VOLATILITY AND RISK MANAGEMENT IN AGRICULTURAL COMMODITY MARKETS

BY

SIYU BIAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Applied Economics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Teresa Serra, Chair and Director of Research
Professor Philip Garcia
Professor Scott Irwin
Assistant Professor Anabelle Couleau, Universidad EAFIT

ABSTRACT

In this thesis we examine agricultural commodity market volatility in modern electronic futures markets from different perspectives. First, we examine the informational shock of news releases in agricultural futures markets by decomposing observed volatility into efficient and noise components. Second, we investigate if machine learning algorithms can capture nonlinear trends in the informational volatility and generate better volatility forecasting results than simple linear models. Third, we assess the liquidity stress in agricultural futures markets induced by both heightened volatility and unexpected price jumps.

The first essay demonstrates the importance of disentangling efficient market price variance from microstructure noise in the context of public announcements. Our contribution to the literature lies in the use of a Markov Chain framework that solves the shortcomings of previously used approaches and the analysis of the impacts of recent public information release policy changes affecting agricultural markets. We show that previous literature, by ignoring noise, has underestimated the informational content of the United States Department of Agriculture (USDA) reports, overlooked the costs of price discovery, and produced biased estimates of the time required for the market to absorb the information. Overall, findings are highly relevant to public policy, particularly USDA, and have implications for market design.

The second essay provides a systematic comparison between linear Heterogeneous Autoregressive type models and nonlinear machine learning-based models for forecasting realized volatility. To further assess the need for nonlinear models to produce better forecasting results, we propose a novel hybrid artificial neural network model based on residual learning to separate forecasts into linear and nonlinear components. We use these models to forecast the daily realized volatility in highly traded commodity and, for comparison purposes, stock futures markets. We

also implement different tests and forecast accuracy measures to compare among the different models. Findings show that the best (worst) out-of-sample forecast performance measures are usually achieved by using linear (nonlinear) models. The decomposition of hybrid neural network models' forecasts shows that the hybrid models essentially rely on their linear components to produce forecasts. Therefore, we recommend linear models over nonlinear machine learning algorithms when forecasting realized volatility.

The third essay evaluates the impact of extreme price volatility on the funding risk faced by agricultural futures traders. Extreme price movements significantly challenge traders' ability to comply with the exchange's margin requirements and impose funding risks on all market participants. We propose a new intra-horizon conditional Value-at-Risk method to measure the funding risk faced by agricultural futures traders. Unlike previous methods, our approach incorporates both asymmetric jumps and seasonal volatility components. Findings suggest that funding risks in agricultural futures markets are notably seasonal and heavily influenced by price jumps. Also, we find that margin requirements in agricultural futures markets are inadequate to cover funding risks faced by clearinghouses in the presence of frequent price return jumps.

ACKNOWLEDGEMENTS

I would like to thank my advisors, Teresa Serra and Philip Garcia, for your guidance during my journey as a Ph.D. student. Your advice helped me grow as a researcher. Thank you for your time and efforts. The thesis also benefited from comments and suggestions made by Scott Irwin and Anabelle Couleau. Thank you for willing to be part of my dissertation committee. I would also like to thank all the faculty members and non-faculty staff who make our working environment efficient and enjoyable.

Completing this work would have been more difficult without the support from my family and friends. Thank you for always cheering me up and encouraging me. Finally, I would like to thank Xinyue, my fiancée, who always stay with me and support me in whatever I pursue. Thank you for your continued encouragement during this journey.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Gaining a deeper understanding on the dynamics and information content of commodity futures market volatility can significantly improve the decision-making process for policy makers, market participants, and exchanges. Several issues are particularly relevant. First, the existence of microstructure noise confounds the responses of prices to information by introducing a noise component in volatility observed at high frequency. Moreover, agricultural futures markets exhibit characteristics that may induce non-linearities in the time-series behavior of volatility, such as seasonality, announcement effects, and time-to-maturity effects. These may impact our ability to predict future volatility. Also, agricultural futures volatility has been shown to contain a jump component induced by large discontinuous price returns movements that may have different implications for risk management than the continuous component of volatility. This dissertation contains three essays that shed light on these three aspects of volatility in modern agricultural futures markets. The first essay investigates the effect of public announcements on the information and noise components of market volatility. The second essay focuses on the role of non-linearities in volatility forecasting. The third essay examines how jumps and seasonality affect funding risks faced by market participants.

The first essay "New evidence on market response to public announcements in the presence of microstructure noise" analyzes the impact on the efficient price return variance of the changes in the release time of public information in the era of automated trading. Due to the contamination of noise, efficient variance must be obtained through the decomposition of observed variance. Previous literature decomposes observed intraday return variance into efficient variance and noise

using a state-space model (Hautsch, Hess and Veredas 2011). Yet, this parametric approach is heavily dependent on the statistical distributional assumptions and fails to address the interdependence between efficient and noise return variance, which leads to biased variance decomposition results (Diebold and Strasser 2013). We solve the two problems by using the Markov Chain (MC) framework proposed by Hansen (2015) which makes minimal assumptions on the underlying return processes and allows for covariance between efficient returns and noise. Through the decomposition results, we study the effects of the United States Department of Agriculture (USDA) public news releases on the efficient and noise volatility components in corn and soybean markets. We compare the effects before and after the agency transitioned from releasing during a futures market trading halt to releasing during normal trading hours. We also demonstrate changes in price discovery speed by comparing efficient variance absorption time between the two release periods.

The second essay "Can machine learning algorithms beat linear models in realized volatility forecasting?" assesses the need of adopting complex and adaptive machine learning (ML) algorithms in realized volatility (RV) forecasting across different markets. The linear Heterogeneous Autoregressive (HAR) model proposed in Corsi (2008) has become the workhorse in RV modeling due to its simplicity and consistent forecasting performance in empirical applications (Qiu et al. 2019). However, numerous forms of nonlinearities in futures markets may only be captured through flexible and adaptive non-linear methods. Prior studies apply ML methods to volatility forecasting and generate superior forecasting performances (Luong and Dokuchaev 2018; Santamaría-Bonfil, Frausto-Solís and Vázquez-Rodarte 2015; D'Ecclesia and Clementi 2019). However, most studies fail to provide systematic performance evaluations of the ability of ML to forecast RV under general conditions. To fill these gaps, we evaluate the

forecasting performance of multiple ML algorithms and linear models across a wide range of futures markets and apply rigorous robustness checks. Besides commonly adopted ML models, such as support vector machine and random forests, we also design a hybrid ANN model using residual connections inspired by He et al. (2015) to show the relevance of nonlinear functional forms.

The third essay "Jumps, seasonality, and funding risks in futures markets" examines the funding risk faced by futures traders under stressful market events in corn, soybean, and wheat futures markets. Funding risks emerge as a consequence of the daily settlement procedure where market participants must comply with margin calls in the presence of unfavorable price movements to maintain their market positions. An accurate measurement of funding risks is essential for traders to better prepare for extreme losses. Also, it can help exchanges and clearinghouses set optimal margin requirements to improve resiliency and systemic stability when facing large price jumps. A concept similar to funding risks, denoted as maximum margin liability, was introduced in previous literature (Shi and Isengildina-Massa 2021). However, it focuses on longer holding periods with weekly returns and suffers from the Peso problem by ignoring possible extreme returns not observed in historical returns. To overcome these shortcomings, we use intra-horizon Conditional Value-at-risk (CVaR) proposed by Farkas, Mathys and Vasiljević (2021) to measure funding risks, allowing for price jumps and seasonal volatility. Not only do we provide with more accurate funding risk estimation results, but we also decompose funding risk results into diffusive, seasonal, and jump components. This allows us to evaluate the impact of each component on funding risks. Moreover, we assess the efficacy of the exchange margin requirements by converting funding risks into dollar values per contract and comparing them against margin requirements in each market.

1.1 References

Corsi, F. 2008. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7(2):174–196. Available at: https://academic.oup.com/jfec/article-lookup/doi/10.1093/jjfinec/nbp001.

D'Ecclesia, R.L., and D. Clementi. 2019. "Volatility in the stock market: ANN versus parametric models." *Annals of Operations Research*. Available at: http://link.springer.com/10.1007/s10479-019-03374-0.

Diebold, F.X., and G. Strasser. 2013. "On the Correlation Structure of Microstructure Noise: A Financial Economic Approach." *Review of Economic Studies* 80(4):1304–1337.

Farkas, W., L. Mathys, and N. Vasiljević. 2021. "Intra-Horizon expected shortfall and risk structure in models with jumps." *Mathematical Finance* 31(2):772–823. Available at: https://onlinelibrary.wiley.com/doi/10.1111/mafi.12302.

Hansen, P.R. 2015. "A Martingale Decomposition of Discrete Markov chains." *Economics Letters* 133:14–18. Available at: http://dx.doi.org/10.1016/j.econlet.2015.04.028.

Hautsch, N., D. Hess, and D. Veredas. 2011. "The impact of macroeconomic news on quote adjustments, noise, and informational volatility." *Journal of Banking and Finance* 35(10):2733–2746. Available at: http://dx.doi.org/10.1016/j.jbankfin.2011.03.004.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Deep Residual Learning for Image Recognition." Available at: http://arxiv.org/abs/1512.03385.

Luong, C., and N. Dokuchaev. 2018. "Forecasting of Realised Volatility with the Random Forests Algorithm." *Journal of Risk and Financial Management* 11(4):61. Available at: http://www.mdpi.com/1911-8074/11/4/61.

Qiu, Y., X. Zhang, T. Xie, and S. Zhao. 2019. "Versatile HAR model for realized volatility: A

least square model averaging perspective." *Journal of Management Science and Engineering* 4(1):55–73. Available at: https://doi.org/10.1016/j.jmse.2019.03.003.

Santamaría-Bonfil, G., J. Frausto-Solís, and I. Vázquez-Rodarte. 2015. "Volatility Forecasting Using Support Vector Regression and a Hybrid Genetic Algorithm." *Computational Economics* 45(1):111–133. Available at: http://link.springer.com/10.1007/s10614-013-9411-x.

Shi, R., and O. Isengildina-Massa. 2021. "Costs of Futtures Hedging in Corn and Soybean Markets." *Journal of Agricultural and Resource Economics* Prepint. Available at: https://ageconsearch.umn.edu/record/313311/files/JARE47.2.10%2CShi%2C390-409S.pdf?version=1.

CHAPTER 2

NEW EVIDENCE ON MARKET RESPONSE TO PUBLIC ANNOUNCEMENTS IN THE PRESENCE OF MICROCTRUCTURE NOISE

2.1 Introduction

The market effects of public news releases are commonly measured by their impact on price returns variance. Differences in how traders interpret the newly released information and their expectations fundamentally drive trading volume (Black 1986). This leads to a price discovery process where market prices change to reflect the value of new information. While informed traders are correct in their expectations on average, noise traders are not and introduce noise into observed prices. This results in noise variance that blurs changes in efficient price returns. Using a dynamic trading model with informed and noise traders, Collin-Dufresne & Fos (2016) show that informed traders' optimal strategy is to trade more aggressively when noise variance is large. The rationale is that noise variance results in price inefficiencies that offer profit opportunities to informed traders. Intensification of informed traders' activity as a response to noise variance accelerates the rate at which information is impounded in prices and thus efficient price discovery. This implies dependency between the efficient price return variance and noise variance that may be particularly relevant when public information releases occur during normal trading hours, allowing for real-time trading of information.

During real-time releases of public information, noise variance in electronically-traded markets is expected to be large and involve substantial profit opportunities. Collin-Dufresne & Fos' (2016) framework, however, requires a more nuanced characterization of market participants in this setting. Recent literature on the release of public reports (Budish, Cramton and Shim 2014;

Foucault, Hombert and Roşu 2016; Foucault, Kozhan and Tham 2017; Roşu 2019) has shown that speed in responding to new information is critical to profit from quotes that have not been updated and still reflect old information. In this context, not only noise traders but also slow informed traders offer attractive profit opportunities to fast informed traders and can promote faster efficient price discovery.

Hautsch, Hess, and Veredas (2011) acknowledged that the effects of public news releases on the efficient and noise-driven variance were widely unexplored. They adopted a state-space model to decompose conditional variance of intraday returns into efficient price return variance and noise return variance, but failed to show how noise return variance can be used to improve policy decisions. Since Hautsch et al. (2011) no research has assessed the impacts of public news releases on the two components of price variance. Not surprisingly, the operations research literature has not incorporated noise variance estimates in its range of analytical tools that can improve decision-making and problem-solving processes. In contrast, it has viewed noise variance as something that needs to be filtered out and has no value in itself. Our article shows, for the first time, the usefulness of noise variance estimates as an analytical tool for policy makers. First, we show how to estimate the cost of market microstructure frictions based on microstructure noise variance estimates. Second, by comparing the efficient price discovery process represented by the efficient price variance with the cost of microstructure frictions, we show how policy makers can draw a better understanding about the impacts of public information releases on electronic markets.

Both the method and the period studied by Hautsch et al. (2011) make the work dated. Their method suffers from two main shortcomings that can be addressed using more state-of-the-art approaches. First, their parametric method is highly dependent on the statistical distributional assumptions made. Second, their model fails to address the interdependence between efficient

price return variance and noise. Ignoring this interdependence leads to biased variance decomposition results (Diebold & Strasser, 2013; Hansen & Lunde, 2006). Increased availability of intraday high frequency financial data over past decades has spurred the development of new methodological approaches that have greatly improved our knowledge of financial price variance and its dynamic properties (Hansen and Lunde 2011). Nonparametric realized variance (RV) measures have opened a new era of modeling and forecasting financial price variance. Key to this progress has been the recognition of the RV failure to converge to the efficient price variance in the presence of microstructure noise, which reflects imperfections and frictions caused by the market structure and the price setting behavior of market participants (Bandi and Russell 2008). Hansen (2015) developed a Markov Chain (MC) framework that decomposes intraday observed price return variance into efficient price and microstructure noise return variance. We use the Markov Chain (MC) framework to provide an innovative way to study how intraday price return variance responds to public information announcements and provide more nuanced conclusions about the speed and the costs of the price discovery process relative to previous methods. The MC framework offers two key advantages relative to the model used by Hautsch et al. (2011). First, assumptions underlying the MC framework are minimal (discrete returns follow an ergodic and homogenous Markov process with certain lag orders). Second, the MC framework allows for the covariance between efficient price return and noise, which solves the interdependency issue in Hautsch et al. (2011).

Our paper is the first to study the impacts of public announcements on price variance using the nonparametric method by Hansen (2015) and to show the usefulness of microstructure noise estimates as an analytical tool for decision marking. We apply this model to study the effects of the United States Department of Agriculture (USDA) public information releases on corn and

soybean futures prices. To provide a benchmark for comparison and augment our methodological contribution, we replicate the variance decomposition model used by Hautsch et al. (2011) using our dataset. In 2012, USDA changed its crop reports' release policy from a market trading halt to a release during normal trading hours (real-time release). By assessing the impacts of this policy change on efficient price variance and microstructure noise, we contribute to current debates on how public information should be released in the context of fast electronic markets (Budish et al. 2014).

Our analysis is based on nearby futures transaction prices (tick data) for corn and soybeans, time-stamped to the nearest second for the period from July 2009 to May 2017. We show the MC framework to be superior to past methods (Hautsch et al. 2011). Results suggest that microstructure noise represents more than 50% of the observed variance on announcement days and roughly 75% on non-announcement days. Hence, microstructure noise is large and important differences in the composition of variance between announcement and non-announcement days exist. We show that by ignoring microstructure noise, previous papers have produced biased estimates of the informational content of the USDA reports. We find USDA releases increase the informational content of variance by 10-16% in the real-time trading era, and 1-2% in the trading halt era relative to non-announcement days.

In terms of incorporating changes in market fundamentals, we demonstrate that a greater amount of information can be absorbed in either the same or less time in the real-time release era relative to the trading halt. However, consistent with Collin-Dufresne & Fos (2016), while real-time release provides the conditions for a faster price discovery, the market bears a cost in the form of microstructure frictions. In the spirit of Campbell & Kyle (1993), we calculate the dollar value of the noise-driven price innovations. We find the relevant increases in the cost of noise between

the two release eras are proportionally large relative to the improvements in the speed of price discovery. Placing the relevance of speed in perspective, trading halts appear to be a more effective market mechanism to trade public information, particularly since public information is released at a very low frequency (e.g. once every month).

2.2 Relevant Literature

In the ideal scenario where intraday prices are observed continuously and free from measurement error, realized variance (RV) yields a perfect estimate of price return variance (Merton 1980). However, the presence of microstructure noise in high frequency financial data complicates and challenges the reliability of RV. Since Zhou (1996), the statistics and financial statistics literature has pioneered the characterization of the empirical properties of microstructure noise, including its time and efficient price dependency, and proposed methods that allow to disentangle noise from RV. A first procedure subsamples and/or pre-averages data that are then used to produce noise-free variance estimates (Bandi & Russell, 2008; Hansen & Lunde, 2006). Another method uses nonparametric variance measures that do not require data pre-filtering to capture the efficient price returns variance (Barndorff-Nielsen et al. 2009). Hansen & Horel (2009) establish a consistent nonparametric estimator of integrated variance (IV—a population measure of return variance) based on the Markov Chain (MC) properties of discrete observed price changes. In a similar fashion, Hansen et al. (2015) and Hansen (2015) propose a MC framework to decompose observed price return variance into its efficient price return variance and noise-related components. The MC framework can be applied to non-filtered discrete (resulting from the tick size) high-frequency data and it is thus useful to investigate questions that cannot be easily examined using subsampling and/or pre-averaging techniques.

As a discipline concerned with quantitative decision problems, the operations research (OR) literature in this area has relied on the advances in RV measurement to improve financial management decision tools. However, progress has been slow and only reached a limited scale Moallemi and Sağlam (2013) consider the optimal execution problem of a risk-neutral human investor who must decide between market orders and limit orders to sell a stock share. They focus on the costs of immediacy of trading that are based on price variance measures noise-filtered in order to represent humans' trading horizon. They use a parametric filtering procedure that ignores both the time and efficient price dependency of microstructure noise. Sévi (2014) provides better RV forecasting models by extending the Heterogeneous Autoregressive (HAR) model (Corsi 2008) to allow for positive and negative semivariances, as well as continuous and discontinuous RV components. To filter RV from microstructure noise Sévi (2014) relies on the two-scales RV approach developed by Zhang, Mykland and Ait-Sahalia (2005), which is based on the unrealistic assumption that microstructure noise is independent from the efficient price. Based on the same two-scales RV approach, Barunik, Krehlik and Vacha. (2016) provide enhanced modeling and forecasting price volatility tools based on GARCH models. Specifically, they propose a new realized GARCH model with multiple time-frequency decomposed realized variance measures, as well as a Jump-GARCH model. Bajgrowicz et al.(2016)develop an improved method to identify discontinuities (jumps) in RV that overcomes multiple testing issues affecting previous research. While allowing for microstructure noise, they rely on the questionable assumption that microstructure noise is identically and independently distributed. More recently, Ho and Xin (2019) propose a Kalman Expectations Conditional Maximization (KECM) covariance estimator for high-frequency data that allows for asynchronous returns, market microstructure noise and jumps. They model microstructure noise under the assumption of independence from the efficient price

variance, which yields to large estimation errors when the assumption is violated. Amaya, Begin and Gautier (2021) are the first to document the use of noise-robust high-frequency options RV estimates. Their approach relies on existing filtering methods that differ in their degree of complexity and include more recent techniques allowing for efficient price dependency.

While recognizing the inconsistency of RV estimators in the presence of noise and the need to filter out the impacts of the latter on the former (Gkillas et al., 2019), the OR literature has remained relatively silent on the economic relevance of microstructure noise per se and how it can be used to improve decision-making. Our paper is pioneer not only in the OR literature, but also in the wider statistics and financial literature by using Hansen et al. (2015) proposal to document the economic relevance of microstructure noise as an estimate of the costs of market microstructure frictions during releases of public information. By comparing the efficient price discovery process represented by the efficient price variance with the market microstructure frictions, we show how policy makers can derive more nuanced conclusions on the impacts of information release policies. Our empirical application focuses on USDA announcements and their effects on corn and soybean markets.

The literature that investigates intraday effects of USDA announcements in agricultural commodity derivatives markets is limited. Lehecka et al. (2014) study announcement effects on observed corn futures price return variation using intraday data (sampled every minute) for a period (July 2009 to May 2012) when reports were released outside of trading hours. They find strongest reactions occur immediately after the market opens and persist for about 10 minutes. Chicago Mercantile Exchange (CME) concerns over losing trading volume to other exchanges, and USDA's desire to release their reports when markets are more liquid led to changes in the release. For decades, reports were released in a morning or overnight trading halt, providing participants

the opportunity to digest the information prior to the open of the market. In May 2012, the trading halt was eliminated by CME and in January 2013, the USDA changed the release time to 11:00 AM Central Time. Joseph and Garcia (2018) study the intraday (minute-by-minute) effects of USDA reports under the new release policy in the soybean futures market. They find that observed soybean price returns variance remains above normal levels for about one hour. Adjemian and Irwin (2018) compare the duration of the USDA announcement effects between different release policies in the corn, soybeans, and wheat markets. Using minute-level data, they find that grain markets absorb 90% of the additional variance brought by the USDA announcements in a little less than an hour under both the trading halt and the real-time releases. They also report that real-time trading increased observed variance by two to three times relative to the halt-period in the first minutes of trading.

These articles derived the intraday effects of USDA announcements by comparing measures of observed variance for specific time intervals on days with and without announcements. Comparisons are based on ratios of observed variance on announcement to non-announcement days. These ratios only allow inference of the effects of public information on the efficient market price in the absence of microstructure noise, or when the composition of observed variance does not change across announcement and non-announcement days. However, this assumption is not likely to hold since microstructure noise in other markets appears to be substantial and also a function of the efficient price return variance (Hansen and Lunde 2006; Hansen et al. 2015). As a result, separating the noise from the efficient price return variance is essential to a better understanding of the magnitude and duration of public announcement effects. While real-time trading of information may result in a quicker and more transparent price discovery process, traders' uncertainty about the implications of the new information may increase pricing mistakes, motivate

run games by algorithmic traders (Budish et al. 2014), and increase transactions costs (Adjemian & Irwin, 2018). Hence, it may result in increased microstructure noise as the imperfections of the trading process arise more intensely during such tumultuous periods. Disentangling efficient price variance from noise may thus be especially relevant in assessing the implications of USDA reports traded in real-time.

2.3 Method

High-frequency intraday price returns are a noisy measure of efficient price returns. The literature has proposed different techniques to separate noise from intraday returns. A common approach is to re-sample the data at a frequency that is low enough to eliminate microstructure noise. An alternative uses statistical models such as autoregressive models, to filter out microstructure noise under the assumption that noise is a stationary process with serial correlation. Neither resampling techniques nor statistical models can accurately capture efficient price variance over short intraday time intervals. For instance, Hansen and Lunde (2006) identify that resampling every 5 minutes is required to effectively filter out microstructure noise, making the method inappropriate to measure efficient price return variance in shorter time windows which are relevant in modern electronic markets. Statistical models require enough data to be estimated, which may not always be available in short intraday intervals. Statistical models also often overlook the time-varying dependence between noise and efficient returns (Bandi and Russell 2008) and generally suffer from endogeneity issues (Hansen and Lunde 2006; Diebold and Strasser 2013).

We use the Markov chain (MC) framework proposed in Hansen et al. (2015) and Hansen (2015) to assess corn and soybean futures price returns variance every minute. The procedure addresses endogeneity and allows analysis at short intraday intervals. It is based on the efficient market theorem which implies efficient price returns are a martingale process (Fama 1970; LeRoy 1989;

McCauley, Bassler and Gunaratne 2008). Assuming observed intraday returns follow a MC and using its transition matrix, the martingale component is estimated by the long-term conditional projection of the intraday return series. Given current returns, the transition matrix reflects the likelihood of future returns and allows for the development of an infinite (long-term) forecast, which represents the efficient price returns. The remaining component of the Markov chain is microstructure noise, which has been shown to be stationary and serially correlated. In effect, the observed return process can be decomposed into the permanent and transitory components. Any transitory price changes such as those related to liquidity-oriented trades or undirected trades will be included in the microstructure noise component.

The procedures to develop the variance estimators are provided in more detail below following a discussion of the MC framework. Within each minute following a USDA report, let X_t be a one-dimensional process of irregularly spaced (tick) transactions futures prices at time *t.* Then, the following martingale decomposition for X_t is assumed:

$$X_t = Y_t + \mu_t + U_t \tag{2.1}$$

where Y_t represents the efficient latent price vector, μ_t is X_t's deterministic mean trend and U_t a stationary, ergodic and bounded process representing noise.[1] Based on the natural filtration $\mathrm{F}_t = \sigma(X_t, X_{t-1}, \dots)$, $Y_t \coloneqq \lim_{h\to\infty} E\,(X_{t+h} - \mu_{t+h}\,|\mathrm{F}_t)$.[2] By taking the first difference of observed prices, the decomposition of observed returns can be expressed as:

$$\Delta X_t = \Delta Y_t + \Delta U_{t,} \tag{2.2}$$

[1] A graphic illustration of hypothetical Y_t and X_t values is presented in figure A.1 in Appendix A.

[2] This is analogous to the Beveridge-Nelson (1981) approach to extract the stochastic trend of the ARIMA processes (Hansen 2015).

where ΔY_t is a martingale process representing the efficient price returns that would prevail in the absence of market friction. The sum of squared ΔY_t can consistently measure the IV within the minute.

The vector of high-frequency observed returns ΔX_t is assumed to be ergodic and distributed as an order $k < \infty$ homogenous MC, with finite $S < \infty$ states. The assumption of homogeneity can be relaxed by adopting a large k which allows an inhomogeneous MC to be approximated by a homogeneous MC. Under this assumption, it can be shown that ΔY_t is a Markov process with S^{k+1} possible states (Hansen 2015). The finite S assumption implies that the returns vector takes a finite number of values at time t. This is compatible with the tick size, which imposes a minimum (discontinuous) price movement on agricultural commodity futures prices, and the daily price limits that restrict the maximum change that prices can experience within a day. Let $\Delta \mathcal{X}_t = \{\Delta X_{t-k+1,\dots}, \Delta X_t\}$ be an ordered set whose possible values are indexed $\boldsymbol{x}_s$, s=1,…,S^k, with $\Delta \mathcal{X}_t = \boldsymbol{x}_s$ being a vector representing the observed state for $\Delta \mathcal{X}$ at time t.

The transition matrix that identifies the probabilities of price returns moving from one state to another can be characterized as a $S^k \times S^k$ matrix P, with each element defined as:

$$P_{r,s} = Pr(\Delta \mathcal{X}_{t+1} = \boldsymbol{x}_s | \Delta \mathcal{X}_t = \boldsymbol{x}_r), \text{for } r, s = 1, \dots S^k. \tag{2.3}$$

The homogenous MC assumption implies that the transition matrix P is stationary within each minute. A discrete, stationary and homogeneous MC framework is justified when the underlying

process (ΔY_t) asymptotically follows a continuous time Markov process (Hansen and Horel 2009).[3] The stationary distribution of *P*, denoted π, is assumed to be uniquely defined as $\pi' P = \pi'$. The matrix containing π in its main diagonal is $\Lambda_\pi = diag(\pi_1, \dots, \pi_{S^k})$. From the transition matrix we can derive the fundamental matrix which identifies the number of times observed returns are in state $s = j$ if they started in state $s = i$. This fundamental matrix permits projections of observed returns to infinity, thus allowing us to derive the efficient price returns. Following Kemeny and Snell (1983), the fundamental matrix (*Z*) can be computed as:

$$Z = \left(I' - P + \Pi\right)^{-1}, \tag{2.4}$$

where $\Pi = l\pi'$, $l = (1, \dots, 1)'$ and π' represents each row of Π. Given observed returns,[4] *P* can be estimated by maximum likelihood (Hansen and Horel 2015) as:

$$\hat{P}_{r,s} = \frac{\sum_{t=1}^{n} 1_{\{\Delta X_t = x_s,\ \Delta X_{t-1} = x_r\}}}{\sum_{t=1}^{n} 1_{\{\Delta X_{t-1} = x_r\}}} \quad r, s = 1, \dots, S, \tag{2.5}$$

where $\sum_{t=1}^{n} 1_{\{\Delta X_{t-1} = \boldsymbol{x}_r\}}$ counts the number of events in the state $\boldsymbol{x}_r$ at time *t*-1, and $\sum_{t=1}^{n} 1_{\{\Delta X_t = \boldsymbol{x}_s,\ \Delta X_{t-1} = \boldsymbol{x}_r\}}$ counts the number of events initially in state $\boldsymbol{x}_r$, transitioning to state $\boldsymbol{x}_s$.

---

[3] Hansen and Horel (2009) show by exhaustive sampling ΔX_t and removing zero returns, the discrete Markov chain with a finite state space is asymptotically equivalent to a continuous time Markov chain. Therefore, in practice, the results of discrete Markov Chain estimators applied on discrete high-frequency returns are asymptotically equivalent to the results based on continuous time settings.

[4] Following Hansen and Horel (2009), we artificially extend the samples with *k* observations. Specifically, if $k = 4$ and the sample of returns is $r_1, \dots, r_n$, we extend the sample with 4 observations and use the sample $r_1, \dots, r_n, r_1, r_2, r_3, r_4$ to estimate the empirical transition matrix of the MC process. This sample extension technique eliminates any absorbing states. Thus, the empirical transition matrix becomes a member of the class of ergodic transition matrices.

$\hat{\pi}$ is obtained as the eigenvector of $\hat{P}$, the maximum likelihood estimator in (2.5), which allows us to derive $\hat{\Pi} = l\hat{\pi}'$, $\hat{\Lambda}_\pi = diag(\hat{\pi}_1, \dots, \hat{\pi}_{S^k})$ and $\hat{Z} = \left(I - \hat{P} + \hat{\Pi}\right)^{-1}$. The empirical estimation of the efficient return variance ($Var^{\#}(\Delta Y_t)$) and the observed return variance ($Var^{\#}(\Delta X_t)$) is presented in equations (2.6) and (2.7), respectively. The difference $Var^{\#}(\Delta X_t) - Var^{\#}(\Delta Y_t)$ in equation (2.8) is our measure of microstructure noise, whose components are the microstructure noise return variance ($Var^{\#}(\Delta U_t)$) and the covariance between efficient and noise returns ($Cov^{\#}(\Delta Y_t, \Delta U_t)$).

$$Var^{\#}(\Delta Y_t) = n \times f'\hat{Z}'\left(\widehat{\Lambda_\pi} - \hat{P}'\widehat{\Lambda_\pi}\hat{P}\right)\hat{Z}f, \tag{2.6}$$

$$Var^{\#}(\Delta X_t) = n \times f'\hat{Z}'\left(I - \hat{P}\right)'\widehat{\Lambda_\pi}\left(I - \hat{P}\right)\hat{Z}f, \tag{2.7}$$

$$Noise = Var^{\#}(\Delta X_t) - Var^{\#}(\Delta Y_t) = 2 \times Cov^{\#}(\Delta Y_t, \Delta U_t) + Var^{\#}(\Delta U_t), \tag{2.8}$$

where *n* is the number of observations and matrix *f* of order $S^k \times 1$ contains the last columns of $\Delta\mathcal{X}$, and where *f* is a matrix whose s-th row (f_s) is the realization of ΔX_t in state S.

The $Var^{\#}$ estimators above provide variance estimates of the non-transformed returns, i.e., the prices in levels (cents/bushel). Since previous research has typically measured variance of log-prices, we use an estimator proposed by Hansen et al. (2015) which is virtually identical to the realized variance of log prices. To derive the transformed return variance, the $Var^{\#}$ estimators in equations (2.6) to (2.8) are divided by the average of observed squared prices. For instance, if $Var(\Delta Y_t)$ is the log-transformed version of $Var^{\#}(\Delta Y_t)$ in equation (2.6), it is expressed as: $Var(\Delta Y_t) = \frac{Var^{\#}(\Delta Y_t)}{\frac{1}{n}\sum_1^n X_{t,i}^2}$. The transformed measures are used in the remainder of the paper and denoted as (2.6') to (2.8').

We use the *Var* estimators to assess how markets respond to major news releases. To evaluate both the magnitude and duration of the informational impact of news releases, we rely on efficient

variance ($Var(\Delta Y_t)$) rather than the noise-polluted observed variance ($Var(\Delta X_t)$). Specifically, we quantify the relative effects of USDA announcements by comparing the informational content of variance on announcement days to non-announcement days. The costs of efficient price discovery in the form of market frictions are measured through $Var(\Delta U_t)$. By comparing these measures across release policies, we provide insights on their market impacts and contribute to current debates on how public news should be released in the framework of fast electronic markets.

2.4 Data

The data are corn and soybean transaction prices traded on the CBOT electronic platform and obtained from the CME for July 2009 to May 2017. We use tick data time-stamped to the nearest second.[5] Following Adjemian and Irwin (2018), we divide the data into two trading periods—a halt period (up to May 2012), and a real-time period (after June 2012). Changes in market conditions over years can cause changes in the news effects and reduce the effectiveness of simple cross-era comparisons. To improve the effectiveness of our comparison, we follow Adjemian and Irwin (2018) and separate within each period USDA reports based on the size of their impacts on the market. A report is considered a large (small) shock if the corresponding day settlement price return is larger (smaller) than the median value of announcement day settlement price returns.

Previous research has identified that the first 60 minutes after announcements experience the most significant market reactions (Joseph and Garcia 2018; Adjemian and Irwin 2018). Erring on the side of caution, we perform the analysis on the first 120 minutes after USDA announcements. The *Var* estimator is calculated for every minute within this period, which permits an assessment of how the impact of USDA announcements changes within the day. To identify the announcement

[5] Following Barndorff-Nielsen et al. (2009), we use pre-processing procedures to remove issues in high-frequency data (e.g., outliers or data errors) that may bias the estimation results.

effect on market price variance more clearly, we examine the differences between announcement and non-announcement days. The 1-minute *Var* estimator is obtained for the same intraday intervals on non-announcement days. The selection of corresponding non-announcement days is discussed below.

While the corn and soybean futures markets are similar in several dimensions, they differ in others. The soybean futures contracts are traded with seven maturities per year: January, March, May, July, August, September, and November. The corn futures market trades five maturities per year: March, May, July, September, and December. For both commodities, we use the first nearby futures contract and rollover to the next nearby at the beginning of the delivery month. Since nearby contracts are usually more heavily traded than deferred contracts, they are likely to respond more quickly and strongly to new information. We follow Lehecka et al. (2014) and consider five major reports: Acreage, Crop Production, Grain Stocks, Prospective Planting, and World Agricultural Supply and Demand Estimates (WASDE). Announcement days are trading days when one or more of these reports are released and non-announcement days are the five consecutive business days before each announcement day. Non-announcement days serve as baseline to identify the magnitude of the increase in variance, relative to nearby days without an information shock. If the five days before announcement contain another announcement day, only the non-announcement days are included. The non-announcement time chosen to compare the announcement and non-announcement days corresponds to the same time as the one used to measure announcement effects: the first 120 minutes of the trading session in the halt era,[6] and the 120 minutes from 11:00 to 13:00 in real-time era.

[6] Following previous literature (Lehecka et al. 2014), in the halt period, the return of the first minute is based on the last price of the previous trading session.

2.5 Empirical Results

This section examines the effects of USDA announcements across different dimensions. First, we derive observed price variance after the announcements and separate the efficient price variance from the market microstructure noise. Second, we provide the dollar value of the noise-driven price innovations. Third, we obtain a relative measure of the USDA impacts by comparing announcement and non-announcement days. Next, we compare the *Var* estimator results with those derived from Hautsch et al. (2011). Finally, we examine the time it takes for USDA reports to be absorbed by the market.

2.5.1 Price variance on USDA announcement and non-announcement days

The lag order (k) influences the *Var* estimator's accuracy. Based on simulation results under different noise assumptions, Hansen and Horel (2009) and Hansen et al. (2015) find that $k = 3$ or $k = 4$ maximizes estimator accuracy. Also, time homogeneity of the Markov process may not hold for high-frequency data, which leads Hansen et al. (2015) to recommend larger lag orders to alleviate this potential difficulty. Accordingly, we set the lag order to $k = 4$.[7] In Appendix B, we examine the effectiveness of the *Var* estimator in measuring relevant variances.

The efficient and observed variances estimated using the *Var* model (equations 6' and 7') for the first 120 minutes following the halt and real-time announcements are presented in figures A.2 and A.3 in Appendix A for corn and soybean markets, respectively. Consistent with previous research, we find responses to USDA announcements are concentrated during the first minutes after the release of information and dissipate after about an hour. As a result, the remainder of the

[7] We check our estimation results with k =1, …, 7. The results suggest that $Var(\Delta X_t)$ is stable across all k values and $Var(\Delta Y_t)$ stabilizes after $k = 4$. To preserve a sufficiently large number of observations, we choose $k = 4$.

analysis focuses on the first 60-minute period following the release. Table 1 presents the average 1-minute $Var(\Delta X_t)$ and $Var(\Delta Y_t)$ estimator results (equations 6' to 7') and the difference between the two representing microstructure noise (equation 8') on announcement and non-announcement days in the halt and real-time periods. Several strong patterns emerge. First, results clearly demonstrate the widespread presence of noise across markets and periods, identifying that market frictions are a large part of observed variance. Regardless the market (corn or soybeans), the time period (halt or real-time), the shock type (large or small shocks), or the intraday time period (first 60 minutes of the trading session at the market open in the halt, or the 60 minutes from 11:00 to 12:00 in the real-time), noise constitutes at least 50% of the observed variance.

Second, observed and decomposed variances for corn and soybeans on announcement days, are larger in real-time than in the halt period. The relatively smaller variances in the halt are likely due to traders having an opportunity to digest the public information prior to the open. On non-announcement days, the pattern is reversed as all variances are larger in the halt. Differences in the timing of the release explain this behavior. In the halt, non-announcement days correspond to the market open which is highly active as market participants incorporate overnight information into prices. In real-time, non-announcement days correspond to the period between 11:00 and 12:00, with muted market trading activity.

Third, differences can be observed in the composition of variances. In both markets, efficient variance increases on announcement days relative to non-announcement days. For corn on announcement days in the halt, efficient variances for both large and small shocks are about 38% of the observed variance compared to 29% on non-announcement days. In real-time, efficient corn variances for large and small shocks represent 49% and 39%, respectively, while barely reaching 14% on non-announcement days. For soybeans on announcement days in the halt, efficient

variances for large and small shocks represent 42 and 46% of the observed variance, respectively compared to 36% on non-announcement days. In real-time on announcement days, the soybean efficient variance represents to 49% and 43% for large and small shocks respectively, which is much larger than the 25% observed on non-announcement days. Notice the soybean market appears to be less affected by microstructure noise which may be due to differences in the relative tick size to contract price between the two markets.

Observed variances in the table allow us to generate variance ratios of announcement to non-announcement days which have been reported in the literature to measure the effects of USDA information. We compare these measures to efficient variance ratios which eliminate microstructure noise and are more in line with the measurement of changing information. Focusing on corn in real-time, the ratio of observed variances on announcement to non-announcement days for large (small) shocks is 4.49 (3.63). This means the announcement effect for large (small) shocks in the real-time period causes more than a fourfold (threefold) increase in observed variance. Similar calculations using efficient variances lead to much larger effects—16.93 (10.04) for large (small) shocks. Calculations in the halt period show much smaller differences between the observed and efficient variance ratios. Observed variance ratios for large (small) shocks are 1.61 (1.45), while efficient variance ratios for large (small) shocks are 1.95 (1.80). These findings highlight differences in the halt and real-time periods and demonstrate that the use of observed variance ratios underestimates substantially the impacts of USDA releases on the efficient market price variance. This occurs because the noise component of observed variance is larger on non-announcement days compared to announcement days, biasing the ratio of observed variances downward. Calculations for soybeans are qualitatively similar and substantiate that the use of observed variances can underestimate the effects of USDA releases.

2.5.2 The economic value of microstructure noise

Table 1 demonstrates the pervasiveness of microstructure noise, with the noise magnitude being especially large on announcement days in real-time relative to the trading halt for both corn and soybeans. For example, on large shock days noise in corn increases from 3.55E-06 to 3.92E-06. Increases in the soybean market are larger and on the order from 9.02E-7 to 1.60E-06, suggesting that market frictions are especially relevant when fundamental information is incorporated into the prices in real-time. These frictions materialize in price returns that do not reveal changes in the efficient price and are likely the result of the bid-ask bounce associated to large number of small trades around announcements, market over or under-reaction to the new information, or the tick size which forces the price to change at least by $12.50 per contract for both corn and soybeans. In this subsection we estimate the dollar cost of microstructure noise for both corn and soybeans.

We approximate the per contract average dollar-value microstructure noise costs (NC) during a minute (t) within a day as $NC_t^i = \sigma\left(\Delta \mathrm{U}_t^i\right) \times 5{,}000 \div 100$, where $\sigma\left(\Delta \mathrm{U}_t^i\right) = \sqrt{Var^{\#}\left(\Delta \mathrm{U}_t^i\right)/n_t^i}$ is the standard deviation of microstructure noise within minute t expressed in cents per bushel, $Var^{\#}\left(\Delta \mathrm{U}_t^i\right)$ is the non-log transformed microstructure noise variance, with n_t being the number of observations during minute t and with i separating the shocks into large and small. Since we want to express the costs on a per contract basis, we multiply the standard deviation by the number of bushels per contract (5,000). Also, since futures prices are expressed in cents, the economic cost measure is divided by 100 to quantify costs in dollars. Table 2 presents average results for the first five minutes and the last five minutes within the one-hour interval studied. The first five minutes after the announcement are of particular interest as they register most of the trade that occurs after announcements (Adjemian and Irwin 2018). The last five minutes help to identify the persistence of noise during announcements.

During the halt, corn (soybean) noise costs on announcement days during the first five minutes of trade are 23-100% (17-53%) larger than noise costs on non-announcement days. These costs are at a maximum during the first minute of trade, reaching $23.93 ($29.72) per corn (soybean) contract on large shock days and $18.60 ($29.31) per corn (soybean) contract on small shock days. Within the last five minutes of the one-hour period examined, noise costs drop to values very close to non-announcement days, around $7-8 and $9-12 per contract for corn and soybeans, respectively which suggests noise persistence is not particularly relevant. The real-time release scenario brings considerably different results with the first five minutes of trade on announcement days generating noise costs 47-266% (141-456%) times larger than non-announcement days' noise costs for corn (soybeans). Noise costs are also largest during the first minute of trading reaching $26.58 ($45.47) per corn (soybean) contract for large shocks and $21.66 ($38.56) for small shocks. By the end of the 60-minute period, noise variances have a value of around $7-8 ($7-12) for corn (soybeans) which are much closer but still larger than non-announcement days, and thus suggesting higher noise persistence during the real-time release era.

Relative cumulative noise costs after announcements are also informative. Here, we use the transactions volume per minute to calculate the cumulative noise costs as follows: $Total\ NC_T^i = \sum_{t=1}^{T} NC_t^i \times Vol_t^i$, where ($NC_t^i$) are the per minute and per contract noise costs, Vol_t^i is the volume in number of contracts in minute *t* and shock *i*, and *T* = 1,..,60. We report the difference between these costs on announcement and non-announcement days $NC_T^i = NC_{T,ann}^i - NC_{T,non}^i$ and present average results in figure 2.1. During the halt release, the average additional cumulative noise cost

per day is \$200,000, increasing close to \$600,000 in the real-time.[8] This suggests that additional noise costs during the real-time are more than 3 times larger. There are 14 announcement days each year for corn and soybean markets. Therefore, the annual cost of real-time releases is about \$8.4 million in both corn and soybean markets, while the same number is less than \$2.8 million during halt releases. Hence, the costs of bid-ask bounce, market over or under-reaction and other market frictions are substantially larger in real-time, which is expected given the nature of the releases. Later, we put these costs into perspective, by comparing them with the reduction in the time of shock absorption under the different release regimes.

2.5.3 Relative USDA announcement effects: Announcement vs non-announcement days

Consistent with previous studies, we measure the relative effects of USDA announcements by comparing the informational content of variance on announcement days to non-announcement days. To reduce the influence of idiosyncratic effects and allow more effective comparisons across trading regimes and markets, we use information ratios (IR) where efficient price variance estimates are standardized by observed variance. IRs are derived for every minute from equations (6') and (7'). Let $IR^i{}_{ann} = \frac{Var(\Delta Y_t)^i{}_{ann}}{Var(\Delta X_t)^i{}_{ann}}$ and $IR^i{}_{non} = \frac{Var(\Delta Y_t)^i{}_{non}}{Var(\Delta X_t)^i{}_{non}}$ measure the informational content of variance on announcement (*ann*) and non-announcement (*non*) days, with *i* separating the shocks into large and small. Then, the informational impact of USDA announcements is the difference between these two IRs, $\Delta IR^i = IR^i{}_{ann} - IR^i{}_{non}$. The difference can be interpreted as

[8] While these measures are relevant to policy markers, they are less so to individual traders. In Appendix C we offer an alternative measure of the value of identifying microstructure noise, by illustrating how individual traders can benefit from using efficient variance as opposed to observed variance in their investment decisions.

the change in the relative importance of efficient variance to observed variance induced by the introduction of new information.

The importance of standardizing the efficient variances emerges if you recall that baselines (i.e., non-announcement days) differ substantially between the halt and real-time periods. For the halt, the one-hour window on baseline days corresponds to the first 60 minutes of the trading session. This window contains the market open which exhibits high observed variance since information accumulated overnight is incorporated into the market. This contrasts with the real-time period baseline that runs from 11:00 to 12:00, a period when the market is more muted due to less trading and variance is small. This results in the composition of variance being different across halt and real-time baselines. Also, standardizing is relevant to increase comparability between the two eras after the technological changes that have affected futures markets.[9]

Results are presented in table 2.3 and show that halt releases lead to an average increase in the informational content of corn (soybean) variance for large and small shocks of 2.28% and 1.96% (1.17% and 1.27%). In real-time, releases result in an average increase in the informational content of corn (soybean) variance of 15.67% and 12.64% (10.2% and 10.47%) for large and small shocks.

[9] Shares of automated trading, which includes algorithmic trading, increased from 39% between 2012-2014 to 55.2% between 2016-2018, in both grain and oilseed markets (Haynes and Roberts 2015; Haynes and Roberts 2019). By using state-of-art machine learning techniques such as natural language processing and sentiment analysis, automated traders can interpret the content of reports very fast after releases (Nam and Seong 2019; Vo et al. 2019). Further, automated order placement systems and low-latency transmission speed are key to submit orders to the market within milliseconds. As a result, changes in the way markets respond to USDA announcements may not only be due to changes in release policies, but also to the adoption of algorithmic trading and artificial intelligence techniques.

