin{aligned}\limsup_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,q^k(dx) &= \int d_{\mathcal{X}}(x,x_0)^p\,\rho(dx) - \liminf_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,\tilde{q}^k(dx)\\ &\le \int d_{\mathcal{X}}(x,x_0)^p\,\rho(dx) - \int d_{\mathcal{X}}(x,x_0)^p\,\tilde{q}(dx)\\ &= \int d_{\mathcal{X}}(x,x_0)^p\,q(dx),\end{aligned}$$

yields $\lim_{k\to\infty}\int d_{\mathcal{X}}(x,x_0)^p\,q^k(dx) = \int d_{\mathcal{X}}(x,x_0)^p\,q(dx)$. □

References

[1] Aurélien Alfonsi, Jacopo Corbetta, and Benjamin Jourdain, *Sampling of probability measures in the convex order by Wasserstein projection*, Ann. Inst. Henri Poincaré Probab. Stat. **56** (2020), no. 3, 1706–1729. MR4116705

[2] Julio Backhoff, Daniel Bartl, Mathias Beiglböck, and Manu Eder, *Adapted Wasserstein distances and stability in mathematical finance*, Finance Stoch. **24** (2020), no. 3, 601–632. MR4118990

[3] Julio Backhoff, Mathias Beiglböck, and Gudmund Pammer, *Existence, duality, and cyclical monotonicity for weak transport costs*, Calculus of Variations and Partial Differential Equations **58** (2019), no. 6, 1–28. MR4029731

[4] J. Backhoff-Veraguas and G. Pammer, *Stability of martingale optimal transport and weak optimal transport*, Ann. Appl. Probab. **32** (2022), no. 1, 721–752. MR4386541

[5] D. Bartl, M. Beiglböck, and G. Pammer, *The Wasserstein space of stochastic processes*, ArXiv e-prints (2021).

[6] M. Beiglböck, P. Henry-Labordère, and F. Penkner, *Model-independent bounds for option prices: A mass transport approach*, Finance Stoch. **17** (2013), no. 3, 477–501. MR3066985

[7] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer, *Approximation of martingale couplings on the line in the adapted weak topology*, Probab. Theory Related Fields **183** (2022), no. 1-2, 359–413. MR4421177

[8] M. Beiglböck and N. Juillet, *On a problem of optimal transport under marginal martingale constraints*, Ann. Probab. **44** (2016), no. 1, 42–106. MR3456332

[9] Mathias Beiglböck, Benjamin Jourdain, William Margheriti, and Gudmund Pammer, *Stability of the weak martingale optimal transport problem*, Ann. Appl. Probab. **33** (2023), no. 6B, 5382–5412. MR4677736

[10] Mathias Beiglböck and Nicolas Juillet, *Shadow couplings*, Trans. Amer. Math. Soc. **374** (2021), no. 7, 4973–5002. MR4273182

[11] Douglas T Breeden and Robert H Litzenberger, *Prices of state-contingent claims implicit in option prices*, The Journal of Business **51** (1978), no. 4, 621–51. MR0538840

[12] Martin Brückerhoff and Nicolas Juillet, *Instability of martingale optimal transport in dimension $d \geq 2$*, Electron. Commun. Probab. **27** (2022), Paper No. 24, 10. MR4416823

[13] Manu Eder, *Compactness in Adapted Weak Topologies*, arXiv e-prints (2019), arXiv:1905.00856.

[14] A. Galichon, P. Henry-Labordère, and N. Touzi, *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, Ann. Appl. Probab. **24** (2014), no. 1, 312–336. MR3161649

[15] Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405. MR3706606

[16] Julien Guyon, Romain Menegaux, and Marcel Nutz, *Bounds for VIX futures given S&P 500 smiles*, Finance and Stochastics **21** (2017), no. 3, 593–630. MR3663638

[17] David Hobson and Dominykas Norgilas, *Robust bounds for the American put*, Finance and Stochastics **23** (2019), no. 2, 359–395. MR3933425

[18] David G Hobson and Dominykas Norgilas, *The left-curtain martingale coupling in the presence of atoms*, The Annals of Applied Probability **29** (2019), no. 3, 1904–1928. MR3914560

[19] B. Jourdain and W. Margheriti, *One dimensional martingale rearrangement couplings*, ESAIM Probab. Stat. **26** (2022), 495–527. MR4525181

[20] B. Jourdain, W. Margheriti, and G. Pammer, *Lipschitz continuity of the wasserstein projections in the convex order on the line*, Electron. Commun. Probab. **28** (2023), no. 18, 13. MR4596533

[21] Benjamin Jourdain and William Margheriti, *A new family of one dimensional martingale couplings*, Electronic Journal of Probability **25** (2020). MR4171389

[22] Nicolas Juillet, *Peacocks parametrised by a partially ordered set*, Séminaire de Probabilités XLVIII, Lecture Notes in Math., vol. 2168, Springer, Cham, 2016, pp. 13–32. MR3618125

[23] Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR1876169

[24] Robert P Kertz and Uwe Rösler, *Complete lattices of probability measures with applications to martingale theory*, Lecture Notes-Monograph Series (2000), 153–177. MR1833858

[25] Paul André Meyer and WA Zheng, *Tightness criteria for laws of semimartingales*, Annales de l'IHP Probabilités et statistiques **20** (1984), no. 4, 353–372 (eng). MR0771895

[26] Gudmund Pammer, *A note on the adapted weak topology in discrete time*, Electron. Commun. Probab. **29** (2024), no. 1, 13. MR4696436

[27] C. Villani, *Optimal transport. old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, 2009. MR2459454

[28] Johannes Wiesel, *Continuity of the martingale optimal transport problem on the real line*, Ann. Appl. Probab. **33** (2023), no. 6A, 4645–4692. MR4674061