We thank the reviewers for their comments to improve the paper. The main contributions have been well identified, i.e. using soft-constraints to control for arbitrage opportunities, and the NN-based correction of a quant finance-based prior.

Broader impact and relevance to the NeurIPS community (R3/R4): We agree that we study a specific problem, but our primary subject area is “Applications: Quantitative Finance and Econometrics” and the concern applies to most applications papers. For the finance industry, second only to tech as sponsor of NeurIPS2019, and in our opinion a significant part of the community (maybe explaining the subject area’s existence), the problem’s relevance cannot be overstated. Most banks and hedge-funds use IV surfaces (IVSs) and need such models. And derivatives are important not only to the financial industry, but to any firm aiming at quantitatively manage risk via hedging, e.g., agricultural commodities for farmers, fuel consumption for airlines, FX rates for companies having multi-currency activities, etc. We will clarify by bringing Appendix E.2 (our current “broader impact section”, not detailed enough) to the main text, expand on it with such a discussion along with numbers relevant to derivatives markets and IVSs.
Additional experiments (R1/R4/R5): We tested, our approach works both in high-vol periods (e.g., 09/2008) and with “W” shapes. We will add figures/tables to the appendix. We will also add computation times, an experiment with spread weighting for $\mathcal{L}_0$, and we are open to suggestions for additional benchmarks beyond SVJ/Bates & SSVI (see Table 5).
Training & data (R4/R5): The information is in Appendix C because of the page limit, but we will add the key details in the main paper, and expand the section on the real data pre-processing. Data is from optionmetrics (option/underlying prices & r_f), historically recording the best 15:59EST bid/ask. We use midquotes everywhere, the put-call parity & r_f to extract div yields, Brent’s method to extract the IV, and CBOE’s S&P500 options are European (i.e., no pricer).

<u>Reviewer #1</u> **Theoretical justification:** Apologies, we do not understand. Economic theory states that put-call prices should be arbitrage free; anomalies would otherwise be exploited until they disappeared. Thus, models like ours, mathematically guaranteeing the absence of arbitrage, albeit via soft-constraints, are theoretically justified.
<u>Reviewer #3</u> **Yu 2018 [60]:** Thanks for letting us know. We cited this PhD thesis for the gated NN ideas published in [59] in which constraints on the price surface are used. But we were not aware of the work on IVS constraints, this part of the thesis was not published. Albeit our work was done independently, we will correct the over-claim and give credit.
Final layer: Good remark, the “1 + ” allows the NN to be initialized at 1 by setting $\alpha = 1$, $W_{n+1} = 0$, $b_{n+1} = 0$, i.e. start from the prior. We tried the sigmoid, and tanh appeared to work better. We will discuss this in the manuscript.
C6: Apologies if we misunderstand, but if $\partial^2\omega/\partial k\partial k = 0$ for large k, then ω is at most linear asymptotically.
SVI not SOTA: Agreed, we actually use the SSVI with power-law parametrization (see Appendix B.2), we will change SVI into SSVI everywhere. The Heston-like parametrization gave worst results and thus was excluded. The eSSVI, i.e. making ρ a function of the ATM total variance, can be included easily (already implemented). But eSSVI and other SVI extensions are, in our opinion, less popular in the industry. We will gladly add results for any SOTA method of interest.
<u>Reviewer #4</u> **Feed-forward NN:** We are not sure whether you mean replacing our NN correction by a basis expansion-based NN correction, or replacing the whole approach altogether by a basis expansion. The latter has been done with cubic splines in [16]. But it doesn’t allow to extrapolate, and the training algorithm requires a fixed grid and has convergence issues when the grid is too dense (i.e., limited to IVSs with few options). For the former, it is surely possible, albeit our final layer has some advantages (see answer to R3). For this application, what matters is mostly model capacity and ease of computing partial derivatives to penalize the NN plugged on top of the prior via soft constraints. And while both shallow and deep NNs can approximate continuous function with arbitrary precision, deep NNs can reach the same accuracy as shallow ones with less neurons (see e.g., Thm 1, Foggio et al., PNAS 2020).
Other NN baselines: We implemented [13] but it didn’t work well on the S&P500 data. We can include these results, as well as any baseline of your choosing. Typically, NN approaches for the price surface struggle with deep OOM options (small price errors become large IV errors). This is not the case with ours which controls the asymptotic IV.
<u>Reviewer #5</u> **Ojective function:** Spread weighting for $\mathcal{L}_0$ is an excellent idea, overlooked for simplicity as it didn’t fundamentally change the results, but we can add a figure. In our experience, vega weighting helps to fit prices, where small errors can lead to large IV errors. Also, not penalizing models for predictions within the spread is important. In our opinion, this idea is similar to that for $\mathcal{L}_{C5}$; extremely relevant in the real world, somewhat beyond the scope of this paper. Regarding $\mathcal{L}_{C6}$, its application in regions without tradeable strikes is extreme by design. Because $C6$ describes the asymptotic behavior, we aimed to regularize/“clip” the NN when $|k| \to \infty$, hence the strikes much larger than observed. We agree that one could be maturity dependent, i.e. use a τ-dependent $\mathcal{K}_6^\tau$, and will update accordingly.
RMSE/MAPE attribution: It is difficult to disentangle the two cases in real data, especially given that we use midquote IVs. We will add to the appendix and experiment where we created synthetic data with a model generating arbitrage opportunities, applied our method, and compared metrics knowing the “true” values for $C_{4,5,6}$.
Results: Thanks for the comment on short maturities, we will add loss statistics as well as a figure to the appendix. Note also that, in the figures about real data, the first maturity (0.02) corresponds to 7 days. We are computing arbitrage losses on the grids $\mathcal{K}_{C_{45}}$ and $\mathcal{K}_{C_6}$. On a grid point, unless it is “close” to a real option, nothing constraints the NN.
Figures 2 & 3: Thanks for noticing, we will update accordingly.
Additional feedback: We highly appreciate the efforts in providing detailed and insightful comments. We drafted point by point detailed answers but unfortunately couldn’t fit them on one page.