These results show that the impacts of the real-time announcements on informational variance are much larger than those in the halt period. It appears that the release of the USDA report before trading hours allowed market participants to develop a better understanding of the impacts of the public information on the efficient market price. This likely produced starting bids and offers on the information that were close to the actual value of the information in the halt era. As a result, the informational content of additional efficient variance in the halt was low. Further, since the halt baseline coincides with the market open, the baseline contains considerable private information accumulated overnight which contributes to the low impact of USDA announcements. These results are consistent with Adjemian & Irwin (2018) who for large shocks in the halt found no statistically significant differences between trade volumes on announcement and non-announcement days.

Figures 2.1 and 2.2 provide our relative measures of the USDA effects per trading minute for the halt and real-time. To refine the comparison between announcement and non-announcement days, we test the null hypothesis that minute-level IR^i on announcement and non-announcement days is identical by using Kruskal-Wallis χ^2 test. Whenever ΔIR^i is not statistically significant, we set the ΔIR^i value in the plot to zero. In the halt, except for a few scattered statistically significant differences, the plot is virtually flat for small and large corn and soybean shocks. This is consistent with the low average information content measures presented earlier. These plots suggest that when information enters the markets, traders respond to it in a discontinuous manner. This is also consistent with Adjemian & Irwin (2018)'s measures of announcement effects in the halt, which intermittently fluctuate from zero to positive values. After cleaning for noise, however, the pattern here is more pronounced.

The plots differ substantially in real-time. For corn in the first 20 minutes after the announcement, the relative importance of the efficient variance for large (small) shocks is intermittently 25-50% (0-50%) larger than on non-announcement days. A notable exception is in the first minute after announcement for large shocks, where the information content is 80% larger than on non-announcement days. In the 20- to 40-minute interval after the announcement, the additional efficient variance fluctuates between 0 and 25%; afterward, it disappears. For soybeans the plots show for small shocks (figure 2.2 bottom panel) the effects fluctuate between 0 and 25% for about 45 minutes and then, except for a few spikes, they virtually disappear. The effects of large shocks fluctuate between 0-50% in the first 40 minutes, then they virtually fade away. Several exceptions close to the announcement time exist in the soybean market when the informational content increases to 65% for large shocks and 31% for small shocks.

Figures 2.1 and 2.2 provide a different picture than the relative intraday effects captured by previous research, which have a clearer 'L' shape with important variance spikes in the first minutes after announcement. Our measures, particularly for corn in real-time, have large spikes in the early minutes after the announcements, some reaching more than a 50% increase relative to non-announcement days. Despite declining much less abruptly following the announcement, they do not extend in a protracted manner as observed in previous research.

2.5.4 State-space model results

We compare the *Var*-based changes in IRs to their counterparts derived from the state-space model used by Hautsch et al. (2011). This model assumes that the latent efficient price return follows a VAR-EGARCH and relies on bid and ask quotes to derive the conditional variance of the efficient price return and the conditional noise variances (the noise in bid and the noise in ask quotes). Details on the method can be found in Hautsch et al. (2011). Here we discuss the implementation

of this model to our data. Following Hautsch et al. (2011), we apply the model to 1-minute sampled log bid and ask returns for corn and soybeans. We conduct model estimation separately for announcement and non-announcement days, for large and small shocks and for each crop. We focus on an interval of 180 minutes surrounding the announcement time, 60 minutes before and 120 after. The estimation process is not re-initialized each day, instead the 180 1-minute returns across days are stacked to form the estimation sample. As explained by Hautsch et al. (2011), treating the return series as continuous requires the use of a buffer period of at least 20 minutes when moving from one day to another to allow the memory from the previous day to vanish. This is the rationale for using 60 minutes before the announcement. The need for the buffer period precludes the use of the state-space model to assess variance during the trading halt era, since this would require the use of overnight quotes that suffer from quality issues due to the lack of transactions.[10] The effectiveness of the state-space model in capturing intraday variance on announcement days can be observed in figures A.4 and A.5 in Appendix A. The figures show the widely reported conditional variance peaks around announcement times, a quick decay afterwards (Joseph and Garcia 2018; Adjemian and Irwin 2018) and a stabilization in about an hour. Hence, in the rest of this section we focus on the first 60 minutes after announcement time. Notice that while our method produces unconditional variance estimates, the method used by (Hautsch et al. 2011) produces conditional variances. While unconditional realized variance estimates are an ex-

[10] In order to compare our MC-based *Var* results to the state-space model a few additional changes are introduced. First, since the liquidity indicators are not part of the MC framework, they are ignored. Also, since we estimate the model separately for large and small shocks, we do not include the announcement surprises as an explanatory variable in the model. The state-space model also precludes consideration of the period from May to December 2012, when the reports were released in the early morning.

post variance measure based on returns observed within the 1-minute measurement interval, conditional variance estimates are based on a modeling approach that accounts for relevant conditioning information and are based on 1 – minute observations. Information ratios allow a robust comparison between the two methods.

Figures 3.1 and 3.2 present the real-time USDA announcement effects on IR changes in the corn and soybean markets ($\Delta IR^i = IR^i{}_{ann} - IR^i{}_{non}$) based on the state-space model. Notice each plot has two measures of the USDA impacts since the information ratios $\left(IR^i = \frac{Var(\Delta Y_t)^i}{Var(\Delta X_t)^i}\right)$ can be computed either based on bid or ask quotes (see Hautsch et al. (2011) for further detail). Differences between the MC framework and state-space models are clear when comparing these figures with figures 2.1 and 2.2. Plots derived from the state-space models are relatively stable over the hour following releases, which is inconsistent with previous research findings and our MC-based *Var* results. Also, the effects of USDA reports during the first minutes after release appear much smaller with the state-space model. More surprising are the persistently negative state-space model announcement effects for corn on large announcement days. These differences can be explained by the features of the different models.

First, in contrast to the MC framework, the state-space approach assumes that the covariance between efficient returns and noise is zero. Figure A.6 in Appendix A shows the covariance estimates from the MC framework ($Cov(\Delta Y_t, \Delta U_t)$) on announcement days. These are strongly negative during the first minutes after the release, especially in the corn market, and stabilize near zero after about 20 minutes. In contrast, covariances on non-announcement days (Figure A.6), are very close to zero. By assuming covariances to be zero, the state-space model leads to biased announcement effect estimates. To understand this, we first change notation used in equation (8) to generalize it beyond the MC framework and solve for the information ratio; $IR^i = \frac{Var(\Delta Y_t)^i}{Var(\Delta X_t)^i} =$

$1 - 2 \times \frac{Cov(\Delta Y_t, \Delta U_t)^i}{Var(\Delta X_t)^i} - \frac{Var(\Delta U_t)^i}{Var(\Delta X_t)^i}$. Notice that a strongly negative covariance on announcement days results in an increase in the information ratio ${IR^i}_{ann}$. As a result, by setting the covariance to zero, the state-space model underestimates ${IR^i}_{ann}$. In contrast, since covariances on non-announcement days are virtually zero (Figure A.6), the amount of bias introduced in ${IR^i}_{non}$ is much less. This results in a downward bias of the informational impact of USDA announcements ($\Delta IR^i = {IR^i}_{ann} - {IR^i}_{non}$) especially during in the first minutes after the release when covariances are highly negative. In the case of corn, this reduction leads to negative USDA effects on large shock days and no USDA effects on small shock days.

Second, the state-space model relies on conditional variances that are fitted to the shocks of a VAR(1) process of the unobservable return components (i.e., efficient price return variance and noise variances). Filtering out serial dependence through a VAR model results in more stable variance processes around announcement time than the MC-based *Var* estimators. Third, the state-space model is estimated under the assumption that both unobserved state variables and parameters are normally distributed. However, the distribution of intraday financial returns is known to be skewed and fat-tailed, especially during major news releases. This is confirmed in Figure A.7 in the Appendix A which presents a Q-Q plot of the efficient price return shocks in the corn market.[11] Thus, the validity of the state-space model results is questionable. Finally, the use of minute-level data following (Hautsch et al. 2011) may also contribute to generating more stable announcement effects by smoothing out the very high-frequency trading activity that occurs right after the information release. In short, our article shows the relevance of properly selecting the methods in order to correctly assess the impacts of news on the market.

[11] The Q-Q plots for soybean efficient returns are similar.

2.5.5 Absorption time of announcement effects

Another dimension of the USDA effects is the time required to absorb the additional variance induced by announcements (Adjemian and Irwin 2018), computed as the difference between the variance on announcement and non-announcement days. Here, we calculate the variance absorption path using a ratio of the cumulative sum of additional observed (efficient) variance at minute *t*, over the total sum of additional observed (efficient) variance for the 60-minute period following the announcement.[12] By distinguishing between observed and efficient variances, the analysis identifies whether microstructure noise delays the time to absorb the informational content of the USDA reports, and whether there are differences between the halt and real-time releases. Results are presented in Table 4 which shows the minutes necessary to reach a pre-specified percent. Figure 4 provides plots of the efficient variance absorption path over the 60-minute period for large shocks in both the corn and soybean markets.

Results differ by commodity, period, and variance measure. First, focus on corn observed and efficient variances presented in table 2.4 (top panel). Most of the adjustment occurs quickly, but it takes a while for the shock to completely disappear. Concavity is generally more pronounced in the efficient variance absorption path than in the observed variance path. This suggests that shocks to the efficient price require less time to decline than the observed prices. For example, Table 4 shows that 90% of the additional observed (efficient) variance generated by large shocks in the corn market in real-time is absorbed in 33 (22) minutes. For soybeans, absorption paths are also concave, but differences in the degree of concavity between observed and efficient variances are less clear-cut. For example, in real-time, 90% of the additional observed (efficient) variance

[12] The ratios are calculated using equations (2.6') and (2.7').

generated by large shocks is absorbed in 33 (29) minutes. However, in the halt, it took longer to absorb the efficient variance than the observed variance (31 compared to 41 minutes). Hence, the use of observed variance to measure the duration of the USDA effect can produce misleading results. In the remaining of this subsection, we focus on the efficient variance absorption paths.

Our findings are consistent with Adjemian & Irwin (2018) since most of the adjustments occur quickly. However, they differ in the time required for absorption. Adjemian & Irwin (2018) identify for corn that 90% of large (small) announcement shocks are absorbed in an hour (a little more than 50 minutes) in real-time, and both in a little less than 50 minutes in the halt.[13] We find the speed of absorption is faster than previously identified. Using efficient variances, our results show for corn that 90% of large (small) announcement shocks are absorbed in 22 (31) minutes in real-time, and 21 (34) minutes in the halt. For soybeans, 90% of large (small) announcement shocks are absorbed in 29 (31) minutes in real-time, and 41 (43) minutes in the halt. For corn, the absorption paths of the efficient variance are more concave in the trading halt, with both paths converging after about 20 minutes following the shock (table 2.4, top panel; fig. 2.5). In the case of soybeans, paths of the efficient variance are more concave in real-time and converge after about 35 minutes, suggesting less time is needed to absorb the magnitude of the shock in real-time than in the trading halt. Differences between corn and soybean markets are compatible with the generally smaller changes in the information ratios between announcement and non-announcement days (Table 3) in the soybean relative to the corn market. Overall, while information is absorbed faster than previously measured, the findings here indicate the change to real-time releases either

[13] See figure 4 in Adjemian and Irwin (2018).

do not improve absorption speed or imply temporal savings for the 90% absorption rate that are on the order of only 10-41% relative to the trading halt.

In a policy context, these findings suggest that in the real-time release, the markets can absorb large amounts of information at approximately the same speed, or even faster than previously. This comes however at the cost of larger microstructure frictions and their dollar value, which we find to be three times larger in the real-time relative to the halt (figure 2.1). The combination of fast price discovery and large microstructure costs is consistent with finance theories (Kyle 1985; Collin-Dufresne and Fos 2016) that have shown that larger noise variance encourages more informed trading and results in faster price discovery. However, on balance, improvements in the speed of incorporating new information are modest and it is difficult to identify how they offset the cost increases resulting from the accompanying market frictions.

2.6 Discussion and Concluding Remarks

Since Hautsch et al. (2011), no other published paper has assessed the impacts of public news releases on price variance by disentangling efficient price variance and microstructure noise. In this article we study the USDA announcement effects on the efficient and noise components of variance in corn and soybean futures markets. With the shift to electronic trading and mid-day USDA announcements, measuring the effects of announcements requires intraday high-frequency data. However, the use of high-frequency data can distort the measurement of informational effects of the releases when the presence of microstructure noise is large and overlooked. Here, we employ a Markov Chain (MC) framework that decomposes observed variance ($Var(\Delta X_t)$) into its efficient variance ($Var(\Delta Y_t)$) and microstructure noise. The flexible *Var* estimator permits a more nuanced assessment of the structure of variance. We use the *Var* estimates and derive information ratios of the efficient to observed variances. The informational impact of USDA announcements is then

measured using differences of information ratios between announcement and non-announcement days. Using absorption paths, we quantify the speed by which additional variance is absorbed in markets. Given the economic relevance of microstructure noise (Bandi & Russell, 2006), we compare its dollar value in the two release eras. By documenting the value of microstructure noise and placing it relative to the speed of price discovery, we provide a path for improved policy making. The analysis is performed with nearby futures transaction prices for corn and soybeans, tick data, time-stamped to the second from July 2009 to May 2017.

This is the first paper to study the impacts of public announcements on variance in any market using the method by Hansen (2015). We find this method performs much better than Hautsch et al. (2011), providing results that are more detailed and consistent with recent theory. We study soybean and corn futures markets that have larger tick sizes than the German Bond futures contracts studied by Hautsch et al. (2011), which is likely to result in larger market frictions. More importantly, recent changes in USDA crop reports' release policy allow us to compare announcement effects and market frictions under two very different release policy regimes (trading halt and real-time).

Previous research has measured the impacts of USDA reports on agricultural commodity futures markets using the ratio of the observed price variance on announcement relative to non-announcement days. We find that more than 50% of the observed variance corresponds to microstructure noise on announcement days and roughly 75% on non-announcement days. Clearly, the ratio of observed variances used by previous research can be influenced by both changes in the magnitude of the information effect and the composition of observed variance. This substantiates the limitations of using observed variances to infer the effects of public information; observed

measures can substantially underestimate the informational impact and bias the estimates of the time for the USDA effect to dissipate.

Examination of the *Var* estimates allows us to identify the effects of USDA releases on efficient variance. Average increases in the informational content of efficient variance on announcement days differ between halt and real-time periods. In the halt period, USDA releases for corn and soybeans produce an average 1-2% increase in the relative importance of efficient to observed variance between non-announcement and announcement days. In the real-time period, the increase is much larger, reaching an average 10-16%. The increase in informational content of variance in the real-time era is not necessarily a reflection that the quality of USDA information or its perceived importance has changed, but rather points more directly to the change in the release policy which generated more variance in the pricing process.

We find that information is incorporated into the market in a non- 'L' shape after the announcement. This differs from previous research in agricultural commodity markets based on observed variance (Joseph and Garcia 2018; Lehecka et al. 2014). Our per-minute measures of the magnitude of the USDA effect, particularly for corn in real-time, have large spikes in the immediate minutes after the announcements, some reaching more than 50%. Despite declining much less abruptly following the announcement, they do not extend in a protracted manner observed in previous research. It appears that traders respond to information in a more discontinuous fashion than previously identified. The extended response to USDA information in previous studies is likely a function of the temporal persistence of microstructure noise which past estimates contained.

Our results also demonstrate the importance of separating noise from efficient price variance when comparing the time to absorb a shock in information. For corn, using a 90% absorption rate

of observed variance that was employed in the literature, large shocks are absorbed more slowly in real-time than in trading halt. However, the efficient variance indicates that information is absorbed at approximately the same speed in both periods. For soybeans, while observed variance is absorbed more slowly in real-time, efficient variance indicates otherwise, with the 90% absorption rate during large shocks being cut by about twelve minutes relative to the trading-halt. The shorter duration of the information effect occurs even though market participants do not have an opportunity to review USDA releases in advance of trading, and variance is larger in this period. In summary, our findings suggest that the real-time either does not improve absorption speed or implies time savings on the order of only 10-41% for the 90% absorption rate. Noting that public information releases occur at a very low frequency (e.g. monthly, quarterly or annually), it is difficult to identify how the relatively modest improvements in speed compensate for the increases in costs of microstructure frictions which are on the order of 300% higher in real time. The evidence points more clearly to the trading halt as a better alternative to release public information than the real-time.

2.7 Conclusion

This research is pioneering in showing how microstructure noise variance can be used to estimate the costs of market frictions generated by public information releases and take better policy decisions. We demonstrate the effects of a change in the release policy on the informational content of USDA announcements and microstructure noise costs. In an era of fast, automated traders and real-time releases, the informational content of variance during USDA announcements is large and does not decline immediately after announcements. We find that the change to real-time releases magnifies efficient variance as market participants interact to discover price with limited or no time to analyze the content of releases. Relative to the trading-halt, the price discovery costs in the

form of microstructure frictions are large, and not likely offset by possible savings in the speed of price discovery. Our results are consistent with calls to modify public information trading mechanisms (Budish et al. 2014).

The findings also have other implications for market design, as the tick size has remained constant in agricultural commodity markets, while traders have adopted technologies that allow for increased speed and precision. This may be one of the reasons why microstructure noise is so relevant in these markets and research may be needed to address in more depth the appropriate tick size in these markets. Also, in the context of the economic crisis and public budget constraints, it seems reasonable to develop methods that better assess the extent to which public expenses to collect and disseminate market information are justified. In this setting, disentangling what portion of the variance corresponds to value of information and what portion to market friction is thus very meaningful. Our paper sets a path for future research on the appropriate assessment of intraday impacts of news releases.

2.8 Tables and Figures

Table 2.1 Summary Statistics - Average MC-based Variances (*Var*) Based on Log-transformed Transactions Prices

		Corn, Announcements				Corn, Non-Announcements			
		Observed	Efficient	Noise	Efficient %	Observed	Efficient	Noise	Efficient %
Halt									
	Large	5.70E-06	2.16E-06	3.55E-06	37.89%	3.53E-06	1.11E-06	2.42E-06	29.82%
	Small	4.71E-06	1.82E-06	2.89E-06	38.64%	3.25E-06	1.01E-06	2.24E-06	28.93%
Real									
	Large	7.72E-06	3.81E-06	3.92E-06	49.35%	1.72E-06	2.25E-07	1.49E-06	13.08%
	Small	6.13E-06	2.39E-06	3.74E-06	38.99%	1.69E-06	2.38E-07	1.45E-06	14.08%
		Soybeans, Announcements				Soybeans, Non-Announcements			
		Observed	Efficient	Noise	Efficient %	Observed	Efficient	Noise	Efficient %
Halt									
	Large	1.55E-06	6.45E-07	9.02E-07	41.61%	9.63E-07	3.80E-07	5.82E-07	35.45%
	Small	1.61E-06	7.46E-07	8.66E-07	46.34%	1.04E-06	4.12E-07	6.27E-07	36.52%
Real									
	Large	3.15E-06	1.54E-06	1.60E-06	48.89%	3.45E-07	8.57E-08	2.59E-07	24.84%
	Small	2.37E-06	1.02E-06	1.34E-06	43.04%	3.23E-07	7.78E-08	2.46E-07	24.09%

Note: Halt and Real are the periods identified in the text. Large represents greater than the median shocks and small represents less than or equal to the median shocks. Observed and Efficient variances are derived from (6') and (7'), and Noise is derived (8'). Efficient % is the percentage that efficient variance represents with respect to observed variance. Results are presented for the first 60 minutes following the USDA report release.

Table 2.2 Average Dollar Value of Microstructure Noise per Minute on Transaction Prices

	Halt				Real			
Corn	Large		Small		Large		Small	
Time	Ann	Non	Ann	Non	Ann	Non	Ann	Non
1	$23.93	$11.99	$18.60	$11.21	$26.58	$7.26	$21.66	$7.73
2	$14.19	$9.30	$13.20	$9.47	$16.66	$6.86	$12.99	$6.91
3	$11.43	$8.71	$12.81	$8.98	$14.94	$7.28	$11.53	$7.05
4	$10.85	$8.70	$12.14	$8.68	$13.00	$7.01	$11.40	$7.34
5	$10.66	$8.69	$10.45	$8.39	$12.65	$7.01	$10.39	$7.07
…	…	…	…	…	…	…	…	…
56	$8.00	$7.65	$6.98	$7.34	$7.71	$7.15	$8.17	$7.36
57	$8.31	$7.82	$8.16	$7.32	$6.96	$7.19	$7.53	$7.02
58	$7.90	$7.71	$7.87	$7.44	$7.22	$6.90	$7.51	$7.13
59	$7.74	$7.78	$7.75	$7.72	$7.56	$6.96	$7.19	$7.26
60	$7.62	$7.65	$7.62	$7.69	$7.49	$6.90	$7.34	$6.87
Soybeans								
1	$29.72	$19.48	$29.31	$19.29	$45.47	$8.18	$38.56	$8.23
2	$20.20	$14.49	$20.08	$15.52	$32.05	$7.97	$26.42	$7.92
3	$20.75	$14.13	$18.21	$13.28	$22.28	$8.18	$22.76	$7.67
4	$16.33	$13.01	$17.07	$11.83	$21.67	$7.59	$19.11	$7.94
5	$14.95	$12.51	$14.58	$12.43	$20.15	$8.34	$19.54	$7.92
…	…	…	…	…	…	…	…	…
56	$10.61	$9.37	$8.72	$9.10	$10.88	$8.34	$8.99	$7.89
57	$10.79	$8.97	$9.60	$9.32	$11.17	$8.53	$9.14	$7.27
58	$11.94	$9.07	$9.40	$9.60	$11.57	$8.93	$11.01	$7.58
59	$10.52	$9.45	$9.60	$9.82	$12.19	$8.16	$10.04	$7.69
60	$10.37	$9.57	$10.35	$9.83	$12.48	$8.26	$9.13	$7.88

Note: Halt and Real are the periods identified in the text. Time shows minutes after USDA announcement releases. Large represents greater than the median shocks and small represents less than or equal to the median shocks. Ann and Non represent announcement and non-announcement days. Average dollar value of microstructure noise is the mean value of each minute within the first 60-minutes after news releases across announcement or non-announcement days. The dollar value in the table represents the prices investors pay for microstructure noise per contract $\left(\sqrt{MC^{\#}(\Delta U_t)/\mathrm{n}} * 5000/100\right)$.

Table 2.3 Changes in the Information Ratios Between Announcement and Non-Announcement Days

Corn Market					
	Max	75%	Median	25%	Mean
Halt					
Large	29.31%	0.00%	0.00%	0.00%	2.28%
Small	53.53%	0.00%	0.00%	0.00%	1.96%
Real					
Large	78.51%	26.44%	9.38%	0.00%	15.67%
Small	54.45%	22.07%	9.95%	0.00%	12.64%
Soybean Market					
	Max	75%	Median	25%	Mean
Halt					
Large	35.25%	0.00%	0.00%	0.00%	1.17%
Small	22.96%	0.00%	0.00%	0.00%	1.27%
Real					
Large	63.23%	14.88%	5.93%	0.00%	10.20%
Small	34.43%	20.28%	7.86%	0.00%	10.47%

Note: Large represents greater than the median shocks and small represents less than or equal to the median shocks. Numbers in the table are the average changes in the informational content of variance between announcement and non-announcement days after USDA announcements. Results are presented for the first 60 minutes following the USDA report release.

Table 2.4 Minutes to Absorb Additional Variance in the First Hour After Announcements

		Corn - Observed		Corn - Efficient	
		Halt	Real	Halt	Real
10%	Large	1	1	1	1
	(Small)	(1)	(1)	(1)	(1)
30%	Large	1	2	1	1
	(Small)	(1)	(4)	(1)	(2)
50%	Large	1	6	1	3
	(Small)	(4)	(12)	(1)	(7)
70%	Large	2	15	2	10
	(Small)	(8)	(19)	(4)	(16)
90%	Large	22	33	21	22
	(Small)	(48)	(40)	(34)	(31)
		Soybeans - Observed		Soybeans - Efficient	
		Halt	Real	Halt	Real
10%	Large	1	1	3	1
	(Small)	(1)	(1)	(1)	(1)
30%	Large	2	2	12	2
	(Small)	(3)	(3)	(1)	(2)
50%	Large	7	6	20	4
	(Small)	(7)	(7)	(3)	(5)
70%	Large	13	16	27	15
	(Small)	(14)	(15)	(11)	(15)
90%	Large	31	33	41	29
	(Small)	(27)	(33)	(43)	(31)

Note: Absorption paths measure the time required to absorb the additional variance induced by the announcements. Large represents greater than the median shocks and small (in parentheses) represents less than or equal to the median shocks. Absorption ratios, expressed in percent, are computed as the cumulative sum of additional variance up to minute t, over the total sum of additional variance for the 60-minute period after the announcement. Numbers in the table are the minutes required for a certain percentage of total variance to be absorbed.

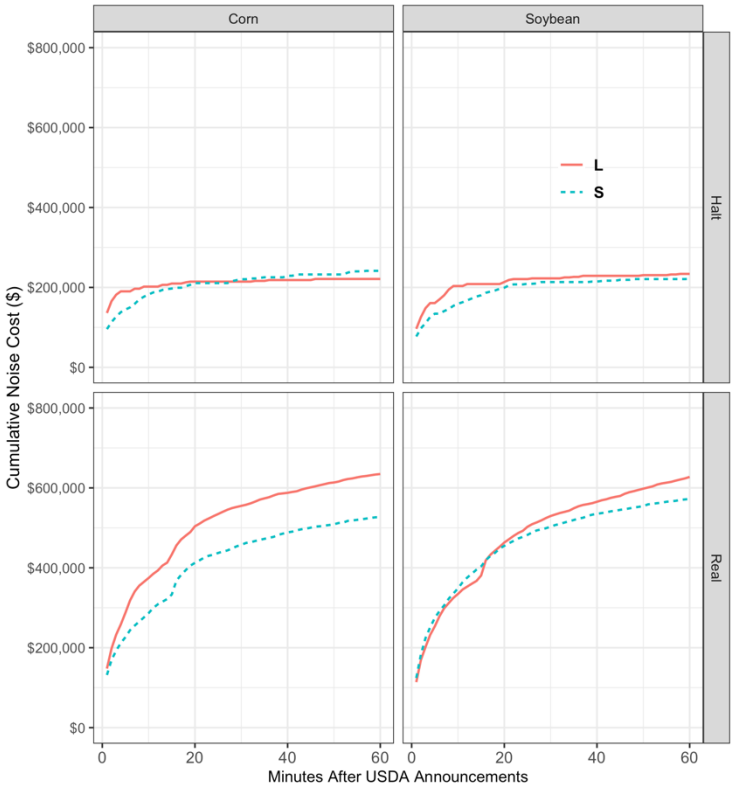

Figure 2.1 Average Cumulative Noise Cost Differences between Announcement and Non-Announcement Days.

Note: The y-axis depicts dollar value of cumulative noise cost differences up to time *t*. Noise cost for each minute is the dollar amount of microstructure noise variance per contract times total transaction volume within each minute. Cumulative noise cost differences are the difference between cumulative noise cost up to *t* minutes after news release time between announcement and non-announcement days. The x-axis represents minutes after news releases. Real and Halt represent real-time and halt release periods. Large (L) represents greater than the median shocks and small (S) represents less than or equal to the median shocks.

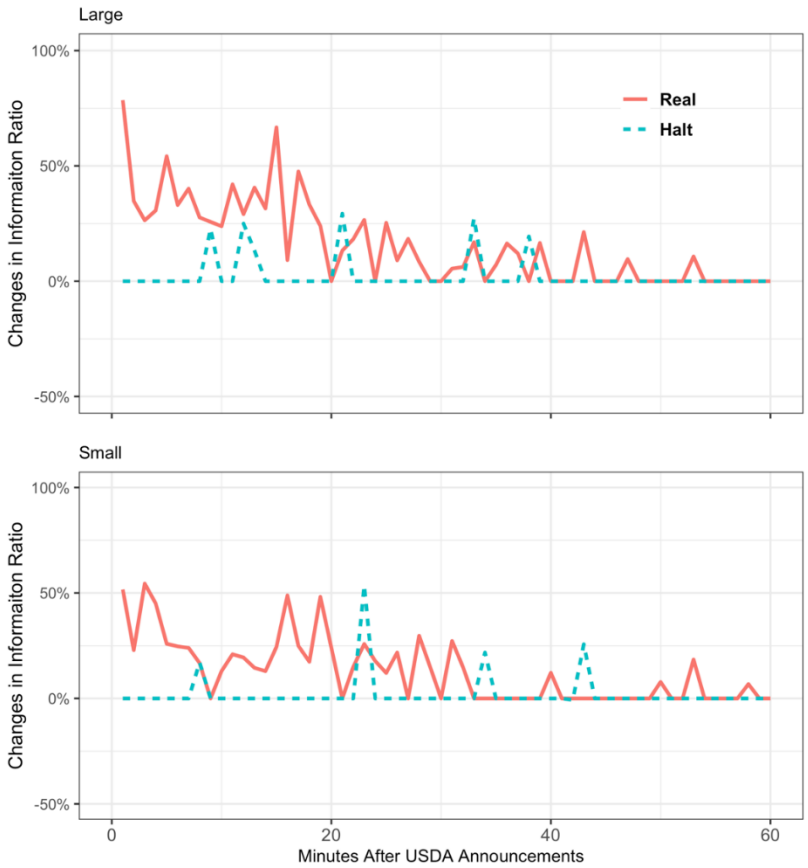

Figure 2.2 USDA Announcement Effects on Changes of Information Ratios in the Corn Market. Note: The y-axis measures the changes in informational content of variance between announcement and non-announcement days; the x-axis represents minutes after USDA announcements. Only the differences that are significant at 5% or lower given the Kruskal-Wallis χ^2 test are plotted. Large represents greater than the median shocks and small represents less than or equal to the median shocks. The red line represents the changes in values during the real-time period, while the dotted blue line represents the changes in values during the halt-release period.

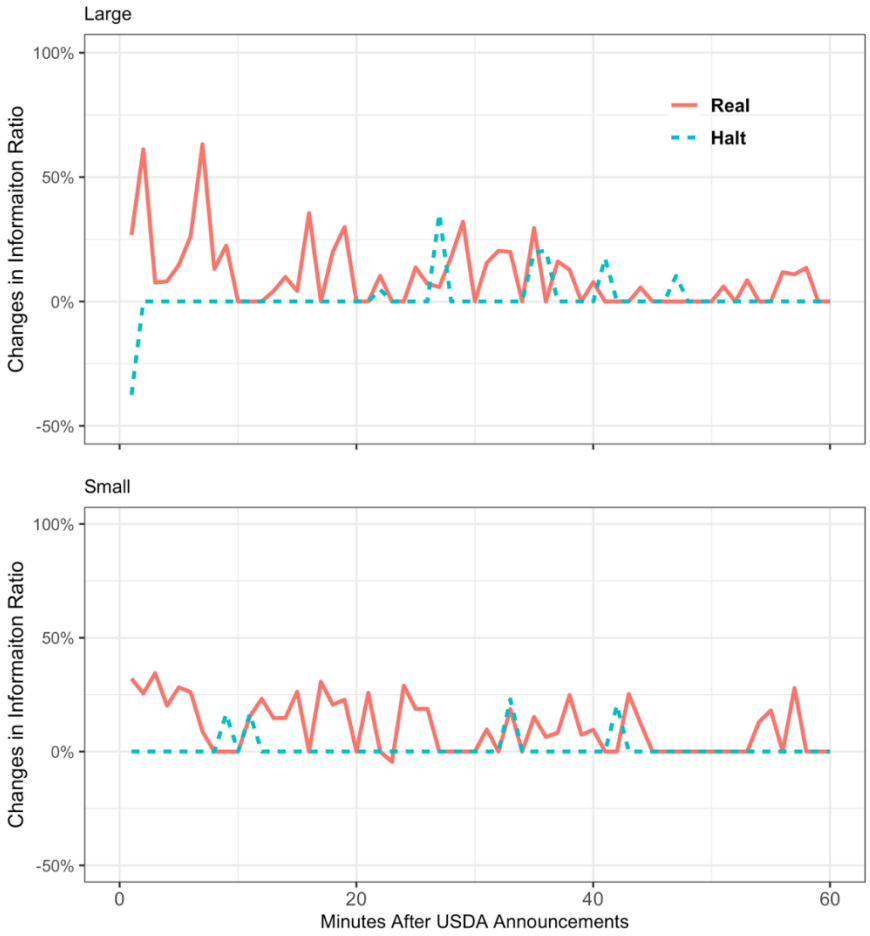

Figure 2.3 USDA Announcement Effects on Changes of Information Ratios in the Soybean Market.

Note: The y-axis measures the changes in informational content of variance between announcement and non-announcement days; the x-axis represents minutes after USDA announcements. Only the differences that are significant at 5% or lower given the Kruskal-Wallis χ^2 test are plotted. Large represents greater than the median shocks and small represents less than or equal to the median shocks. The red line represents the changes in values during the real-time period, while the dotted blue line represents the changes in values during the halt-release period.

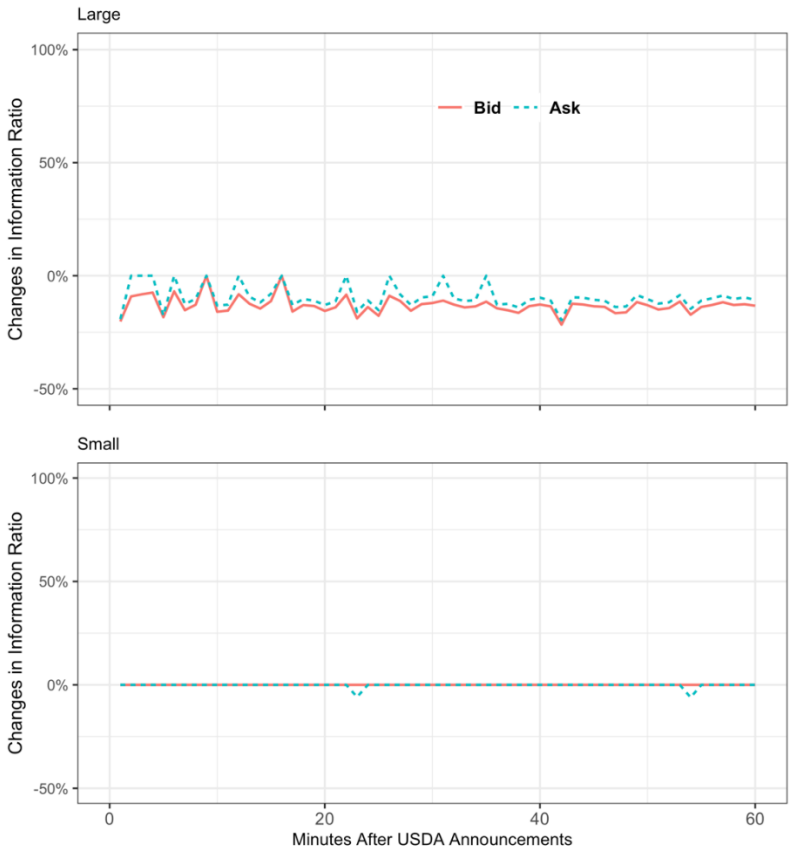

Figure 2.4 USDA Announcement Effects on Changes of VAR-EGARCH Information Ratios in the Corn Market.

Note: The y-axis measures the changes in informational content of variance between announcement and non-announcement days; the x-axis represents minutes after USDA announcements. Only the differences that are significant at 5% or lower given the Kruskal-Wallis χ^2 test are plotted. Large represents greater than the median shocks and small represents less than or equal to the median shocks. The red line represents the changes in values estimated by bid quote returns, while the dotted blue line represents the changes in values estimated by ask quote returns. Both changes are estimated during the real-time period.

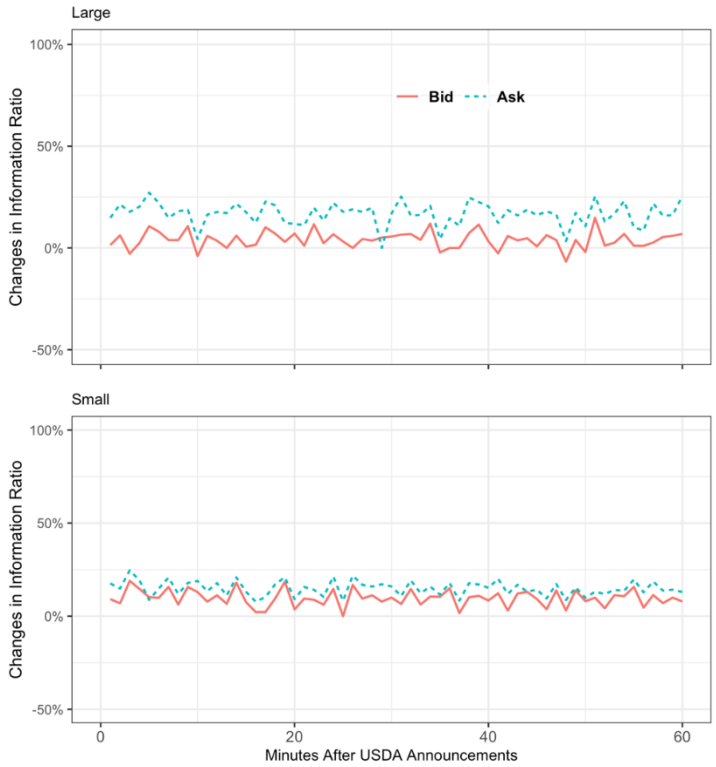

Figure 2.5 USDA Announcement Effects on Changes of VAR-EGARCH Information Ratios in the Soybean Market.

Note: The y-axis measures the changes in informational content of variance between announcement and non-announcement days; the x-axis represents minutes after USDA announcements. Only the differences that are significant at 5% or lower given the Kruskal-Wallis χ^2 test are plotted. Large represents greater than the median shocks and small represents less than or equal to the median shocks. The red line represents the changes in values estimated by bid quote returns, while the dotted blue line represents the changes in values estimated by ask quote returns. Both changes are estimated during the real-time period.

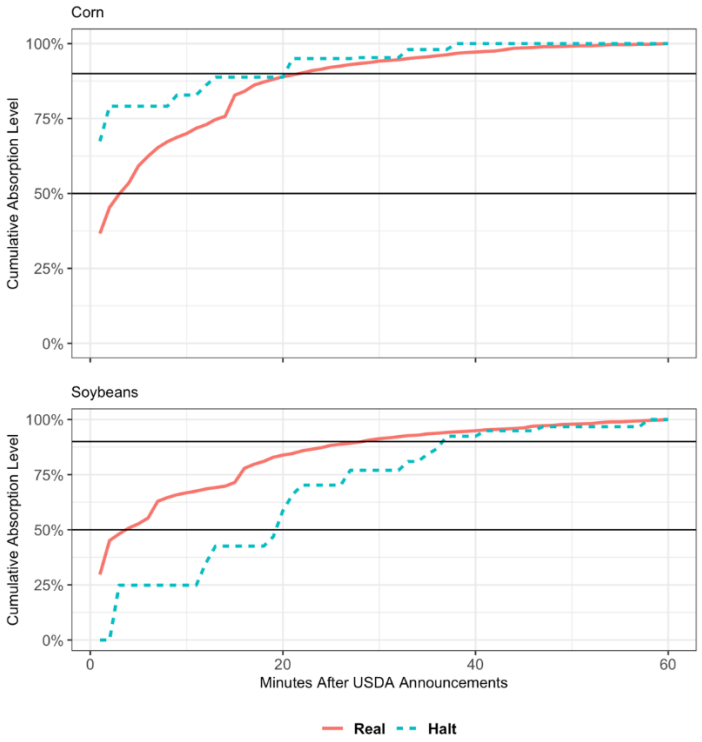

Figure 2.6 Efficient Variance Absorption Paths for Large Shocks in the Corn and Soybean Markets.

Note: The y-axis represents additional variance absorption measured by the cumulative sum of additional efficient variance measures up to a particular minute over the total sum of additional efficient variance measures for the first 60 minutes after USDA announcements. The x-axis represents minutes after USDA announcements. Red and blue lines represent corresponding absorption paths during real-time and halt period. The two horizontal dark lines represent 50% and 90% absorption level.

2.9 References

Adjemian, M.K., and S.H. Irwin. 2018. "USDA announcement effects in real-time." *American Journal of Agricultural Economics* 100(4):1151–1171. Available at: https://doi.org/10.1093/ajae/aay018.

Amaya, D., J.-F. Bégin, and G. Gauthier. 2021. "The Informational Content of High-Frequency Option Prices." *Management Science*:1–19. Available at: http://pubsonline.informs.org/doi/10.1287/mnsc.2020.3949.

Bajgrowicz, P., O. Scaillet, and A. Treccani. 2016. "Jumps in High-Frequency Data: Spurious Detections, Dynamics, and News." *Management Science* 62(8):2198–2217. Available at: http://pubsonline.informs.org/doi/10.1287/mnsc.2015.2234.

Bandi, F.M., and J.R. Russell. 2008. "Microstructure Noise, Realized Variance, and Optimal Sampling." *Review of Economic Studies* 75(2):339–369. Available at: https://academic.oup.com/restud/article/75/2/339/1620899.

Bandi, F.M., and J.R. Russell. 2006. "Separating microstructure noise from volatility." *Journal of Financial Economics* 79(3):655–692. Available at: https://www.sciencedirect.com/science/article/pii/S0304405X05001534.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2009. "Realized Kernels In Practice: Trades and Quotes." *The Econometrics Journal* 12(3).

Barunik, J., T. Krehlik, and L. Vacha. 2016. "Modeling and forecasting exchange rate volatility in time-frequency domain." *European Journal of Operational Research* 251(1):329–340. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0377221715011194.

Beveridge, S., and C.R. Nelson. 1981. "A New Approach to Decomposition of Economic Time Series Into Permanent Attention to Measurement of The 'Business Cycle.'" *Journal of*

Monetary Economics 7:151–174.

Black, F. 1986. “Noise.” *The Journal of Finance* XLI(3):528–543.

Budish, E., P. Cramton, and J. Shim. 2014. “Implementation Details for Frequent Batch Auctions: Slowing Down Markets to the Blink of an Eye.” *American Economic Review* 104(5):418–424.

Campbell, J.Y., and A.S. Kyle. 1993. “Smart money noise trading and stock price behaviour.” *Review of Economic Studies* 60(1):1–34.

Collin-Dufresne, P., and V. Fos. 2016. “Insider Trading, Stochastic Liquidity, and Equilibrium Prices.” *Econometrica* 84(4):1441–1475.

Corsi, F. 2008. “A Simple Approximate Long-Memory Model of Realized Volatility.” *Journal of Financial Econometrics* 7(2):174–196. Available at: https://academic.oup.com/jfec/article-lookup/doi/10.1093/jjfinec/nbp001.

Diebold, F.X., and G. Strasser. 2013. “On the Correlation Structure of Microstructure Noise: A Financial Economic Approach.” *Review of Economic Studies* 80(4):1304–1337.

Fama, E.F. 1970. “Efficient Capital Markets : A Review of Theory and Empirical Work.” *The Journal of Finance* 25(2):28–30.

Foucault, T., J. Hombert, and I. Roşu. 2016. “News Trading and Speed.” *Journal of Finance* 71(1):335–382.

Foucault, T., R. Kozhan, and W.W. Tham. 2017. “Toxic arbitrage.” *Review of Financial Studies* 30(4):1053–1094.

Gkillas, K., D. Vortelinos, C. Floros, A. Garefalakis, and N. Sariannidis. 2019. “Greek sovereign crisis and European exchange rates: effects of news releases and their providers.” *Annals of Operations Research*. Available at: https://doi.org/10.1007/s10479-019-03199-x.

Hansen, P.R. 2015. “A Martingale Decomposition of Discrete Markov chains.” *Economics Letters* 133:14–18. Available at: http://dx.doi.org/10.1016/j.econlet.2015.04.028.

Hansen, P.R., and G. Horel. 2015. “Limit Theory for the Long Run Variance of Finite Markov Chains.” CREATES Research Papers, , Department of Economics and Business Economics, Aarhus University

Hansen, P.R., and G. Horel. 2009. “Quadratic Variation by Markov Chains.”

Hansen, P.R., G. Horel, A. Lunde, and I. Archakov. 2015. “A Markov Chain Estimator of Multivariate Volatility from High Frequency Data.” In *The Fascination of Probability, Statistics and their Applications*. pp. 361–394.

Hansen, P.R., and A. Lunde. 2011. *Forecasting Volatility Using High-Frequency Data*. Oxford University Press. Available at: http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780195398649.001.0001/oxfordhb-9780195398649-e-20.

Hansen, P.R., and A. Lunde. 2006. “Realized Variance and Market Microstructure Noise.” *Journal of Business and Economic Statistics* 24(2):127–161.

Hautsch, N., D. Hess, and D. Veredas. 2011. “The impact of macroeconomic news on quote adjustments, noise, and informational volatility.” *Journal of Banking and Finance* 35(10):2733–2746. Available at: http://dx.doi.org/10.1016/j.jbankfin.2011.03.004.

Haynes, R., and J.S. Roberts. 2015. “Automated Trading in Futures Markets.”

Haynes, R., and J.S. Roberts. 2019. “Automated Trading in Futures Markets - Update #2.” CFTC.

Ho, M., and J. Xin. 2019. “Sparse Kalman filtering approaches to realized covariance estimation from high frequency financial data.” *Mathematical Programming* 176(1–2):247–278.

Available at: http://link.springer.com/10.1007/s10107-019-01371-6.

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter. 2009. "Microstructure noise in the continuous case: The pre-averaging approach." *Stochastic Processes and their Applications* 119(7):2249–2276. Available at: http://dx.doi.org/10.1016/j.spa.2008.11.004.

Joseph, K., and P. Garcia. 2018. "Intraday market effects in electronic soybean futures market during non-trading and trading hour announcements." *Applied Economics* 50(11):1188–1202. Available at: https://www.tandfonline.com/doi/full/10.1080/00036846.2017.1355542.

Kemeny, J.G., and J.L. Snell. 1983. *Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix."* Springer-Verlag.

Kyle, A.S. 1985. "Continuous Auctions and Insider Trading." *Econometrica* 53(6):1315–1335. Available at: http://www.jstor.org.ezproxy.eafit.edu.co/stable/1913210.

Lehecka, G. V., X. Wang, and P. Garcia. 2014. "Gone in Ten Minutes: Intraday Evidence of Announcement Effects in the Electronic Corn Futures Market." *Applied Economic Perspectives and Policy* 36(3):504–526.

LeRoy, S. 1989. "Efficient Capital Markets and Martingales." *Journal of Economic literature* 27(4):1583–1621.

McCauley, J.L., K.E. Bassler, and G.H. Gunaratne. 2008. "Martingales, detrending data, and the efficient market hypothesis." *Physica A: Statistical Mechanics and its Applications* 387(1):202–216.

Merton, R.C. 1980. "On estimating the expected return on the market." *Journal of Financial Economics* 8(4):323–361. Available at: https://linkinghub.elsevier.com/retrieve/pii/0304405X80900070.

Moallemi, C.C., M. Sağlam, and M. Sağlam. 2013. "OR forum-the cost of latency in high-

frequency trading." *Operations Research* 61(5):1070–1086. Available at: http://pubsonline.informs.org/doi/abs/10.1287/opre.2013.1165.

Nam, K.H., and N.Y. Seong. 2019. "Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market." *Decision Support Systems* 117(April 2018):100–112. Available at: https://doi.org/10.1016/j.dss.2018.11.004.

Roşu, I. 2019. "Fast and slow informed trading." *Journal of Financial Markets* 43:1–30.

Sévi, B. 2014. "Forecasting the volatility of crude oil futures using intraday data." *European Journal of Operational Research* 235(3):643–659. Available at: https://linkinghub.elsevier.com/retrieve/pii/S037722171400040X.

Vo, N.N.Y., X. He, S. Liu, and G. Xu. 2019. "Deep learning for decision making and the optimization of socially responsible investments and portfolio." *Decision Support Systems* 124(February):113097. Available at: https://doi.org/10.1016/j.dss.2019.113097.

Zhang, L., P.A. Mykland, and Y. Aït-Sahalia. 2005. "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data." *Journal of the American Statistical Association* 100(472):1394–1411. Available at: http://www.jstor.org/stable/27590680?seq=1#page_scan_tab_contents.

Zhou, B. 1996. "High-Frequency Data and Volatility in Foreign-Exchange Rates." *Journal of Business and Economic Statistics* 14(1):45–52. Available at: http://www.tandfonline.com/doi/abs/10.1080/07350015.1996.10524628.

CHAPTER 3

CAN MACHINE LEARNING ALGORITHMS BEAT LINEAR MODELS IN REALIZED VOLATILITY FORECASTING?

3.1 Introduction

Linear models have been widely used in realized volatility (RV) forecasting. The Heterogeneous Autoregressive (HAR) model (Corsi 2008) forecasts daily RV as a linear function of past RVs averaged at daily, weekly and monthly horizons. The HAR model captures the interrelations of volatility aggregated over different frequencies. Short-term traders will take the level of long-term volatility into consideration. However, long-term traders will ignore the level of short-term volatility when making decisions. It has become the workhorse in RV modeling due to its simplicity and consistent forecasting performance in empirical applications (Qiu et al. 2019). Due to the possibility of a time varying lag structure, several studies use more adaptive linear models to improve the selection of explanatory variables within the HAR model. These include the use of linear models with variable selection capabilities such as Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge (Audrino and Knaus 2016; Ding, Kambouroudis and McMillan 2021; Audrino, Huang and Okhrin 2019). All these linear models yield forecasting accuracy similar to HAR.

Besides the time varying lag structure, numerous forms of nonlinearities in financial markets also pose challenges to linear models. These nonlinear patterns are related to a wide array of causes such as the release of government agency reports (Bian, Serra, Garcia, & Irwin, 2021; Du, Fung, & Loveland, 2018; Miao, Ramchander, Wang, & Yang, 2018), strong seasonality (Arismendi et al. 2016; Egelkraut, Garcia and Sherrick 2007), time-to-delivery effects (Karali, Dorfman and

Thurman 2010; Anderson and Danthine 1983), changes in macroeconomic conditions (Karali and Power 2013; Beltratti and Morana 2006), or price jumps and leveraged volatilities (Patton and Sheppard 2015). These effects create nonlinearities of unknown functional forms that may only be captured through flexible and adaptive nonlinear methods.

Several existing studies apply machine learning (ML) methods to volatility forecasting and generate superior forecasting performances. Researchers show that Random Forests (RF) and Support Vector Regression (SVR) models can improve volatility forecasting performance over traditional volatility models in the Australian stock market (Luong and Dokuchaev 2018) and currency markets (Santamaría-Bonfil et al. 2015; Peng et al. 2018). The popularity of Artificial Neural Networks (ANN) has surged with growing evidence of their accuracy in approximating nonlinear functions and ability to provide better forecasts relative to alternative methods (Franses and van Dijk 2000). In several previous studies, ANN models have achieved better volatility forecasting results over traditional volatility models across different stock markets (D'Ecclesia and Clementi, 2019; Donaldson and Kamstra, 1997; Gonzalez Miranda and Burgess, 1997). Some authors also combine ANN models with resampling techniques, such as bagging (Inoue and Kilian 2008) and model combination methods. They find these approaches can significantly improve volatility forecasting accuracy in financial markets (McAleer and Medeiros 2011) and energy markets (Baruník and Křehlík 2016).

Despite multiple successful attempts in the previous literature, most studies fail to provide systematic performance evaluations that support the ability of ML to forecast RV under more diverse conditions, such as different markets, loss functions, forecasting horizons and forecasting methods. Therefore, the outstanding performance of ML algorithms from previous studies may be of little generalizability. First, because most studies only cover one or two ML algorithms in a

limited number of markets. The performance of ML algorithms depends on the problem they try to solve, which is known as the "no free lunch theorem" (Wolpert and Macready 1997). In other words, tailoring ML algorithms to different forecasting problems should result in better model performance than applying one ML algorithm to all cases. Therefore, researchers should not expect to identify an overall best performing model for all situations and applications. Instead, researchers should test a wide selection of ML algorithms for RV forecasting to conclude which model has the best forecasting capability over another in different markets. Also, as discussed, the forms of nonlinearities vary from one market to another. It is thus important to investigate whether the winning algorithm in one market is still the best choice in another market. In addition to the insufficient choices of models and markets, previous literature also lacks rigorous robustness checks for their forecasting results. Forecasting results can change under different forecasting methods (Marcellino, Stock and Watson 2006; Ing 2003). There are two forecasting methods regularly adopted by researchers: the iterative and the direct method. The iterative method uses the forecasting results as new inputs and iteratively generates new forecasting outputs. The direct method includes only observed values as inputs and directly forecasts the mean of h-step out-of-sample outputs. To ensure the robustness of ML algorithms' forecasting performances, researchers should compare forecasting results from both methods to identify whether they differ. Furthermore, different market participants care about RV forecasting accuracy at different horizons. Nonlinearities, such as news announcements and seasonality, also emerge at different time frequencies. Therefore, researchers should also evaluate forecasting performances at different out-of-sample forecasting steps.

To fill the gaps discussed above, we evaluate the forecasting performance of multiple ML algorithms and linear models across a wide range of financial markets and rigorous robustness

checks. The linear models include HAR, HAR-PC, LASSO and Ridge. The nonlinear models include popular ML algorithms such as ANN, RF, SVR and Gradient Boosting Regression Tree (GBRT). Unlike previous volatility studies using ANN algorithms (D'Ecclesia and Clementi, 2019; Donaldson and Kamstra, 1997; Gonzalez Miranda and Burgess, 1997), we follow Gu et al. (2020) and adopt modern deep learning practices such as batch normalization, L1/L2 regularization and learning-rate decay to prevent model overfitting and improve performance stability. We select a wide range of highly traded futures markets representing agricultural and energy commodities and the stock market: corn, soybeans, wheat, crude oil and S&P 500 E-mini observed from January 2009 through May 2018. All these futures are heavily traded and their volatilities are affected by different sources of nonlinearities. To check if results are robust under different forecasting methods, we follow Marcellino, Stock, and Watson (2006) and use both the direct and the iterative method. We also choose 1-day, 5-day and 22-day out-of-sample forecasts as suggested in Andersen, Bollerslev, and Diebold (2007). To measure the forecasting performance of the different models and forecasting methods, we use the mean squared errors (MSE) and the QLIKE loss functions. To assess the statistical relevance of performance differences across models we use the Modified Diebold-Mariano (MDM) and the Model Confidence Set (MCS) tests. We identify the optimal RV forecasting models by taking all the methods and measurements into consideration.

We also design a hybrid ANN model to show the relevance of nonlinear functional forms. Unlike the naïve ANN model, the design of our hybrid ANN model based on residual learning proposed in He, Zhang, Ren, and Sun (2015a) has never been used in the RV forecasting literature. Our hybrid ANN model establishes a direct linear connection between outputs and inputs and first fits the linear component. A stack neural network then separates the nonlinear part from the residual. The design of this hybrid ANN model separates linear from nonlinear forecast

components which allows us to assess the need for nonlinearities in the forecasting model. If nonlinear components are redundant after fitting linear components, the hybrid ANN model shrinks the kernel weights of nonlinear components to zero via the L1 regularization. McAleer and Medeiros (2011) propose a similar model. However, their ANN uses a now outdated design that lacks regularization, is prone to overfitting and impractical for determining the importance of nonlinear components in RV forecasting.

Our results show that linear models consistently dominate nonlinear and hybrid ANN models in RV forecasting accuracy. The best (worst) out-of-sample forecast performance measures are usually achieved through linear (nonlinear) models. The MCS and MDM tests show that the superior performance of linear specifications is robust across different forecasting methods and out-of-sample forecasting steps. The MCS test always selects linear and hybrid models, but less so purely nonlinear models. Hybrid ANN models' decomposition shows that they essentially rely on their linear components to produce forecasts, with the nonlinear component representing, on average, less than 3.27% of the total forecast across markets. Consistently, the kernel weights of the nonlinear component are set to zero under most cases. When the forecasting horizon expands to 22 days, nonlinear kernel weights marginally increase but remain close to zero.

In conclusion, we recommend that researchers and market participants should choose linear models over complex, nonlinear ML algorithms when forecasting RV. The phenomena of linear models beating nonlinear models in terms of out-of-sample forecasting has been documented by previous forecasting literature (Stock and Watson 2004; Blanc and Setzer 2020). Adaptive nonlinear models, such as ML algorithms, are prone to overfitting with finite samples and introduce estimation uncertainty that in turn magnifies out-of-sample forecasting errors. Certain non-linear patterns like news-related structural breaks captured by ML algorithms do not

generalize well to future RV. While hybrid ANN models may become more useful with longer forecasting horizons, improvements in forecasting performance over linear specifications are limited and may not be worth the computational burden of these models.

3.2 Relevant Literature

The forecasting of RV during the last decade has heavily relied on linear models. Among them, the HAR model proposed in Corsi (2008) has become the benchmark model for RV forecasting. The HAR model includes past daily, weekly, and monthly average RVs as inputs. However, the optimal number of lagged RVs can be time varying. To select them, the RV forecasting literature has widely used penalized linear models such as LASSO or Ridge Regression as extensions of the HAR model. Audrino and Knaus (2016) and Zhang et al. (2019) combine shrinkage models such as LASSO regression and elastic net with lagged RV to forecast volatility. Audrino and Knaus (2016) conclude that LASSO regression shows no better forecast accuracy than the standard HAR model at a one-day forecast horizon using selected U.S. stock data. Based on oil market data, Zhang et al. (2019) find a less than 2% forecast improvement with the elastic net over the HAR model. In addition, they do not find significantly different forecasting accuracy between the HAR model and LASSO.

Nonlinearities in financial markets' volatility are also of concern to volatility forecasting. Financial markets' volatility is usually characterized by nonlinear patterns caused by an array of events such as scheduled releases of public information, seasonality, delivery date, or macroeconomic conditions. Bian et al. (2021) find USDA reports induce short-lived bursts of RVs in corn and soybean markets. The same effects are also identified by Miao et al. (2018) in the crude oil market and Du et al. (2018) in the equity market with the release of the Energy Information Administration (EIA) inventory reports and the Federal Open Market Committee (FOMC) reports,

respectively. Seasonality is another source of nonlinearity that creates cyclical effects. Egelkraut et al. (2007) show that both realized and implied volatilities reach peaks during the corn growing season. Arismendi et al. (2016) incorporate a seasonally varying long-run mean-variance process into the Heston (1993) stochastic volatility model for natural gas and corn options. The seasonal pattern significantly improves pricing accuracy for both options contracts. Futures contract price volatilities also increase as the delivery date approaches. Anderson and Danthine (1983) use a theoretical model to show that futures prices become more volatile when most uncertainty is resolved. Karali et al. (2010) document this effect in corn, soybeans and oats futures. Macroeconomic conditions can also affect volatility. Beltratti and Morana (2006) associate structural breaks in the stock market volatility process with the Federal funds rate and M1 growth. Karali and Power (2013) jointly test the impact of macroeconomic factors on agricultural, energy and metal futures volatilities. They find agriculture and crude oil markets are affected by changes in exchange rates, industrial production activities, and the spread between the 10-year and 2-year Treasury rates. Patton and Sheppard (2015) suggest that signed jumps and leveraged volatilities can increase RV forecasting accuracy.

To capture nonlinearities in financial markets' volatility process, researchers propose highly adaptive and flexible nonlinear models. As a result, ML algorithms such as RF, GBRT and SVR have become popular in the recent volatility forecasting literature. RF combines the regression tree model and bootstrapping methods (Breiman 2001). RF lies on the notion that while a simple regression tree model is a weak learner that generates noisy predictions, the average output of uncorrelated simple trees yields higher predictive accuracy. Luong and Dokuchaev (2018) use RF to predict the direction of future realized volatility. The predicted directions are included in the HAR model to further improve RV forecasting accuracy. In contrast to RF that builds parallel

uncorrelated trees, GBRT grows trees sequentially (Friedman 2001) by following the steepest gradient descent in a cost function, such as the mean square error (MSE) function. Yet, GBRT has received limited attention in RV modeling despite its reputation in other forecasting areas. Unlike RF and GBRT, SVR is based on kernel functions and identifies nonlinear patterns in features spaces. Vapnik et al. (1997) introduce the $\epsilon-$ insensitive loss function for SVR estimation and generate robust prediction results. Recent studies adopt SVR models for volatility forecasting in different financial markets. Santamaría-Bonfil et al. (2015) use SVR to predict the volatilities of market indexes in ASEAN and Latin American counties. Their results suggest that SVR outperforms the GARCH (1, 1) model in these markets. Peng et al. (2018) propose a hybrid model between SVR and GARCH models. The hybrid model generates superior forecasting results in currency and cryptocurrency markets than other traditional methods such as GARCH, EGARCH and GJR-GARCH models.

The ANN model can approximate any functional form through an arbitrarily large number of neurons (Cybenko 1989) and has also been considered to forecast RV. McAleer and Medeiros (2011) conclude that the ANN-based HAR model with bagging surpasses the linear HAR model without bagging. Bagging relies on an algorithm that combines predictions of multiple time series models and has been shown by Inoue and Kilian (2008) to significantly improve predictive accuracy. As acknowledged by the authors, the superior forecasting results are generated by the bagging technique rather than forecasting capabilities of the underlying models themselves. In energy futures markets, the empirical results from Baruník and Křehlík (2016) suggest that the forecasting accuracy from combining the HAR and ANN models surpasses both the HAR and ANN model individually. Gonzalez Miranda and Burgess (1997) conclude that the ANN model can generate better volatility forecasting results than linear models in the Ibex35 options market

due to its flexibility to capture complex nonlinear patterns. Similar results are also identified by D'Ecclesia and Clementi (2019) and Donaldson and Kamstra (1997) in major stock market indexes.

Two patterns emerge from previous studies. First, ML algorithms can beat traditional models in conditional variance forecasting. As suggested by Andersen et al. (2003), GARCH-family models rely on slowly decaying moving averages to forecast future volatility, which makes them less responsive to sharp changes in the recent volatility process. In contrast, ML algorithms are free of parametric assumptions and can utilize nearby volatility spikes to generate more accurate forecasting results. Second, the ML-boosted HAR-type model can consistently beat other models, including the HAR model. Previous studies suggest that the ML component of the hybrid HAR-type model capture nonlinearities and boost forecasting performance of the HAR model (McAleer and Medeiros 2011; Baruník and Křehlík 2016). Yet, these studies did not explain how the hybrid HAR-type model utilize linear and nonlinear components when forecasting RV in different markets.

3.3 Methods

To forecast volatility, we first compare the forecasting accuracy between linear and ML algorithms to illustrate whether flexible nonlinear models perform better than linear models. To further the comparison, we design a hybrid ANN model and decompose its volatility forecasts into linear and nonlinear components. Unlike ANN models in the previous literature, our hybrid ANN model is inspired by He et al. (2015a) and utilizes a residual connection with a linear activation function to create a shortcut connection between inputs and outputs. The shortcut connection forces the stacked neural network structure to learn nonlinear patterns from the residuals of the linear connection. The novel design helps us identify the contribution of each component (linear and nonlinear) in volatility forecasting. We compare the hybrid-ANN with the naïve ANN model,

which does not comprise the residual connection. All models are evaluated under both direct and iterative forecasting methods. The iterative forecasting method forecasts one step at a time (h=1) and reuses RV forecasting results as inputs to move into the future and generate the h-step out-of-sample forecasting results. The direct method uses the average of the next h-step out-of-sample RV as its forecasting target. For convenience, forecasting methods are discussed in the empirical application section.

3.3.1 Linear HAR Model

Andersen et al. (2003) show that RV approximates integrated volatility (IV) better than other parametric conventional models. Let $RV_t = \sqrt{\sum_{\mathrm{n_t}=0}^{N_t}\left[p_{n_t+1,t} - p_{\mathrm{n_t},t}\right]^2} = \sqrt{\sum_{n_t=1}^{N_t} {r_{n_t,t}}^2}$, where $p_{n_t,t}$ is the n_tth intraday log-price on day t, $r_{n_t,t}$ is the n_tth intraday log-return, and N_t is the number of returns within day t. Intraday returns are plagued with microstructure noise that induces return autocorrelation (Bandi and Russell 2006; Hansen and Lunde 2006) and renders RV a biased estimator of integrated volatility. Hence, we use a consistent daily IV estimator ($\widehat{IV}_t^d$) that corrects for autocorrelation in returns caused by microstructure noise (Christensen, Oomen and Podolskij 2014; Jacod et al. 2009). Details are in Appendix A. Let $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\mathrm{d}} = \{\widehat{IV}_{i,t}^d\}_{i=1}^N$ be the 1-dimensional output space and $\widehat{\boldsymbol{IV}} = \{\widehat{IV}_{i,t-1}^d, \ldots, \widehat{IV}_{i,t-22}^d\}_{i=1}^N$ be the P-dimensional input space (P=22). Let $N < N_t$ denote the number of daily observations left after cleaning for microstructure noise. Weekly and monthly IVs are defined in equations (3.1) and (3.2), respectively,

$$\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{w}} = \frac{1}{5}\sum_{p=1}^{5} \widehat{\boldsymbol{IV}}_{\boldsymbol{t-p}}^{\boldsymbol{d}} \tag{3.1}$$

$$\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{m}} = \frac{1}{22}\sum_{p=1}^{22} \widehat{\boldsymbol{IV}}_{\boldsymbol{t-p}}^{\boldsymbol{d}} \tag{3.2}$$

where $\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}$ and $\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}$ denote the average lagged weekly and monthly RV[14], respectively. The simplest version of the HAR model is defined as

$$\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}} = \boldsymbol{\alpha_0} + \boldsymbol{\alpha_1}\left(\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{d}}\right) + \boldsymbol{\alpha_2}\left(\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}\right) + \boldsymbol{\alpha_3}\left(\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}\right) + \boldsymbol{u_t} \tag{3.3}$$

where $u_{i,t} \sim N(0, \sigma_u^2)$. The HAR model can be estimated by ordinary least squares (OLS).

3.3.2 Linear HAR-PC Model

Long memory in volatility is captured in the HAR model through the autocovariance among lagged IVs. The HAR model relies on one-day lag and weekly and monthly average lags. However, equally weighted averages of lagged IVs do not necessarily represent the most persistent component among all possible linear combinations of lags. To allow flexibility in capturing persistent weekly (monthly) volatility components, we replace the HAR weekly (monthly) averages with the first principal component of the set comprising the daily IV lags from 1 to 5 (1 to 22) which we denote as $\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}$ ($\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}$). The HAR-PC model is then defined as

$$\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}} = \boldsymbol{\alpha_0} + \boldsymbol{\alpha_1}\left(\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{d}}\right) + \boldsymbol{\alpha_2}\left(\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}\right) + \boldsymbol{\alpha_3}\left(\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}\right) + \boldsymbol{u_t} \tag{3.4}$$

where $u_{i,t} \sim N(0, \sigma_u^2)$. The HAR-PC model is estimated by ordinary least squares (OLS). Further details of this model are presented in Appendix B.

3.3.3 Penalized Regression Models: Ridge and LASSO Regression

A fixed number of lagged daily IVs may not be an optimal solution over time. Penalized regression models can help find the optimal dynamic lag structure for volatility forecasting. We consider both Ridge and LASSO regression. Ridge regression shrinks the coefficients of lagged daily IVs via L2-norms and is defined as:

[14] In the rest of the article, the term RV refers to the realized integrated volatility (free of microstructure noise).

$$\min_{\boldsymbol{\beta}} \sum_{i}^{N} \left(\widehat{IV}_{i,t}^{d} - \beta_0 - \sum_{\boldsymbol{p=1}}^{\boldsymbol{22}} \widehat{IV}_{i,t-p}^{d} \cdot \beta_p \right)^2 + \lambda \sum_{p=1}^{22} \beta_p^2 \ , \tag{3.5}$$

where λ is the complexity parameter controlling shrinkage. The LASSO regression relies on the L1-norm penalty and is defined as:

$$\min_{\boldsymbol{\beta}} \frac{1}{2} \sum_{i}^{N} \left(\widehat{IV}_{i,t+1}^{d} - \beta_0 - \sum_{\boldsymbol{p=1}}^{\boldsymbol{22}} \widehat{IV}_{i,t-p}^{d} \cdot \beta_p \right)^2 + \lambda \sum_{p=1}^{22} |\beta_p| . \tag{3.6}$$

The only difference between equations (3.5) and (3.6) is the penalty term and has implications for the number of lags considered in the model. While the L1-norm tends to shrink the input space to achieve higher forecasting accuracy by reducing coefficients to zero, the L2-norm does not. To optimally select λ in both models, we use the growing window cross-validation method described later.

3.3.4 Tree-Based Models: RF and GBRT

A regression tree model is a non-linear ML method that predicts $\boldsymbol{\widehat{IV}_t}$ by splitting the data into different subsets by using binary questions (splits) based on the features of the data (the input space). In a sequential process, the tree compares one feature $\widehat{IV}_{i,t-p}^{d}$ at a time against a threshold value and classifies observations into two groups based on a binary classification rule such as $I\left(\widehat{IV}_{i,t-p}^{d} < t_p^s\right)$, where $I(.)$ is the indicator function and t_p^s is the threshold value for $\widehat{IV}_{i,t-p}^{d}$ at the s-th split point, s= 1,…, S. The objective of the sequential classification process is to cleanly divide the data into final classes (nodes) that allow formulating a prediction. The final M data subsets $\{R_m\}_{m=1}^{M}$, are called terminal nodes, while the intermediate nodes are called internal nodes. Each value $\widehat{IV}_{i,t}^{d}$ falls exactly into one terminal node. The predicted outcome for each terminal node is the mean of the $\widehat{IV}_{i,t}^{d}$ values in the same node.

For instance, a regression tree model may divide the data into the following 4 non-overlapping subsets $\{R_m\}_{m=1}^4$ based on the first two elements in the input space (Figure 3.1)

$$\{R_m\}_{m=1}^4 := \begin{cases} R_1 = \{\widehat{IV}_{i,t}^d | \widehat{IV}_{i,t-1}^d \geq t_1^1, \widehat{IV}_{i,t-2}^d \geq t_2^3\} \\ R_2 = \{\widehat{IV}_{i,t}^d | \widehat{IV}_{i,t-1}^d < t_1^1, \widehat{IV}_{i,t-2}^d \geq t_2^2\} \\ R_3 = \{\widehat{IV}_{i,t}^d | \widehat{IV}_{i,t-1}^d \geq t_1^1, \widehat{IV}_{i,t-2}^d < t_2^3\} \\ R_4 = \{\widehat{IV}_{i,t}^d | \widehat{IV}_{i,t-1}^d < t_1^1, \widehat{IV}_{i,t-2}^d < t_2^2\} \end{cases} \tag{3.7}$$

This would yield the following regression tree

$$f\left(\widehat{IV}_{i,t}^d\right) = \sum_{m=1}^{4} c_m \cdot I\left(\widehat{IV}_{i,t}^d \in R_m\right) \tag{3.8}$$

where $c_m = argmin \sum (\widehat{IV}_{i,t}^d - f\left(\widehat{IV}_{i,t}^d\right))^2 = avg\left(\widehat{IV}_{i,t}^d \middle| \widehat{IV}_{i,t}^d \in R_m\right)$.

The regression tree model's prediction error consists of a bias error and a variance error. The bias error originates from erroneous assumptions of the learning methods (underfitting). The variance error arises from the sensitivity of the prediction to changes in the training dataset (overfitting). A tree regression with too many split points will overfit the training data. Intuitively, the tree could keep growing until there is a single observation left in each leaf node, which would eliminate bias at the cost of increasing variance. We rely on cost-complexity-pruning (CCP) to determine the number of optimal split points. Under CCP, the researcher pre-determines the minimum number of observations (n) in each terminal node that stops the growth of the tree structure and results in S_0 split points. Then, a subtree with $S < S_0$ splits is selected by pruning the tree model to minimize the complexity cost function as follows:

$$C_\alpha(T) = \sum_{m=1}^{|S|} \sum_{\widehat{IV}_{i,t}^d \in R_m} (\widehat{IV}_{i,t}^d - f\left(\widehat{IV}_{i,t}^d\right))^2 + \alpha|S| \tag{3.9}$$

where α is the complexity-punishment coefficient. The larger the α, the smaller the optimal number of splits.

A fully-grown tree has low bias but large variance. To reduce variance without increasing the bias, Breiman (2001) proposed the Random Forest (RF) model based on hundreds of individual trees trained on different bootstrapped samples. The final RF predictions are derived by averaging the predictions from each individual tree. The bootstrap method chooses only a limited number of features $P'<P$ to grow the tree. The internal split points of the trees grow until each terminal subspace contains no less than the pre-specified minimum number of observations (n). To optimally select the number of trees (B) to grow and number of features to use (P'), we use the growing window cross-validation.

Unlike the RF model, the Gradient Boosting Regression Trees (GBRT) model depends on shallow trees instead of fully-grown trees. A shallow tree has low variance but high bias. To reduce the bias, the GBRT model iteratively grows shallow trees to fit residuals of existing trees. Starting with a shallow tree (e.g. split points S=1), the GBRT model provides biased predictions for the first step. In the next step, another shallow tree with S splits is grown to fit the prediction residuals of the first tree. The results of these two trees are combined to create a more accurate predictor. During the additive step, a penalty parameter $\eta \in (0,1)$ is applied to the second tree. The same process is repeated until B trees have been created. For the GBRT model, we also use the growing window cross-validation method to select the set (B, S, η) with the smallest validation error.

3.3.5 Support Vector Machine for Regression

Through inner products of nonlinear basis functions, the SVR model captures nonlinear trends in the feature space. Given the input space $\widehat{\boldsymbol{IV}} = \{\widehat{IV}_{i,t-1}^{d}, \dots, \widehat{IV}_{i,t-22}^{d}\}_{i=1}^{N}$, a set of nonlinear basis functions $\boldsymbol{h}(\widehat{\boldsymbol{IV}}) = \{h_1(\widehat{\boldsymbol{IV}}), \dots, h_M(\widehat{\boldsymbol{IV}})\}$ transform the P-dimensional input space into an M-dimensional space, where M can be larger than P, that makes linear separation possible. The approximate regression function with the nonlinear basis functions is defined as follows:

$$\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\mathrm{d}} = f(\widehat{\boldsymbol{IV}}) = \sum_{m=1}^{M} \beta_m h_m\big(\widehat{\boldsymbol{IV}}\big) + \beta_0 \tag{3.10}$$

To estimate $= (\beta_0, \beta_1, \dots, \beta_M)$, the SVR model minimizes the cost function:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} V\left(\widehat{IV}_{i,t}^{d} - f\big(\widehat{\boldsymbol{IV}}_i\big)\right) + \frac{\lambda}{2}\|\boldsymbol{\beta}\|^2 \tag{3.11}$$

where λ is the inverse of the cost parameter and is also selected by the growing window cross-validation method. According to Vapnik et al. (1997), the loss function, $V\left(\widehat{IV}_{i,t}^{d} - f\big(\widehat{\boldsymbol{IV}}_i\big)\right)$, should be insensitive to small changes so that the optimization procedure can focus on significant deviations. As a result, the authors propose the use of the following ϵ-insensitive loss function:

$$V_{\epsilon}\left(\widehat{IV}_{i,t}^{d} - f\big(\widehat{\boldsymbol{IV}}_i\big)\right) = \begin{cases} 0, & \left|\widehat{IV}_{i,t}^{d} - f\big(\widehat{\boldsymbol{IV}}_i\big)\right| - \epsilon < 0 \\ \left|\widehat{IV}_{i,t}^{d} - f\big(\widehat{\boldsymbol{IV}}_i\big)\right| - \epsilon, & \left|\widehat{IV}_{i,t}^{d} - f\big(\widehat{\boldsymbol{IV}}_i\big)\right| - \epsilon \geq 0 \end{cases} \tag{3.12}$$

where $\epsilon > 0$. The regression model in equation (10) with the optimal $\widehat{\boldsymbol{\beta}}$ obtained in equation (3.11) can also be expressed as follows:

$$\hat{f}\big(\widehat{\boldsymbol{IV}}\big) = \boldsymbol{h}\big(\widehat{\boldsymbol{IV}}\big)'\widehat{\boldsymbol{\beta}} = \sum_{i=1}^{N} (\hat{\alpha}_i^* - \hat{\alpha}_i)\, \langle \boldsymbol{h}(\widehat{\boldsymbol{IV}}), \boldsymbol{h}(\widehat{\boldsymbol{IV}}_{\boldsymbol{i}}) \rangle \tag{3.13}$$

where $\langle \boldsymbol{h}(\widehat{\boldsymbol{IV}}), \boldsymbol{h}(\widehat{\boldsymbol{IV}}_{\boldsymbol{i}}) \rangle$ is the inner product of two basis functions. The coefficients $\hat{\alpha}_i^*$ and $\hat{\alpha}_i$ are positive and solve the quadratic programming problem as below:

$$\begin{gathered} \min_{\alpha_i, \alpha_i^*} \epsilon \sum_i^N (\alpha_i^* + \alpha_i) - \sum_{i=1}^N \widehat{IV}_{i,t}^{d} (\alpha_i^* - \alpha_i) + \frac{1}{2} \sum_{i,j=1}^N (\alpha_i^* - \alpha_i)\big(\alpha_j^* - \alpha_j\big) \langle \boldsymbol{h}(\widehat{\boldsymbol{IV}}_{\boldsymbol{j}}), \boldsymbol{h}(\widehat{\boldsymbol{IV}}_{\boldsymbol{i}}) \rangle \\ 0 \leq \alpha_i, \alpha_i^* \leq \frac{1}{\lambda} \\ \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0 \\ \alpha_i^* \alpha_i = 0 \end{gathered} \tag{3.14}$$

To solve for the inner products in equation (3.13) and (3.14), we can rely on general kernel functions that help us avoid explicitly estimating the nonlinear basis function $\boldsymbol{h}(\cdot)$. There are several popular choices for the kernel functions. We use the radial basis kernel function which is

denoted as $K(x, y) = exp\{-\gamma\|x - y\|^2\}$. Then, the inner product in equation 13 can be replaced by the radial basis kernel function. Equation (3.13) becomes:

$$\hat{f}(\widehat{\boldsymbol{IV}}) = \sum_{i=1}^{N} (\hat{\alpha}_i^* - \hat{\alpha}_i)\, K(\widehat{\boldsymbol{IV}}, \widehat{\boldsymbol{IV}}_i) \tag{3.15}$$

3.3.6 Hybrid Neural Network Models

Applications of the ANN model to nonlinear time series analysis have been discussed in Franses and van Dijk (2000). In this section, we introduce a hybrid ANN model which can separate linear components from nonlinear components. The ANN-based HAR model proposed by McAleer and Medeiros (2011) is similar to the ANN-based HAR model used in this paper. However, their approach uses a logistic function with an arbitrary number of inputs to capture nonlinearities without any regularization methods. This approach is prone to overfitting and unable to determine the significance of the nonlinear term relative to the linear component.

Instead of using a single layer with a large number of neurons, modern neural network models consist of multiple stacked layers with a smaller number of neurons in each layer. The building structure of a stacked ANN model has a significant impact on the efficiency and stability of forecasting results by reducing vanishing/exploding gradients during the phase of model training. Following He et al. (2015a), we use the residual connection building structure to skip certain middle layers if they fail to improve forecasting results. The layers of the ANN model with residual connections are called residual blocks. Each residual block has the same design as figure 3.2A which, as an example, assumes a residual block with 4 neurons and thus the input and the output of each residual block are (4 x 1) vectors. The identity layer establishes the residual connection by copying the result from the previous layer and adding it back before entering the activation function. The circle with a cross means the addition of results from two layers. If the forecasting

results are not improved after passing middle layers, the optimization procedure will penalize the kernel weights of the 4 neurons via L1 or L2 regularization methods (Appendix B) and only feed the results through the identity layer to the ReLU activation function in figure 3.2A.

We adopt a funnel shape network with residual blocks as it is the common design pattern of modern neural networks (Gu et al. 2020). The actual numbers of neurons in each residual block varies by total number of layers and the distance between each layer and the input layer, which is the first layer in Figure 2B. If there is a total *J* layers in the ANN model, then the $l-th$ layer following the input layer will have 2^{J-l+1} neurons. For example, if the total number of layers is 4 and J^5 is denoted as the final output layer, the number of neurons for each layer before the final output layer are $\{J^1, J^2, J^3, J^4\} = \{2^{4-1+1}, 2^{4-2+1}, 2^{4-3+1}, 2^{4-4+1}\} = \{16,8,4,2\}$. Between any two successive layers, a linear transformation kernel, whose dimension is $J^{l-1} \times J^l$, is applied to the output from the previous layer $(l-1)$ to align its dimension with the residual connection at the current layer (l) (He et al. 2015a).

Residual connections also help in decomposing the results of our ANN models. As presented in Figure 2B, our neural network model contains two parts: a linear component and a nonlinear component. The linear component of $\widehat{IV}_{i,t}^{d}$ consists of a simple linear function as follows:

$$\widehat{IV}_{i,t}^{d,L} = \beta_0^1 + \beta_1^1 \widehat{IV}_{i,t-1}^{d} + \cdots + \beta_{22}^1 \widehat{IV}_{i,t-22}^{d} \tag{3.16}$$

where *L* denotes linear. The linear component is captured by a direct linear connection between input and output layers as shown in Figure 2B. The nonlinear component contains multiple stacked residual blocks that learn complex nonlinear interactions among $\widehat{IV}_{i,t-1}^{d}, \ldots \widehat{IV}_{i,t-22}^{d}$ from the residual, $\tilde{H}\left(\widehat{IV}_{i,t-1}^{d}, \ldots, \widehat{IV}_{i,t-22}^{d}\right)$. The output from the last residual block is $\widehat{IV}_{i,t}^{d,NL}$, with *NL* denoting nonlinear. Then, the prediction of volatility, $\widehat{IV}_{i,t}^{d}$, is expressed as follows:

$$\widehat{IV}_{i,t}^{d} = \widehat{IV}_{i,t}^{d,L} + \widehat{IV}_{i,t}^{d,NL} = \left(\beta_0^1 + \beta_1^1 \widehat{IV}_{i,t-1}^{d} + \cdots + \beta_{22}^1 \widehat{IV}_{i,t-22}^{d}\right) +$$

$$\widetilde{H}\left(\widehat{IV}_{i,t-1}^{d}, \ldots, \widehat{IV}_{i,t-22}^{d}\right) \tag{3.17}$$

where $\widetilde{H}(\cdot)$ is a deeply nested function approximated by stacked residual blocks on the right-hand side of figure 3.2B.

In summary, the building structure of the hybrid ANN model has three main advantages. First, stacked residual blocks are more stable than stacked neural layers without residual connections and can prevent non-convergence issues induced by vanishing/exploding gradients in the deep network structure. Second, the direct linear connections between input and output layers in Figure 2B force stacked residual blocks to focus on nonlinear components. If a linear trend is dominant in volatility forecasting, it will be captured by the linear component in our ANN model. Regularization methods will act as a variable selection process and turn off neurons in residual blocks by shrinking kernel weights to zero for all i. Third, we can decompose the volatility forecast by extracting the linear and nonlinear components separately. Decomposition results allow us to assess the source of the ANN model forecast accuracy and whether nonlinearity increases the out-of-sample forecast accuracy of volatility in different markets. For comparison purposes, we also consider the naïve ANN model which ignores the direct linear connection between input and output layers. This helps us assess the role of the direct linear connection in out-of-sample forecasting.

Besides residual connections, we incorporate several common design patterns in hybrid and naïve ANN models to prevent overfitting and improve numerical stability as well. We use the "He normal initialization" which initializes the weights based on the normal distribution centered at zero allowing for a more diversified initialization among neurons (He et al. 2015b). We also use L1/L2-regularization and batch normalization (Ioffe and Szegedy 2015) to prevent overfitting and reduce the variability of inputs across different layers. We rely on the rectifier activation function

(ReLU) which replaces the widely used sigmoid function for sparse activation and better gradient propagation (Appendix C). Before estimating any ANN models, we also make several choices of hyperparameters. We offer detailed discussions on these choices in Appendix B.

3.3.7 Cross-validation Methods and Hyperparameter Tunning

Hyperparameters of ML models, except ANN models, are tuned with the growing window cross-validation method and the grid search approach. ANN models are manually tuned with the fixed window validation method. We make the exception for ANN models because of both complexity and training time requirements. In contrast to other ML methods, the number of hyperparameters in ANN models is large, which makes the grid search method tedious and non-economical. As result, we make the hyperparameter choice by following previous literature's suggestions and repeated experiments. Also, the growing window cross-validation method leaves too few observations at the beginning of the validation procedure making the performance of the ANN model unreliable. Thus, we separate each training dataset into a training and validation window with a ratio of 80-20. We select the first 80% of each training window as the training data set and the last 20% as the validation data set. In this way, we preserve the time dependency of the observations. Also, this prevents ANN models from overfitting by relying on more recent observations. For each sample period, the ANN models with one to four layers are estimated separately. The model with the lowest validation error is used to generate out-of-sample predictions.

As discussed, for the other ML algorithms we implement a growing window cross-validation scheme. We select the first 20% of the training data as the initial training data and the immediate 200 observations following the initial training data as the initial validation data. Then, we create the new training data by appending the initial validation data to the end of the initial training data

and select the consecutive 200 observations as the new validation data for the next iteration. We continue the process until exhausting the data. Within each iteration, we use the grid search approach to test possible combinations of hyperparameters and calculate forecast errors for each set of hyperparameters. The grid resolution and possible combinations of hyperparameters are presented in Appendix B. Then, we choose the set of hyperparameters with the minimum average forecast error across all iterations as the optimal hyperparameters.

3.4 Empirical Application

In this paper, we focus on five futures contracts representative of commodity and stock markets. These include the three most liquid agricultural commodity futures: corn (C), soybeans (S) and Soft Red Winter wheat (WS); the most liquid energy futures: crude oil (CL); and the most liquid stock market futures: S&P 500 E-mini (ES). All futures transactions prices are time-stamped to the nearest second and obtained from the CME Group's Time-and-Sales database from January 2, 2009, to May 22, 2018. We roll from the nearby to the next deferred contract when the trading volume of the deferred contract is higher than the nearby contract. As illustrated in Table 1, average daily trading volume fluctuates between 60,569 and 1,897,394 contracts across markets, and the daily open interest between 201,012 and 2,780,643 contracts, with ES being the most liquid contract, followed at a distance by crude oil and corn. RV based on equation (B.1) in Appendix B is computed using log transaction price returns for the day trading session hours.[15] The intraday

[15] For agricultural commodities, these vary during our sample period as follows, before May 21, 2012, from 9:30 to 13:15; May 21-December 31, 2012: from 7:30 to 14:00; January 2, - April 5, 2013, from 9:30 to 14:00; since April 8, 2013, from 8:30 to 13:15, and since July 6, 2015, from 8:30 to 13:20. For crude oil and S&P 500 E-mini futures, the markets are active over the entire business day. Therefore, we use all the observations during each day for realized volatility calculation.

data cleaning procedure follows Barndorff-Nielsen et al. (2009). Limit price move days spanning almost all day are excluded as their integrated volatility ($\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$) is close to zero, which does not allow log transformation.[16] The number of days in the sample for each market can be found in Table 1.

3.4.1 Empirical model settings

To identify the relevance of nonlinearities in volatility forecasting, we estimate the different models described in the methods section. Three hybrid ANN models with residual connections are estimated with distinct inputs: ANN-IV (first 22 lags of $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$); ANN-CumSum (cumulative averages of first 22 lags of $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$); and ANN-HAR ($\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{d}}$, $\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}$ and $\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}$). To further investigate the relevance of nonlinear models in forecasting RV, we also extract the linear components ($\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d,L}}$) from all ANN models and compare them against their hybrid counterparts. As mentioned previously, a naïve ANN model without residual connections, denoted as ANN-Naïve (first 22 lags of $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$), is also included for comparison. We use the comparison results between ANN-Naïve and the other hybrid ANNs to show the relevance of the shortcut connection between input and output layers. Hyperparameter settings for selected models are presented in Appendix C and D. Table 2 summarizes the input set for each model. All models have the same output variables across different out-of-sample forecast scenarios. We elaborate on these scenarios in the next subsection.

[16] We exclude 2 observations in the corn market and 1 observation in the soybeans market. These three observations belong to two announcement days, which are Oct. 08, 2010 and Mar. 31, 2011. The RVs on these days are zero.

3.4.2 Out-of-sample forecast evaluation methods

We now turn to the out-of-sample model predictability. To capture any time-varying dynamics in the volatility process, we implement the expanding window sampling approach suggested by Gu et al. (2020)[17]. This approach should not be confused with the growing window cross-validation scheme mentioned earlier. The expanding window sampling approach generates both training and test sample before each iteration of estimation. The growing window cross-validation scheme is for cross-validation purposes and only applied to the training sample. At the start, an initial training dataset of size w_0 = 800 (≈35% of the total observations) and a non-overlapping testing dataset of size t_0 = 22 (≈ one month) are selected. Each model is trained on the training dataset and generates prediction results for the next month. During the training, the growing window cross-validation scheme is applied to the initial training dataset of size w_0 = 800 as described before. Then, the testing dataset is appended to the end of the training dataset and form a new training dataset of size w_1=822. Simultaneously, a new non-overlapping test dataset of size t_1=22 is generated. A new model is trained on the new training dataset and tested on the new testing dataset. The growing window cross-validation scheme is also applied with the new training dataset with 822 observations. The same procedure repeats itself until exhausting all the inputs.

[17] We also produce the out-of-sample forecasts following the widely used fixed rolling window approach (Baruník and Křehlík 2016; Fernandes, Medeiros and Scharth 2014). To determine how results vary by window size, two training samples with sizes w=800, and w=1,200 are assessed. We find time varying volatility patterns in all markets since the estimation period contains significant market events, such as financial crisis and extreme weather conditions. Overall, the rankings of forecast accuracy of the different models are similar between fixed rolling and growing window approaches. The forecasting results between w=800 and w=1200 are similar for all models.

We use $h = 1, 5$, and 22 out-of-sample forecast steps to measure each model's forecast accuracy. The one-step-ahead forecast is equivalent to the fitted value of the estimated model. Multi-step ahead forecasts are computed using both the iterative forecasting method (i.e. by re-inserting the forecast back into the forecasting model) and the direct forecasting method (i.e. by directly estimating the mean of the h steps ahead output variable). For the iterative forecasting, we define the forecast error for each step as $\boldsymbol{e_h = \frac{1}{h}\sum_{s=0}^{h-1} \widehat{IV}_{t+s} - \widehat{IV}_{t+s|t-1}}$, where $\boldsymbol{\widehat{IV}_{t+s|t}}$ is the out-of-sample forecast of $\boldsymbol{\widehat{IV}_{t+s}}$ based on $\mathcal{F}_t$, the information set available at time t, and h is the forecast horizon. For the direct forecasting, we directly forecast the average of the next h-step out-of-sample RV. We generate three forecast errors as below:

$$\begin{cases} \boldsymbol{e_1 = \widehat{IV}_t^d - \widehat{IV}_{t|t-1}^d} \\ \boldsymbol{e_5 = \widehat{IV}_t^w - \widehat{IV}_{t|t-1}^w} \\ \boldsymbol{e_{22} = \widehat{IV}_t^m - \widehat{IV}_{t|t-1}^m} \end{cases}, \tag{3.18}$$

one for each of the three direct forecasts ($\boldsymbol{\widehat{IV}_{t|t-1}^d, \widehat{IV}_{t|t-1}^w}$ and $\boldsymbol{\widehat{IV}_{t|t-1}^m}$) where $\boldsymbol{\widehat{IV}_t^w}$ and $\boldsymbol{\widehat{IV}_t^m}$ are defined as follows:

$$\begin{cases} \boldsymbol{\widehat{IV}_t^w} = \dfrac{1}{5}\displaystyle\sum_{p=0}^{4} \boldsymbol{\widehat{IV}_{t+p}^d} \\ \boldsymbol{\widehat{IV}_t^m} = \dfrac{1}{22}\displaystyle\sum_{p=0}^{21} \boldsymbol{\widehat{IV}_{t+p}^d} \end{cases} \tag{3.19}$$

Notice that $\boldsymbol{\widehat{IV}_{t|t-1}^d}$, $\boldsymbol{\widehat{IV}_{t|t-1}^w}$, and $\boldsymbol{\widehat{IV}_{t|t-1}^m}$ are the direct estimate of corresponding h-day moving averages in the future instead of the cumulative averages of h 1-day ahead output estimates.

Out-of-sample forecast accuracy is computed using the following forecast loss functions:

$$MSE_h = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} e_{i,h}^2 , h = 1, \dots, 22 \\ \frac{1}{N}\sum_{i=1}^{N} \widehat{IV}_{i,t}^{j} - \widehat{IV}_{i,t|t-1}^{j} , j = d, w, m \end{cases}$$

$$QLIKE_h = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} \widehat{IV}_{i,t+s|t-1}^{d} + exp\{\widehat{IV}_{i,t+s}^{d} - \widehat{IV}_{i,t+s|t-1}^{d}\}, \ \ s = 0, \dots, 21 \\ \frac{1}{N}\sum_{i=1}^{N} \widehat{IV}_{i,t|t-1}^{j} + exp\{\widehat{IV}_{i,t}^{j} - \widehat{IV}_{i,t|t-1}^{j}\}, \ \ j = d, w, m \end{cases} \quad (3.20)$$

where N is the number of out-of-sample predictions. The MSE stands for mean squared errors. The QLIKE loss function requires exponential transformation due to the negative values of log volatility estimates. In equation (3.20), we define the MSE and QLIKE loss functions separately for the iterative (top equation) and the direct (bottom function). As suggested in Patton (2011), the QLIKE and MSE loss functions are robust to noise in the proxy for volatility.

3.5 Results

3.5.1 Forecasting results across different scenarios

Out-of-sample loss function values for the linear, hybrid and nonlinear models are provided in table 3.3, with table 3.3A (3.3B) presenting results for the iterative (direct) forecasting method. The group of linear models includes HAR, HAR-PC, LASSO, and Ridge. For hybrid models, we choose ANN-IV, ANN-CumSum, and ANN-HAR. Finally, nonlinear models include RF, SVR, and GBRT. The ANN-Naïve model is excluded from the nonlinear models in table 3.3 because of its inferior forecasting performance relative to all the other models. We consider the ANN-Naïve model in table 3.4 and compare its performance to other hybrid ANN models. Notice that for each model specification, 60 forecast accuracy measures are computed in total: 5 markets (C, S, WS, CL and ES) x 2 loss functions (MSE and QLIKE) x 3 steps (h = 1, 5, 22) x 2 forecast methods (indirect and direct).

Results suggest that linear models consistently dominate nonlinear and hybrid models in RV forecasting. In tables 3.3A and 3.3B, if nonlinearities played an important role in volatility forecasting, hybrid and nonlinear models should outperform linear models. Out of the 60 forecast accuracy measures per model specification, ANN-CumSum is the best model in five occasions, with ANN-IV, ANN-HAR and RF being best model on one occasion each. Hence, jointly nonlinear and hybrid models lead to minimum loss function values in only 8 occasions. Most minimum loss function values are related to HAR (32), followed at a distance by Ridge (9). In contrast, all the maximum loss function values are from either nonlinear or hybrid models, especially from SVR. These results suggest that nonlinear and hybrid models are usually unnecessary for RV forecasting and lead to worse forecasting performance.

To further compare across different models, we employ the model confidence set (MCS) by Hansen et al. (2011). Intuitively, the MCS test builds a group of models that contains the best forecasting model given a confidence level. The details of the MCS method are presented in Appendix E. MCS results for each forecast evaluation measurement are integrated into Table 3. The cells with two (one) asterisks and bold (regular) font represent the optimal model set given $\alpha = 0.25$ ($\alpha = 0.10$), with α being the significance level. These correspond to 75% (MCS_{75}) and 90% (MCS_{90}) confidence levels, respectively. We follow previous literature and focus on MCS_{75} for result interpretation. Several patterns emerge from the MCS_{75} test results. First, the MCS_{75} test results in a disproportionate selection that favors linear models over other specifications. While linear models represent 40% of all candidate models, 69.45% (63.25%) of the optimal model set consists of linear models under the iterative (direct) method. Notice that the proportion of linear models in the optimal set is larger under the iterative forecasting where forecast bias is more prone to occur. Second, nonlinear models, which represent 30% of all candidate models, are much less

likely to be selected by the MCS_{75} test. Only 6.61% (11.45%) of the optimal models selected by the MCS_{75} are nonlinear under the iterative (direct) method. Hence, the odds that the MCS_{75} chooses nonlinear models test are much lower than linear models' odds. More nonlinear models are selected under the direct method than the indirect method. Despite nonlinear models being unlikely to pass the MCS_{75} test, when they do, their mean loss function values are considerably larger than their linear competitors. Taking the wheat market as an example, the RF model passes the MCS_{75} test for $h = 22$ under the MSE measure and direct forecasting (Table 3.3B). However, its MSE equals 0.0177, which is 14% higher than the HAR model in the same comparison group. These results strongly suggest that linear models surpass nonlinear models in volatility forecasting. The portion of hybrid models passing the MCS_{75} test is 23.94% (25.3%) under the iterative (direct) method. Since hybrid models constitute 30% of all candidate models for the MCS_{75} test, hybrid models pass the MCS_{75} test more frequently than nonlinear models, but much less frequently than linear specifications. However, in some instances, hybrid models have lower loss measures than linear models. Table 3.3 shows that more hybrid ANN models pass the MCS_{75} test in soybeans and S&P 500 E-mini markets when $h = 5$ and 22. In Table 3.3A, for example, the ANN-CumSum model carries the minimum MSE and QLIKE measures in the soybeans market for $h = 5$ and $h = 22$.

We also compare loss function values between the iterative and direct methods to assess the robustness of the results across forecasting methods. First, linear models' loss function values are similar for both iterative and direct methods. Differences in forecasting accuracy between the two methods are however larger for nonlinear and hybrid models. If a one-period ahead model is correctly specified, then its loss function under the iterative method should be no more than under the direct method. Otherwise, we should expect to see significant loss function reductions under

the direct forecasting method (Marcellino et al. 2006). Taking the corn market as an example, when $h = 22$, the RF model's MSE is 0.0246 under the direct method and 0.0335 under the iterative method. The same measurements for the HAR model are identical and smaller. Therefore, the HAR model is better specified as a one-step-ahead forecasting model than the RF model. We test the differences between iterative and direct methods for each model under $h = 5$ and 22 across different markets using the Modified Diebold-Mariano (MDM) test. We omit h =1 because the forecast results are almost identical between iterative and direct methods. MDM test details are presented in Appendix B. Table B.1 presents the p-values of the MDM test. Results suggest that for the linear models' group, there are no statistically significant differences between the iterative and direct forecasting methods. Some statistically significant differences are however identified in hybrid and nonlinear models, almost all of them in ANN-HAR and RF, which suggests one-step ahead misspecification of these models and favors the direct over the iterative method.

3.5.2 The comparison of forecasting results between hybrid ANN models and their linear components

To shed further light on the forecasting role of nonlinear functional forms, we assess whether nonlinear components contribute significantly to the good performance of hybrid ANN models. We do so by separating the linear components ($\widehat{IV}_{i,t}^{j,L}$) from the total volatility prediction ($\widehat{IV}_{i,t}^{j}$) as shown in equation (3.17), where $j = d, w, m$. In Table 3.4A and 3.4B, we present the loss measures of the hybrid ANN models and their linear components for the iterative and direct method, respectively. We also consider the ANN-Naïve to help illustrate the relevance of linear components in the hybrid ANN model. We apply the MCS test on the seven models[18] to identify

[18] We also conduct the MCS test without the ANN-Naïve model. The results remain the same.

whether linear components are more likely to be included in the final set, which would signal the irrelevance of nonlinearities. We also apply the MDM test between each pair of hybrid models and their respective linear components in Table B.2A and B.2B in Appendix B[19]. We focus the discussion on the MCS_{75} test because the MDM test results are qualitatively similar.

The ANN-Naïve model generates the worst forecasting results and it is never selected by the MCS_{75} test. The only difference between the ANN-Naïve model and other hybrid models is the missing linear component. Thus, the comparison results between the ANN-Naïve model and other ANN hybrid models emphasize the importance of linear components in those models. Its associated MSE and QLIKE measures are much worse than any hybrid ANN model (Table 3.4A and 3.4B). For instance, the MSE measure for the ANN-Naïve model in the corn market for $h = 1$ in Table 4A is 0.0739. The second highest MSE value is 0.062, which belongs to the linear component of the ANN-IV model. Hence, the MSE value for the ANN-Naïve model is at least 20% larger than the rest of hybrid models, which leads the MCS_{75} test to never select ANN-Naïve models. As mentioned earlier, the ANN-Naïve model has the same design as the ANN-IV model but excludes the linear connection in Figure 2B. Given the performance gap between the ANN-Naïve and ANN-IV model, researchers should consider adding linear connections when using neural networks to forecast volatility in futures markets. Due to its significantly worse forecasting results, we ignore the ANN-Naïve model from the following discussions.

[19] We realize that the hybrid ANN model and its linear component are nested models. For nested models, the Clark-McCracken test (Clark and McCracken 2001) is more appropriate. However, the MDM test is still suitable for the purpose of comparing forecasting accuracy between two models. Besides, Clark and McCracken (2001) test relies on the assumption that the parameters of the test are estimated by OLS, which is not satisfied with the hybrid ANN model.

Under the iterative method, hybrid models can only beat their linear components in one-day ahead RV forecasting. In Table 3.4A and for $h = 1$, loss function values are smaller for the hybrid models than their corresponding linear components. For $h = 5$ the differences in loss values between hybrid models and their linear components become very small, but the balance starts to tip in favor of the latter. For $h = 22$, differences in loss functions grow again with linear components prevailing over the full hybrid model. For example, the MSE values of the ANN-IV model and its linear component are 0.0611 (0.027) [0.0248] and 0.062 (0.0269) [0.0219] in the corn market when $h = 1\ (5)\ [22]$. These results are consistent with the MCS_{75} test results. As discussed in Patton (2011), MSE is more sensitive to outliers than QLIKE, which has implications for our MCS_{75} results. While for $h = 1$ the MSE-based MCS_{75} test does not differentiate between hybrid models and their linear components, QLIKE-based MCS_{75} test suggests hybrid models are generally better than their linear components. For example, all hybrid models and their linear components in corn, soybean, and S&P 500 E-mini markets are selected by the MSE-based MCS_{75} test. Instead, the QLIKE-based MCS_{75} test only chooses the hybrid specifications. Consistent with loss function values pointing at hybrid linear components doing a better job than the overall hybrid models for $h = 5$ and 22, the MCS_{75} tests based either on QLIKE or MSE measures usually fail to differentiate the forecasting performances between the two models. Based on these results, we conclude that hybrid models can beat their linear components when $h = 1$ under the iterative method. However, the edge of hybrid models disappears over a longer forecasting horizon.

Under the direct method, hybrid models yield lower forecasting errors than their linear components. Table 3.4B shows that linear components' loss function values are systematically worse than the hybrid models in all markets except for S&P 500 E-mini market at $h = 5$ and 22. For instance, the MSE values for the ANN-IV model and its linear component in the corn market

are 0.0612 (0.0269) [0.0235] and 0.0615 (0.0274) [0.0276] when $h = 1$ (5) [22] under the direct method. For the same reason mentioned before, the MSE-based MCS_{75} test still struggles to differentiate forecasting performances between hybrid models and their linear components, especially for longer forecasting horizons. In contrast, the QLIKE-based MCS_{75} tests generally suggest that hybrid models are significantly better than their linear components for forecast horizons $h = 1$ and $h = 5$. For $h = 22$, however, the same test chooses both the hybrid models and their linear components in all markets except for the soybeans market. Hence, we conclude that hybrid models can outperform their linear components when $h = 1$ and 5 under the direct method.

3.5.3 Nonlinear components of hybrid ANN models

To better understand how hybrid ANN models utilize nonlinear components in RV forecasting, we explicitly calculate the nonlinear component $\widehat{IV}_{i,t}^{j,NL}$ (equation (17)) for the ANN-IV model across different markets and out-of-sample forecasting steps. We select the ANN-IV model because it uses the 22 lagged RVs separately. Thus, we can examine how hybrid models utilize different lagged inputs to generate the nonlinear component. Also, the results of the ANN-IV model are similar with the other hybrid models. We use the ratio of the absolute value of the nonlinear component over total volatility (we refer to them as $\widehat{IV}_{i,t}^{j,NL}$ ratios) to evaluate the relative importance of the nonlinear component on RV forecasting. We choose the direct method because it allows for more differences in both the absolute magnitude of $\widehat{IV}_{i,t}^{j,NL}$ and the $\widehat{IV}_{i,t}^{j,NL}$ ratios across different forecasting horizons. This is caused by the direct method involving a different model for each different forecasting horizon. As mentioned earlier, different nonlinearities emerge over different periods. Thus, both the absolute magnitude and the relative ratios of $\widehat{IV}_{i,t}^{j,NL}$ can behave differently under the direct method. Unlike the direct method, under the iterative method, $\widehat{IV}_{i,t}^{j,NL}$

and $\widehat{IV}_{i,t}^{j,NL}$ ratios remain unchanged across forecast horizons due to the same model being iteratively reused 22 times to produce the forecasts.

We compute the ratio for each market and out-of-sample forecasting step and present summary statistics in table 3.5. The $\widehat{IV}_{i,t}^{j,NL}$ ratios are less than 3.27% on average across futures markets and forecast horizons and remain below 8.55% under all scenarios. The $\widehat{IV}_{i,t}^{j,NL}$ ratio grows as the number of out-of-sample forecasting steps increases. The table shows that mean $\widehat{IV}_{i,t}^{j,NL}$ ratios increase from a range between 0.53% and 1.24% for $h = 1$ to a range between 1.87% and 3.27% for $h = 22$. These results can be attributed to the time required for nonlinear patterns such as seasonality or scheduled releases of public information to emerge. As suggested by median values in Table 5, more than half of $\widehat{IV}_{i,t}^{j,NL}$ ratios over the entire forecasting period are less than 0.5% to 1% for $h = 1$, 0.95% to 1.51% for $h = 5$, and 1.8% to 2.62% for $h = 22$. Thus, we conclude that the relative influence of nonlinear components on hybrid models is very small.

The small $\widehat{IV}_{i,t}^{j,NL}$ ratios are the result of the reduced kernel weights that hybrid models attribute to their nonlinear components. From equation (3.17), and under the assumption that $\widetilde{H}$ contains only one layer, we can express $\widehat{IV}_{i,t}^{j,NL}$ as follows:

$$\widehat{IV}_{i,t}^{j,NL} = \widetilde{H}\left(\widehat{IV}_{i,t-1}^{d}, \dots, \widehat{IV}_{i,t-22}^{d}\right) = \widetilde{H}\left(\sum_{m=1}^{22} W_m^1 \cdot \widehat{IV}_{i,t-\mathrm{m}}^{d} + W_0^1\right) \tag{3.21}$$

where W_m^1 is the kernel of the m-th lagged daily IV value in the first layer of the nonlinear component in Figure 3.2B and W_0^1 is the bias term of the first layer. Whenever the kernels for lagged IVs are set to zero, $\widehat{IV}_{i,t}^{j,NL}$ becomes a function of the bias term W_0^1 and incorporates no dynamics through information on lagged IVs. Hence, $\widehat{IV}_{i,t}^{j,NL}$ becomes a constant value over each out-of-sample forecasting window until the next iteration. Although $\widetilde{H}$ may contain more than one

layer, if the kernels become zero in the first layer, they force the outcomes of the ANN model to be zeros regardless of the kernel values in subsequent layers. This results in $\widehat{IV}_{i,t}^{j,NL}$ displaying a stair-like time series plot such as the one observed in Figure B.1 in Appendix B for the wheat market.[20] The stair-like shape of the plot suggests that $\widehat{IV}_{i,t}^{j,NL}$ is constant, with its specific value being adjusted each time the model is re-estimated.

To illustrate kernel sizes, we continue focusing on the ANN-IV model and calculate the kernel weights of the first layer for each lagged IV. For example, if there are four neurons in the first layer of the ANN-IV model, then the size of $W^1 = [W_1^1, \dots, W_{22}^1]$ is 4×22 with each W_m^1 being a 4×1 vector. We calculate the absolute average of the four coefficients in W_m^1 and use it as a summary statistic of the kernel weighting the m-th lagged IV. Kernel weights are re-calculated once the ANN-IV model is re-estimated. We also include the 22 coefficients of lagged IVs in $\widehat{IV}_{i,t}^{j,L}$ from equation (3.17) as the comparison group and call these kernel weights of the linear component.

To illustrate the heterogeneity of kernel weights for linear and nonlinear components, we separate all kernels into daily, weekly, and monthly lag groups and present summary statistics of kernel weights within each group for linear and nonlinear components separately in Table 3.6. The daily (weekly) [monthly] lag group "L1" ("L2-5") ["L6-22"] contains the 1st (2nd to the 5th) [6th to the 22nd] lagged IV kernel weights. We ignore $h = 5$ case in Table 3.6 because it has similar patterns as $h = 1$ and $h = 22$. The average linear (nonlinear) kernel values when $h = 5$ are lower (higher) than $h = 1$ but higher (lower) than $h = 22$. Thus, adding $h = 5$ to Table 3.6 will not alter the conclusion based on $h = 1$ and 22. For the same reason mentioned before, we only present the

[20] The results for other markets are qualitatively similar.

results under the direct method. Since weights are very small, we multiply all values by 100 to facilitate reading the results.

Several patterns can be observed in table 3.6. First, the mean values of nonlinear kernel weights are between 0 and 0.01 (0.01 and 0.09) at $h = 1\ (22)$, while the same values are between 2.74 and 60.49 (1.61 and 33.08) for corresponding linear kernel weights. The differences between linear and nonlinear kernel weights support the significantly smaller ratios of nonlinear components in table 3.5. Second, the kernel weights change along with out-of-sample forecasting steps. For linear components, the kernel weights shrink as the number of step increases. However, the kernel weights of nonlinear components increase as the number of step increases. For example, in the crude oil market, the mean daily, weekly, and monthly kernel weights of nonlinear components are 0.01 (0.03), 0.01 (0.03), and 0.01 (0.04) at $h = 1\ (22)$. For linear components, the same numbers are 59.79 (29.94), 6.06 (5.51), and 2.74 (2.04) at $h = 1\ (22)$. This result indicates that the relative ratios of nonlinearity grow along with the number of out-of-sample forecasting steps, which can also be observed in table 3.5. As we mentioned earlier, some forms of nonlinearities, such as seasonality, only emerge at the weekly or even monthly forecasting frequency. Therefore, hybrid models may be useful over the medium- and long-term forecasting horizons. We also include a heatmap for the wheat market's linear and nonlinear kernel weights for $h = 22$ in Figure B.2[21] in the Appendix B. The heatmap presents more granular information on how kernel weights of linear and nonlinear components change over time. Overall, the results from the heatmap are largely consistent with results in Table 6 and with Audrino and Knaus (2016).

[21] The results for other markets are qualitatively similar.

Based on the findings in table 3.5 and 3.6, we can conclude that hybrid models tend to shrink nonlinear components to constant values and use them to adjust the intercepts in the linear components. Although hybrid models act like linear models, their optimization procedure is less efficient than simple linear models. Therefore, the loss function measures of hybrid models cannot surpass linear models in Tables 3.3A and 3.3B, especially for one-day-ahead RV forecasting.

3.6 Concluding remarks

Increased availability of high-frequency intraday financial data over the past decade led to the introduction of realized volatility (RV) measures, which initialized a new era of modeling and forecasting financial volatility. Linear models, such as the Heterogeneous Autoregressive (HAR) model and its extensions, became the pillar of the RV-based volatility modeling and forecasting due to its simplicity and consistent performance across different empirical applications. However, nonlinearities in volatility driven by factors, such as seasonality, time-to-delivery effects, or scheduled releases of public information, have imposed challenges on linear models. Recent applications of machine learning (ML) algorithms suggest that ML models are particularly suited to capture unknown forms of nonlinearities. Despite the methodological improvements in ML that have become available in the last decade, the literature has paid only limited attention to using these developments in the context of RV forecasting.

This research provides a systematic comparison of the forecasting performance of linear HAR-type and nonlinear ML-based models. We choose HAR, HAR-Principal Components (HAR-PC), Lasso Regression (LASSO) and Ridge regression (Ridge) as linear specifications. Nonlinear models considered are popular ML algorithms such as the Random Forest (RF), Gradient Boosting Regression Trees (GBRT) and Support Vector Machine for Regression (SVR). To further assess the need for nonlinearities to produce better forecasts, we consider a hybrid Artificial Neural

Network (ANN) model with a novel design based on residual learning which allows us to separate forecasts into linear and nonlinear components. We forecast the noise-free daily RV in highly traded futures markets including corn, soybeans, wheat, crude oil and S&P 500 E-mini observed from January 2009 through May 2018. We use different tests and forecast accuracy measures to compare the different models. For each model, we also implement both direct and iterative forecasting methods.

The main finding is that linear models generally produce better RV forecasting results than nonlinear models in the futures markets. The best out-of-sample forecast performance measures are usually achieved by linear models, especially by the simple HAR model. In contrast, the worst performance measures are attributed to nonlinear models, especially SVR. Further, the Model Confidence Set (MCS) is less likely to choose nonlinear models than linear and hybrid models. The underperformance of RF, GBRT and SVR models indicates the disadvantages of using highly nonlinear models in this context. The decomposition results of the hybrid ANNs suggest that these models heavily rely on their linear components to capture volatility dynamics in different markets. Also, hybrid ANN models reduce nonlinear components to constant values most of the time by setting the kernel weights inside the nonlinear components to values close to zero. As a result, hybrid ANN models serve as linear models during most forecasting periods. This results in the nonlinear component representing, on average, less than 3.27% of the overall hybrid model volatility forecast.

Our results help researchers and investors make more informed decisions when searching for better volatility forecasting models in futures markets. Instead of considering more flexible nonlinear models to fit the ever-changing volatility process, researchers can boost forecast accuracy by finding linear specifications that can efficiently capture long memory in volatility

such as HAR models. While nonlinear functions may have greater explanatory power of the in-sample volatility process (Andersen et al. 2007), nonlinear models do not improve the forecasting accuracy over linear models (Degiannakis et al., 2020; Sévi, 2014). Our paper uses the hybrid ANN models to further show that the linear component plays a dominant role in RV forecasting when both linear and nonlinear components are considered. Even stacked neural networks could not separate meaningful nonlinear trends from the residual of linear component forecasting results. To conclude, we recommend researchers and market participants to choose linear models even in the presence of non-linear dimensions in volatility. For hybrid models, researchers can use them as alternatives to linear models for medium- and long-term RV forecasting in soybeans and S&P 500 E-mini markets based on our empirical findings. However, the gains may not compensate the computational burden for training these models.

3.7 Tables and Figures

Table 3.1: Summary of Futures Contracts

Name	Symbol	Exchange	Contract Months	Avg. Daily Volume	Avg. Daily Open Interest	# of Days
Corn	C	XCBT	Mar, May, July, Sep, Dec	159052	551001	2360
Soybeans	S	XCBT	Jan, Mar, May, July, Aug, Sep, Nov	107965	274971	2361
SRW Wheat	WS	XCBT	Mar, May, July, Sep, Dec	60569	201012	2362
S&P 500 E-mini	ES	XCME	Mar, Jun, Sep, Dec	1897394	2780643	2373
Crude Oil	CL	XNYM	Every Month	388371	315200	2379

Notes: XCBT, XCME, and XNYM are acronyms for Chicago Board of Trade, Chicago Mercantile Exchange, and New York Mercantile Exchange. Our sample period is from January 2, 2009 to May 22, 2018.

Table 3.2: Model Summary

Model	Type	Input Set
ANN-IV	Hybrid	First 22 lags of $\widehat{IV}_t^d$
ANN-CumSum	Hybrid	Cumulative averages of first 22 lags of $\widehat{IV}_t^d$
ANN-HAR	Hybrid	$\widehat{IV}_{t-1}^d$, $\widehat{IV}_{t-1}^w$ - equation (1) and $\widehat{IV}_{t-1}^m$ - equation (2)
Linear-ANN-IV	Linear	First 22 lags of $\widehat{IV}_t^d$
Linear-ANN-CumSum	Linear	Cumulative averages of first 22 lags of $\widehat{IV}_t^d$
Linear-ANN-HAR	Linear	$\widehat{IV}_{t-1}^d$, $\widehat{IV}_{t-1}^w$ - equation (1) and $\widehat{IV}_{t-1}^m$ - equation (2)
ANN-Naïve	Non-linear	First 22 lags of $\widehat{IV}_t^d$
RF	Non-linear	First 22 lags of $\widehat{IV}_t^d$
GBRT	Non-linear	First 22 lags of $\widehat{IV}_t^d$
SVR	Non-linear	First 22 lags of $\widehat{IV}_t^d$
LASSO	Linear	First 22 lags of $\widehat{IV}_t^d$
Ridge	Linear	First 22 lags of $\widehat{IV}_t^d$
HAR-PC	Linear	First lag of $\widehat{IV}_t^d$, and first principal component of the set of $\widehat{IV}_t^d$ lags from 1 to 5, and the set of $\widehat{IV}_t^d$ lags from 1 to 22
HAR	Linear	$\widehat{IV}_{t-1}^d$, $\widehat{IV}_{t-1}^w$ - equation (1) and $\widehat{IV}_{t-1}^m$ - equation (2)

Notes: $\widehat{IV}_t^d$ ($\widehat{IV}_t^w$) [$\widehat{IV}_t^m$] is the consistent daily (weekly) [monthly] IV estimator in Appendix A (equation 1) and [equation 2]

Table 3.3A: Out-of-sample forecasting evaluation (Iterative forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
HAR	1	**0.0598****	**-3.4719****	0.0519*	-3.7247	**0.0427****	**-3.3081****	**0.0251****	**-3.1190****	**0.0490****	-4.0133*
HAR-PC	1	**0.0598****	**-3.4719****	0.0519*	-3.7247	**0.0428****	**-3.3081****	0.0251*	**-3.1190****	**0.0490****	**-4.0134****
LASSO	1	**0.0599****	**-3.4721****	**0.0515****	**-3.7250****	**0.0430****	**-3.3081****	0.0251*	**-3.1190****	**0.0490****	**-4.0133****
Ridge	1	**0.0600****	**-3.4720****	**0.0513****	**-3.7252****	0.0432*	**-3.3080****	**0.0250****	**-3.1190****	**0.0494****	-4.0130*
ANN-IV	1	0.0611	**-3.4714****	0.0522	-3.7246	0.0443	-3.3075	0.0256	-3.1187	**0.0494****	-4.0131*
ANN-CumSum	1	**0.0601****	**-3.4718****	0.0519*	-3.7247*	0.0435*	**-3.3080****	0.0255	-3.1188	**0.0493****	-4.0132*
ANN-HAR	1	0.0601	**-3.4719****	0.0520*	-3.7246	**0.0429****	**-3.3081****	0.0252*	-3.1189	**0.0492****	-4.0133*
RF	1	0.0608	**-3.4721****	0.0533	**-3.7247****	0.0474	-3.3062	0.0282	-3.1176	0.0538	-4.0114
SVR	1	0.0622	-3.4699	0.0547	-3.7230	0.0483	-3.3053	0.0285	-3.1173	0.0560	-4.0101
XGB	1	0.0606	**-3.4718****	0.0540	-3.7240	0.0445*	**-3.3076****	0.0283	-3.1175	0.0547	-4.0108
HAR	5	**0.0257****	**-3.4961****	**0.0212****	**-3.7439****	**0.0193****	**-3.3223****	**0.0203****	**-3.1211****	**0.0488****	**-4.0136****
HAR-PC	5	**0.0258****	**-3.4961****	**0.0212****	**-3.7439****	**0.0193****	**-3.3223****	0.0203*	-3.1211*	**0.0488****	**-4.0136****
LASSO	5	**0.0260****	**-3.4960****	**0.0212****	**-3.7440****	0.0198	-3.3222	0.0207	-3.1209	**0.0489****	**-4.0137****
Ridge	5	**0.0260****	**-3.4960****	**0.0211****	**-3.7439****	0.0197	-3.3221	0.0206	-3.1210	**0.0489****	**-4.0135****
ANN-IV	5	0.0270*	-3.4956*	**0.0212****	**-3.7439****	0.0212	-3.3215	0.0215	-3.1205	**0.0490****	**-4.0137****
ANN-CumSum	5	0.0267*	-3.4956*	**0.0210****	**-3.7440****	0.0209	-3.3217	0.0216	-3.1205	**0.0491****	**-4.0135****
ANN-HAR	5	0.0304	-3.4942	0.0251	-3.7422	0.0225	-3.3209	0.0222	-3.1202	0.0508*	**-4.0131****
RF	5	0.0290*	-3.4947*	**0.0236****	**-3.7429****	0.0238	-3.3204	0.0252	-3.1188	0.0574	-4.0103
SVR	5	0.0283*	-3.4943*	0.0245	-3.7420	0.0255	-3.3192	0.0267	-3.1177	0.0579	-4.0093
XGB	5	**0.0272****	**-3.4954****	0.0233	-3.7430	0.0214	-3.3213	0.0245	-3.1191	0.0554	-4.0108

Note: The model column contains the name of each model; h indicates the number of steps for out-of-sample forecasting; MSE and QLIKE represent the values of each forecasting evaluation measurements; * means the model pass MCS_{90} test; ** and bold fonts indicate the model pass MCS_{75} test.

Table 3.3A (Cont.): Out-of-sample forecasting evaluation (Iterative forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
HAR	22	**0.0226****	**-3.5008****	**0.0216****	**-3.7427****	**0.0158****	**-3.3239****	**0.0275****	**-3.1161****	**0.0606****	**-4.0084****
HAR-PC	22	0.0228*	**-3.5007****	**0.0215****	**-3.7428****	0.0160*	-3.3238*	0.0278*	-3.1159	**0.0610****	**-4.0084****
LASSO	22	0.0237*	**-3.5004****	**0.0223****	**-3.7425****	0.0169*	-3.3235*	0.0287	-3.1155	0.0630*	**-4.0077****
Ridge	22	0.0230*	**-3.5006****	**0.0219****	**-3.7426****	0.0164*	-3.3236*	0.0278	-3.1159	**0.0607****	**-4.0084****
ANN-IV	22	0.0248	-3.4998	**0.0217****	**-3.7427****	0.0197*	-3.3223*	0.0316	-3.1139	**0.0620****	**-4.0084****
ANN-CumSum	22	0.0251*	**-3.4996****	**0.0215****	**-3.7429****	0.0205*	-3.3219*	0.0310	-3.1143	**0.0610****	**-4.0088****
ANN-HAR	22	0.0321	-3.4967	0.0250	-3.7414	0.0223	-3.3208	0.0347	-3.1124	0.0688	-4.0056*
RF	22	0.0335	-3.4958	0.0298	-3.7390	0.0226	-3.3208	0.0362	-3.1113	0.0787	-4.0010
SVR	22	0.0291	-3.4966	0.0292	-3.7381	0.0267	-3.3177	0.0398	-3.1085	0.0760*	-3.9987*
XGB	22	0.0288	-3.4978	0.0256	-3.7409	0.0208	-3.3214	0.0320	-3.1138	0.0714	-4.0025*

Note: The model column contains the name of each model; h indicates the number of steps for out-of-sample forecasting; MSE and QLIKE represent the values of each forecasting evaluation measurements; * means the model pass MCS_{90} test; ** and bold fonts indicate the model pass MCS_{75} test.

Table 3.3B: Out-of-sample forecasting evaluation (Direct forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
HAR	1	**0.0597****	**-3.4768****	0.0518*	-3.7253	**0.0426****	**-3.3067****	**0.0252****	**-3.1210****	**0.0489****	-4.0116*
HAR-PC	1	**0.0598****	**-3.4768****	0.0518*	-3.7253	**0.0426****	**-3.3067****	0.0253*	**-3.1210****	**0.0489****	**-4.0117****
LASSO	1	**0.0598****	**-3.4771****	**0.0514****	**-3.7256****	0.0429*	**-3.3067****	0.0253*	**-3.1210****	**0.0489****	**-4.0117****
Ridge	1	**0.0599****	**-3.4770****	**0.0512****	**-3.7258****	0.0431	**-3.3066****	**0.0252****	**-3.1210****	**0.0493****	-4.0113*
ANN-IV	1	0.0612	**-3.4763****	0.0523	-3.7251	0.0443	-3.3061	0.0258	-3.1208*	**0.0491****	-4.0115*
ANN-CumSum	1	**0.0600****	**-3.4768****	**0.0517****	-3.7254*	0.0435	**-3.3065****	0.0258	-3.1207	**0.0492****	-4.0115*
ANN-HAR	1	**0.0599****	**-3.4768****	0.0520*	-3.7252	0.0429*	**-3.3067****	0.0254*	**-3.1209****	0.0493	-4.0115*
RF	1	**0.0610****	**-3.4770****	0.0541	-3.7247	0.0479	-3.3048	0.0283	-3.1197	0.0534	-4.0098
SVR	1	0.0622	-3.4749	0.0546	-3.7236	0.0481	-3.3040	0.0287	-3.1193	0.0557	-4.0084
XGB	1	**0.0606****	**-3.4768****	0.0539	-3.7246	0.0443	**-3.3062****	0.0280	-3.1197	0.0520	-4.0105
HAR	5	**0.0256****	**-3.5002****	**0.0211****	**-3.7443****	**0.0194****	**-3.3211****	**0.0204****	**-3.1226****	**0.0486****	**-4.0122****
HAR-PC	5	**0.0256****	**-3.5002****	**0.0211****	**-3.7443****	**0.0194****	**-3.3211****	0.0204*	**-3.1226****	**0.0486****	**-4.0123****
LASSO	5	**0.0258****	**-3.5000****	**0.0210****	**-3.7443****	**0.0197****	**-3.3210****	0.0207	-3.1225	**0.0486****	**-4.0123****
Ridge	5	**0.0259****	**-3.5000****	**0.0209****	**-3.7443****	**0.0196****	**-3.3210****	0.0206	-3.1225*	**0.0489****	**-4.0121****
ANN-IV	5	0.0269	-3.4996	**0.0209****	**-3.7444****	0.0214*	-3.3203*	0.0212	-3.1222	**0.0490****	**-4.0121****
ANN-CumSum	5	0.0266	-3.4997	**0.0212****	**-3.7442****	0.0217	-3.3201	0.0211	-3.1223	**0.0489****	**-4.0123****
ANN-HAR	5	0.0261*	**-3.5000****	**0.0211****	**-3.7443****	0.0207*	**-3.3206****	0.0208	-3.1224*	**0.0492****	**-4.0119****
RF	5	**0.0261****	**-3.4999****	**0.0224****	**-3.7436****	0.0216	-3.3201	0.0259	-3.1200	0.0555	-4.0095
SVR	5	0.0281	-3.4989	0.0245	-3.7427	0.0258	-3.3183	0.0254	-3.1199	0.0556	-4.0092
XGB	5	0.0271*	-3.4994	0.0238	-3.7430	0.0221	-3.3198	0.0239	-3.1209	**0.0514****	**-4.0115****

Note: The model column contains the name of each model; h indicates the number of steps for out-of-sample forecasting; MSE and QLIKE represent the values of each forecasting evaluation measurements; * means the model pass MCS_{90} test; ** and bold fonts indicate the model pass MCS_{75} test.

Table 3.3B (Cont.): Out-of-sample forecasting evaluation (Direct forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
HAR	22	**0.0226****	**-3.5008****	**0.0219****	**-3.7426****	**0.0155****	**-3.3240****	**0.0276****	**-3.1160****	**0.0596****	**-4.0087****
HAR-PC	22	**0.0226****	**-3.5008****	**0.0219****	**-3.7426****	**0.0156****	**-3.3240****	**0.0277****	**-3.1159****	**0.0597****	**-4.0087****
LASSO	22	**0.0228****	**-3.5007****	**0.0226****	**-3.7423****	**0.0158****	**-3.3239****	0.0278*	-3.1159*	**0.0602****	**-4.0083****
Ridge	22	**0.0228****	**-3.5007****	**0.0221****	**-3.7425****	**0.0156****	**-3.3240****	**0.0277****	**-3.1159****	**0.0601****	**-4.0083****
ANN-IV	22	**0.0235****	**-3.5004****	**0.0222****	**-3.7424****	0.0208	-3.3217	0.0304	-3.1146	**0.0620****	**-4.0081****
ANN-CumSum	22	**0.0236****	**-3.5003****	**0.0222****	**-3.7425****	0.0200	-3.3221	0.0317	-3.1141	**0.0629****	**-4.0076****
ANN-HAR	22	**0.0233****	**-3.5005****	**0.0221****	**-3.7425****	**0.0193****	-3.3224*	0.0321	-3.1139	**0.0604****	**-4.0088****
RF	22	**0.0246****	-3.4998*	**0.0253****	**-3.7410****	**0.0177****	-3.3230*	0.0333	-3.1130	0.0676	-4.0054*
SVR	22	0.0263*	-3.4989	0.0262*	**-3.7408****	0.0208	-3.3216	0.0332	-3.1128	**0.0670****	**-4.0048****
XGB	22	0.0267	-3.4985	0.0269*	**-3.7404****	0.0182	-3.3227	0.0335	-3.1129	0.0652*	-4.0062*

Note: The model column contains the name of each model; h indicates the number of steps for out-of-sample forecasting; MSE and QLIKE represent the values of each forecasting evaluation measurements; * means the model pass MCS_{90} test; ** and bold fonts indicate the model pass MCS_{75} test.

Table 3.4A: Out-of-sample forecasting evaluation for ANN Models (Iterative forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
ANN-IV	1	**0.0611****	**-3.4714****	**0.0522****	**-3.7246****	0.0443	-3.3075	0.0256	-3.1187*	**0.0494****	**-4.0131****
Linear-ANN-IV	1	**0.0620****	-3.4687	**0.0526****	-3.7235	0.0449	-3.3058	0.0260	-3.1183	**0.0495****	-4.0124
ANN-CumSum	1	**0.0601****	**-3.4718****	**0.0519****	**-3.7247****	0.0435*	**-3.3080****	0.0255	-3.1188*	**0.0493****	**-4.0132****
Linear-ANN-CumSum	1	**0.0607****	-3.4696	**0.0524****	-3.7235	0.0440	-3.3064	0.0259	-3.1183	**0.0494****	-4.0125
ANN-HAR	1	**0.0601****	**-3.4719****	**0.0520****	**-3.7246****	**0.0429****	**-3.3081****	**0.0252****	**-3.1189****	**0.0492****	**-4.0133****
Linear-ANN-HAR	1	**0.0608****	-3.4697	**0.0525****	-3.7234	0.0446	-3.3060	0.0270	-3.1176	**0.0570****	-4.0096
ANN-Naïve	1	0.0739	-3.4666	0.0609*	-3.7217	0.0594	-3.3009	0.0566	-3.0992	0.0843	-3.9974
ANN-IV	5	**0.0270****	**-3.4956****	**0.0212****	**-3.7439****	**0.0212****	**-3.3215****	**0.0215****	**-3.1205****	**0.0490****	**-4.0137****
Linear-ANN-IV	5	**0.0269****	**-3.4949****	**0.0212****	**-3.7436****	**0.0204****	**-3.3213****	**0.0216****	**-3.1202****	**0.0484****	**-4.0133****
ANN-CumSum	5	**0.0267****	**-3.4956****	**0.0210****	**-3.7440****	**0.0209****	**-3.3217****	**0.0216****	**-3.1205****	**0.0491****	**-4.0135****
Linear-ANN-CumSum	5	**0.0264****	**-3.4951****	**0.0212****	**-3.7436****	**0.0199****	**-3.3216****	**0.0217****	**-3.1202****	**0.0484****	**-4.0132****
ANN-HAR	5	0.0304	-3.4942	0.0251	-3.7422	0.0225	-3.3209*	**0.0222****	**-3.1202****	0.0508	**-4.0131****
Linear-ANN-HAR	5	0.0299	-3.4938	0.0252	-3.7418	0.0229	-3.3201	**0.0236****	**-3.1191****	0.0583	**-4.0096****
ANN-Naïve	5	0.0363	-3.4915	0.0287*	-3.7408*	0.0325	-3.3164	0.0553*	-3.0991*	0.0786	-4.0004
ANN-IV	22	**0.0248****	**-3.4998****	**0.0217****	**-3.7427****	0.0197*	-3.3223*	**0.0316****	**-3.1139****	0.0620	**-4.0084****
Linear-ANN-IV	22	**0.0219****	**-3.5007****	**0.0211****	**-3.7427****	**0.0150****	**-3.3241****	**0.0311****	**-3.1138****	0.0592*	**-4.0090****
ANN-CumSum	22	**0.0251****	**-3.4996****	**0.0215****	**-3.7429****	0.0205*	-3.3219*	**0.0310****	**-3.1143****	0.0610	**-4.0088****
Linear-ANN-CumSum	22	**0.0222****	**-3.5005****	**0.0208****	**-3.7429****	0.0156*	-3.3238*	**0.0304****	**-3.1142****	**0.0577****	**-4.0096****
ANN-HAR	22	0.0321	-3.4967	0.0250	-3.7414	0.0223	-3.3208*	0.0347	-3.1124	0.0688	-4.0056
Linear-ANN-HAR	22	**0.0268****	**-3.4986****	0.0237	-3.7417	0.0194*	-3.3218*	0.0353	-3.1117	0.0750	**-4.0030****
ANN-Naïve	22	0.0356	-3.4948	0.0309	-3.7390	0.0342	-3.3156	0.0903	-3.0756	0.1126	-3.9881

Note: The model column contains the name of each model; Linear- means the linear components of corresponding ANN models; h indicates the number of steps for out-of-sample forecasting; MSE and QLIKE represent the values of each forecasting evaluation measurements; * means the model pass MCS_{90} test; ** and bold fonts indicate the model pass MCS_{75} test.

Table 3.4B: Out-of-sample forecasting evaluation for ANN Models (Direct forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
ANN-IV[1]	1	0.0612*	**-3.4763****	**0.0523****	-3.7251	0.0443	-3.3061	0.0258	-3.1208	**0.0491****	**-4.0115****
Linear-ANN-IV	1	0.0615*	-3.4745	0.0527*	-3.7239	0.0448	-3.3044	0.0260	-3.1204	**0.0491****	-4.0108
ANN-CumSum	1	**0.0600****	**-3.4768****	**0.0517****	**-3.7254****	0.0435*	**-3.3065****	0.0258	-3.1207	**0.0492****	**-4.0115****
Linear-ANN-CumSum	1	**0.0605****	-3.4747	**0.0521****	-3.7242	0.0439*	-3.3049	0.0262	-3.1203	**0.0497****	-4.0105
ANN-HAR	1	**0.0599****	**-3.4768****	**0.0520****	-3.7252	**0.0429****	**-3.3067****	**0.0254****	**-3.1209****	**0.0493****	**-4.0115****
Linear-ANN-HAR	1	**0.0606****	-3.4749	**0.0525****	-3.7241	0.0438*	-3.3048	0.0259	-3.1203	**0.0514****	-4.0100
ANN-Naïve	1	0.0752	-3.4712	0.0647	-3.7198	0.0657	-3.2975	0.0563	-3.1053	0.0772	-3.9976
ANN-IV	5	0.0269	-3.4996	**0.0209****	**-3.7444****	0.0214*	-3.3203*	**0.0212****	**-3.1222****	**0.0490****	**-4.0121****
Linear-ANN-IV	5	0.0274	-3.4984	**0.0221****	-3.7433	0.0220*	-3.3191	**0.0230****	-3.1209	**0.0488****	-4.0108*
ANN-CumSum	5	0.0266	-3.4997	**0.0212****	-3.7442*	0.0217*	-3.3201	**0.0211****	**-3.1223****	**0.0489****	**-4.0123****
Linear-ANN-CumSum	5	0.0275	-3.4983	0.0229	-3.7428	0.0226*	-3.3187	**0.0223****	-3.1212*	**0.0483****	-4.0112*
ANN-HAR	5	**0.0261****	**-3.5000****	**0.0211****	**-3.7443****	**0.0207****	**-3.3206****	**0.0208****	**-3.1224****	**0.0492****	**-4.0119****
Linear-ANN-HAR	5	0.0269	-3.4987	**0.0224****	-3.7431	0.0224*	-3.3189	**0.0299****	-3.1183	0.0500*	-4.0102
ANN-Naïve	5	0.0419	-3.4930	0.0324	-3.7389	0.0413	-3.3113	0.0409	-3.1124	0.0941	-3.9936
ANN-IV	22	**0.0235****	**-3.5004****	**0.0222****	**-3.7424****	**0.0208****	**-3.3217****	**0.0304****	**-3.1146****	**0.0620****	**-4.0081****
Linear-ANN-IV	22	**0.0276****	**-3.4968****	**0.0258****	-3.7396*	**0.0209****	**-3.3204****	**0.0375****	**-3.1093****	**0.0590****	**-4.0055****
ANN-CumSum	22	**0.0236****	**-3.5003****	**0.0222****	**-3.7425****	**0.0200****	**-3.3221****	**0.0317****	**-3.1141****	**0.0629****	**-4.0076****
Linear-ANN-CumSum	22	**0.0271****	**-3.4970****	**0.0272****	-3.7388*	**0.0236****	-3.3189*	**0.0377****	**-3.1092****	**0.0595****	**-4.0053****
ANN-HAR	22	**0.0233****	**-3.5005****	**0.0221****	**-3.7425****	**0.0193****	**-3.3224****	**0.0321****	**-3.1139****	**0.0604****	**-4.0088****
Linear-ANN-HAR	22	**0.0281****	**-3.4971****	**0.0277****	-3.7386*	0.0282	-3.3167	**0.0392****	**-3.1084****	**0.0580****	**-4.0062****
ANN-Naïve	22	0.0362	-3.4947	0.0309	-3.7382	0.0325	-3.3165	0.0517	-3.1047	0.1011	-3.9920

Note: The model column contains the name of each model; Linear- means the linear components of corresponding ANN models; h indicates the number of steps for out-of-sample forecasting; MSE and QLIKE represent the values of each forecasting evaluation measurements; * means the model pass MCS_{90} test; ** and bold fonts indicate the model pass MCS_{75} test.

Table 3.5: Summary Statistics for Ratios Between Nonlinear Components and Total Prediction With the ANN-IV Model

h	Statistic	Corn	Soybean	Wheat	Crude Oil	S&P 500 E-mini
1	Min	0.53%	0.10%	0.59%	0.24%	0.27%
	Median	1.00%	0.62%	0.97%	0.52%	0.50%
	Mean	1.01%	0.63%	1.24%	0.55%	0.53%
	Max	1.83%	1.06%	2.92%	1.28%	1.47%
5	Min	0.98%	0.41%	0.63%	0.60%	0.66%
	Median	1.51%	0.95%	1.48%	1.12%	1.02%
	Mean	1.54%	0.97%	1.91%	1.17%	1.08%
	Max	2.62%	1.44%	4.48%	3.67%	2.03%
22	Min	1.53%	0.00%	0.64%	0.18%	0.26%
	Median	2.62%	1.80%	2.47%	2.42%	2.47%
	Mean	2.76%	1.87%	3.27%	2.51%	2.70%
	Max	8.09%	3.53%	7.07%	5.71%	8.55%

Note: h represents the number of steps for out-of-sample forecasting; Statistic contains the names of statistics provided; the percentages are calculated as $\left|\widehat{IV}_t^{d,NL} / \widehat{IV}_t^{d}\right|$, where $\widehat{IV}_t^{d,NL}$ and $\widehat{IV}_t^{d}$ are out-of-sample predictions for nonlinear components and total volatility values, respectively (equation 19).

Table 3.6: Average Kernel Weights of ANN-IV Model (Linear vs. Non-Linear)

h	Levels	Type	Corn			Soybean			Wheat			Crude Oil			S&P 500 E-mini		
			L1	L2-5	L6-22	L1	L2-5	L6-22	L1	L2-5	L6-22	L1	L2-5	L6-22	L1	L2-5	L6-22
1	Min	Non-Linear	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Linear	28.66	7.33	2.38	26.62	8.63	2.55	23.46	9.18	2.24	56.36	5.25	1.89	56.81	6.00	2.38
	Median	Non-Linear	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Linear	30.14	8.59	3.46	29.27	9.65	3.22	25.26	9.79	2.74	59.91	5.92	2.65	60.17	7.36	2.99
	Mean	Non-Linear	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01
		Linear	31.02	8.48	3.52	30.95	9.77	3.23	25.29	9.96	2.80	59.79	6.06	2.74	60.49	7.55	3.05
	Max	Non-Linear	0.06	0.05	0.04	0.10	0.06	0.06	0.07	0.07	0.16	0.10	0.13	0.12	0.11	0.20	0.34
		Linear	36.19	10.04	4.74	37.54	11.02	3.86	27.12	11.80	3.45	62.59	8.18	4.05	64.64	11.25	6.36

Note: h represents the steps of out-of-sample forecasts; Statistic contains the names of statistics; Type indicates linear and nonlinear kernel separately. L1 column represents the kernel weights of IV_{t-1}^{d}; L2-5 column represents the average kernel weights of IV_{t-i}^{d}, where $i = 2, \dots, 5$; L6-22 column represents the average kernel weights of IV_{t-i}^{d}, where $i = 6, \dots, 22$. We omit h = 5 because it is similar to h = 1 and 22; All results are multiplied by 100 for illustration purposes.

Table 3.6 (Cont.): Average Kernel Weights of ANN-IV Model (Linear vs. Non-Linear)

h	Levels	Type	Corn			Soybean			Wheat			Crude Oil			S&P 500 E-mini		
			L1	L2-5	L6-22	L1	L2-5	L6-22	L1	L2-5	L6-22	L1	L2-5	L6-22	L1	L2-5	L6-22
22	Min	Non-Linear	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		Linear	12.75	4.58	1.52	15.45	6.38	1.26	13.35	5.83	0.84	23.21	3.80	1.54	29.83	4.77	1.04
	Median	Non-Linear	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
		Linear	13.88	6.46	1.94	16.50	7.53	1.73	14.24	7.31	2.17	29.64	5.67	1.93	32.88	6.10	1.59
	Mean	Non-Linear	0.01	0.01	0.01	0.03	0.02	0.02	0.01	0.01	0.01	0.03	0.03	0.04	0.08	0.06	0.09
		Linear	14.07	6.35	1.93	16.73	7.48	1.76	14.33	7.21	1.82	29.94	5.51	2.04	33.08	6.05	1.61
	Max	Non-Linear	0.09	0.10	0.30	0.31	0.11	0.10	0.12	0.24	0.33	0.56	0.29	0.50	1.01	0.63	1.28
		Linear	16.25	7.10	2.53	18.45	8.40	2.22	15.37	8.22	2.33	34.22	7.29	3.81	36.33	7.72	2.38

Note: h represents the steps of out-of-sample forecasts; Statistic contains the names of statistics; Type indicates linear and nonlinear kernel separately. L1 column represents the kernel weights of IV_{t-1}^{d}; L2-5 column represents the average kernel weights of IV_{t-i}^{d}, where $i = 2, \dots, 5$; L6-22 column represents the average kernel weights of IV_{t-i}^{d}, where $i = 6, \dots, 22$. We omit h = 5 because it is similar to h = 1 and 22; All results are multiplied by 100 for illustration purposes.

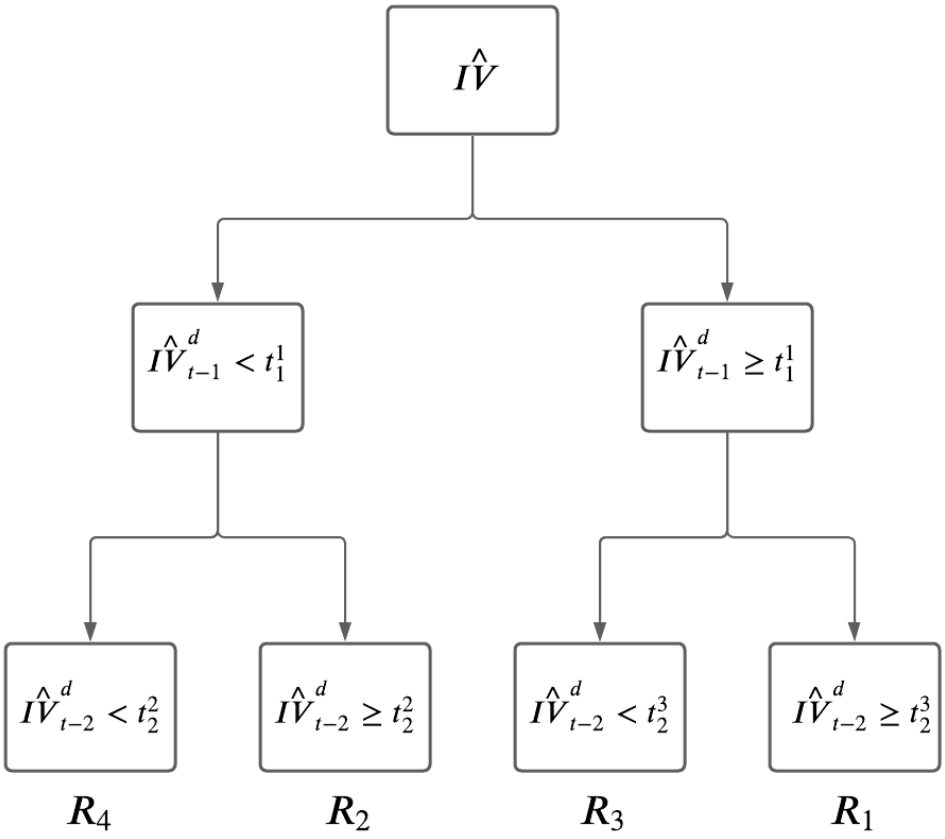

Figure 3.1: The graphic illustration of the tree model

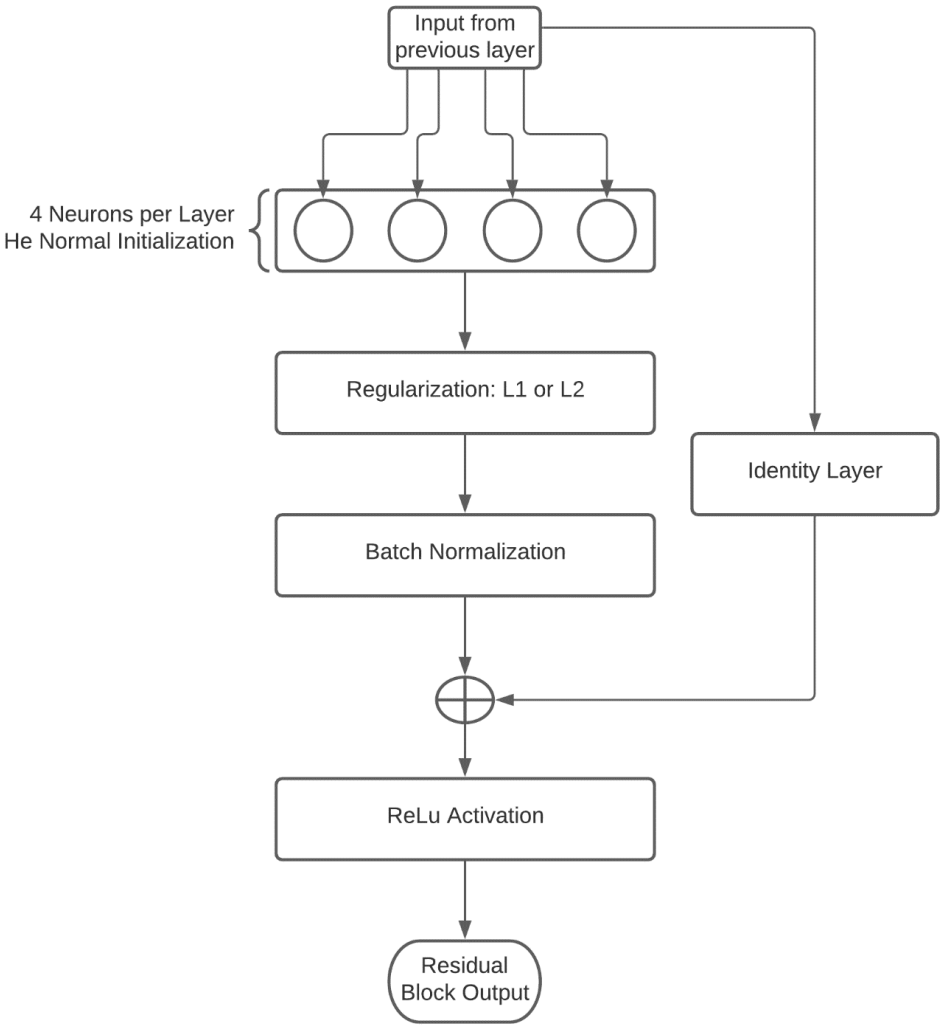

Figure 3.2A: The design of a residual block

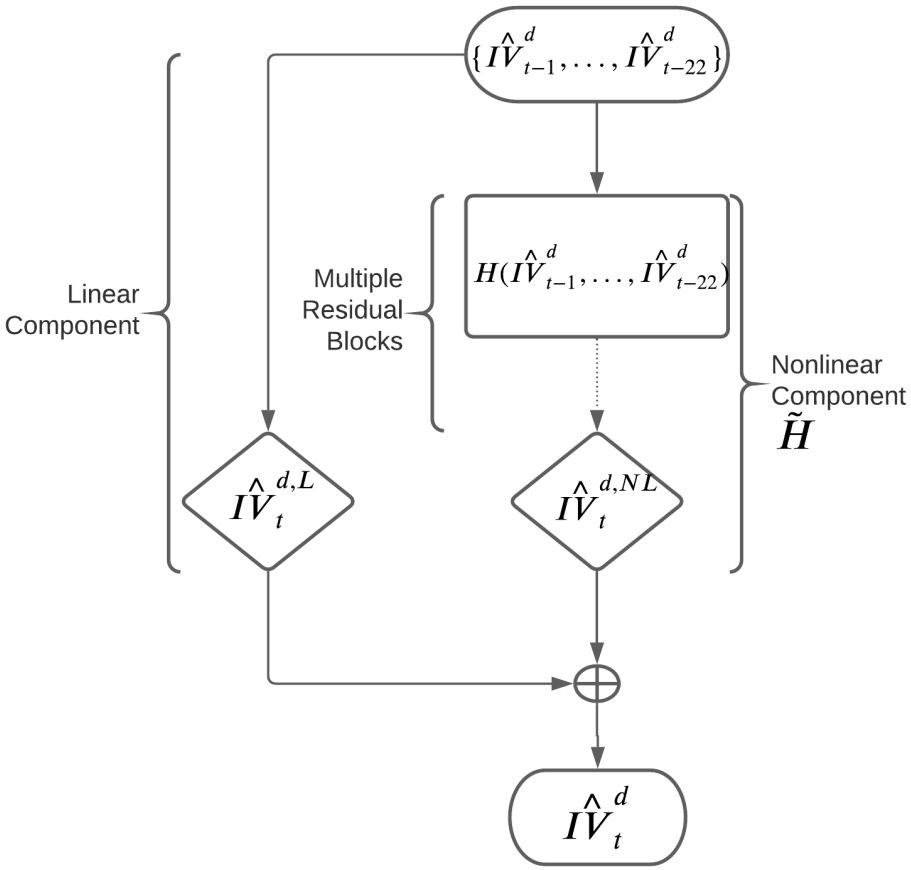

Figure 3.2B: The design of the ANN model with the residual block and linear/nonlinear components

3.8 References

Andersen, T.G., T. Bollerslev, and F.X. Diebold. 2007. "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility." *Review of Economics and Statistics* 89(4):701–720. Available at: http://www.mitpressjournals.org/doi/10.1162/rest.89.4.701.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2003. "Modeling and Forecasting Realized Volatility." *Econometrica* 71(2):579–625. Available at: http://doi.wiley.com/10.1111/1468-0262.00418.

Anderson, R.W., and J.-P. Danthine. 1983. "The Time Pattern of Hedging and the Volatility of Futures Prices." *The Review of Economic Studies* 50(2):249. Available at: https://academic.oup.com/restud/article-lookup/doi/10.2307/2297415.

Arismendi, J.C., J. Back, M. Prokopczuk, R. Paschke, and M. Rudolf. 2016. "Seasonal Stochastic Volatility: Implications for the pricing of commodity options." *Journal of Banking & Finance* 66:53–65. Available at: https://linkinghub.elsevier.com/retrieve/pii/S037842661600039X.

Audrino, F., C. Huang, and O. Okhrin. 2019. "Flexible HAR model for realized volatility." *Studies in Nonlinear Dynamics & Econometrics* 23(3). Available at: https://www.degruyter.com/document/doi/10.1515/snde-2017-0080/html.

Audrino, F., and S.D. Knaus. 2016. "Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics." *Econometric Reviews* 35(8–10):1485–1521. Available at: http://dx.doi.org/10.1080/07474938.2015.1092801.

Bandi, F.M., and J.R. Russell. 2006. "Separating microstructure noise from volatility." *Journal of Financial Economics* 79(3):655–692. Available at:

https://www.sciencedirect.com/science/article/pii/S0304405X05001534.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2009. "Realized Kernels In Practice: Trades and Quotes." *The Econometrics Journal* 12(3).

Baruník, J., and T. Křehlík. 2016. "Combining high frequency data with non-linear models for forecasting energy market volatility." *Expert Systems with Applications* 55:222–242.

Beltratti, A., and C. Morana. 2006. "Breaks and persistency: macroeconomic causes of stock market volatility." *Journal of Econometrics* 131(1–2):151–177. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304407605000096.

Bian, S., T. Serra, P. Garcia, and S. Irwin. 2021. "New evidence on market response to public announcements in the presence of microstructure noise." *European Journal of Operational Research*. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0377221721006251.

Blanc, S.M., and T. Setzer. 2020. "Bias–Variance Trade-Off and Shrinkage of Weights in Forecast Combination." *Management Science* 66(12):5720–5737. Available at: http://pubsonline.informs.org/doi/10.1287/mnsc.2019.3476.

Breiman, L. 2001. "Random forests." *Machine Learning* 45:5–32.

Christensen, K., R.C.A. Oomen, and M. Podolskij. 2014. "Fact or friction: Jumps at ultra high frequency." *Journal of Financial Economics* 114(3):576–599. Available at: http://dx.doi.org/10.1016/j.jfineco.2014.07.007.

Clark, T.E., and M.W. McCracken. 2001. "Tests of equal forecast accuracy and encompassing for nested models." *Journal of Econometrics* 105(1):85–110. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304407601000719.

Corsi, F. 2008. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7(2):174–196. Available at: https://academic.oup.com/jfec/article-

lookup/doi/10.1093/jjfinec/nbp001.

Cybenko, G. 1989. “Approximation by superpositions of a sigmoidal function.” *Mathematics of Control, Signals, and Systems* 2(4):303–314. Available at: http://link.springer.com/10.1007/BF02551274.

D’Ecclesia, R.L., and D. Clementi. 2019. “Volatility in the stock market: ANN versus parametric models.” *Annals of Operations Research*. Available at: http://link.springer.com/10.1007/s10479-019-03374-0.

Degiannakis, S., G. Filis, T. Klein, and T. Walther. 2020. “Forecasting realized volatility of agricultural commodities.” *International Journal of Forecasting*. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0169207019302353.

Ding, Y., D. Kambouroudis, and D.G. McMillan. 2021. “Forecasting realised volatility: Does the LASSO approach outperform HAR?” *Journal of International Financial Markets, Institutions and Money* 74:101386. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1042443121001050.

Donaldson, R.G., and M. Kamstra. 1997. “An artificial neural network-GARCH model for international stock return volatility.” *Journal of Empirical Finance* 4(1):17–46. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0927539896000114.

Du, B., S. Fung, and R. Loveland. 2018. “The informational role of options markets: Evidence from FOMC announcements.” *Journal of Banking & Finance* 92:237–256. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0378426618301134.

Egelkraut, T.M., P. Garcia, and B.J. Sherrick. 2007. “The Term Structure of Implied Forward Volatility: Recovery and Informational Content in the Corn Options Market.” *American Journal of Agricultural Economics* 89(1):1–11.

Fernandes, M., M.C. Medeiros, and M. Scharth. 2014. “Modeling and predicting the CBOE market volatility index.” *Journal of Banking & Finance* 40:1–10. Available at: http://dx.doi.org/10.1016/j.jbankfin.2013.11.004.

Franses, P.H., and D. van Dijk. 2000. *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press. Available at: http://ebooks.cambridge.org/ref/id/CBO9780511754067.

Friedman, J.H. 2001. “Greedy function approximation: A gradient boosting machine.” *The Annals of Statistics* 29(5):1189–1232. Available at: http://projecteuclid.org/euclid.aos/1013203451.

Gonzalez Miranda, F., and N. Burgess. 1997. “Modelling market volatilities: the neural network perspective.” *The European Journal of Finance* 3(2):137–157. Available at: https://www.tandfonline.com/doi/full/10.1080/135184797337499.

Gu, S., B. Kelly, and D. Xiu. 2020. “Empirical Asset Pricing via Machine Learning” A. Karolyi, ed. *The Review of Financial Studies* 33(5):2223–2273. Available at: https://academic.oup.com/rfs/article/33/5/2223/5758276.

Hansen, P.R., and A. Lunde. 2006. “Realized Variance and Market Microstructure Noise.” *Journal of Business and Economic Statistics* 24(2):127–161.

Hansen, P.R., A. Lunde, and J.M. Nason. 2011. “The Model Confidence Set.” *Econometrica* 79(2):453–497.

He, K., X. Zhang, S. Ren, and J. Sun. 2015a. “Deep Residual Learning for Image Recognition.” Available at: http://arxiv.org/abs/1512.03385.

He, K., X. Zhang, S. Ren, and J. Sun. 2015b. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification.” Available at:

http://arxiv.org/abs/1502.01852.

Heston, S.L. 1993. “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options.” *Review of Financial Studies* 6(2):327–343. Available at: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/6.2.327.

Ing, C.-K. 2003. “MULTISTEP PREDICTION IN AUTOREGRESSIVE PROCESSES.” *Econometric Theory* 19(02):254–279. Available at: http://www.journals.cambridge.org/abstract_S0266466603192031.

Inoue, A., and L. Kilian. 2008. “How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation.” *Journal of the American Statistical Association* 103(482):511–522. Available at: https://www.tandfonline.com/doi/full/10.1198/016214507000000473.

Ioffe, S., and C. Szegedy. 2015. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.” Available at: http://arxiv.org/abs/1502.03167.

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter. 2009. “Microstructure noise in the continuous case: The pre-averaging approach.” *Stochastic Processes and their Applications* 119(7):2249–2276. Available at: http://dx.doi.org/10.1016/j.spa.2008.11.004.

Karali, B., J.H. Dorfman, and W.N. Thurman. 2010. “Delivery horizon and grain market volatility.” *Journal of Futures Markets*:n/a-n/a. Available at: https://onlinelibrary.wiley.com/doi/10.1002/fut.20449.

Karali, B., and G.J. Power. 2013. “Short- and Long-Run Determinants of Commodity Price Volatility.” *American Journal of Agricultural Economics* 95(3):724–738.

Luong, C., and N. Dokuchaev. 2018. “Forecasting of Realised Volatility with the Random Forests Algorithm.” *Journal of Risk and Financial Management* 11(4):61. Available at:

http://www.mdpi.com/1911-8074/11/4/61.

Marcellino, M., J.H. Stock, and M.W. Watson. 2006. “A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series.” *Journal of Econometrics* 135(1–2):499–526. Available at: https://linkinghub.elsevier.com/retrieve/pii/S030440760500165X.

McAleer, M., and M.C. Medeiros. 2011. “Forecasting realized volatility with linear and nonlinear univariate models.” *Journal of Economic Surveys* 25(1):6–18.

Miao, H., S. Ramchander, T. Wang, and J. Yang. 2018. “The impact of crude oil inventory announcements on prices: Evidence from derivatives markets.” *Journal of Futures Markets* 38(1):38–65. Available at: https://onlinelibrary.wiley.com/doi/10.1002/fut.21850.

Patton, A.J. 2011. “Volatility forecast comparison using imperfect volatility proxies.” *Journal of Econometrics* 160(1):246–256. Available at: https://linkinghub.elsevier.com/retrieve/pii/S030440761000076X.

Patton, A.J., and K. Sheppard. 2015. “Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility.” *Review of Economics and Statistics* 97(3):683–697. Available at: https://direct.mit.edu/rest/article/97/3/683-697/58249.

Peng, Y., P.H.M. Albuquerque, J.M. Camboim de Sá, A.J.A. Padula, and M.R. Montenegro. 2018. “The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression.” *Expert Systems with Applications* 97:177–192. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0957417417308163.

Qiu, Y., X. Zhang, T. Xie, and S. Zhao. 2019. “Versatile HAR model for realized volatility: A least square model averaging perspective.” *Journal of Management Science and*

Engineering 4(1):55–73. Available at: https://doi.org/10.1016/j.jmse.2019.03.003.

Santamaría-Bonfil, G., J. Frausto-Solís, and I. Vázquez-Rodarte. 2015. "Volatility Forecasting Using Support Vector Regression and a Hybrid Genetic Algorithm." *Computational Economics* 45(1):111–133. Available at: http://link.springer.com/10.1007/s10614-013-9411-x.

Sévi, B. 2014. "Forecasting the volatility of crude oil futures using intraday data." *European Journal of Operational Research* 235(3):643–659. Available at: https://linkinghub.elsevier.com/retrieve/pii/S037722171400040X.

Stock, J.H., and M.W. Watson. 2004. "Combination forecasts of output growth in a seven-country data set." *Journal of Forecasting* 23(6):405–430. Available at: https://onlinelibrary.wiley.com/doi/10.1002/for.928.

Vapnik, V., S.E. Golowich, and A. Smola. 1997. "Support vector method for function approximation, regression estimation, and signal processing." In *Advances in Neural Information Processing Systems*. pp. 281–287.

Wolpert, D.H., and W.G. Macready. 1997. "No free lunch theorems for optimization." *IEEE Transactions on Evolutionary Computation* 1(1):67–82. Available at: http://ieeexplore.ieee.org/document/585893/.

Zhang, Yaojie, Y. Wei, Yi Zhang, and D. Jin. 2019. "Forecasting oil price volatility: Forecast combination versus shrinkage method." *Energy Economics* 80:423–433. Available at: https://doi.org/10.1016/j.eneco.2019.01.010.

CHAPTER 4

JUMPS, SEASONALITY, AND FUNDING RISK

4.1. Introduction

Futures contracts are mark-to-market derivatives that require daily gains and losses to be realized through the daily settlement procedure. At the end of the day, the exchange determines the settlement price of the assets covered in the futures contract and based on this price, the daily settlement occurs where traders' losses (profits) experienced during the day are deducted from (deposited into) their margin accounts. In the event of relevant losses, traders may face margin calls if their accounts lack sufficient funds to cover the required margins to hold their positions. The requirement to realize intra-horizon losses poses funding risks to market participants, since inability to comply with margin calls due to insufficient funds may lead to pre-mature liquidation of their positions. Moreover, such liquidations in certain circumstances may push the price even further in the tail of the distribution, generate additional margin calls and result in a liquidity spiral (Brunnermeier and Pedersen 2009). As a result, it is crucial for market participants and exchanges to develop tools that can help anticipate the future level of funding risks in the futures market.

Over the past two decades agricultural commodity markets have witnessed extreme price volatility that poses funding risks to both traders and the futures clearinghouses. Given their speculative nature, non-commercial traders are inherently vulnerable to extreme volatility and unfavorable price movements.[22] However, commercial traders who hedge their cash positions in

[22] Westgaard, Frydenberg and Mohanty (2021) study 14 historical commodity trading disasters and find that at least 4 cases are directly induced by failing to meet margin requirements. Speculators, such as

the futures market are not immune to funding risks, especially those who have limited access to cash or credit lines (Westgaard, Frydenberg and Mohanty 2021; Charupat and Deaves 2003). For example, during the Agricultural Markets Roundtable meeting organized by the Commodity Futures Trading Commission (CFTC) on April, 2008, representatives of multiple national associations of grain farmers and co-operatives complained about excessive margin calls induced by continuously rising grain prices since 2006 (CFTC 2008). Extreme volatility can challenge futures traders' ability to comply with margin requirements, which in turn may reduce the clearinghouses' financial resources and lower their resiliency in times of market turbulence. The ability to produce accurate predictions of traders' maximum losses, especially during highly volatile periods, is thus important for both clearinghouses and traders. Clearinghouses can use these predictions to guide the setting of margin requirements and avoid liquidity stress that may compromise the futures market integrity (Fishe and Goldberg 1986; Day and Lewis 2004). This information can also help traders to manage their risk exposures, adjust trading and hedging strategies, and prepare additional funds to prevent forced liquidation of their positions.

Nevertheless, there are few existing studies that help provide insights into anticipating funding risks in agricultural futures markets. Previous studies typically use hypothetical costs to represent funding constraints when evaluating the role of funding risks on hedging performance (Lien 2003; Dahlgran 2005). Shi and Isengildina-Massa (2021) measure the cost of hedging using the maximum cumulative gains (losses) plus the initial margin over pre-specified holding periods. While useful to capture realized funding risks, this approach cannot be used for forecasting, as measures based on realized returns fail to incorporate the likelihood of extreme events not observed

Amaranth Advisors and Optionssellers.com, can fail to pay their margin requirement due to accumulated speculative futures positions and unfavorable price movements.

in a particular historical period. This issue is commonly referred to as the Peso problem and may lead to significant underestimation of the impact of infrequent large price movements on funding risks faced by market participants.

In this paper, we build on the intra-horizon Conditional Value-at-Risk (CVaR) model originally developed by Farkas, Mathys and Vasiljević (2021) to measure funding risks in corn, soybeans, and wheat futures markets. The CVaR measure estimates the conditional expectation of losses in the tail of the return distribution beyond a cut-off level. The CVaR model, however, focuses on the losses at the end of a predetermined period (terminal risk), which renders it inappropriate to measure funding risks. The latter depend on the daily returns over the entire holding period and in addition to the terminal risk are also concerned with large losses at an intermediate date (intra-horizon risk). Losses in the middle of the holding period do not represent the terminal risk as they may be compensated by traders' later gains. Unlike traditional CVaR measures, the intra-horizon CVaR, denoted as MaxCVaR, measures the conditional expected losses that occur within the entire holding period. Based on the MaxCVaR, we avoid the Peso problem presented in previous studies via Monte Carlo simulations that allow incorporating extreme events beyond pre-determined thresholds, even if they are unobservable in the historical period.

When calculating MaxCVaR, we consider two salient features of the futures price dynamics that may be particularly relevant to funding risks in grains futures markets, namely asymmetric price jumps and seasonal volatility. Jumps are large discontinuous price movements that are primarily induced by the arrival of significant information such as public announcements, with positive and negative jumps exhibiting different behaviors due to the asymmetric role of inventories (Christensen et al. 2014; Couleau et al. 2020; Aravindhakshan and Brorsen 2011;

Degiannakis et al. 2020). Besides price jumps, grains futures prices usually display pronounced seasonal volatility patterns that are closely associated with the different growing stages of the underlying crops (Sørensen 2002; Karali and Power 2013). To properly measure these two salient features of agricultural futures prices, we extend Farkas, Mathys and Vasiljević (2021)'s asymmetric jump model to consider seasonal volatility. Accounting for price jumps and seasonal volatility in the calculation of MaxCVaR not only helps produce realistic forecasts, but also allows us to gauge their relative importance in determining the expected level of funding risk in agricultural futures markets.

We use our MaxCVaR model to estimate funding risks under extreme conditions, and to examine whether margin payments are sufficient to cover funding risk faced by market participants in agricultural futures markets. Currently, the Chicago Mercantile Exchange (CME) margin model uses a combination of quantitative and qualitative metrics to determine margin requirements. Numerous quantitative factors, such as volatility and seasonality, and qualitative components, like relevant political elections or referendums, are considered when setting margin requirements for each product everyday (CME 2022). The margin model creates margin requirements to cover maximum potential losses that traders can incur from holding a position over one-day period, if the price returns move beyond 99% and -99% of the extremes of the distribution. The specifications of the quantitative elements of CME margin model are not published by the exchange. Also, the qualitative components of margin requirements are modified by experts based on their interpretation of related events. Therefore, the margin model cannot be directly used by traders to estimate their future funding risks. Instead, CME, as many other exchanges, implements a system called the Standard Portfolio Analysis of Risk (SPAN) to inform market participants of expected margin payments in the future.

SPAN is not part of the CME margin model, but relies on margin requirements produced by the margin model to help traders anticipate expected margin payments in the future. SPAN considers a few fixed price movement scenarios represented in terms of percentages of current margin requirements and generates an expected margin payment being equal to the maximum loss among all scenarios. While different in the case of portfolios integrated by different financial products where portfolio margin risks may differ from individual assets' financial risk, in the case of futures outright positions SPAN sets the expected margin payment exactly equal to the futures' contract margin requirement (CME 2019). As a result, if the futures' contract margin requirement is insufficient to cover funding risks, the expected margin payment calculated by SPAN will also underestimate funding risks faced by relevant parties.

We find that funding risks are significant in corn, soybean, and wheat futures markets between 2000 and 2021. Traders holding a long (short) position for a month, are required to prepare on average between 14.9% and 22.8% (18.1% and 29.1%) of the contract value to cover the 99-percentile funding risk. Price jumps and seasonal volatility have a significant impact on the level and temporal variation of funding risk, respectively. Compared to a no-jump scenario, jumps in the futures price increase the average level of funding risk by up to 46.8% across different markets. Further, allowing for seasonal volatility patterns leads to funding risk being around 9.17% higher during the growing season than during the pre-planting season. Using the backtesting procedure proposed in Righi and Ceretta (2015), we find supporting evidence that the model with jumps and seasonal volatility offers better protection against realized extreme losses than models without jumps, seasonal volatility, or both. Finally, we show that expected margin payments estimated by CME fall short of expected coverages of the 99-percentile funding risks after accounting for the

possibility of price jumps in the market. We find that funding risks are consistently greater than the actual margin requirements by as much as 15.9%.

This paper makes several contributions. First, our method measures intra-period losses and avoids the Peso problem affecting previous research via Monte Carlo simulations. As a result, funding risks estimated by our MaxCVaR method are significantly higher than those in Shi and Isengildina-Massa (2021). Second, we provide novel evidence on the impact of different features of the price dynamics on funding risks in agricultural futures markets. Our empirical findings reveal the importance of jumps in determining the overall level of funding risks faced by market participants, consistent with the notion that jumps contribute to excess kurtosis and fat-tails of the price distribution (Kou 2002; Kou and Wang 2004). Our findings on seasonal volatility also point to the need to allow for seasonal variations in funding risks. The backtesting results suggest that these two features significantly affect the ability of the model to accurately anticipate funding risks, thus providing valuable information to traders and clearinghouses who need to evaluate and monitor funding risks in the market. Finally, our results have implications for the maintenance margin requirements established by CME in grains futures markets. They indicate that current margin requirements may lead to insufficient protection from market losses in the event of significant price jumps. Overall, our findings have the potential to help provide informed risk management, facilitate efficient allocation of financial resources, and assist the evaluation of exchange policies that aim at maintaining futures market integrity.

4.2. Methodology

Traditional CVaR and VaR measurements are of limited value to evaluate margin funding risks in positions held longer than one day, because they only depict expected losses at the end of the holding period (terminal risks). Intra-horizon losses in mark-to-market financial instruments will affect investors through the margin requirement and daily settlement procedure and will be relevant in the presence of large price movements in the middle of the holding period. To overcome the shortcoming of traditional VaR methods, Boudoukh et al. (2004) define the intra-horizon VaR as the loss in investment that will be exceeded with a given probability on or before the pre-determined holding period. Furthermore, Rossello (2008) replaces the assumption of the geometric Brownian motion in Boudoukh et al. (2004) with the double-exponential jump model proposed by Kou (2002). Later, Bakshi and Panayotov (2010) estimate intra-horizon VaR with other popular jump models across different financial markets. Leippold and Vasiljević (2020) derive the intra-horizon VaR from options prices. However, intra-horizon VaRs do not provide a meaningful estimate of funding risks because they do not assess the magnitude of the loss when a VaR breach occurs. Rather, intra-horizon CVaR should be used because it depicts the conditional expectation of intra-horizon losses given the occurrences of VaR breaches. Farkas, Mathys and Vasiljević (2021) develop the intra-horizon CVaR, the MaxCVaR measurement for general Levy processes. However, they do not consider seasonality in their study. Both asymmetric jumps and seasonal volatility patterns are essential factors for modeling grain futures price dynamics. Motivated by Farkas, Mathys and Vasiljević (2021), we propose the MaxCVaR method with a seasonally-adjusted double-exponential jump model. We use crude Monte Carlo simulation to assess the relevance of the different components of volatility (jumps and seasonality) on funding risk.

Assume X_t is the daily settlement price with $t \in [0,T]$ and $X_{min}^{0,T}$ ($X_{max}^{0,T}$) is the minimum (maximum) value of X_t over the holding period $[0,T]$, where t is measured in years. Following Rossello (2008), the MaxVaR of the cumulative return $R_t = \frac{X_t}{X_0} - 1$ over a pre-determined holding period $[0,T]$ with confidence level $1-\alpha$ is the quantile x, which satisfies

$$F_\alpha^T(x) = \begin{cases} \boldsymbol{P}(R_{min}^T \leq x) = \alpha \\ \boldsymbol{P}(R_{max}^T \geq x) = \alpha \end{cases} \tag{4.1}$$

where $R_{min}^T = \frac{X_{min}^{0,T}}{X_0} - 1$ and $R_{max}^T = \frac{X_{max}^{0,T}}{X_0} - 1$; $F_\alpha^T(x)$ is the cumulative density function (cdf) of R_{min}^T and R_{max}^T, with x being the cumulative return level corresponding to α. Hedgers who typically hold short positions, may face margin calls during sudden price increases. Therefore, hedgers will focus on the right tail events where $R_{max}^T \geq x$. In contrast, traders who act as the counterparty for hedgers and hold long positions are more sensitive to the situations where $R_{min}^T \leq x$. Although the MaxVaR can assign confidence levels (α) to each cumulative return level (x), it fails to inform investors of the expected losses beyond x. Therefore, we should use a different risk measure to estimate funding risk.

4.2.1. The Definition of MaxCVaR

Motivated by the traditional CVaR from Rockafellar and Uryasev (2000) and the intra-horizon risk measures recently proposed by Boudoukh et al. (2004), Rossello (2008), and Farkas, Mathys and Vasiljević (2021), we use the MaxCVaR to measure funding risk of holding a financial instrument over a pre-determined period. We define the MaxCVaR as follows:

$$MaxCVaR^T(\alpha) = \begin{cases} MaxCVaR^T(\alpha_D) = \dfrac{1}{1+\alpha}\displaystyle\int_0^{1+\alpha} x dF_{1+\alpha}^T(x) & \alpha = -95\%, -97.5\%, -99\% \\ MaxCVaR^T(\alpha_U) = \dfrac{1}{1-\alpha}\displaystyle\int_\alpha^1 x dF_\alpha^T(x) & \alpha = 95\%, 97.5\%, 99\% \end{cases} \tag{4.2}$$

where α is the pre-specified threshold of tail risks with α_D and α_U denoting lower and upper tail risks, respectively and $dF_{1+\alpha}^T(x)$ ($dF_{\alpha}^T(x)$), is the probability density function (pdf) of R_{min}^T (R_{max}^T) derived from equation (4.1) over the period $[0, T]$. As shown in equation (4.2), the $MaxCVaR^T(\alpha)$ is the expected loss of tail events at the pre-specified confidence level α over the period $[0, T]$. We tailor the functional form of $dF_{\alpha}^T(x)$ ($dF_{1+\alpha}^T(x)$) to the characteristics of agricultural commodity returns.

To estimate (2) we follow Rossello (2008)'s procedure and generate the empirical distribution $dF_{1+\alpha}^T(x)$ ($dF_{\alpha}^T(x)$) for R_{min}^T (R_{max}^T) with a Monte Carlo simulation method in three steps:

1. First, we estimate a parametric model of returns by maximum likelihood. The model allows for both jumps and seasonality. The specification of this parametric model is presented in subsection 4.2.2.
2. Second, we use a Monte Carlo simulation method to generate 10,000 return sequences[23] for every *t*; each sequence is denoted $R_t \ \forall \ t \in [0, T]$.
3. Third, we choose the maximum and minimum value from each sequence $R_t \ \forall \ t \in [0, T]$, denote them as R_{max}^T and R_{min}^T and form the empirical distributions of $dF_{\alpha}^T(x)$ and $dF_{1+\alpha}^T(x)$. This allows deriving a MaxCVaR for each *t*. Steps 2 and 3 are described in subsection 4.2.3.
4. To identify whether MaxCVaR values are effective in predicting extreme price movements, we use the backtesting method described in subsection 4.2.4. Intuitively, the backtesting

[23] We also try simulation results with 100,000 and 1,000,000 return sequences, but MaxCVaR estimation results remain virtually the same. Hence, to avoid excessive computational burdens, we set the number to be 10,000.

compares the MaxCVaR for each *t* against the extreme returns that occur over the following days to assess the ability of the MaxCVaR to predict funding risk.

4.2.2. Theoretical Models for Daily Returns

Denote futures returns as dX_t at time $t \in [0, T]$. If X_t follows a Geometric Brownian Motion (GBM), then its dynamic path can be expressed as:

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t \tag{4.3}$$

This is a linear stochastic differential equation with a constant drift term μ and volatility $\sigma > 0$; dW_t denotes the increments of a Brownian motion. Under specification (3), X_t follows a lognormal distribution that has been questioned by previous studies as a useful tool to model agricultural commodity returns. Empirical findings document both skewed and fat-tailed distributions of futures returns (Hilliard and Reis 1999; Koekebakker and Lien 2004). These in turn lead to volatility clustering in agricultural commodity prices. In addition, under the influence of seasonal supply and demand, agricultural futures' volatility also exhibits strong and diverse seasonal patterns across different markets (Karali and Power 2013). To address both the asymmetric leptokurtic feature and the seasonality in agricultural futures volatility, we propose the Double-Exponential Jump-diffusion Model with Seasonal volatility (DEJ-Seasonal). This is a novel hybrid model that combines the Double-Exponential Jump-diffusion (DEJ) model by Kou (2002) and the seasonal volatility process based on a truncated Fourier series by Koekebakker and Lien (2004).

The DEJ model proposed in Kou (2002) captures the asymmetric leptokurtic feature by adding to the original GBM model in equation (4.3) two exponential processes for positive and negative jumps separately. The DEJ model can be expressed as follows:

$$\frac{dX_t}{X_t} = \mu dt + \sigma dW_t + d\left(\sum_{i=1}^{N_t}(J_i - 1)\right) \tag{4.4}$$

where N_t is a homogenous Poisson process generating the jumps with average jump intensity equal to λ which depicts the number of jumps per year. In equation (4.4), J_i denotes the jump process and represents the magnitude of price deviations from the constant drift term μ. Here, we assume J_i is an i.i.d nonnegative random variable and the log-transformation of the jump process $\ln(J_i)$ is defined as:

$$\ln(J_i) = X_u \cdot \mathbb{I}_{\{x \geq 0\}} + X_d \cdot \mathbb{I}_{\{x<0\}} \tag{4.5}$$

where X_u and X_d represent positive and negative jumps that follow exponential distributions with mean values being $\frac{1}{\eta_1}$ and $\frac{1}{\eta_2}$, respectively. Therefore, $\ln(J_i)$ follows a double exponential distribution with asymmetric density as:

$$f_{\ln(J_i)}(x) = p \cdot \eta_1 \exp(-\eta_1 x)\, \mathbb{I}_{\{x \geq 0\}} + (1-p) \cdot \eta_2 \exp(\eta_2 x)\, \mathbb{I}_{\{x<0\}} \tag{4.6}$$

where $\eta_1 > 1$ and $\eta_2 > 0$. The probability $p = P(\ln(J_i) = X_u)$ measures the chance of upward jumps. Hence, the probability of downward jumps is $(1-p) = P(\ln(J_i) = X_d)$. Thus, the expected value of J_i can be expressed as:

$$E(J_i) = p \cdot \frac{\eta_1}{\eta_1 - 1} + (1-p) \cdot \frac{\eta_2}{\eta_2 + 1} \tag{4.7}$$

In the DEJ model, W_t, N_t, and $\ln(J_i)$ are independent processes which allows Kou and Wang (2004) to derive an explicit expression for the pdf of the cumulative return $R_{t,t+\Delta} = \frac{X_{t+\Delta}}{X_t} - 1$ over a time interval Δ:

$$f^{DEJ}_{R_{t,t+\Delta}}(x) = \lambda\Delta\left\{p\eta_1 \exp\left(\frac{\sigma^2\eta_1^2\Delta}{2} - (x - \mu\Delta)\eta_1\right) \cdot \Phi\left(\frac{x - \mu\Delta - \sigma^2\eta_1\Delta}{\sigma\sqrt{\Delta}}\right)\right.$$

$$\left. + (1-p)\eta_2 exp\left(\frac{\sigma^2\eta_2^2\Delta}{2} + (x - \mu\Delta)\eta_2\right) \cdot \Phi\left(\frac{-x + \mu\Delta - \sigma^2\eta_2\Delta}{\sigma\sqrt{\Delta}}\right)\right\} +$$

$$\frac{1-\lambda\Delta}{\sigma\sqrt{\Delta}} \cdot \frac{d}{dx}\Phi\left(\frac{x-\mu\Delta}{\sigma\sqrt{\Delta}}\right). \tag{4.8}$$

Here, $R_{t,t+\Delta}$ is the cumulative return over the period $[t, t+\Delta]$, $\Phi(\cdot)$ is the cdf of the lognormal distribution. Given that t is measured in years, we set $\Delta = \frac{1}{252}$ to obtain the pdf of daily returns for the DEJ model. As shown in equation (4.8), the first two terms within brackets describe the discontinuous price changes motivated by rare events, while the last term relates to the continuous price movements under normal market conditions. Rare events, such as geopolitical conflicts and extreme weather conditions, are sparsely observed over time but have significant impact on the expected return of investments.

In addition to jumps, we incorporate seasonality into the volatility process, following Koekebakker and Lien (2004). Specifically, we replace the constant volatility rate σ in (8) with the following truncated Fourier series:

$$\sigma_t = \sigma + \sum_{s=1}^{S}(\alpha_s \cdot \sin(2\pi st\Delta) - \beta_s \cdot \cos(2\pi st\Delta)) \tag{4.9}$$

where $\sigma > 0$ and α_s and β_s are the seasonality coefficients. As suggested by Koekebakker and Lien (2004), we include three seasonal terms ($S = 3$) in the stochastic volatility specification. The parameters pertaining to jumps, seasonality, and diffusive process are jointly estimated by maximum likelihood using the log likelihood function derived from equations (8) and (9).

To identify the impacts of jumps and seasonality on funding risks in agricultural commodity markets, we compare $MaxCVaR^T(\alpha)$ results across different model specifications which include the GBM model, the DEJ model and the GBM and DEJ models with seasonal volatility, denoted GBM-Seasonal and DEJ-Seasonal, respectively. The DEJ-Seasonal model is a general model which encompasses the other models by considering both jump and seasonal components. To derive the other models, we set certain coefficients of the DEJ-Seasonal model to zero. For the

DEJ model, we set the estimated seasonal coefficients, α_s and β_s in equation (4.9) to zero; for the GBM-Seasonal model, we set the estimated jump intensity, λ in equation (4.8), to zero. The GBM model requires setting α_s, β_s, and λ in equations (8) and (9) to zero. Comparing the MaxCVaR associated with the GBM-Seasonal and DEJ-Seasonal models allows us assessing the impact of jumps on funding risks. The pair composed by DEJ-Seasonal and the DEJ (the GBM) models can be used to assess the impact of seasonal volatility (seasonal volatility and jumps) on funding risks.

4.2.3. *Generating a daily $MaxCVaR$ with the Monte Carlo Simulation Results*

As discussed, funding risk is associated to holding a position in the market for a certain period of time which we express in number of days as $H=T\Delta$. To overcome the Peso problem affecting previous research, we generate a distribution of returns over the holding period. First, we estimate the DEJ-Seasonal model and use it to simulate daily returns over the *H*-day holding period. This is achieved using a fixed rolling window estimation that contains 1,000 daily returns. We choose this window size to have enough observations for estimation of the DEJ-Seasonal model. The number of increments between successive rolling windows is one day, such that the first window contains observations from day 1 to day 1,000, the second window encompasses observations from day 2 to 1,001, and so on. This generates a set of parameter estimations for each day $t = 1{,}000\Delta$ to $N\Delta$, where N is the total number of sample days and *t* is measured in years. Second, we use each set of parameters to generate 10,000 iterations of *H* daily returns using the Monte Carlo method. This allows us to generate MaxCVaR values for sample days from $t = 1{,}000\Delta$ to $N\Delta$. To examine the impacts of jumps and seasonality on funding risk, the 10,000 Monte Carlo iterations are generated under each of the four model specifications by placing the corresponding parameter constraints on the DEJ-Seasonal model to derive the GBM, DEJ and GBM-Seasonal models. If jumps (seasonal volatility) [jumps and seasonal volatility] were insignificant in agricultural

commodity markets, we should expect to see similar MaxCVaR results for the DEJ-Seasonal model and the GBM-Seasonal (DEJ) [GBM] model.

Let each sequence of H daily returns be denoted by $\left\{R_{t+(h-1)\cdot\Delta,t+h\cdot\Delta}\right\}_{h=1}^{H}$. We convert the sequence of daily returns into cumulative returns by summing up returns up to time $t+h\cdot\Delta$. Thus, the new sequence of cumulative returns can be denoted as $\left\{\sum_{d=1}^{h} R_{t+(d-1)\cdot\Delta,t+d\cdot\Delta}\right\}_{h=1}^{H}$. With the cumulative returns, we can determine $R_{min,t,j}$ and $R_{max,t,j}$ for the sequence with respect to H as:

$$\begin{cases} R_{min,t,j}^{H} \coloneqq min\left\{\sum_{d=1}^{h} R_{t+(d-1)\cdot\Delta,t+d\cdot\Delta}\right\}_{h=1}^{H} \\ R_{max,t,j}^{H} \coloneqq max\left\{\sum_{d=1}^{h} R_{t+(d-1)\cdot\Delta,t+d\cdot\Delta}\right\}_{h=1}^{H} \end{cases} \tag{4.10}$$

Equation (4.10) is calculated 10,000 times, one for each Monte Carlo sequence denoted as $j = 1$ to $10{,}000$. In this study, we assess funding risks for $H = 1$ to 66. We group all $R_{min,t,j}^{H}$ ($R_{max,t,j}^{H}$) for each *t* and form empirical distributions of $dF_{1+\alpha}^{H}(x)$ ($dF_{\alpha}^{H}(x)$) for each value of H. Specifically, we convert the MaxCVaR definition in equation (4.2) into a discretized approximation as below:

$$MaxCVaR_t^H(\alpha) \approx \begin{cases} MaxCVaR_t^H(\alpha_D) = \frac{1}{1+\alpha}\sum_{j=1}^{10,000} R_{min,t,j}^{H} \cdot \mathbb{I}_{\left\{F_{1+\alpha}^{H}\left(R_{min,t,j}^{H}\right)\le 1+\alpha\right\}} & \alpha = -95\%, -97.5\%, -99\% \\ MaxCVaR_t^H(\alpha_U) = \frac{1}{1-\alpha}\sum_{j=1}^{10,000} R_{max,t,j}^{H} \cdot \mathbb{I}_{\left\{F_{\alpha}^{H}\left(R_{max,t,j}^{H}\right)\ge \alpha\right\}} & \alpha = 95\%, 97.5\%, 99\% \end{cases} \tag{4.11}$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator function and equals 1 (0) if the condition in brackets is (not) satisfied. Notice (11) is derived for each model specification. While (11) generates the empirical distribution of tail returns for each *t*, investors fear only those returns exceeding (falling short of) the expected size of right (left) returns, as these are difficult to anticipate and thus to prepare for. We use a backtesting method to shed light on this.

4.2.4. Backtesting

We assess whether the *ex-ante* funding risk levels are a good forecast of the *ex-post* tail losses in agricultural commodity markets. We define *ex-ante* funding risk levels as the $MaxCVaR^H(\alpha)$ levels based on equation (4.11). We define *ex-post* tail losses as the maximum and minimum cumulative returns over the holding period derived as in equation (4.10) but computed based on realized sample returns rather than simulated returns. To distinguish from simulated returns, we denote them as $R^{H*}_{max,t}$ and $R^{H*}_{min,t}$ Following Righi and Ceretta (2015), the comparison is established through a t-statistic of standardized extreme returns. The test determines whether realized (*ex-post*) extreme returns $R^{H*}_{max,t}$ $(R^{H*}_{min,t})$ are statistically larger (smaller) than *ex-ante* levels. To standardize, we calculate the square root of the truncated variance conditional on the probability α over the next H days and denote it as $SD^H_{\alpha,t}$:

$$SD^H_{t,\alpha} = \begin{cases} SD^H_{t,\alpha_D} = \left(\frac{1}{1+\alpha} \int_0^{1+\alpha} \left(F^H_{t,s}(x)^{-1} - MaxCVaR^H_t(\alpha) \right)^2 ds \right)^{1/2} & \alpha = -95\%, -97.5\%, -99\% \\ SD^H_{t,\alpha_U} = \left(\frac{1}{1-\alpha} \int_\alpha^1 \left(F^H_{t,s}(x)^{-1} - MaxCVaR^H_t(\alpha) \right)^2 ds \right)^{1/2} & \alpha = 95\%, 97.5\%, 99\% \end{cases} \tag{4.12}$$

where $F^H_{t,s}(x)^{-1} = MaxVaR^H_t(s)$ is the quantile of the cdf in equation (4.1) over the holding period H; the confidence level s is within the range $[0, 1+\alpha]$ for left-tail events or $[\alpha, 1]$ for right-tail events. $SD^H_{t,\alpha}$ in equation (4.12) measures the conditional dispersion of large price movements over or under the α-percentile threshold of the return distribution. As in Righi and Ceretta (2015) we use $SD^H_{t,\alpha}$ to standardize extreme returns as follows:

$$\begin{aligned} r^H_{t,\alpha_D} &= \begin{cases} \left(SD^H_{t,\alpha_D}\right)^{-1} \left(R^{H*}_{min,t} - MaxCVaR^H_t(\alpha_D) \right), & R^{H*}_{min,t} \le F^H_{\alpha_D}(x)^{-1} \\ 0, \quad otherwise \end{cases} \\ r^H_{t,\alpha_U} &= \begin{cases} \left(SD^H_{t,\alpha_U}\right)^{-1} \left(R^{H*}_{max,t} - MaxCVaR^H_t(\alpha_U) \right), & R^{H*}_{max,t} \ge F^H_{\alpha_U}(x)^{-1} \\ 0, \quad otherwise \end{cases} \end{aligned} \tag{4.13}$$

where $\alpha_U = 95\%, 97.5\%, 99\%$ and $\alpha_D = -99\%, -97.5\%, -95\%$; r^H_{t,α_D} and r^H_{t,α_U} are the standardized left-tail and right-tail returns respectively and are also calculated on daily basis. Following Righi and Ceretta (2015) and McNeil and Frey (2000), we conduct one-sided t-tests on standardized returns. For example, the null hypothesis for the r^H_{t,α_U} t-test is $E\left[r^H_{t,\alpha_U}\right] = 0$ against the alternative that $E\left[r^H_{t,\alpha_U}\right] > 0$. This is relevant to investors who hold short positions in futures, as they should fear the positive price spikes exceeding the conditional expected right-tail returns over the holding period *H*. Thus, the t-test statistics for r^H_{t,α_U} and r^H_{t,α_D} are defined as below:

$$t^H_\alpha = \frac{\overline{r^H_\alpha} - 0}{sd(r^H_\alpha)}, \quad \alpha = \alpha_U, \alpha_D \tag{4.14}$$

where t^H_α is the t-test statistic for standardized tail returns at the α confidence level and *H*-day holding period; $\overline{r^H_\alpha}$ and $sd(r^H_\alpha)$ are the sample average and estimated standard error of standardized tail returns, both calculated from $t = 1{,}000\Delta$ to $\mathrm{N}\Delta$ standardized returns. The distribution of t^H_α can be found using a standard bootstrap simulation[24]. We conduct the t-test in (14) for each different model to assess which models perform better in predicting funding risks.

4.3. Empirical Results

4.3.1. Data

We calculate $MaxCVaR^H(\alpha)$ for corn, soybeans, and wheat futures contracts traded in the CME. We use daily settlement prices from Bloomberg. The futures dataset covers the period from

[24] The bootstrap technique corresponds to Efron and Tibshirani (1994). We draw 10,000 random samples with replacement from the sample observations. Each sample has the same size as the number of observations. After obtaining all samples, we calculate the t-statistics as suggested in equation (4.14) for each sample and create the empirical distributions of the t-statistics.

January 3rd, 2000 to October 15th, 2021. Corn, soybeans, and wheat futures are the three agricultural futures contracts with the largest trading volume at CME. Our analysis focuses on the nearby contract series since most trading occurs in the first to expire contract. The nearby contract is rolled over to the next contract when trading volume in the nearby is lower than the trading volume in the next delivery contract. To avoid artificial price jumps induced by the rollover method, we calculate daily log returns within contracts and thus we exclude the roll yields (Paschke, Prokopczuk and Wese Simen 2020). We then re-construct the price levels from the log returns of the most-traded contact series. Price levels are presented in Figure 1 which shows all three commodities experience highly volatile periods. The first one is between 2006 and 2008 and comprises the commodity price boom and the financial crisis. The second is between 2011 and 2014, which was driven by severe drought conditions in the Midwest causing significant yield reductions. Additionally, the covid-19 outbreak also induced extreme price movements in corn and soybean markets between 2020 and 2021.

4.3.2. Empirical findings

In Table 1, we present the mean, variance, skewness, and excess kurtosis of the annualized daily returns for the sample. Annualized daily returns are derived by multiplying daily log price returns by 252. As shown in Tabel 4.1, excess kurtosis ranges between 2.830 and 3.526 with the return distributions for corn and soybeans having thicker tails than for wheat. Corn and wheat futures returns have positive skewness, with soybean returns being characterized by negative skewness. Jointly, excess kurtosis and skewness suggest that return distributions of grain futures are leptokurtic and asymmetric. The Jarque-Bera test allows us to formally test whether returns are normally distributed. The test rejects the null hypothesis that the log return distribution is normal in all the markets. This is consistent with the empirical findings in numerous studies (Hilliard and

Reis 1999; Koekebakker and Lien 2004; Morgan, Cotter and Dowd 2012; Xouridas 2015) and justifies the use of a DEJ-Seasonal model.

4.3.3. Parameter Estimation Results

We use the maximum likelihood method to estimate the DEJ-Seasonal model parameters using annualized daily returns for each agricultural commodity. As explained in the methodology section, parameters are obtained for each fixed rolling window comprising 1,000 days. This allows us to produce 4,466 sets of coefficient estimates whose summary statistics are presented in Tabel 4.2 separated into three groups by the horizontal lines. We present the mean and median values of estimated parameters for each market over the entire estimation period. We also include the proportion of significant estimations across the 4,466 iterations given the t-test results (we use 5% as the significance level). The top segment contains all seasonal coefficients ($\{\alpha_s\}_{s=1}^{3}$ and $\{\beta_s\}_{s=1}^{3}$) in equation (4.9) which suggest that different markets display different seasonal volatility patterns. In the corn market, β_1 is the most influential seasonal factor with the largest absolute mean and median values (0.054 and 0.057, respectively). At the same time, the percentage of statistically significant estimations over the total number of estimations for β_1 is 89.7%. While β_1 is also the most influential seasonal factor in both soybean and wheat markets, the magnitude and significance of β_1 are lower in these two markets relative to the corn market. Thus, seasonal volatility patterns in wheat and soybean markets should be less strong than in the corn market. We confirm this finding later in Figures 2 and 3. The remaining seasonal parameters are less relevant in determining seasonal volatility in the three markets, both in terms of significance and magnitude.

The middle segment of Tabel 4.2 contains jump-related parameters, such as jump sizes (η_1^{-1} and η_2^{-1}), jump intensity (λ) and the probability of upward jumps (p). As discussed earlier, η_1^{-1} and η_2^{-1} are the expected upward and downward jump size expressed in annualized returns,

respectively. Results suggest that jump magnitudes are more symmetric in the corn market than in the other two markets. Average jump sizes in corn futures are 1.5% for both positive and negative jumps. The same parameters in the soybean (wheat) market are 1.2% (2%) and 1.4% (1.2%) for positive and negative jumps, respectively. Unlike corn and soybean markets, with 99-100% of jump size parameters being statistically significant, only 90% and 51% of positive and negative wheat jump sizes are statistically significant. Thus, jump sizes are less significant in the wheat market than in the other two markets. This is consistent with jump intensity estimates (λ). The average number of jumps per year in the wheat market is 51, which contrasts with corn (73) and soybean (94) markets. Also, the average percentage of significant estimations of the jump intensity in the wheat market is 97%, slightly lower than the same results in corn and soybean markets (100%). The parameters measuring the probability of positive jumps, p, also tell the same story with mean values (percent of statistically significant estimations) being 0.531 (76%), 0.525 (84%), and 0.598 (27%) in corn, soybean, and wheat markets, respectively. Thus, we conclude that the jump process in the wheat market is less strong than in the corn and soybean markets. This is also consistent with the excess kurtosis results in Tabel 4.1, where corn and soybean markets have higher excess kurtosis than the wheat market.

Drift terms and constant volatility rates, both annualized, also vary considerably across different markets. The average values for drift terms vary between -0.138 and 0.159. However, the percentage of significant estimations for μ are close to zero in all markets. The mean values for σ range between 0.174 to 0.269 across different markets with virtually 100% of parameters being significant. The wheat market has the highest annualized continuous volatility rate, while the soybean market has the lowest.

4.3.4. Funding Risks for Market Participants

Tabel 4.3 presents a summary of the $MaxCVaR_t^H(\alpha)$ results generated from the Monte Carlo method for $t = 1{,}000\Delta$ to $N\Delta$, over the holding periods $H = 22$, 44, and 66, and for each alternative model specification (GBM, GBM-Seasonal, DEJ and DEJ-Seasonal).[25] The values in Tabel 4.3 are averages of daily $MaxCVaR_t^H(\alpha)$ results over *t*. As discussed, the $MaxCVaR_t^H(\alpha)$ value is the ex-ante expectation of extreme returns given certain confidence levels and number of holding days. Confidence levels α with positive and negative signs refer to the right and left tails of the distribution, respectively. For example, $\alpha = -99\%$ and $\alpha = 99\%$ measure the $MaxCVaR_t^H(\alpha)$ corresponding to the 1% extreme tails of the distribution which generate funding losses for long (short) positions. All $MaxCVaR_t^H(\alpha)$ results in Tabel 4.3 are expressed in proportions. For example, the GBM model $MaxCVaR_t^{22}(\alpha)(-99\%)$ = -0.167 for corn represents an average -16.7% expected loss for a long corn futures position at the 1% quantile over a 22-day holding period.

For simplicity, we only illustrate our findings with the DEJ-Seasonal model in the corn market. The results for the other models and markets are qualitatively the same. We will pay attention to the other models and markets when assessing the relative importance of jumps and seasonal volatility. Average $MaxCVaR_t^H(\alpha)$ values increase with both the absolute value of α and H. This implies that to increase the confidence level (α) against more extreme situations, investors are expected to hold more cash in hand. For example, in the corn market, the magnitude of

[25] The initial 1,000 observations of the full sample from January 3, 2000 are lost to the rolling window process. We will pay attention to $H = 1$ in section 4.3.7. where we assess the efficacy of margin requirements.

$MaxCVaR_t^H(\alpha)$ when $H = 22$ increases from 20.4% (-17.1%) to 25.6% (-20.3%) when α changes from 95% (-95%) to 99% (-99%). Second, absolute $MaxCVaR_t^H(\alpha)$ values also increase when the number of holding days H increases. With longer holding periods, investors are expected to experience higher aggregated volatility levels. Therefore, the $MaxCVaR_t^H(95\%)$ increases from 20.4% to 38.2% in the corn market when H increases from 22 days to 66 days. Third, absolute $MaxCVaR_t^H(\alpha)$ values in the positive tails are always larger than in the negative tails with the same absolute confidence levels. For instance, the $MaxCVaR_t^H(97.5\%)$ values for $H = 22$, 44 and 66 are 22.6%, 33.4%, and 42.3%, respectively, while their counterparts $MaxCVaR_t^H(-97.5\%)$ are only -18.5%, -25.3%, and -30.2%. As shown in Tabel 4.2, both mean and median values of p are higher than 0.5, which indicates that positive jumps are more likely to occur than negative jumps. This relates to our sample period containing proportionally more extreme positive price movements, such as severe yield reductions and commodity price booms. Thus, we can conclude that during the period studied, funding risk was higher for market participants holding short positions than long positions in agricultural commodity futures markets.

Our funding risk estimates based on the $MaxCVaR_t^H(\alpha)$ model are higher than the maximum margin liability derived by Shi and Isengildina-Massa (2021) for corn and soybean markets and $H = 22$. Shi and Isengildina-Massa (2021) report maximum margin liability in the corn market to average 11% and 15% for long and short positions between 2004 and 2018. This amount includes the initial margin payment, which is part of hedging costs, but not of futures trading risks. Despite consideration of initial margins, the Peso problem leads Shi and Isengildina-Massa's (2021) estimates to be substantially below the mean values of our DEJ-Seasonal $MaxCVaR_t^H(\alpha)$ in the

corn market which are 21% and 26.9% for long and short positions over the same period.[26] Similar results can also be found in the soybean market. As mentioned earlier, the maximum margin liability understates funding risks in agricultural futures markets by ignoring infrequent and unprecedented price movements, such as jumps. In the next subsection we compare across different model specifications to assess the relevance that different volatility components have on funding risk.

4.3.5. MaxCVaR Decomposition Results: Jumps vs. Seasonal Volatility

As discussed, different $MaxCVaR_t^H(\alpha)$ values across various models illustrate the contribution of jump and seasonal volatility components to funding risk in agricultural commodity markets. We test the differences in mean $MaxCVaR_t^H(\alpha)$ between each pair of models using the paired Wilcoxon signed rank test. Results are exhibited in Tabel 4.4 where for ease of interpretation we present average ratios of $MaxCVaR_t^H(\alpha)$ of the models compared (rather than the differences we actually test). If jumps (seasonal volatility) play an important role in average funding risk levels, then differences in $MaxCVaR_t^H(\alpha)$ values between the DEJ and GBM (GBM and GBM-Seasonal) models should be statistically significant. Comparison between DEJ-Seasonal and GBM-Seasonal (DEJ-Seasonal and DEJ) models can further inform us if jumps (seasonal volatility) are still necessary after taking seasonal volatility (jumps) into consideration.

Results in Tabel 4.4 suggest that jumps clearly dominate seasonal volatility in funding risks. The $MaxCVaR_t^H(\alpha)$ values from the DEJ model are statistically larger than the GBM model at the 99% level across all markets and confidence levels. These ratios represent the importance of funding risks driven by jumps relative to the continuous diffusive volatility process. It is worth

[26] The mean values here are different from Table 3 since it only covers a subperiod (from 2004 to 2021).

mentioning that higher ratios on one side of the return distribution, either positive or negative, only imply that jumps are more relevant factors of funding risks particularly on this side of the distribution. Average ratios between DEJ and GBM models fluctuate between 1.238 and 1.464 in the soybean market, with jumps having the strongest impact on traders' long positions. The same ratios are only between 1.150 (1.056) and 1.387 (1.288) in the corn (wheat) market with jumps having the strongest impact on short positions. As suggested in Tabel 4.1 and 4.2, the average values of annualized daily returns and drift terms are positive in the soybean market and negative in the corn and wheat markets. Therefore, funding risks of traders who hold positions benefiting from current market trends (price increases in soybeans and declines in corn and wheat) are more relevant under jumps in the opposite than in the same market direction. By taking soybeans as an example, this results in soybean traders holding long positions being more likely to suffer from extreme losses due to price jumps than traders holding short positions. The ratios in Tabel 4.4 imply that jumps increase funding risk in the soybean market between 23.8% and 46.4%, above the values for corn (wheat) between 15% (5.6%) and 38.7% (28.8%).

The $MaxCVaR_t^H(\alpha)$ ratios between DEJ-Seasonal and GBM-Seasonal models indicate that jumps still affect the mean funding risk after filtering for seasonal patterns in returns. All differences are statistically significant at the 99% level and have the same patterns as the differences between DEJ and GBM, with jumps having the strongest impact on long (short) positions in the soybean (corn and wheat markets). Therefore, jumps must be considered when calculating $MaxCVaR_t^H(\alpha)$ values.

In contrast to jumps, seasonal volatility components have limited influence on average $MaxCVaR_t^H(\alpha)$ values. As demonstrated by the ratios between the GBM and GBM-Seasonal models, seasonal volatility components have no statistically significant impacts on the average

levels of funding risk in any of the three markets, except for the negative tails in the soybean market. Although these differences are significant at the 99% level, the increased funding risk magnitudes are less than 0.2%. Comparison between the DEJ-Seasonal and DEJ models suggests that seasonal volatility components become irrelevant in estimating average funding risk levels after taking jumps into consideration. Test results indicate that none of the differences in $MaxCVaR_t^H(\alpha)$ values between these two models is significant at the 99% level. The results mentioned above do not overturn previous findings in terms of the importance of seasonality in agricultural commodity markets. In fact, seasonal volatility defined in equation (4.9) can only alter the temporal distribution of funding risks rather than change average funding risks levels. Thus, we conclude that jumps are the main source of average funding risk levels.

Agricultural commodity price volatility is subject to seasonal demand and supply patterns (Aravindhakshan and Brorsen 2011; Sørensen 2002; Karali and Power 2013). Hence, although seasonal volatility components do not affect average levels of funding risk, they may alter the temporal distribution of funding risk across different months. As a result, seasonal volatility components may enable models with stochastic volatility to capture higher (lower) funding risk during peak (off-peak) seasons. To illustrate the effects of seasonal volatility components on the temporal distribution of funding risk, we plot the absolute magnitude of monthly average $MaxCVaR_t^H(\alpha)$ estimated with the four models across all years. We set $\alpha = -99\%, 99\%$ and $H = 22$. The results for other combinations of α and H have the same patterns. Yet, the differences between maximum and minimum monthly average funding risks increase with both α and H.

As depicted in Figure 2, funding risks derived from GBM-Seasonal and DEJ-Seasonal models exhibits strong seasonality. Funding risk reaches peak levels during June and July and touches

bottom levels in December and January. The maximum monthly average $MaxCVaR_t^H(\alpha)$ values estimated from the DEJ-Seasonal model in the corn market with $\alpha = 99\%$ and -99% are 9.17% and 6.07% larger than the minimum values. The seasonal funding risk pattern coincides with the grow-harvest cycle of corn in the U.S. Midwest. As suggested in Tabel 4.2, seasonal volatility components are less relevant in the wheat and soybean markets relative to the corn market. Therefore, the seasonal funding risk pattern is also less pronounced in these two markets than in the corn market. Maximum monthly average funding risks derived from the DEJ-Seasonal model with α =99% and -99% are 5.95% (5.54%) and 3.83% (3.69%) higher than the minimum value in the wheat (soybean) market.

4.3.6. Backtesting Results for MaxCVaR Estimations

As discussed, the exchange clearinghouse and futures market participants' funding risks will be especially relevant when extreme returns exceed ex-ante expectations for which clearinghouses and traders can prepare. Here, we investigate to what extent the $MaxCVaR_t^H(\alpha)$ derived based on data from the previous 1,000 days constitutes a good ex-ante prediction of future funding risk. To answer this question, we backtest $MaxCVaR_t^H(\alpha)$ values from all four models using the t-test in equation (4.14). A model is viewed as sufficient for managing funding risk if the ex-ante $MaxCVaR_t^H(\alpha)$ is larger (smaller) than any realization of extreme positive (negative) returns over the subsequent *H* holding period. For example, if the $MaxCVaR_t^H(\alpha)$ equals 15% and the realized maximum return is 10% (20%), then the ex-ante $MaxCVaR_t^H(\alpha)$ estimation is (not) sufficient to cover funding risks for traders who hold short futures positions. Backtesting $MaxCVaR_t^H(\alpha)$

values using all observations from 2004 to 2021 together can be misleading.[27] As we can see in Figure 1, explosive price movements tend to cluster around certain periods, such as during the financial crisis or extreme weather conditions. Funding risks during these periods are likely to be unprecedented and exceed the threshold estimated using the previous 1000-day historical returns. This may result in rejection of the t-test when conducted on all years. Hence, we also conduct tests for each year separately. Given the role of jumps and, to a lesser extent, seasonal volatility in explaining tail returns, we expect DEJ, DEJ-Seasonal and GBM-Seasonal to result in lower rejection rates of the null hypothesis that realized returns do not exceed the ex-ante $MaxCVaR_t^H(\alpha)$ than the GBM model. Also, we expect the DEJ-Seasonal model to outperform the rest.

Results show lower backtesting rejection rates for the DEJ-Seasonal model in all markets for $H = 22$ and $\alpha = 99\%, -99\%$ (Tabel 4.5A and 4.5B). This implies the DEJ-Seasonal model can better anticipate extreme price movements in agricultural commodity markets than other models. The results with $H = 44$ and 66 are similar to the case with $H = 22$ and not presented to preserve space. Backtesting results using all available observations between 2004 and 2021, generates 6 scenarios (3 markets $\times$ 2 αs) in total. All four models are rejected as useful in predicting funding risk by the t-test at either a p-value ≤ 0.05 or ≤ 0.01. As discussed, extreme turbulent sub-periods are likely to be driving these results. Instead of relying on the results over the entire period, we apply the backtesting results by year. The annual backtesting method expands the number of scenarios from 6 to 108 (3 markets $\times$ 2 αs $\times$ 18 years). Out of the 108 cases and relying on a p-

[27] Recall that while our sample starts in 2000, the first 1,000 observations are lost to the rolling window process.

value ≤ 0.01, the GBM and GBM-Seasonal models are rejected 45 and 46 times, respectively. Thus, adding seasonal volatility without jumps cannot improve protection against funding risks. The same numbers for the DEJ and DEJ-Seasonal models are only 26 and 20, respectively. These findings are consistent with Table 4, which shows the relevance of allowing for jumps in explaining funding risk. Seasonal volatility components can also offer marginal improvements after taking jumps into consideration. Furthermore, jumps and seasonal volatility components provide uneven levels of improvements across different markets. In the corn (soybean) [wheat] market, the numbers of rejections are 19 (18) [9] and 13 (20) [13] for the GBM and GBM-Seasonal model, respectively. The same numbers are only 12 (10) [4] and 8 (9) [3] for the DEJ and DEJ-Seasonal model. Thus, seasonal volatility components alone can only offer meaningful improvements in anticipating funding risk levels in the corn market. However, after taking jumps into consideration, seasonal volatility components can reduce rejection rates in all markets. To summarize, rejection rates of the DEJ-Seasonal model are lower than rejection rates of the other models across all markets. This allows concluding that among all models considered, the DEJ-Seasonal model is the optimal model to anticipate funding risks in all three markets. Relative to the GBM model, the DEJ-Seasonal model reduces funding risk by at least 50%. Therefore, exchanges and market participants should consider both jumps and seasonal volatility patterns when estimating funding risk levels in agricultural commodity markets.

4.3.7. *Are Margin Payments Sufficient to Cover Funding Risks?*

As discussed early, SPAN generates expected margin payments of futures contracts that are identical to the futures contract margin requirements. Thus, if margin requirements are not sufficient to cover funding risks in the first instance, SPAN expected payments will also underestimate funding risks faced by futures traders. To assess if CME margin requirements are

sufficient to cover funding risks, we convert $MaxCVaR_t^H(\alpha)$ results, which are expressed in percent returns, into dollar value per contract with $\alpha = 99\%$ and -99% and $H = 1$, and compare them against margin requirements. We choose $H = 1$, as margin requirements are established daily. We calculate the absolute dollar values of funding risk per contract as $\left|MaxCVaR_t^H(\alpha) \times X_{t,0} \times contract\ size\right|$, where $X_{t,0}$ is the price at the beginning of each holding period. We use DEJ-Seasonal and GBM-Seasonal models to improve comparability with the CME margin model that allows for seasonality (CME 2022). Results are presented in Tabel 4.6, where column 2 presents the daily average values of CME margin requirements in dollars per contract over the sampling period. Columns 4 (6) and 5 (7) contain average funding risks derived from the DEJ-Seasonal (GBM-Seasonal) model.

The gaps between margin requirements and our estimated absolute dollar value of funding risks per contract under $H = 1$ in Tabel 4.6 suggest that CME's margin requirements are generally insufficient to cover funding risks in grain futures markets in the presence of price jumps. The average margin requirements per contract are \$1,100, \$2,311, and \$1,615 in corn, soybean, and wheat markets, respectively. While in most cases, these requirements are above risks derived from the GBM-Seasonal model, they are below the DEJ-Seasonal model. Specifically, margin requirements in corn (soybean) [wheat] are up to 14.2% (15.9%) [10.9%] below funding risks based on the DEJ-Seasonal model. The only exception exists in the wheat market with $\alpha = -99\%$, where average margin requirements (\$1,615) exceed DEJ-Seasonal funding risks per contract (\$1,601). As shown in Tabel 4.4, relative to corn and soybeans, jumps are less relevant in funding risks in the wheat market with $\alpha = -99\%$. Therefore, margin requirements are adequate to cover funding risks of long positions in the wheat market because of smaller jump ratios. In Figure 3, we plot monthly average funding risks and margin requirements between 2004 and 2021. As is the

case with funding risks, margin requirements also exhibit seasonal patterns. Also, margin requirements consistently fall between funding risks derived from DEJ-Seasonal and GBM-Seasonal models across all markets, except for long positions in the wheat market. In summary, while margin requirements are sufficient to cover funding risks generated by continuous price movements and seasonality, they are insufficient in the event of price jumps.

4.4. Conclusion

Commercial traders can use futures contracts to hedge their cash positions to manage price risks. However, increased volatility in agricultural commodity markets during the past two decades challenges traders' ability to meet the exchange margin requirements posing funding risks for both traders and the clearinghouse. In this study, we use a new MaxCVaR method to measure funding risks in CME agricultural futures markets for the first time. Specifically, we extend the MaxCVaR method by Farkas, Mathys and Vasiljević (2021) to consider seasonal volatility. We apply the model to estimate funding risk in corn, soybean, and wheat futures markets. We concentrate on the extremes of the return distribution as these are the ones used by the exchange to define margin requirements. We assume the return process follows the DEJ model proposed in Kou (2002) and model large positive and negative jumps in agricultural futures markets separately. Also, we follow Koekebakker and Lien (2004) and capture seasonal volatility patterns in our markets with a truncated Fourier series. We denote the new DEJ model with seasonal volatility patterns as the DEJ-Seasonal model. We estimate the parameters of the DEJ-Seasonal model with the maximum likelihood method and then generate MaxCVaR values with a Monte Carlo simulation. By comparing these values with those generated using models that either do not consider or partially consider jumps and seasonality and by backtesting the models, we shed light on how jumps and seasonality contribute to funding risk and which models can help better prepare for this risk.

We find that funding risks impose significant financial burden on market participants. For example, holding a long (short) position over a month in the corn futures market results in 20.3% (25.6%) losses in the lower (upper) 1-percentile extremes of the return distribution. Losses increase with the length of the holding period in all markets. Moreover, traders holding short positions in agricultural futures are generally subject to higher funding risks than their counterparties. The decomposition results of MaxCVaR values show that jumps and seasonality play important roles in the determination of funding risk. Jumps alone significantly increase MaxCVaR values by up to 46.8% across different markets. Despite of not being able to change average funding risk levels, seasonal volatility components alter the temporal distribution of funding risk across different months. The timing of the highest and lowest MaxCVaR values coincide with the growing and pre-planting seasons for the underlying crops. Funding risks during the growing season can be 9.17% higher than during the pre-planting season in grain markets over a one-month holding period. Therefore, market participants will face different funding risks depend upon the seasonality in underlying markets.

We also backtest the MaxCVaR values estimated through different models against realized extreme price movements. We use the method developed by Righi and Ceretta (2015). Testing results suggest that MaxCVaR values estimated with jumps and seasonal volatility components over the past 1,000 days are significantly useful in anticipating extreme price movements and thus funding risks in all three markets. In comparison, MaxCVaR values estimated with seasonal volatility components alone can only offer marginal improvements. Thus, we recommend the DEJ-Seasonal model over alternative models for estimating funding risk.

Finally, we convert funding risks into absolute dollar values per contract and compare them against margin requirements. The comparison results suggest that margin requirements established

by CME are insufficient to cover funding risks induced by price jumps. Specifically, they are up to 15.9% below actual funding risks on average. Thus, the exchange should consider either expanding margin requirements to allow for futures price jumps or adding more extreme scenarios to the SPAN model.

Our empirical findings are useful for both academic research and real-world applications. Researchers can incorporate funding risk into the existing hedging literature to understand the impact of liquidity constraints on hedging decisions. Also, researchers can develop new strategies to help market participants manage funding risk effectively. As for policy makers, funding risk provides an important tool to better understand market liquidity under stressful conditions. Especially, the share of jumps in funding risks can inform policy makers of the magnitude of liquidity pressure, a measure of market fragility (Brunnermeier and Pedersen 2009), induced by unexpected market events. For market participants, funding risk measured as the MaxCVaR method estimated by the DEJ-Seasonal model, is a reliable tool to help them determine the necessary level of liquid assets they may want to withhold before taking any futures positions. For exchanges and clearinghouses, the DEJ-Seasonal model can help them adjust margin requirements accordingly and reduce liquidity stress in the presence of price jumps.

4.5 Tables and Figures

Table 4.1: Summary Statistics for Daily Returns Across Different Markets

Market	Mean	Variance	Skewness	Excess.Kurtosis	JB.statistics	JB.p.value	N
Corn	-0.040	17.948	0.010	3.526	1,452.957	0	5,466
Soybean	0.057	13.479	-0.207	3.476	1,435.548	0	5,465
Wheat	-0.080	22.439	0.120	2.830	775.597	0	5,466

Note: the market column contains names for each market; the annualized daily returns are generated by multiplying log returns by 252; the first four columns from the left to right exhibit the mean, variance, skewness, and kurtosis of the log return distribution in each market; the columns of JB.statistics and JB.p.value represent the test statistics and p-values of the Jarque-Bera normality test; the null hypothesis under the Jarque-Bera test is a joint hypothesis of both skewness and excess kurtosis being zero; the column N contains the number of observations for each market.

Table 4.2: Summary Statistics for Parameters Across Markets

Param	Corn			Soybean			Wheat		
	Mean	Median	Perc.Sig	Mean	Median	Perc.Sig	Mean	Median	Perc.Sig
a_1	-0.010	-0.009	0.412	-0.020	-0.021	0.518	0.010	0.014	0.459
a_2	-0.005	-0.008	0.141	0.007	0.004	0.170	0.004	0.003	0.162
a_3	0.002	0.001	0.050	-0.003	-0.002	0.103	0.001	0.001	0.054
β_1	0.054	0.057	0.897	0.023	0.024	0.633	0.028	0.025	0.529
β_2	-0.009	-0.015	0.412	-0.008	-0.010	0.299	0.007	0.006	0.315
β_3	0.011	0.010	0.221	0.010	0.009	0.248	0.012	0.011	0.219
η_1^{-1}	0.015	0.014	0.997	0.012	0.011	0.988	0.020	0.015	0.901
η_2^{-1}	0.015	0.015	0.997	0.014	0.014	0.995	0.012	0.012	0.512
λ	73.333	64.286	0.999	93.816	75.171	0.999	50.686	22.040	0.970
p	0.531	0.529	0.755	0.525	0.516	0.837	0.598	0.539	0.266
μ	-0.066	-0.077	0.009	0.159	0.142	0.031	-0.138	-0.151	0.015
σ	0.215	0.207	1.000	0.174	0.176	0.999	0.269	0.253	1.000

Note: All parameters are estimated by maximum likelihood with the density function in equation (4.8) and the seasonal volatility rate in equation (4.9); $\{\alpha_s\}_{s=1}^3$ and $\{\beta_s\}_{s=1}^3$ are seasonal coefficients in equation (4.9); η_1^{-1} and η_2^{-1} are annualized mean jump magnitudes for positive and negative jumps; λ is the average number of jumps per year; p is the probability of positive jumps; μ and σ are annualized drift and volatility rates; each parameter set is estimated for each rolling window of fixed size 1,000, with 4,466 being the total number of rolling windows; the Mean and Median columns present the mean and median values of all parameters; the Perc.Sig column represents the percentage of significant estimations with p.value = 5% across all estimation iterations; the number of observations is 4,467 for corn and wheat markets and 4,466 for the soybean market.

Table 4.3: Average Daily $MaxCVaR_t^H(\alpha)$ Estimation Results Across Markets: 2004 - 2021

Market	α	GBM			GBM-Seasonal			DEJ			DEJ-Seasonal		
		22	44	66	22	44	66	22	44	66	22	44	66
Corn	-99%	-0.167	-0.232	-0.280	-0.167	-0.233	-0.280	-0.202	-0.273	-0.325	-0.203	-0.275	-0.326
	-97.5%	-0.153	-0.214	-0.260	-0.153	-0.215	-0.260	-0.183	-0.251	-0.300	-0.185	-0.253	-0.302
	-95%	-0.142	-0.201	-0.245	-0.143	-0.202	-0.246	-0.170	-0.235	-0.282	-0.171	-0.236	-0.283
	95%	0.151	0.219	0.275	0.152	0.222	0.278	0.202	0.298	0.377	0.204	0.301	0.382
	97.5%	0.165	0.241	0.303	0.167	0.245	0.307	0.223	0.329	0.418	0.226	0.334	0.423
	99%	0.185	0.272	0.342	0.187	0.276	0.347	0.253	0.373	0.474	0.256	0.378	0.481
Soybean	-99%	-0.122	-0.166	-0.196	-0.123	-0.166	-0.196	-0.179	-0.240	-0.283	-0.180	-0.240	-0.283
	-97.5%	-0.111	-0.151	-0.179	-0.111	-0.151	-0.179	-0.161	-0.218	-0.258	-0.162	-0.219	-0.259
	-95%	-0.102	-0.139	-0.166	-0.102	-0.139	-0.166	-0.148	-0.202	-0.240	-0.149	-0.203	-0.241
	95%	0.139	0.214	0.280	0.140	0.216	0.282	0.180	0.270	0.347	0.181	0.272	0.349
	97.5%	0.150	0.231	0.302	0.151	0.233	0.304	0.197	0.296	0.381	0.199	0.298	0.383
	99%	0.166	0.255	0.333	0.167	0.257	0.336	0.221	0.332	0.427	0.223	0.335	0.430
Wheat	-99%	-0.207	-0.288	-0.347	-0.207	-0.288	-0.346	-0.228	-0.310	-0.369	-0.228	-0.310	-0.369
	-97.5%	-0.191	-0.267	-0.324	-0.191	-0.267	-0.323	-0.209	-0.287	-0.343	-0.209	-0.287	-0.343
	-95%	-0.179	-0.252	-0.307	-0.179	-0.252	-0.306	-0.194	-0.270	-0.324	-0.195	-0.270	-0.324
	95%	0.185	0.267	0.331	0.186	0.268	0.332	0.229	0.336	0.423	0.230	0.338	0.425
	97.5%	0.204	0.295	0.368	0.204	0.297	0.369	0.254	0.373	0.471	0.256	0.376	0.474
	99%	0.229	0.334	0.418	0.230	0.336	0.420	0.289	0.425	0.539	0.291	0.428	0.542

Note: The Market and α columns represent different markets and confidence levels; GBM, GBM-Seasonal, DEJ, and DEJ-Seasonal represent results for different models; 22, 44, and 66 are numbers of holding days (H); All values are average $MaxCVaR_t^H(\alpha)$ values between 2004 and 2021; for corn and wheat markets, the numbers of observations for $H = 22, 44, 66$ are 4,446, 4,424, and 4,402; for the soybean market, the numbers of observations are 4,445, 4,423, 4,401.

Table 4.4: Average Ratios for $MaxCVaR_t^H(\alpha)$ Across Models

Market	α	DEJ/GBM			DEJ-Seasonal/DEJ			GBM-Seasonal/GBM			DEJ-Seasonal/GBM-Seasonal		
		22	44	66	22	44	66	22	44	66	22	44	66
Corn	-99%	**1.212***	**1.178***	**1.159***	1.007	1.006	1.004	1.003	1.002	1.000	**1.217***	**1.182***	**1.164***
	-97.5%	**1.201***	**1.172***	**1.155***	1.007	1.006	1.004	1.003	1.002	1.000	**1.205***	**1.176***	**1.159***
	-95%	**1.193***	**1.167***	**1.150***	1.007	1.005	1.004	1.004	1.003	1.001	**1.196***	**1.170***	**1.154***
	95%	**1.340***	**1.357***	**1.373***	1.011	1.012	1.012	1.011	1.013	1.013	**1.340***	**1.356***	**1.373***
	97.5%	**1.351***	**1.364***	**1.379***	1.012	1.013	1.013	1.011	1.014	1.013	**1.352***	**1.363***	**1.379***
	99%	**1.367***	**1.374***	**1.387***	1.012	1.014	1.014	1.012	1.015	1.014	**1.367***	**1.373***	**1.387***
Soybean	-99%	**1.464***	**1.446***	**1.442***	1.005	1.004	1.003	**1.002***	**1.000***	**1.000***	**1.468***	**1.451***	**1.446***
	-97.5%	**1.456***	**1.449***	**1.448***	1.004	1.003	1.003	**1.002***	**1.001***	**1.000***	**1.460***	**1.452***	**1.451***
	-95%	**1.451***	**1.451***	**1.452***	1.004	1.003	1.003	**1.002***	**1.001***	**1.001***	**1.454***	**1.453***	**1.455***
	95%	**1.292***	**1.261***	**1.238***	1.007	1.007	1.006	1.005	1.006	1.006	**1.294***	**1.263***	**1.239***
	97.5%	**1.310***	**1.280***	**1.258***	1.007	1.007	1.007	1.006	1.006	1.006	**1.313***	**1.281***	**1.259***
	99%	**1.333***	**1.301***	**1.281***	1.008	1.008	1.008	1.006	1.007	1.007	**1.336***	**1.302***	**1.282***
Wheat	-99%	**1.098***	**1.076***	**1.063***	1.001	1.001	1.000	1.000	0.999	0.998	**1.099***	**1.078***	**1.065***
	-97.5%	**1.093***	**1.072***	**1.060***	1.002	1.001	1.000	1.000	0.999	0.998	**1.094***	**1.075***	**1.062***
	-95%	**1.088***	**1.069***	**1.056***	1.002	1.002	1.001	1.000	0.999	0.998	**1.090***	**1.071***	**1.058***
	95%	**1.239***	**1.260***	**1.279***	1.006	1.006	1.005	1.004	1.005	1.004	**1.242***	**1.262***	**1.280***
	97.5%	**1.248***	**1.265***	**1.282***	1.006	1.007	1.005	1.004	1.005	1.004	**1.250***	**1.266***	**1.283***
	99%	**1.260***	**1.272***	**1.288***	1.007*	1.008	1.005	1.005	1.006	1.005	**1.263***	**1.274***	**1.288***

Note: The Market and α columns represent different markets and confidence levels; Model A/Model B measures the mean of ratios of $MaxCVaR_t^H(\alpha)$ between Model A and Model B between 2004 and 2021; 22, 44, and 66 are numbers of holding days (H); All results are tested with the paired Wilcoxon signed rank sum test; the asterisk symbol ("*") and bold fonts indicate p-values ≤ 0.05 and 0.01, respectively; for corn and wheat markets, the numbers of observations for $H = 22, 44, 66$ are 4,446, 4,424, and 4,402; for the soybean market, the numbers of observations are 4,445, 4,423, 4,401.

Table 4.5A: Bootstrap t-test Results for $MaxCVaR_t^H(\alpha)$ Estimations with H = 22

Model	Year	Corn		Soybean		Wheat	
		-99%	99%	-99%	99%	-99%	99%
GBM	2004	**0.000***	0.665	**0.000***	**0.005***	1.000	0.919
	2005	0.085	**0.000***	0.863	**0.000***	1.000	0.017*
	2006	1.000	**0.000***	1.000	**0.000***	1.000	**0.000***
	2007	0.048*	0.930	1.000	0.971	1.000	**0.000***
	2008	**0.000***	**0.000***	**0.000***	**0.000***	**0.002***	**0.009***
	2009	**0.000***	0.991	**0.000***	0.926	0.312	0.918
	2010	0.012*	0.076	**0.000***	0.465	1.000	**0.000***
	2011	**0.005***	0.045*	**0.000***	0.210	0.116	1.000
	2012	0.916	**0.000***	**0.006***	**0.001***	1.000	**0.000***
	2013	**0.000***	0.153	**0.000***	0.030*	1.000	1.000
	2014	**0.001***	0.993	**0.000***	0.208	0.982	0.998
	2015	0.822	**0.000***	0.918	1.000	0.915	**0.000***
	2016	**0.000***	0.081	0.288	0.996	0.919	1.000
	2017	1.000	0.972	1.000	0.435	0.991	**0.000***
	2018	0.474	1.000	0.130	0.991	1.000	0.987
	2019	**0.000***	**0.000***	0.919	0.999	0.917	0.195
	2020	0.942	**0.000***	0.923	**0.000***	1.000	1.000
	2021	**0.000***	**0.000***	**0.000***	**0.000***	1.000	0.998
	All	**0.000***	**0.000***	**0.000***	**0.000***	0.014*	**0.000***
GBM-Seasonal	2004	0.209	**0.000***	**0.000***	**0.000***	1.000	0.995
	2005	1.000	**0.000***	0.985	**0.000***	1.000	**0.000***
	2006	1.000	**0.000***	0.919	**0.001***	1.000	**0.000***
	2007	0.157	**0.000***	1.000	1.000	1.000	**0.000***
	2008	**0.000***	**0.000***	**0.000***	**0.000***	**0.003***	**0.000***
	2009	**0.001***	1.000	**0.000***	**0.006***	0.422	0.919
	2010	0.082	0.914	**0.000***	1.000	0.987	**0.000***
	2011	0.096	0.964	**0.000***	0.961	0.295	1.000
	2012	0.033*	0.012*	**0.000***	**0.000***	1.000	**0.000***
	2013	0.147	0.078	**0.000***	**0.006***	0.022*	0.922
	2014	1.000	0.081	**0.000***	0.963	**0.001***	**0.000***
	2015	1.000	0.201	1.000	1.000	0.169	**0.002***
	2016	0.998	0.995	0.996	0.070	1.000	1.000
	2017	1.000	1.000	1.000	0.991	1.000	**0.008***
	2018	1.000	0.922	0.836	0.978	1.000	0.992
	2019	0.943	**0.000***	0.198	0.493	0.998	0.714
	2020	**0.000***	**0.000***	0.091	**0.000***	1.000	0.998
	2021	**0.000***	**0.000***	**0.000***	**0.000***	1.000	1.000
	All	0.012*	**0.000***	**0.000***	**0.000***	**0.000***	**0.000***

Note: The model column indicates names of corresponding models; for each year, we present the p-value of backtesting results for $MaxCVaR_t^H(\alpha)$ with $H = 22$; the All row in the year column indicates the backtesting results with observations across all years between 2004 and 2021; -99% and 99% represent different α values; the bold fonts and asterisks denote p-values of $MaxCVaR_t^H(\alpha)$ backtesting results less than 0.01 and 0.05, respectively; for corn and wheat markets, the numbers of observations for $H = 22, 44, 66$ are 4,446, 4,424, and 4,402; for the soybean market, the numbers of observations are 4,445, 4,423, 4,401.

Table 4.5B: Bootstrap t-test Results for $MaxCVaR_t^H(\alpha)$ Estimations with H = 22

Model	Year	Corn		Soybean		Wheat	
		-99%	99%	-99%	99%	-99%	99%
DEJ	2004	0.168	1.000	**0.000***	0.392	1.000	1.000
	2005	0.348	0.999	1.000	**0.000***	1.000	0.998
	2006	1.000	**0.000***	1.000	0.921	1.000	**0.000***
	2007	0.633	1.000	1.000	1.000	1.000	0.675
	2008	**0.000***	**0.000***	**0.000***	0.207	0.011*	0.389
	2009	0.998	1.000	**0.000***	1.000	1.000	1.000
	2010	0.012*	0.083	1.000	0.082	1.000	**0.000***
	2011	**0.005***	1.000	0.083	0.082	0.083	1.000
	2012	1.000	**0.002***	**0.010***	**0.000***	1.000	0.054
	2013	**0.000***	1.000	0.022*	0.985	1.000	1.000
	2014	1.000	1.000	**0.000***	1.000	1.000	1.000
	2015	1.000	0.095	1.000	1.000	0.923	0.999
	2016	0.057	1.000	0.641	0.977	1.000	1.000
	2017	1.000	1.000	1.000	1.000	0.999	**0.007***
	2018	1.000	1.000	1.000	1.000	1.000	0.917
	2019	0.997	**0.000***	1.000	1.000	0.913	0.993
	2020	1.000	**0.002***	1.000	0.014*	1.000	1.000
	2021	**0.000***	**0.000***	**0.000***	0.256	1.000	0.996
	All	**0.000***	**0.000***	**0.000***	**0.000***	0.019*	**0.000***
DEJ-Seasonal	2004	0.998	1.000	**0.000***	0.015*	1.000	1.000
	2005	1.000	1.000	1.000	**0.000***	1.000	1.000
	2006	1.000	**0.000***	1.000	0.988	1.000	**0.000***
	2007	0.994	0.920	1.000	1.000	1.000	0.784
	2008	**0.000***	0.130	**0.000***	0.642	0.053	0.090
	2009	0.988	1.000	**0.009***	1.000	1.000	1.000
	2010	0.083	0.081	1.000	0.922	1.000	**0.000***
	2011	0.068	1.000	1.000	0.076	0.079	1.000
	2012	1.000	0.270	0.058	0.090	1.000	0.030*
	2013	0.998	1.000	0.173	1.000	1.000	1.000
	2014	1.000	1.000	**0.000***	1.000	0.920	0.981
	2015	1.000	0.999	1.000	1.000	1.000	0.999
	2016	0.989	1.000	0.919	0.995	1.000	1.000
	2017	1.000	1.000	1.000	1.000	1.000	0.037*
	2018	1.000	1.000	0.999	1.000	1.000	1.000
	2019	0.983	**0.003***	0.923	1.000	0.966	0.922
	2020	0.997	**0.000***	0.920	**0.000***	1.000	1.000
	2021	**0.000***	**0.000***	**0.000***	0.464	1.000	0.920
	All	**0.001***	**0.000***	**0.000***	**0.000***	0.047*	**0.000***

Note: The model column indicates names of corresponding models; for each year, we present the p-value of backtesting results for $MaxCVaR_t^H(\alpha)$ with $H = 22$; the All row in the year column indicates the backtesting results with observations across all years between 2004 and 2021; -99% and 99% represent different α values; the bold fonts and asterisks denote p-values of $MaxCVaR_t^H(\alpha)$ backtesting results less than 0.01 and 0.05, respectively; for corn and wheat markets, the numbers of observations for $H = 22, 44, 66$ are 4,446, 4,424, and 4,402; for the soybean market, the numbers of observations are 4,445, 4,423, 4,401.

Table 4.6: Funding Risks per Contract vs. Margin Requirements

Market	Margin	DEJ-Seasonal		GBM-Seasonal	
		99%	-99%	99%	-99%
Corn	$1,100	$1,282	$1,221	$807	$779
Soybean	$2,311	$2,548	$2,746	$1,512	$1,472
Wheat	$1,615	$1,811	$1,601	$1,300	$1,250

Note: The Market column indicates names of markets. All funding risks per contract are generated with respect to 1-day holding periods; the Margin column measures the daily average values of CME margin requirements per contract; columns 4 and 5 (6 and 7) measure average funding risk levels measured by DEJ-Seasonal (GBM-Seasonal) for $\alpha = 99\%$ and -99% levels; the number of observations for corn and wheat markets is 4467; for the soybean market, the number of observations is 4466.

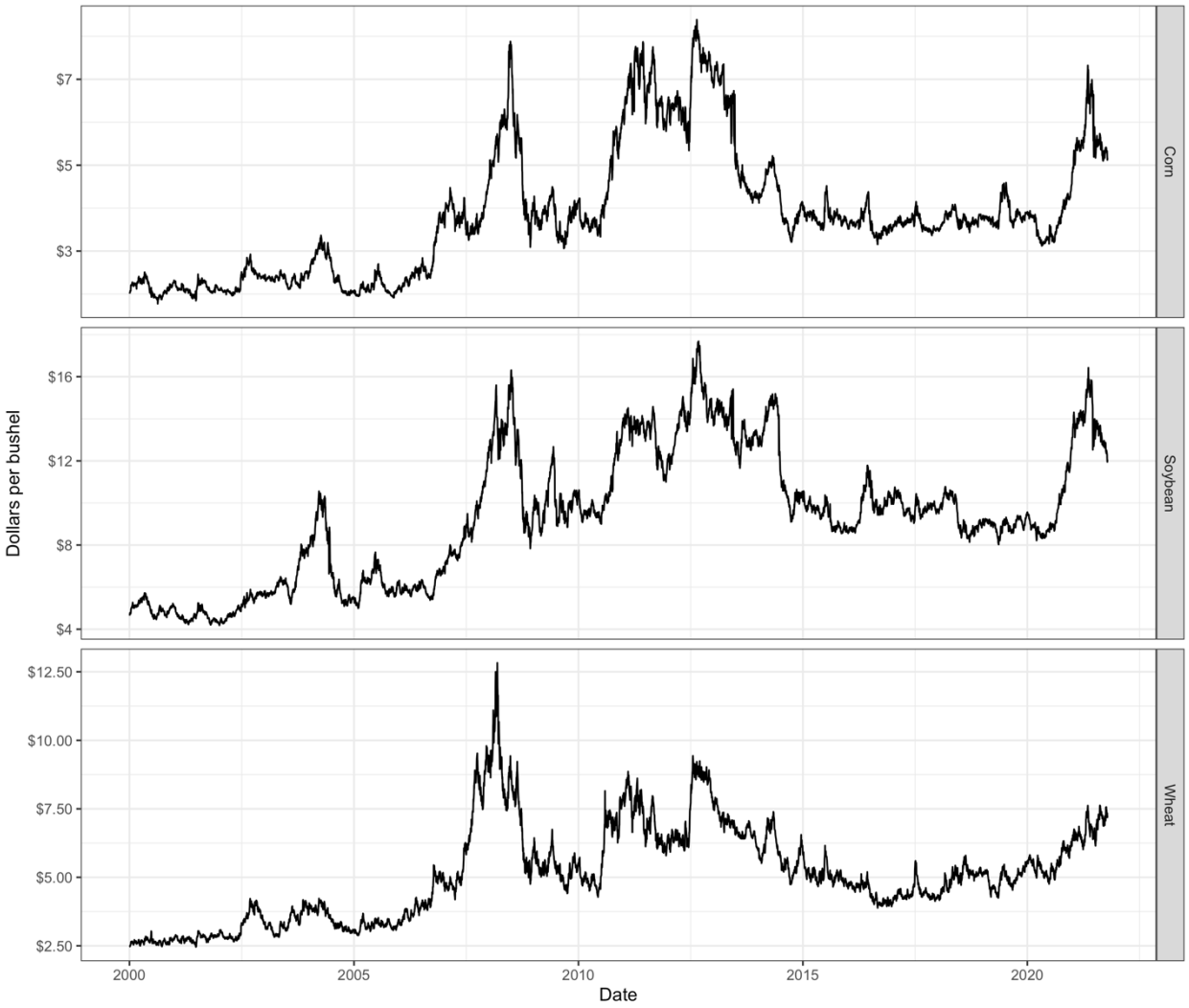

Figure 4.1. Daily corn, soybean, and wheat futures prices based on the most-traded rollover method

Note: the y-axis represents prices measured by dollars per bushel; the x-axis represents dates.

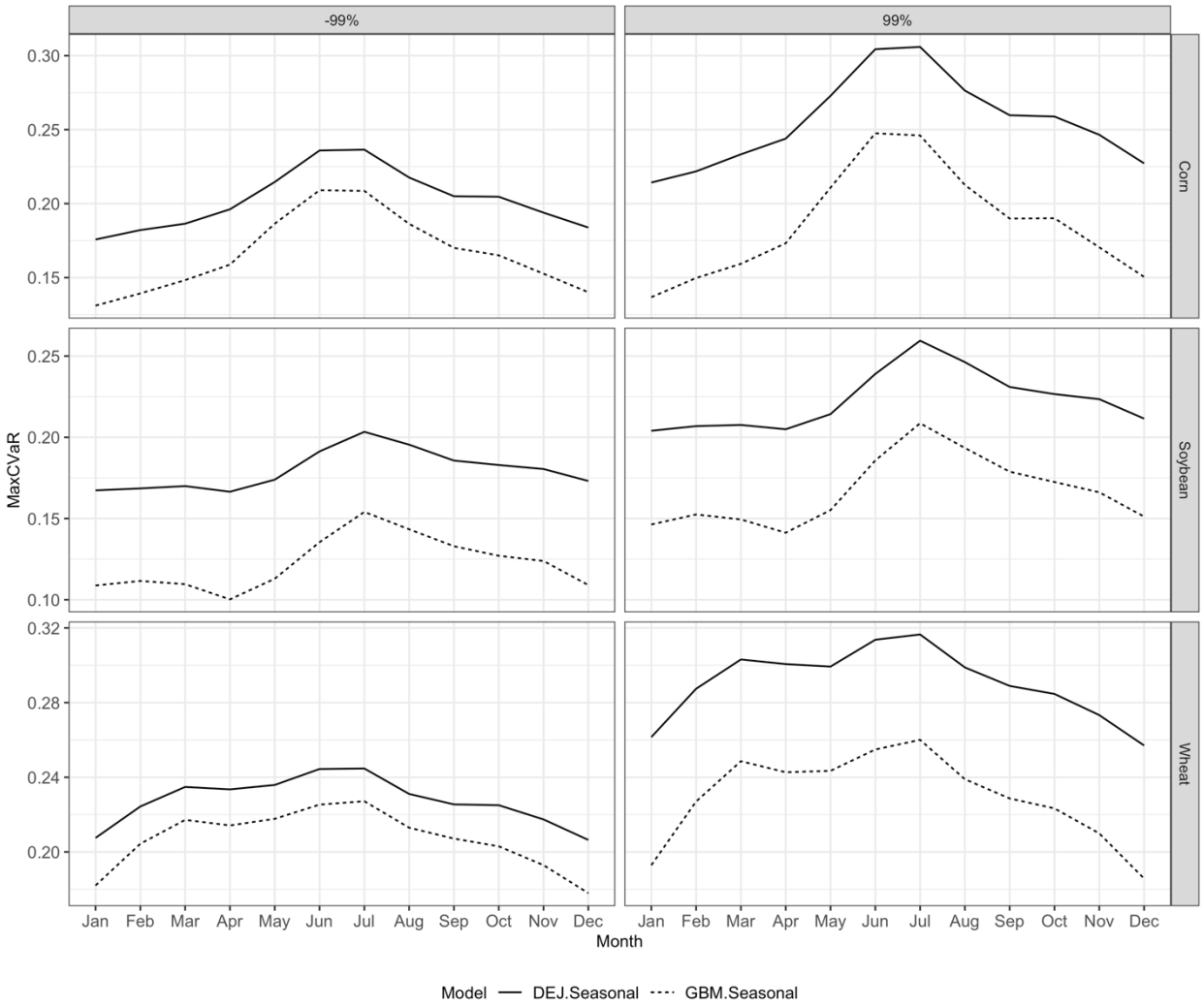

Figure 4.2. Monthly average MaxCVaR values for different models
Note: the x-axis denotes months; the y-axis depicts the absolute magnitudes of monthly average MaxCVaR values. Each line represents a model as shown in the legend at the bottom. The average number of observations per month is 1,482.

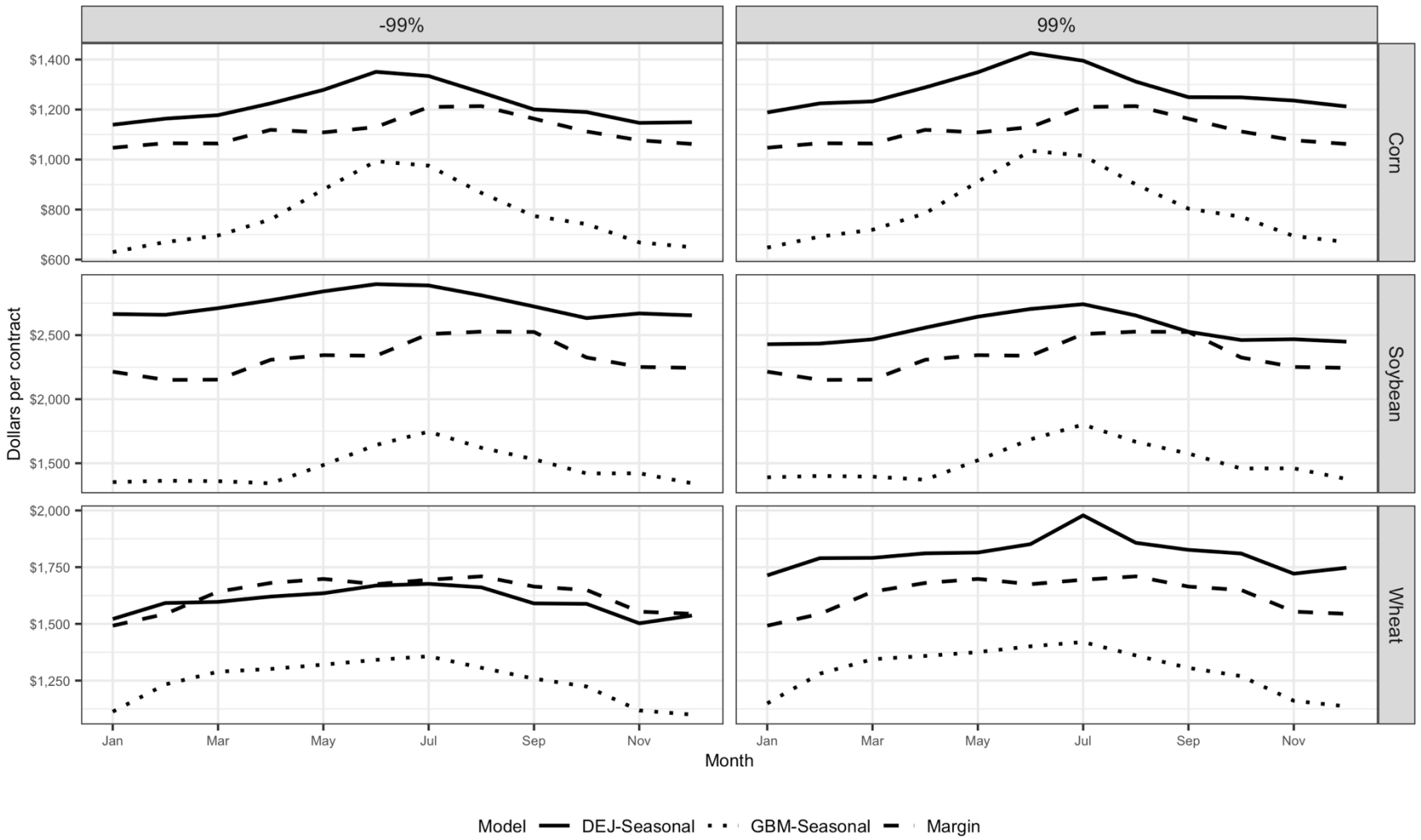

Figure 4.3. Monthly Average Withholdings per Contract vs. Margin Requirements
Note: the y-axis denotes dollars per contract; the x-axis represents different months; the solid and dotted lines represent the monthly average dollar values of $MaxCVaR^{H}(\alpha)$ per contract estimated by the DEJ-Seasonal (GBM-Seasonal) model with respect to $H = 1$ and the monthly average values of maintenance margin requirements per contract; the dashed lines represent monthly average margin requirements per contract; -99% and 99% represent different α values; The average number of observations per month is 1,482

4.6 References

Aravindhakshan, S.C., and B.W. Brorsen. 2011. "Identifying Jumps and Systematic Risk in Futures." In *NCCC - 134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis, MO.

Bakshi, G., and G. Panayotov. 2010. "First-passage probability, jump models, and intra-horizon risk." *Journal of Financial Economics* 95(1):20–40. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304405X09001500.

Boudoukh, J., M. Richardson, R. Stanton, and R.F. Whitelaw. 2004. "MaxVaR: Long-horizon Value at Risk in A Mark-to-Market Environment." *Journal of Investment Management* 2(3):1–6. Available at: https://joim.com/downloads/maxvar-long-horizon-value-risk-mark-market-environment/.

Brunnermeier, M.K., and L.H. Pedersen. 2009. "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22(6):2201–2238. Available at: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhn098.

CFTC. 2008. "Agricultural Markets Roundtable." Available at: https://www.cftc.gov/sites/default/files/idc/groups/public/@newsroom/documents/file/transcript042208.pdf.

Charupat, N., and R. Deaves. 2003. "Backwardation in Energy Futures Markets: Metallgesellschaft Revisited." *Energy Studies Review* 12(1):1–26.

Christensen, K., R.C.A. Oomen, and M. Podolskij. 2014. "Fact or friction: Jumps at ultra high frequency." *Journal of Financial Economics* 114(3):576–599. Available at: http://dx.doi.org/10.1016/j.jfineco.2014.07.007.

CME. 2019. "CME SPAN - Standard Portfolio Analysis of Risk." Available at:

https://www.cmegroup.com/clearing/files/span-methodology.pdf.

CME. 2022. “Futures and Options Margin Model.” Available at: https://www.cmegroup.com/clearing/risk-management/futures-and-options-margin-model.html.

Couleau, A., T. Serra, and P. Garcia. 2020. “Are Corn Futures Prices Getting ‘Jumpy’?” *American Journal of Agricultural Economics* 102(2):569–588.

Dahlgran, R.A. 2005. “Transaction frequency and hedging in commodity processing.” *Journal of Agricultural and Resource Economics* 30(3):411–430.

Day, T.E., and C.M. Lewis. 2004. “Margin Adequacy and Standards: An Analysis of the Crude Oil Futures Market.” *The Journal of Business* 77(1):101–135. Available at: https://www.jstor.org/stable/10.1086/379863.

Efron, B., and R.J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Chapman and Hall/CRC. Available at: https://www.taylorfrancis.com/books/9781000064988.

Farkas, W., L. Mathys, and N. Vasiljević. 2021. “Intra-Horizon expected shortfall and risk structure in models with jumps.” *Mathematical Finance* 31(2):772–823. Available at: https://onlinelibrary.wiley.com/doi/10.1111/mafi.12302.

Fishe, R.P.H., and L.G. Goldberg. 1986. “The effects of margins on trading in futures markets.” *Journal of Futures Markets* 6(2):261–271. Available at: https://onlinelibrary.wiley.com/doi/10.1002/fut.3990060208.

Hilliard, J.E., and J.A. Reis. 1999. “Jump Processes in Commodity Futures Prices and Options Pricing.” *American Journal of Agricultural Economics* 81(2):273–286.

Karali, B., and G.J. Power. 2013. “Short- and Long-Run Determinants of Commodity Price Volatility.” *American Journal of Agricultural Economics* 95(3):724–738.

Koekebakker, S., and G. Lien. 2004. “Volatility and Price Jumps in Agricultural Futures Prices - Evidence From Options.” *American Journal of Agricultural Economics* 86(4):1018–1031.

Kou, S.G. 2002. “A Jump-Diffusion Model for Option Pricing.” *Management Science* 48(8):1086–1101. Available at: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.48.8.1086.166.

Kou, S.G., and H. Wang. 2004. “Option Pricing Under a Double Exponential Jump Diffusion Model.” *Management Science* 50(9):1178–1192. Available at: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1030.0163.

Leippold, M., and N. Vasiljević. 2020. “Option-Implied Intrahorizon Value at Risk.” *Management Science* 66(1):397–414. Available at: http://pubsonline.informs.org/doi/10.1287/mnsc.2018.3157.

Lien, D. 2003. “The effect of liquidity constraints on futures hedging.” *Journal of Futures Markets* 23(6):603–613. Available at: https://onlinelibrary.wiley.com/doi/10.1002/fut.10075.

McNeil, A.J., and R. Frey. 2000. “Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach.” *Journal of Empirical Finance* 7(3–4):271–300. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0927539800000128.

Morgan, W., J. Cotter, and K. Dowd. 2012. “Extreme Measures of Agricultural Financial Risk.” *Journal of Agricultural Economics* 63(1):65–82. Available at: https://onlinelibrary.wiley.com/doi/10.1111/j.1477-9552.2011.00322.x.

Paschke, R., M. Prokopczuk, and C. Wese Simen. 2020. “Curve momentum.” *Journal of Banking & Finance* 113:105718. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0378426619302912.

Righi, M.B., and P.S. Ceretta. 2015. “A comparison of Expected Shortfall estimation models.” *Journal of Economics and Business* 78:14–47. Available at: https://linkinghub.elsevier.com/retrieve/pii/S014861951400068X.

Rockafellar, R.T., and S. Uryasev. 2000. “Optimization of conditional value-at-risk.” *The Journal of Risk* 2(3):21–41. Available at: http://www.risk.net/journal-of-risk/technical-paper/2161159/optimization-conditional-value-risk.

Rossello, D. 2008. “MaxVaR with non-Gaussian distributed returns.” *European Journal of Operational Research* 189(1):159–171. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0377221707004948.

Shi, R., and O. Isengildina-Massa. 2021. “Costs of Futtures Hedging in Corn and Soybean Markets.” *Journal of Agricultural and Resource Economics* Prepint. Available at: https://ageconsearch.umn.edu/record/313311/files/JARE47.2.10%2CShi%2C390-409S.pdf?version=1.

Sørensen, C. 2002. “Modeling seasonality in agricultural commodity futures.” *Journal of Futures Markets* 22(5):393–426. Available at: https://onlinelibrary.wiley.com/doi/10.1002/fut.10017.

Westgaard, S., S. Frydenberg, and S.K. Mohanty. 2021. “Fourteen large commodity trading disasters: What happened and what can we learn?” *Journal of Commodity Markets*:100221. Available at: https://linkinghub.elsevier.com/retrieve/pii/S2405851321000544.

Xouridas, S. 2015. “Agricultural Financial Risks Resulting from Extreme Events.” *Journal of Agricultural Economics* 66(1):192–220. Available at: https://onlinelibrary.wiley.com/doi/10.1111/1477-9552.12083.

CHAPTER 5

CONCLUSIONS

Agricultural commodity markets have experienced several volatile periods in the past two decades. This dissertation focuses on providing a better understanding of market volatility by investigating three issues related to price volatility that have not been examined in previous literature. These issues are: (1) responses of information and noise components of volatility to USDA announcements in grains futures markets and their implications for the release policy; (2) the ability of state-of-the-art machine learning techniques to forecast volatility and the relative importance of linear and non-linear features in volatility forecasting; and (3) the impact of jump volatility and seasonality volatility on the liquidity pressure faced by agricultural futures market participants and their implications for margin requirements set by the exchange. The dissertation consists of three chapters that shed light on each of issues.

In the first chapter, we examine the behavior of efficient and noise volatility around USDA announcements under different release policies in corn and soybean futures markets. Using the method by Hansen (2015), we employ a Markov Chain (MC) framework that decomposes observed variance into its efficient variance and microstructure noise. We find that the two components respond differently to real-time versus trading halt release of USDA grains reports. During the period when reports were released during a trading halt, USDA releases for corn and soybeans resulted in an average 1-2% increase in the relative importance of efficient to observed variance between non-announcement and announcement days. In the real-time period, the increase is much larger, reaching an average 10-16%. Our results also demonstrate the importance of separating noise from efficient price variance when comparing the time to absorb a shock in

information. Our findings suggest that the real-time either does not improve absorption speed or implies time savings on the order of 10-41% for the 90% absorption rate of efficient variance in corn and soybean markets. Noting that public information releases occur at a very low frequency (e.g. monthly, quarterly or annually), it is difficult to identify how the relatively modest improvements in speed compensate for the increases in costs of microstructure frictions which are on the order of 300% higher in real time.

The second essay provides a systematic comparison of the forecasting performance of linear HAR-type and nonlinear machine learning (ML) models. We adopt a wide range of popular ML algorithms to forecast noise-free daily realized volatility (RV) in highly traded futures markets including corn, soybeans, wheat, crude oil and S&P 500 E-mini observed from January 2009 through May 2018. Besides off-the-shelf ML models, we also design a novel hybrid Artificial Neural Network (ANN) model based on residual learning which allows us to separate forecasts into linear and nonlinear components. The main finding suggests that the best out-of-sample forecast performance measures are usually achieved by linear models, especially by the simple HAR model. In contrast, the worst performance measures are attributed to nonlinear models, especially support vector regression (SVR). Furthermore, the decomposition results of the hybrid ANNs suggest that the nonlinear component represents on average, less than 3.27% of the overall hybrid model volatility forecast.

The third essay assesses funding risks faced by agricultural futures traders given asymmetric price jumps, and seasonal volatility patterns. We extend the MaxCVaR model by Farkas, Mathys and Vasiljević (2021) to allow for seasonal volatility and measure funding risks in agricultural futures markets for the first time. We apply the model to estimate funding risk in corn, soybean, and wheat futures markets. Our findings suggest that funding risks impose significant financial

burden on market participants. For example, holding a long (short) position over a month in the corn futures market results in 20.3% (25.6%) losses in the lower (upper) 1-percentile extremes of the distribution. Jumps alone significantly increase MaxCVaR values by up to 46% in agricultural futures markets. Seasonal volatility components alter the temporal distribution of the funding risk across different months instead of increasing average MaxCVaR levels. Using the backtesting method developed by Righi and Ceretta (2015), we show that MaxCVaR values estimated with jumps and seasonal volatility components over the past 1,000 days are significantly useful in anticipating extreme price movements and thus funding risks in all three markets.

The thesis findings have relevant policy implications. Results in the first essay find that relative to the trading-halt, the price discovery costs in the form of microstructure frictions are large, and not likely offset by possible savings in the speed of price discovery. Our results are consistent with calls to modify public information trading mechanisms (Budish et al. 2014). The findings also indicate that the fixed tick size may be one of the reasons why microstructure noise is so relevant in these markets and research may be needed to address in more depth the appropriate tick size in these markets. Findings in the second essay help researchers and investors make more informed decisions when searching for better volatility forecasting models in futures markets. Instead of considering more flexible nonlinear models to fit the ever-changing volatility process, researchers can boost forecast accuracy by finding linear specifications that can efficiently capture long memory in volatility such as HAR models. According to the findings in the third essay, the funding risk provides an important tool to better understand market liquidity conditions under stressful conditions. Specifically, the share of jumps in funding risks can inform policy makers of the magnitude of liquidity pressure induced by unexpected market events, which is considered as a measure of market fragility (Brunnermeier and Pedersen 2009). For exchanges, the funding risk

measured as the MaxCVaR method estimated by the DEJ-Seasonal model, is a reliable tool to help them determine the necessary margin levels to protect clearinghouses from excess losses during volatile periods, especially when facing frequent price jumps.

5.1 References

Brunnermeier, M.K., and L.H. Pedersen. 2009. "Market Liquidity and Funding Liquidity." *Review of Financial Studies* 22(6):2201–2238. Available at: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhn098.

Budish, E., P. Cramton, and J. Shim. 2014. "Implementation Details for Frequent Batch Auctions: Slowing Down Markets to the Blink of an Eye." *American Economic Review* 104(5):418–424.

Farkas, W., L. Mathys, and N. Vasiljević. 2021. "Intra-Horizon expected shortfall and risk structure in models with jumps." *Mathematical Finance* 31(2):772–823. Available at: https://onlinelibrary.wiley.com/doi/10.1111/mafi.12302.

Hansen, P.R. 2015. "A Martingale Decomposition of Discrete Markov chains." *Economics Letters* 133:14–18. Available at: http://dx.doi.org/10.1016/j.econlet.2015.04.028.

Righi, M.B., and P.S. Ceretta. 2015. "A comparison of Expected Shortfall estimation models." *Journal of Economics and Business* 78:14–47. Available at: https://linkinghub.elsevier.com/retrieve/pii/S014861951400068X.

APPENDIX A

SUPPLEMENTARY RESULTS ON HIGH-FREQEUCNY VARIANCE ESTIMATORS

A.1 The effectiveness of the MC-based variance estimator in measuring relevant variances

In this appendix, we examine the effectiveness of the MC estimator in measuring relevant variances. Hansen and Lunde (2006) suggest that realized variance (RV)[28] based on prices sampled every 300 seconds (RV_{300}) can be used as a proxy for the efficient price variance. In contrast, RV based on prices sampled at higher sampling frequency (such as 1 second - RV_1) can only be regarded as observed variance. For each announcement and non-announcement day, we compute these two RV measures using the data for the entire day, which ensures a sufficient number of observations when sampling every 300 seconds. We then compare the MC-based variance measures to their RV counterparts:$Var(\Delta Y_t)$ to RV_{300}, and $Var(\Delta X_t)$ to RV_1. Fig. B.1 and B.2. present the results on announcement and non-announcement days for corn and soybeans, respectively.

$Var(\Delta X_t)$ and RV_1 are virtually identical for both announcement and non-announcement days. $Var(\Delta Y_t)$ and RV_{300} are also very similar, but differences emerge, especially during variance spikes. This suggests that the use of a RV_{300} sampling technique makes it difficult to capture the efficient variance during turbulent periods, and it also may lead to identifying false spikes due to the discontinuities in the data created by resampling every 300 seconds. Consistent with this

[28] RV is defined as $RV_s = \sum_{i=1}^{m} r_i^2$ where m is the number of returns over a time period $t \in [a, b]$ and s is the length of an equidistant time interval such that $s = (b-a)/m$. The intraday return is $r_i = \log(X_{t_i}) - \log(X_{t_{i-1}})$, where $t_i = t_{i-1} + s$. If X_{t_i} is missing, we replace X_{t_i} with the last available observation.

hypothesis, we find the differences between $Var(\Delta Y_t)$ and RV_{300} are much smaller on non-announcement days when variance spikes are less frequent.

A.2 The economic benefits of variance decomposition in agricultural commodity markets

This appendix illustrates how individual traders can benefit from using efficient price returns variance as opposed to observed price returns variance in their investment decisions. Variance constitutes a primary risk measure in modern finance driving several key decisions such as hedging, portfolio management or asset pricing. Here, we study the investor's utility derived from two types of portfolio, one based on efficient and the other on observed variance.

We assume a risk-averse investor with a mean-variance utility function who invests a fraction of her wealth on a risky asset (corn or soybeans futures) and the rest on a risk-free asset (3-month Treasury Bill), based on the strategy proposed by Fleming, Kirby and Ostdiek (2003) and Bandi and Russell (2006). All portfolios rely on the same asset return expectations, so that the derived utility exclusively depends on the investor's ability to predict future (efficient and observed) variance based on past variance. The difference in investor's utility between the two portfolios represents her willingness-to-pay for non-observable efficient variance estimates. We assume the portfolio can be rebalanced daily and thus the relevant time interval is one day.

The investor's conditional mean-variance utility function is as follows:

$$E_t\left(R^p_{t,t+1}\right) - \frac{\lambda}{2} V_t\left(R^p_{t,t+1}\right) \qquad \text{(A. 1)}$$

where the constant λ measures the investor's degree of risk aversion, with a larger λ indicating higher risk-aversion. $R^p_{t,t+1}$ is the return on a portfolio that invests a share (ρ_t) of wealth in the risky asset (i.e. corn or soybeans futures) between day *t* and *t* +1 and can be expressed as follows:

$$R^p_{t,t+1} = R^f_{t,t+1} + \rho_t\left(R_{t,t+1} - R^f_{t,t+1}\right) \qquad \text{(A. 2)}$$

where $\rho_t = (E_t(R_{t,t+1} - R^f_{t,t+1}))/(\lambda V_t(R_{t,t+1}))$. $R_{t,t+1}$ and $R^f_{t,t+1}$ are returns on the risky asset and a risk-free asset, respectively. $E_t(\cdot)$ and $V_t(\cdot)$ are expectations of future returns and variance at time *t*. To solely focus on alternate variance estimates, we set $E_t(R^f_{t,t+1}) = \frac{1}{m}\sum_{t=1}^{m} R^f_{t,t+1}$ and $E_t(R_{t,t+1}) = \frac{1}{m}\sum_{t=1}^{m} R_{t,t+1}$, which are respectively the unconditional average of daily risk-free interest rates and futures price returns over the forecasting period of *m* days.

To forecast the daily conditional variance $V_t(R_{t,t+1})$, we use the heterogeneous autoregression (HAR) model proposed by Corsi (2009) based on intraday data and a rolling estimation window of *n* = 261 days. We produce the forecasts using both intraday efficient and observed price return variances, which allows us to assess which variance measure has the largest predictive power of its own future path. The HAR based on efficient price returns is built as follows:

$$Var(\Delta Y_{t+1}) = \alpha + \beta_D \cdot Var(\Delta Y_t) + \beta_W \cdot \overline{Var(\Delta Y_t)}_W + \beta_M \cdot \overline{Var(\Delta Y_t)}_M + I_{ann,t} + \epsilon_t \quad (A.3)$$

where $\overline{Var(\Delta Y_t)}_W = \frac{1}{5}\sum_{d=4}^{0} Var(\Delta Y_{t-d})$ and $\overline{Var(\Delta Y_t)}_M = \frac{1}{22}\sum_{d=21}^{0} Var(\Delta Y_{t-d})$, with *d* denoting the number of lags. $I_{ann,t}$ is a dummy variable that takes the value of 1 if a USDA announcement is released on day *t*. The HAR model based on observed price returns is:

$$Var(\Delta X_{t+1}) = \alpha + \beta_D \cdot Var(\Delta X_t) + \beta_W \cdot \overline{Var(\Delta X_t)}_W + \beta_M \cdot \overline{Var(\Delta X_t)}_M + I_{ann,t} + \epsilon_t \quad (A.4)$$

The HAR models are estimated via OLS and are more parsimonious than other methods, such as the ARFIMA model used in Bandi and Russell (2006).

Notice that the daily observed and efficient variances are based on market trading hours. However, portfolio returns (A.2) are based on daily return variances. We thus apply the bias-correction measure used by Bandi and Russell (2006) and multiply the MC-based variance estimates by a factor that is defined as:

$$\xi_Y = \frac{\frac{1}{n}\sum_{t=1}^{n} R_{t,t+1}^2}{\frac{1}{n}\sum_{t=1}^{n} Var(\Delta Y_t)} \tag{A.5}$$

and

$$\xi_X = \frac{\frac{1}{n}\sum_{t=1}^{n} R_{t,t+1}^2}{\frac{1}{n}\sum_{t=1}^{n} Var(\Delta X_t)} \tag{A.6}$$

with *n* representing the number of days in the rolling window. The bias-corrected conditional variance measures, $\xi_Y Var(\Delta Y_{t+1})$ and $\xi_X Var(\Delta X_{t+1})$ then coincide with the variance of the daily returns.

We then estimate the optimal share of total wealth ρ_t to be invested in corn and soybean futures. We evaluate ρ_t for $t = 1, \dots, m$, where $m = N - n$, with *N* being the sample size. We follow Bandi and Russell (2006) and consider three conventional values (2, 7, and 10) for the parameter of risk aversion (λ). Thus, the estimated shares of the risky asset ρ_t at time *t* are:

$$\rho_t^Y = \left(E_t\left(R_{t,t+1} - R_{t,t+1}^f\right)\right) / \left(\lambda \xi_Y Var(\Delta Y_{t+1})\right) \tag{A.7}$$

and

$$\rho_t^X = \left(E_t\left(R_{t,t+1} - R_{t,t+1}^f\right)\right) / \left(\lambda \xi_X Var(\Delta X_{t+1})\right) \tag{A.8}$$

The corresponding portfolio returns are:

$$R_{t,t+1}^{p,i} = R_{t,t+1}^f + \rho_t^i\left(R_{t,t+1} - R_{t,t+1}^f\right) \tag{A.9}$$

where $i = X, Y$. Therefore, the investor's long-run utility on each portfolio is:

$$AU^i = \overline{R^{p,i}} - \frac{\lambda}{2} \cdot \frac{1}{m}\sum_{t=1}^{m}\left(R_{t,t+1}^{p,i} - \overline{R^{p,i}}\right)^2 \tag{A.10}$$

where $\overline{R^{p,i}} = \frac{1}{m}\sum_{t=1}^{m}\left(R_{t,t+1}^{p,i}\right)$. The difference between two portfolios ($AU^Y - AU^X$) indicates the investor's willingness-to-pay for the unobservable efficient variance data. For better illustration, we express the difference in basis points.

To empirically estimate the willingness to pay for efficient variance data we produce daily estimates of efficient and observed variances based on the MC estimators. The intraday seasonality of variance may pose challenges for the MC homogeneity property. Hence, we set the MC lag order (k) to $k = 6$ because most of the daily MC estimates stabilize after $k = 5$. As shown in Table C.1, portfolios based on efficient variances yield better performance than portfolios based on observed variances in both the corn and soybean markets. This is an indicator that efficient price return variance is easier to predict than the observed price variance affected by microstructure noise. In the corn market, the willingness to pay for information on efficient variances ranges from 74 basis points for the most risk averse investors ($\lambda = 10$) to 372 basis points for the least risk averse ($\lambda = 2$). Similarly, in the soybean market, the willingness to pay for efficient variance information ranges from 59 basis points ($\lambda = 10$) to 295 basis points ($\lambda = 2$). The smaller values in the soybean market compared to the corn market are anticipated because observed variances in the soybean market contain less microstructure noise than observed variances in the corn market. Thus, investors in the soybean market gain less by switching from observed variances to efficient variances.

A.3 Table and Figures

Table A.1 Differences Between Investor's Utility (in Basis Points) Derived from Portfolios Based on Efficient and Observed Variances

λ	2	7	10
Corn	372	106	74
Soybeans	295	84	59

Note: The number of observations is 907; λ is the risk aversion measure; numerical values in the second and third rows are the differences in utilities between portfolios based on efficient and observed variances in each market measured in basis points; 1 basis point = 0.01%.

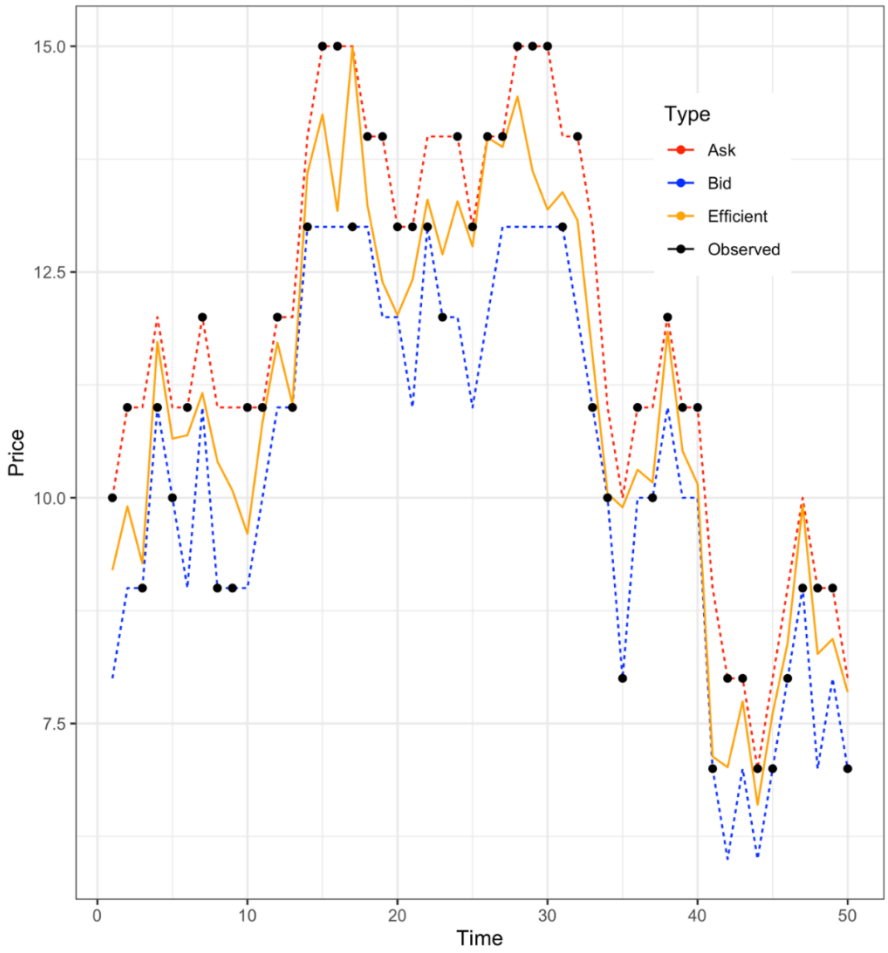

Figure A.1 Simulated Efficient and Observed Prices.
Note: The y-axis measures price levels and the x-axis measures time (t) in seconds. The blue and red dashed lines represent bid and ask prices. The orange solid line depicts unobserved efficient prices (Y_t). The black dots represent observed transaction prices (X_t). This figure helps to understand and appreciate the value of the proposed method. Bid, ask and transaction prices have a discrete nature and are observed from market data. The transaction price coincides with either the bid or the ask price, depending on whether the transaction aggressor is a seller or a buyer. The continuous efficient price is not observable and is free from microstructure frictions. The variance based on the unobserved efficient price reflects the true market turbulence in the absence of noise. The latter is derived through the Markov Chain (MC) variance estimator proposed in Hansen et al. (2015) and Hansen (2015) to assess corn and soybean futures price variance every minute.

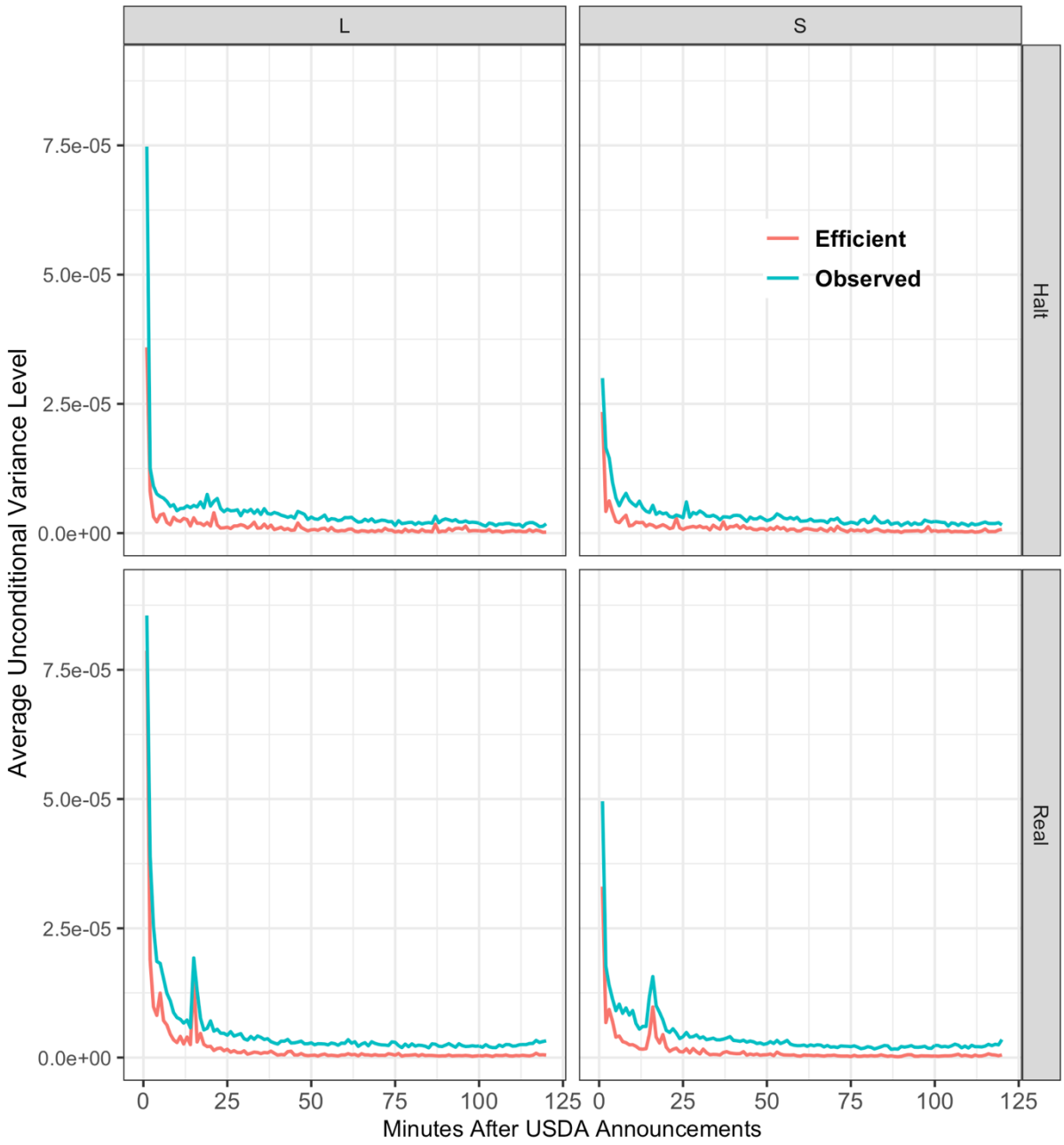

Figure A.2 USDA Announcement Day Unconditional Variance in the Corn Market.
Note: The y-axis measures levels of unconditional variance. The x-axis measures numbers of minutes after news releases. Observed and Efficient represent the 1-minute unconditional variance estimations for observed ($Var(\Delta X_t)$) and efficient returns ($Var(\Delta Y_t)$). Large shock ("L") represents greater than the median shocks and small shock ("S") represents less than or equal to the median shocks. Real and Halt represent real-time and halt release periods.

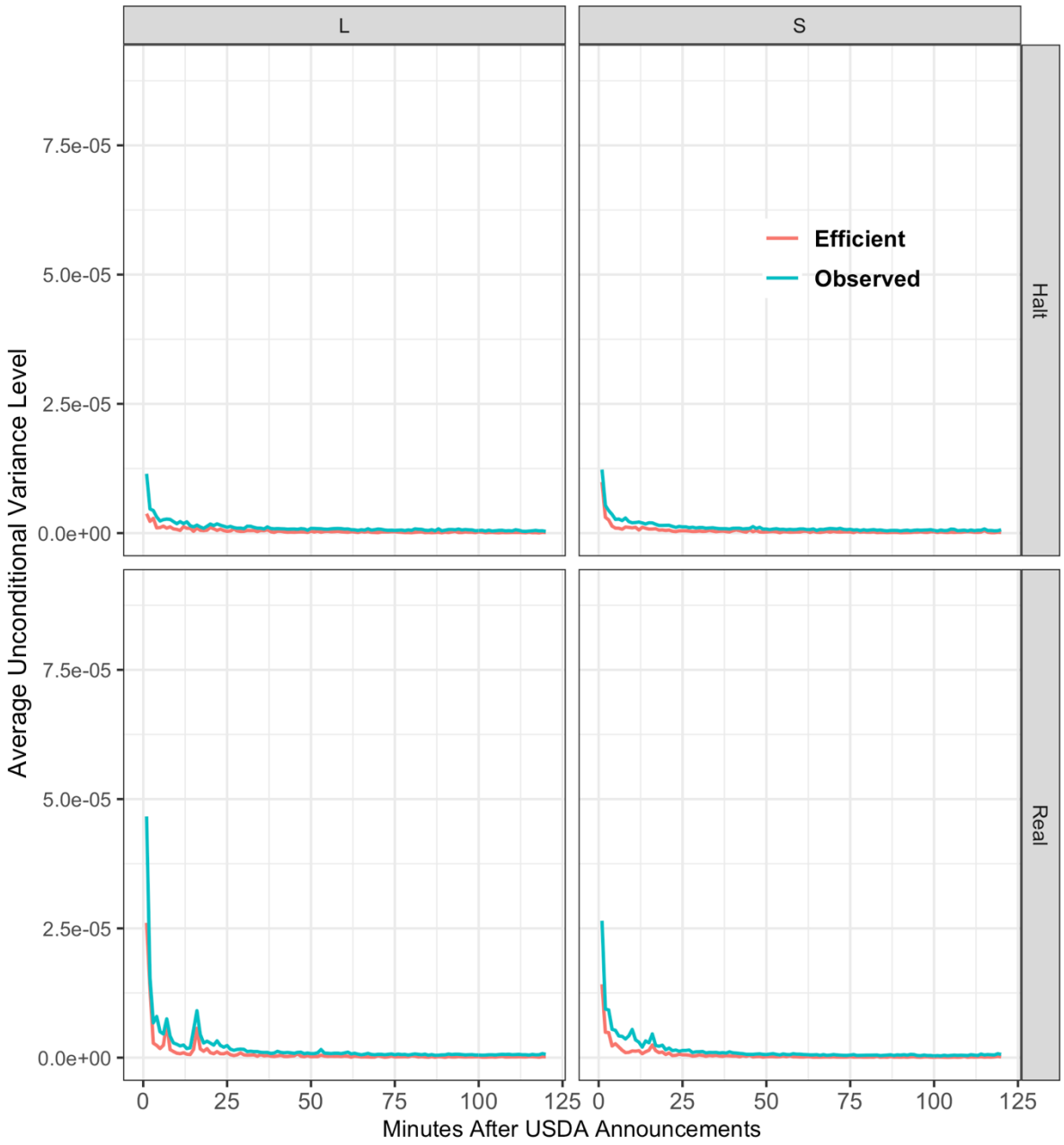

Figure A.3 USDA Announcement Day Unconditional Variance in the Soybean Market.
Note: The y-axis measures levels of unconditional variance. The x-axis measures numbers of minutes after news releases. Observed and Efficient represent the 1-minute unconditional variance estimations for observed ($Var(\Delta X_t)$) and efficient returns ($Var(\Delta Y_t)$). Large shock ("L") represents greater than the median shocks and small shock ("S") represents less than or equal to the median shocks. Real and Halt represent real-time and halt release periods.

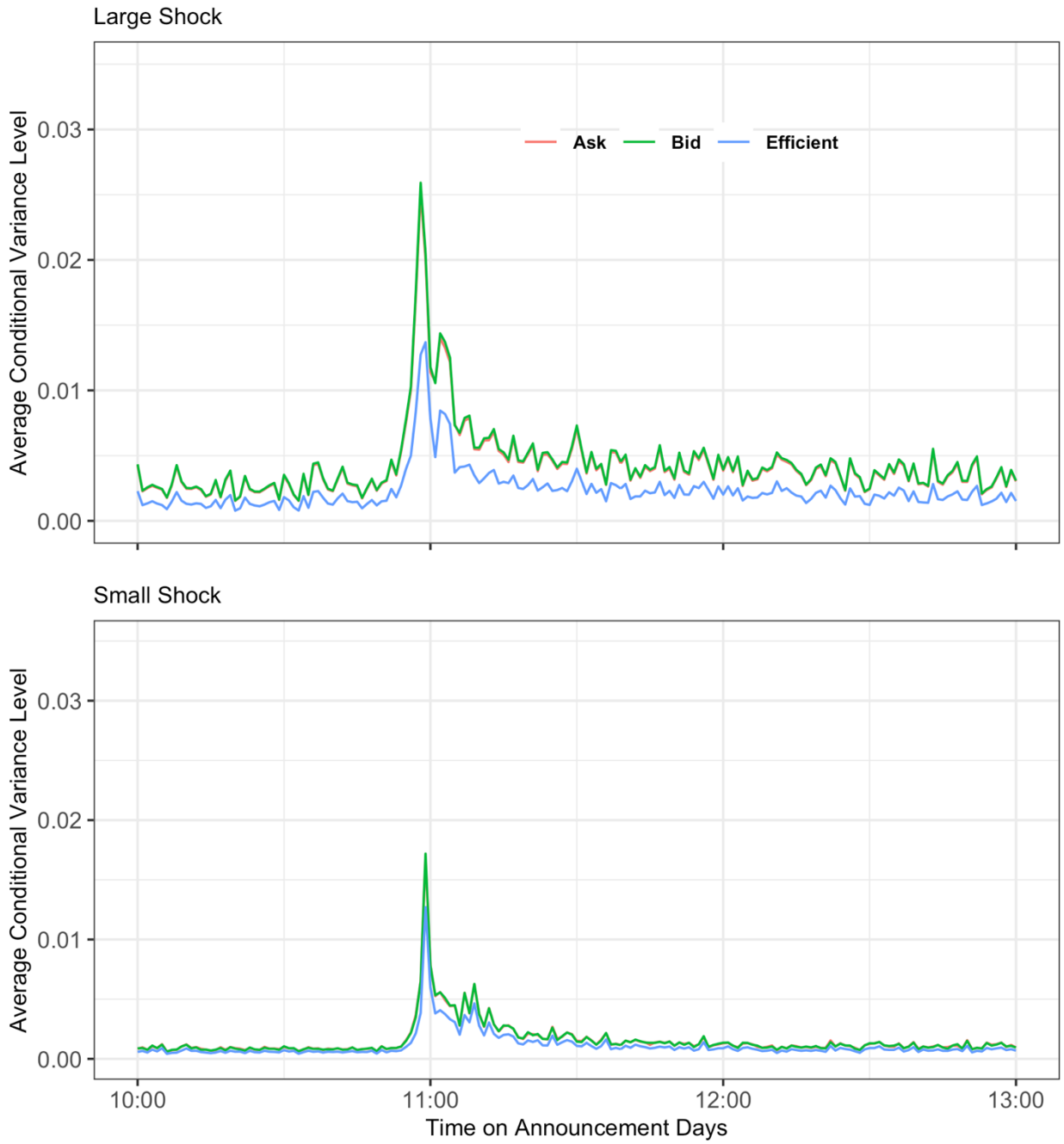

Figure A.4 USDA Announcement Day Conditional Variance in the Corn Market Estimated by the State-Space VAR-EGARCH Model.
Note: The y-axis measures levels of conditional variance, and x-axis time of day. Ask, Bid, and Efficient represent the 1-minute conditional variance estimations for ask quotes, bid quotes and efficient price returns. Bid and Ask conditional variance estimations are identical most of the time. Large shock represents greater than the median shocks and small shock represents less than or equal to the median shocks.

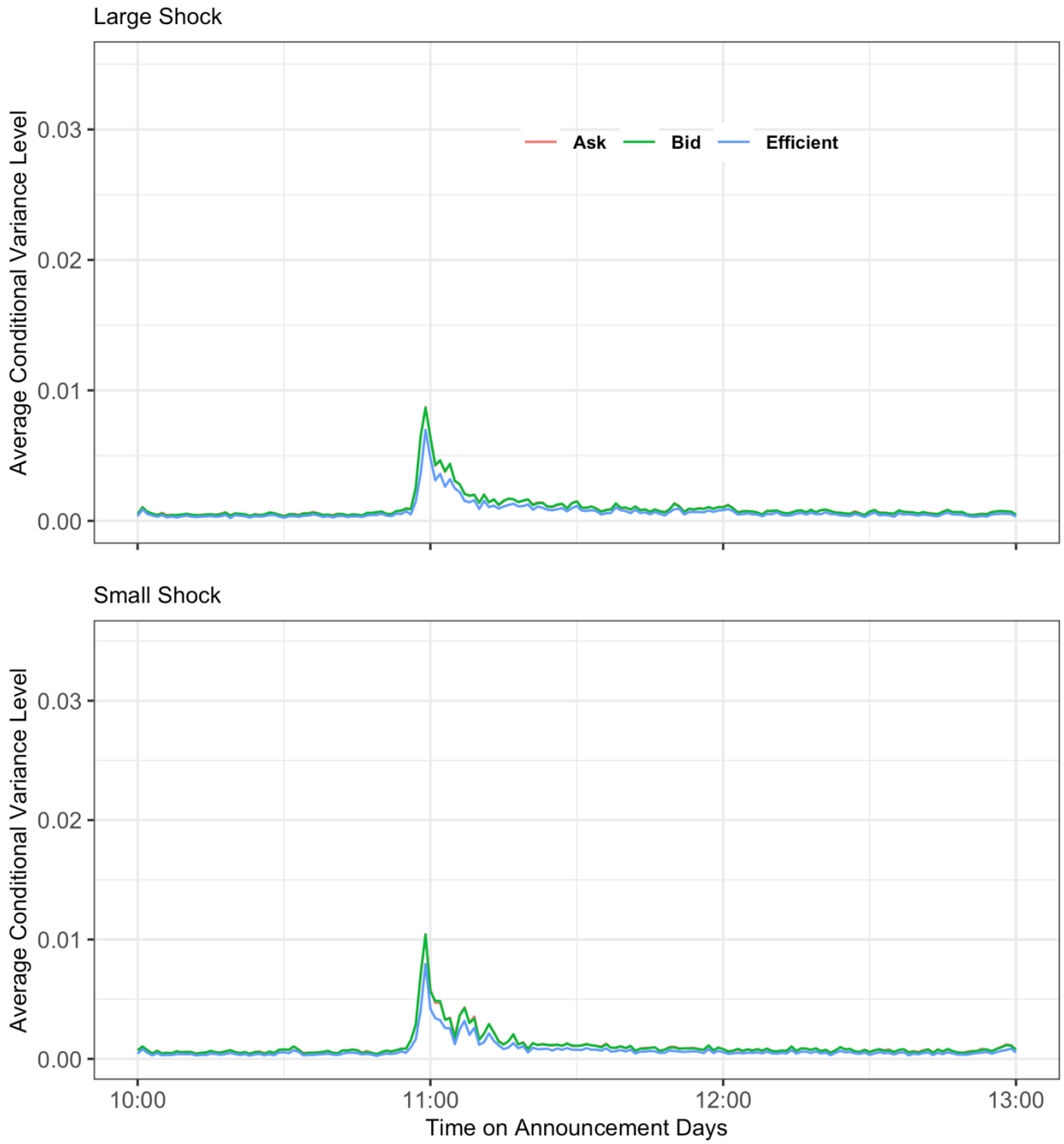

Figure A.5 USDA Announcement Day Conditional Variance in the Soybean Market Estimated by the State-Space VAR-EGARCH Model.
Note: The y-axis measures levels of conditional variance, and x-axis time of day. Ask, Bid, and Efficient represent the 1-minute conditional variance estimations for ask quotes, bid quotes and efficient price returns. Bid and Ask conditional variance estimations are identical most of the time. Large shock represents greater than the median shocks and small shock represents less than or equal to the median shocks.

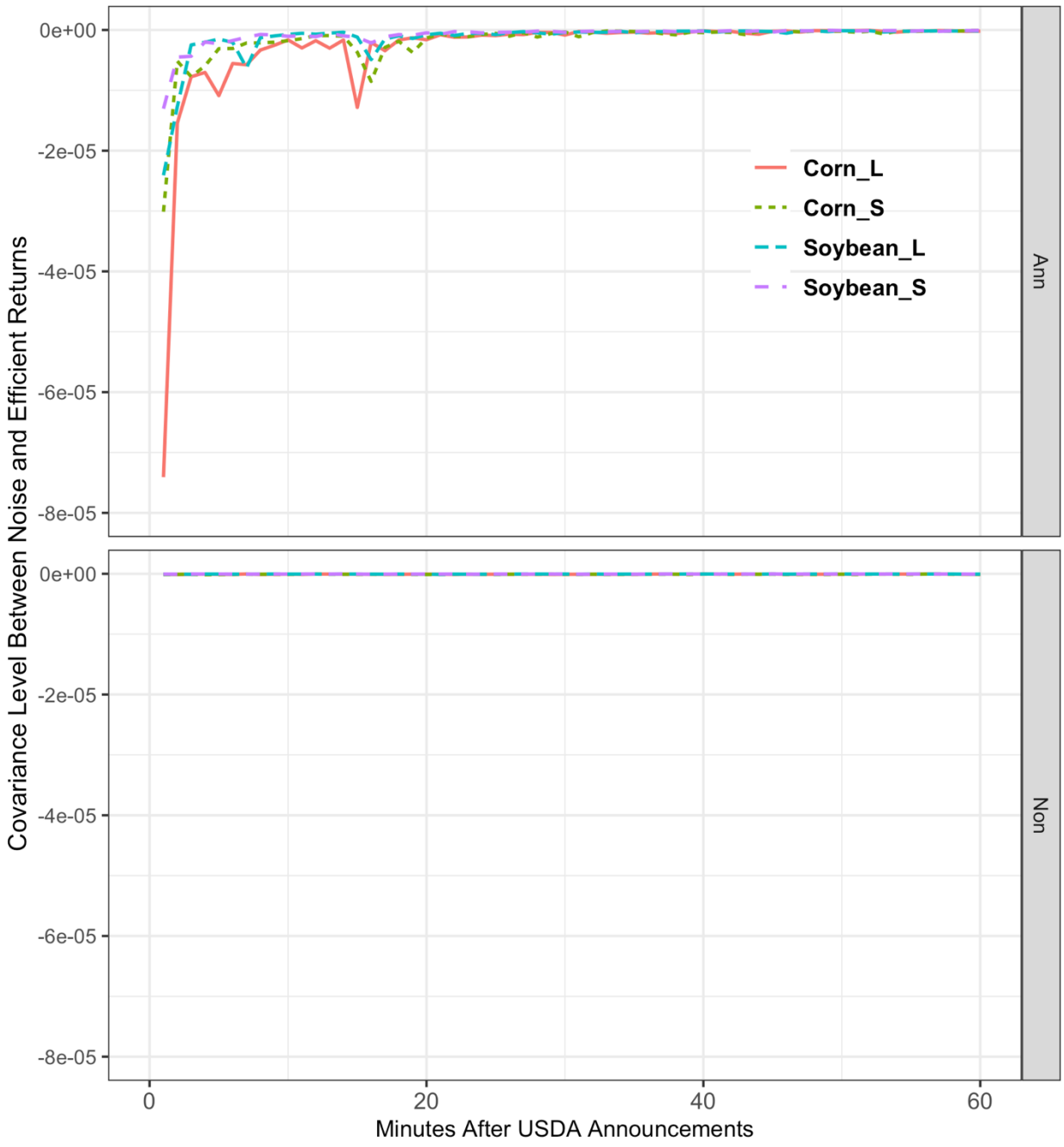

Figure A.6 Covariance Estimation Results Between Efficient Return and Noise on Announcement (Ann) and Non-Announcement (Non) Days.
Note: The y-axis measures levels of covariance and x-axis the minutes after USDA news releases. All covariances are estimated based on MC methods using transaction price returns. All measures are calculated at 1-minute frequency. Corn and soybean represent estimations in corn and soybean market. Large shock ("L") represents greater than the median shocks and small shock ("S") represents less than or equal to the median shocks.

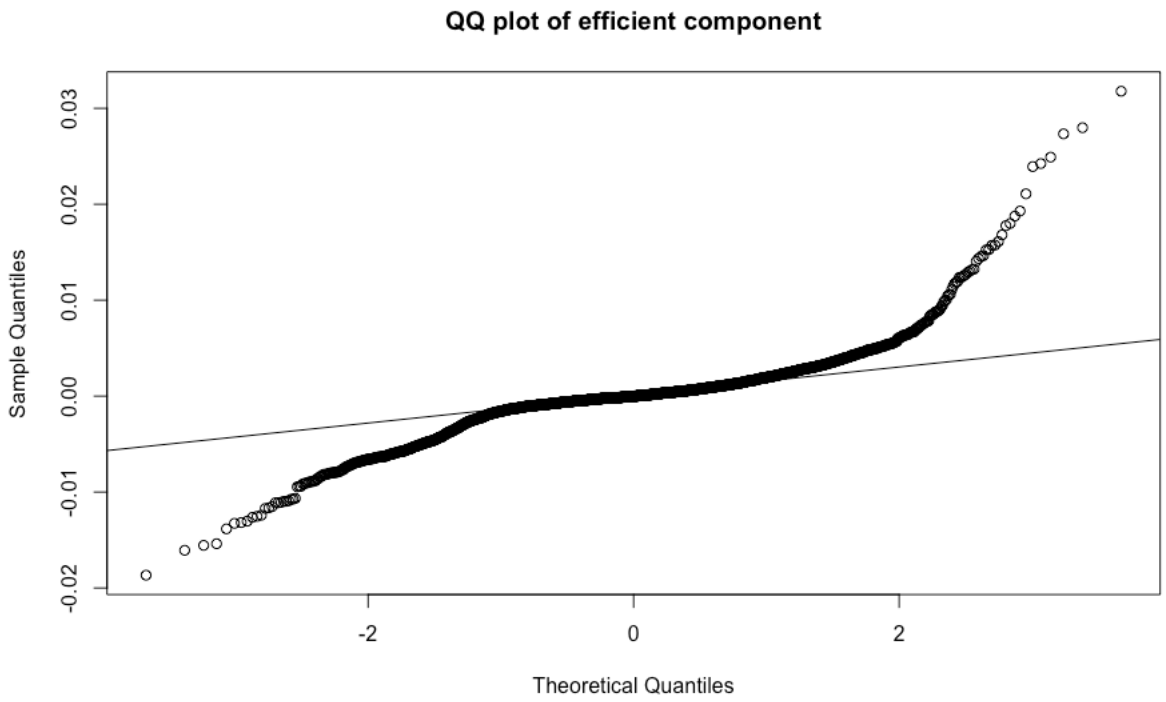

Figure A.7 The Q-Q Plot for Corn Efficient Return Innovations Under the State-Space Model for Large shock days.
Note: Theoretical quantiles on the x-axis refer to quantiles based on the normal distribution. Sample quantiles on the y-axis refer to quantiles based on the residuals from the state-space model for large shock days in the corn market.

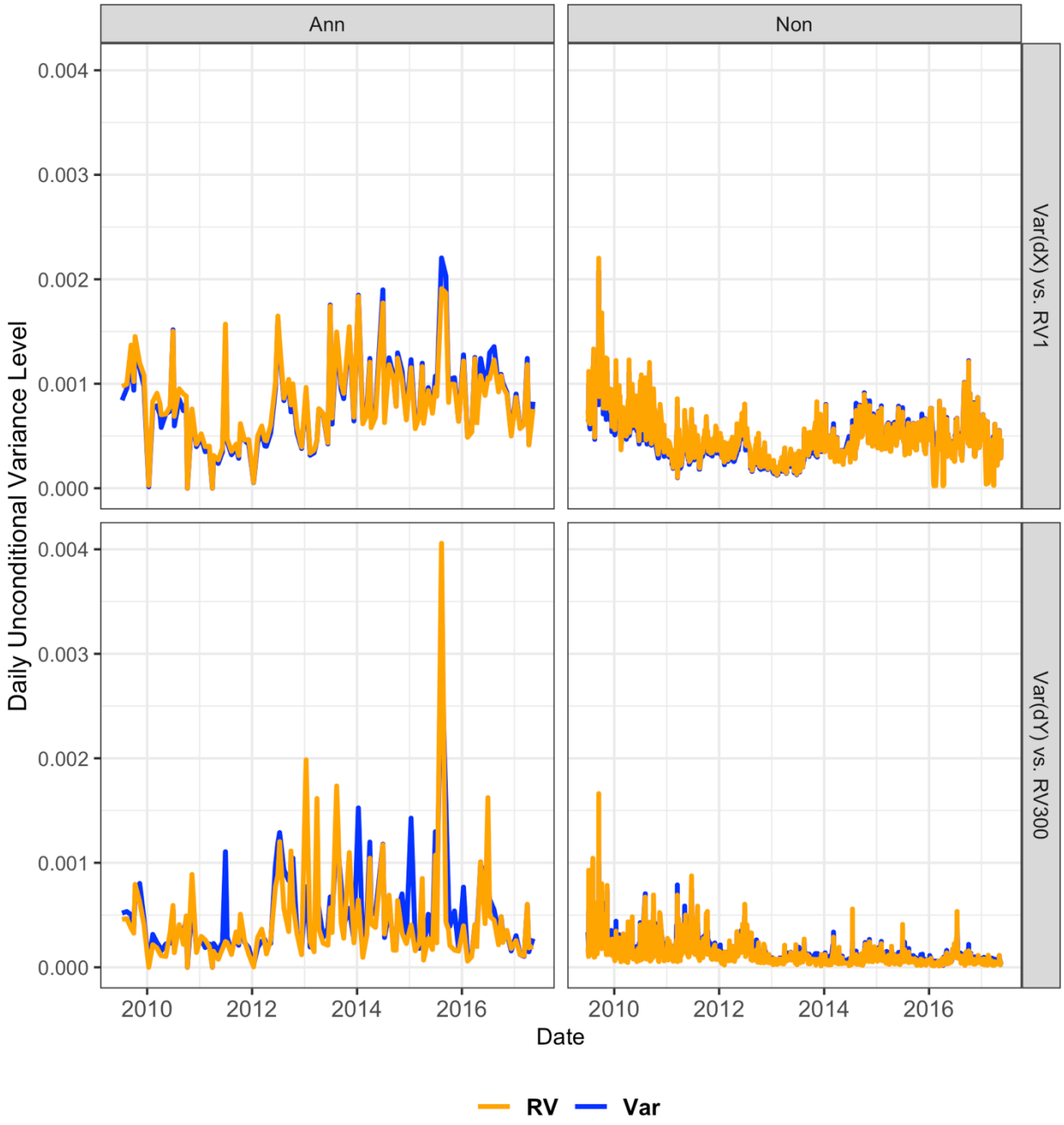

Figure A.8 Unconditional Observed and Efficient Variance Measures in the Corn Market.
Note: Ann and Non stand for announcement and non-announcement days. The y-axis measures daily unconditional variance; the x-axis measures the dates for which the variance measures are computed. Measures are computed based on data available for the entire day. The blue lines are the observed and efficient return variances estimated using (7') and (6') from the paper and located in the top and bottom plots, respectively. The orange lines are the realized variances measured sampling every 1- and 300-seconds in the top and bottom plots, respectively.

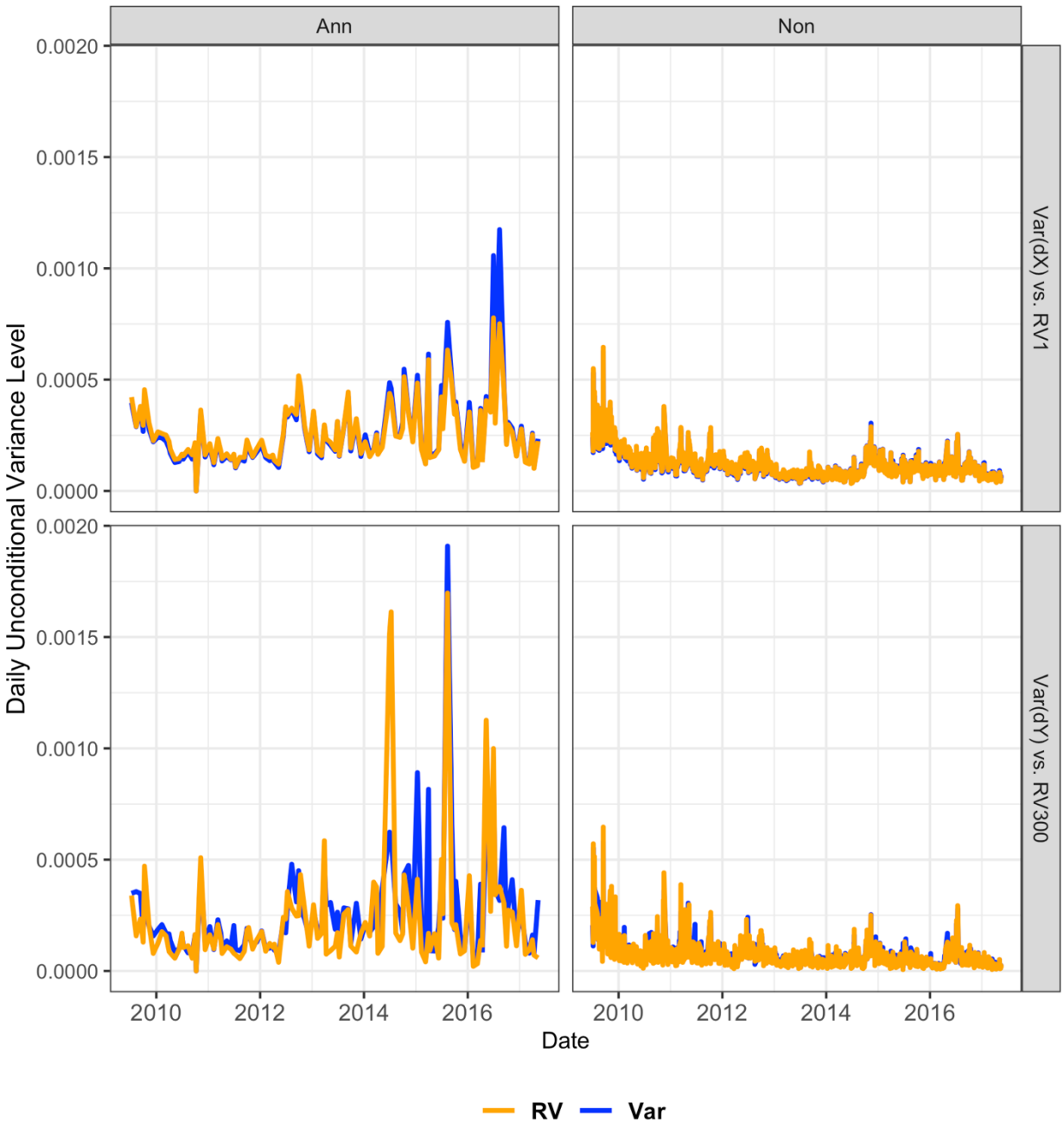

Figure A.9 Unconditional Observed and Efficient Variance Measures in the Soybean Market. Note: Ann and Non stand for announcement and non-announcement days. The y-axis measures daily unconditional variance; the x-axis measures the dates for which the variance measures are computed. Measures are computed based on data available for the entire day. The blue lines are the observed and efficient return variances estimated using (7') and (6') from the paper and located in the top and bottom plots, respectively. The orange lines are the realized variances measured sampling every 1- and 300-seconds in the top and bottom plots, respectively.

A.4 References

Audrino, F., and S.D. Knaus. 2016. "Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics." *Econometric Reviews* 35(8–10):1485–1521. Available at: http://dx.doi.org/10.1080/07474938.2015.1092801.

Bandi, F.M., and J.R. Russell. 2006. "Separating microstructure noise from volatility." *Journal of Financial Economics* 79(3):655–692. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304405X05001534.

Christensen, K., R.C.A. Oomen, and M. Podolskij. 2014. "Fact or friction: Jumps at ultra high frequency." *Journal of Financial Economics* 114(3):576–599. Available at: http://dx.doi.org/10.1016/j.jfineco.2014.07.007.

Corsi, F. 2008. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7(2):174–196. Available at: https://academic.oup.com/jfec/article-lookup/doi/10.1093/jjfinec/nbp001.

Couleau, A., T. Serra, and P. Garcia. 2020. "Are Corn Futures Prices Getting 'Jumpy'?" *American Journal of Agricultural Economics* 102(2):569–588. Available at: https://onlinelibrary.wiley.com/doi/10.1002/ajae.12030.

Fleming, J., C. Kirby, and B. Ostdiek. 2003. "The economic value of volatility timing using 'realized' volatility." *Journal of Financial Economics* 67(3):473–509. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304405X02002593.

Glorot, X., and Y. Bengio. 2010. "Understanding the difficulty of training deep feedforward neural networks." In *The 13th International Conference on Artificial Intelligence and Statistics*. pp. 249–256.

Gonçalves, S., and H. White. 2005. "Bootstrap Standard Error Estimates for Linear Regression."

Journal of the American Statistical Association 100(471):970–979. Available at: http://www.tandfonline.com/doi/abs/10.1198/016214504000002087.

Greene, H.W. 2002. *Econometric Analysis* 5th ed. Prentice Hall PTR.

Hansen, P.R. 2015. "A martingale decomposition of discrete Markov chains." *Economics Letters* 133:14–18. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0165176515001895.

Hansen, P.R., G. Horel, A. Lunde, and I. Archakov. 2015. "A Markov Chain Estimator of Multivariate Volatility from High Frequency Data." In *The Fascination of Probability, Statistics and their Applications*. pp. 361–394.

Hansen, P.R., and A. Lunde. 2006. "Realized Variance and Market Microstructure Noise." *Journal of Business & Economic Statistics* 24(2):127–161. Available at: http://www.tandfonline.com/doi/abs/10.1198/073500106000000071.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification."

Ioffe, S., and C. Szegedy. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift."

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter. 2009. "Microstructure noise in the continuous case: The pre-averaging approach." *Stochastic Processes and their Applications* 119(7):2249–2276. Available at: http://dx.doi.org/10.1016/j.spa.2008.11.004.

Kingma, D.P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization."

Newey, W.K., and K.D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3):703.

Oomen, R.C.A. 2006. "Comment." *Journal of Business & Economic Statistics* 24(2):195–202. Available at: http://www.tandfonline.com/doi/abs/10.1198/073500106000000125.

Rossi, B., J. Stock, and M. Wolf. 2011. “The Model Confidence Set.” *Econometrica* 79(2):453–497. Available at: http://doi.wiley.com/10.3982/ECTA5771.

Sévi, B. 2014. “Forecasting the volatility of crude oil futures using intraday data.” *European Journal of Operational Research* 235(3):643–659.

APPENDIX B

SUPPLEMENTARY RESULTS ON VOLATILITY FORECASTING

B.1 Consistent daily IV estimator

Our consistent daily IV estimator that corrects for autocorrelations in returns caused by microstructure noise (Christensen, Oomen and Podolskij 2014; Jacod et al. 2009) is defined as

$$\widehat{IV}_{i,t}^{d} = \log\left(\sqrt{\frac{N_t}{N_t - K + 2}\frac{1}{K\varphi_K}\sum_{n_t=1}^{N_t-2K-1}\left|r_{n_t,K}^{*}\right|^2 - \frac{\widehat{\omega}^2}{\theta^2\varphi_K}}\right) \qquad (B.1)$$

where returns are calculated on smoothed log-price series that consist of the original log-prices pre-averaged in a local neighborhood of K observations: $r_{n_t,K}^{*} = \frac{1}{K}\left(\sum_{j=K/2}^{K-1} p_j - \sum_{j=0}^{K/2-1} p_j\right)$, with $K = \theta\sqrt{N_t} + o\left(N_t^{-1/4}\right)$, $\varphi = \frac{\left(1+2K^{-2}\right)}{12}$, $j = 0, \ldots, K-1$, $i = 1, \ldots, N$, and θ is a parameter whose value is chosen following (Couleau, Serra and Garcia 2018). The last term on the right hand side in (A1) is a bias correction term that serves the purpose of removing residual noise $\widehat{\omega} = -\frac{1}{N_t-1}\sum_{n_t=2}^{N_t}\left|r_{n_t}^{*}\right|\left|r_{n_t-1}^{*}\right|$ (Oomen 2006).

B.2 HAR-PC Model

A reason underlying the success of the HAR model in volatility forecasting is that the model includes weekly and monthly averages of $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$ that capture volatility persistence. The persistence of time series data is measured by the level of its autocovariance. Therefore, finding a simple but efficient component which captures the most of the autocovariance in the volatility series may further improve volatility forecast performance. A singular value decomposition (SVD) is useful in this regard: by applying SVD to the first 22 lags of $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$ and taking the score corresponding to the largest eigenvalue, we find the most persistent (or informative) linear component among the

22 lagged $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$. Such a score has the largest variance among all possible linear combinations of the 22 lags. We use this score instead of the monthly average in the HAR model. The score is also called the first principal component of 22 lags of $\widehat{\boldsymbol{IV}}_{\boldsymbol{t}}^{\boldsymbol{d}}$. Then, we repeat the same procedure with only the first 5 lags of $\widehat{IV}_t^d$ and replace the weekly average term in the HAR model with the largest eigenvalue score.

By using principal component analysis on the (N x p) matrix $\widehat{\boldsymbol{IV}}^{\mathbf{p}} = \{\widehat{IV}_{i,t-1}^{d}, \dots, \widehat{IV}_{i,t-p}^{d}\}_{i=1}^{N} = \left[\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{d}}, \dots, \widehat{\boldsymbol{IV}}_{\boldsymbol{t-p}}^{\boldsymbol{d}}\right]$, where N is the number of observations and $p \in \{2, \dots, 22\}$, $\widehat{\boldsymbol{IV}}^{\boldsymbol{p}}$, we can decompose $\widehat{\boldsymbol{IV}}^{\boldsymbol{p}}$ with SVD as follows:

$$\widehat{\boldsymbol{IV}}^{\boldsymbol{p}} = \boldsymbol{U}^{\boldsymbol{p}} \boldsymbol{D}^{\boldsymbol{p}} \boldsymbol{V}^{\boldsymbol{p}\prime} \qquad (B.2)$$

Here, $\boldsymbol{U}^{\boldsymbol{p}}$ is a $N \times p$ orthogonal matrix with columns u_j called the left singular vectors; $\boldsymbol{D}^{\boldsymbol{p}}$ is a $p \times p$ diagonal matrix ($\boldsymbol{D}^{\boldsymbol{p}\prime}\boldsymbol{D}^{\boldsymbol{p}} = \boldsymbol{I}^{\boldsymbol{p}}$) with diagonal values $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ called the singular values; and $\boldsymbol{V}^{\boldsymbol{p}}$ is a $p \times p$ orthogonal matrix ($\boldsymbol{V}^{\boldsymbol{p}\prime}\boldsymbol{V}^{\boldsymbol{p}} = \boldsymbol{I}^{\boldsymbol{p}}$). For each rank q, the first column of normalized $\boldsymbol{V}^{\boldsymbol{p}}$, $\boldsymbol{v}^{\boldsymbol{p}}$, creates the weight of the linear combination of columns in $\widehat{\boldsymbol{IV}}^{\boldsymbol{p}}$ which achieves highest variance. The weekly and monthly factors selected by principal component analysis are denoted as

$$\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{w}} = \widehat{\boldsymbol{IV}}^{\boldsymbol{5}}\boldsymbol{v}^{\boldsymbol{5}} = \sum_{p=1}^{5} v_{p,1}^{5} * \widehat{\boldsymbol{IV}}_{\boldsymbol{t-p}}^{\boldsymbol{d}} \qquad (B.3)$$

and

$$\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{m}} = \widehat{\boldsymbol{IV}}^{\boldsymbol{22}}\boldsymbol{v}^{\boldsymbol{22}} = \sum_{p=1}^{22} v_{p,1}^{22} * \widehat{\boldsymbol{IV}}_{\boldsymbol{t-p}}^{\boldsymbol{d}} \qquad (B.4)$$

where $\sum_{p=1}^{5} v_{p,1}^{5} = \sum_{p=1}^{22} v_{p,1}^{22} = 1$. By replacing $\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}$ and $\widehat{\boldsymbol{IV}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}$ by $\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{w}}$ and $\widehat{\boldsymbol{PC}}_{\boldsymbol{t-1}}^{\boldsymbol{m}}$, we obtain a new HAR model with first principal components of both 5-day and 22-day periods as in equation (4) in the main article. In effect, this framework uses the structure of the HAR, but allows

the weights of the lagged 5- and 22-day persistent effects to change by selecting the most informative rather than assigning equal weights to the lagged volatilities.

B.3 The implementations of ANN models

B.3.1 Data normalization

The behavior of neural networks is sensitive to the scale of the variables. To improve the estimation of the neural network, we follow previous research and scale each variable in the output and input set to be between [0,1] as follows: $Y_t^{scaled} = \frac{Y_t - \min(Y_t)}{\max(Y_t) - \min(Y_t)}$ and $X_t^{scaled} = \frac{X_t - \min(X_t)}{\max(X_t) - \min(X_t)}$, where Y and X represent the output and input vector, respectively. The advantage of this normalization is that it makes the distribution of variables approximately uniform. By ameliorating the differences in distributions across all inputs, the ANN model avoids converging to suboptimal solutions. Once the ANN model is trained, the forecasted values are unscaled by taking, $Y_t^{scaled} * (\max(Y_t) - \min(Y_t)) + \min(Y_t)$.

B.3.2 Rectified linear units (ReLU)

By definition, a rectifier activation function in neural networks is implemented as follows:

$$f(x) = \max(0, x) \qquad (B.5)$$

A rectifier activation function has two advantages over traditional sigmoid activation functions. First, the calculation of a rectifier activation function is simpler than a sigmoid activation function. The output of a rectifier function equals the input if the input was positive and zero otherwise. Second, unlike sigmoid activation functions, a rectifier activation function will not have vanishing gradient if $x = \epsilon$ where ϵ is a very small number. Modern ANN models are mostly built with neuron units with rectifier activations. Such units are commonly referred to as rectified linear units (ReLU).

B.3.3 Kernel and bias initialization

Proper kernel initialization will speed up convergence in stacked ANN models (Glorot and Bengio 2010). The optimal choice of kernel initialization depends on the type of activation function used. With ReLU neurons, He et al. (2015b) prove that the proper initialization of the kernels in each layer should be drawn from the following normal distribution:

$$W_i^l = N\left(0, \frac{2}{J^l}\right) \tag{B.6}$$

where $\boldsymbol{W^l} = [W_1^l, \dots, W_J^l]$ is the vector of kernels of layer l, and J^l is the number of neurons in layer l. As for the bias term in each layer, the authors suggest setting the initial value to zero.

B.3.4 The number of epochs

The number of epochs refers to the number of times that an entire dataset is being passed through ANN models with the feedforward-backpropagation training process. If the number of epochs is too large, ANN models are likely to be overfitted. If the number of epochs is too small, ANN models risk being underfitted. Based on the results of our trials, we fix the number of epochs in all our ANN models to 50.

B.3.5 The L1/L2 regularization

In addition to the small and fixed number of epochs, we also introduce the L1 and L2 kernel regularizations in our ANN models. Assume the output before entering the rectifier activation function (ReLU) is defined as follows:

$$\boldsymbol{Y^l} = \boldsymbol{W^l X^l} \tag{B.7}$$

where $\boldsymbol{X^l}$ and $\boldsymbol{Y^l}$ are the input and the output vector (before entering rectifier functions) of layer l, and $\boldsymbol{W^l}$ is the vector of kernels of layer l. Then, the L1 regularization, which is known as least absolute shrinkage and selection operator (LASSO), is specified for $\boldsymbol{W^l}$ as:

$$min_{\boldsymbol{W^l}\in\mathbb{R}^J}\frac{1}{n}\left\|\boldsymbol{W^l X^l}-\boldsymbol{Y^l}\right\|^2+\lambda\left\|\boldsymbol{W^l}\right\| \qquad (B.8)$$

where λ is the shrinkage parameter. The L1 regularization acts as a variable selector and induces sparsity. In the design of our ANN model, the L1 regularization is implemented in the first layer after the input layer for nonlinear components in order to control the number of inputs entering the nonlinear components in our ANN models. In the layers between the first and the final output layer, the L2 regularization is used to prevent overfitting during training. As defined below, the L2 regularization relies upon the Euclidean norm:

$$min_{\boldsymbol{W^l}\in\mathbb{R}^J}\frac{1}{n}\left\|\boldsymbol{W^l X^l}-\boldsymbol{Y^l}\right\|^2+\lambda\left\|\boldsymbol{W^l}\right\|.^2 \qquad (B.9)$$

As a result, the L2 regularization shrinks the size of the vector of kernel $\boldsymbol{W^l}$ instead of forcing $\boldsymbol{W^l}$ = $\mathbf{0}$ for some variables. We set $\lambda = 0.01$ for both L1 and L2 regularization methods.

B.3.6 Batch normalization

Unlike data normalization which is conducted on raw output and inputs, during the pre-training process batch normalization is the normalization step between the L1/L2 regularization and the rectifier activation function of each layer. In a stacked ANN model, the outputs of each intermediate layer depend on the kernels and bias from previous layers. The inter-layer dependence shifts the distribution of the intermediate outputs through the training process, which is known as internal covariate shifts. As a result, deep ANN models converge slower than shallow ANN models. Ioffe and Szegedy (2015) propose to reduce internal covariate shifts by normalizing (whitening) the outputs of each layer before entering the activation functions. Since the normalization is done after every batch (subsample) of inputs, the technique is also called batch normalization. For each batch, all intermediate outputs are normalized as follows:

$$\widehat{X_B^l} = \frac{X_B^l - E\left[X_B^l\right]}{\sqrt{Var[X_B^l] + \epsilon}} \sim N(0,1) \qquad (B.10)$$

where X_B^l is the vector of intermediate outputs of layer l of batch B. The variance of X_B^l is estimated by $Var\left[X_B^l\right] = \frac{m}{m-1} E\left[\sigma_{X_B^l}^2\right]$, where m is the size of B. With the help of batch normalization, stacked ANN models converge to optimal solution with fewer training steps. Also, batch normalization reduces the need of careful initialization and allows larger learning rates for stacked ANN models.

B.3.7 Learning rate optimization

ANN models are optimized via the gradient descent algorithm. The algorithm adjusts all parameters according to both the gradients of loss functions given each parameter and a pre-specified learning-rate parameter. Assume x and y are the inputs and outputs of a neural network with a loss function $J(\boldsymbol{\Gamma}, x, y)$. Then, the gradient descent algorithm adjusts the parameter set Γ as follows:

$$\boldsymbol{\Gamma_s} = \boldsymbol{\Gamma_{s-1}} - \eta \nabla_{\boldsymbol{\Gamma}} J(\boldsymbol{\Gamma_{s-1}}, x, y) = \boldsymbol{\Gamma_{s-1}} - \eta \cdot g_s(\boldsymbol{\Gamma}) \qquad (B.11)$$

where η is the learning rate and $g_s(\boldsymbol{\Gamma}) = \nabla_{\Gamma} J(\boldsymbol{\Gamma_{s-1}}, x, y)$ is the gradient of loss function J with $\boldsymbol{\Gamma}$ being the vector of parameters from the *s*-1th iteration. Here, the number of iterations is defined as the number of batches of inputs used for parameter updates. Choosing a proper η can be challenging. Also, the stochastic nature of ANN models may require different learning rates at different stages of training process. Researchers in the artificial intelligence research community have proposed several learning-rate searching algorithms which can adaptively adjust learning rates. In this paper, we choose the Kingma and Ba's (2014) Adam optimizer. By computing the exponentially smoothed estimator of $g(\boldsymbol{\Gamma})$ and $g(\boldsymbol{\Gamma})^2$ as below, we can track the decay of both the mean and (uncentered) variance of $g(\boldsymbol{\Gamma})$:

$$m_s = \beta_1 \cdot m_{s-1} + (1-\beta_1) g_s(\boldsymbol{\Gamma}) \quad (B.12)$$

$$v_s = \beta_2 \cdot v_{s-1} + (1-\beta_2) g_s(\boldsymbol{\Gamma})^2 \quad (B.13)$$

To correct the bias in m and v, the following equations are applied after each round of calculation:

$$\widehat{m}_s = \frac{m_s}{(1-\beta_1^s)} \quad (B.14)$$

$$\hat{v}_s = \frac{v_s}{(1-\beta_2^s)} \quad (B.15)$$

Then, the gradient descent process revised by Adam optimization algorithm is defined as follows:

$$\boldsymbol{\Gamma_s} = \boldsymbol{\Gamma_{s-1}} - \frac{\widehat{m}_s}{\sqrt{\hat{v}_s} + \epsilon} \cdot \eta \quad (B.16)$$

where ϵ is a very small positive number the prevents a zero in the denominator. To apply the Adam optimizer, β_1, β_2, and ϵ need to be pre-specified. As suggested by Kingma and Ba (2014), we set $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. We set the learning rate η to 0.1 because of the batch normalization layers we included in our model. Also, to prevent large learning rates to destabilize final outputs, we also adopt learning-rate decay mechanism besides the Adam algorithm. Learning-rate decay mechanism will monitor validation losses continuously and shrink η accordingly. In our model, if validation losses do not decrease more than 10^{-4} within 10 epochs, then $\eta_{new} = \eta * 0.1$.

B.4 Statistical tests for model forecast performance

B.4.1 The model confidence set

The model confidence set (MCS) goes beyond pairwise model testing and proposes a selection procedure. Intuitively, the MCS procedure consists of selecting the set of 'best' models (M^*) given a collection of candidate forecast models (M_0), where 'best' is defined by an evaluation forecast measure selected by the modeler. The procedure sequentially identifies $M^* \subset M_0$ given a confidence interval α. In practice, an equivalence test is applied to the collection of models in M_0.

If the null hypothesis is rejected, there is evidence that the objects in M_0 are not equally "good" and an elimination rule is used to remove from M_0 the model with the poorest sample performance. This procedure is repeated until the null is accepted and the MCS is defined by a set of "surviving" models. We define a set of models M_0 and sequentially test the null hypothesis of equal prediction accuracy, i.e.

$$H_0: E\left(d_{z_1 z_2,t}\right) = 0 \; \forall z_1, z_2 \in M_0, \qquad (B.17)$$

where $d_{z_1 z_2,t} = g\left(e_{z_1,t}\right) - g\left(e_{z_2,t}\right)$ is the loss differential between models z_1 and z_2 in the collection of models and g(.) is the MSE function. Following Hansen et al. (2011), two types of statistics are used to test the null hypothesis, the range statistics T_R, and the semi-quadratic statistics T_{SQ}, both relying on the following t-statistic: $t_{z_1 z_2} = \frac{\overline{d_{z_1 z_2}}}{\sqrt{\widehat{var(\overline{d_{z_1 z_2}})}}}$ for $z_1, z_2 \in M_0$, with $\overline{d_{z_1 z_2}} = \frac{1}{H}\sum_{t=1}^{H} d_{z_1 z_2,t}$. The t-statistic, $t_{z_1 z_2}$, provides scaled information on the average difference in the point forecast quality of models z_1 and z_2. Additionally, $\widehat{var\left(\overline{d_{z_1 z_2}}\right)}$ is an estimate of $var\left(\overline{d_{z_1 z_2}}\right)$, obtained by using the block bootstrap of Gonçalves and White (2005) as proposed by Hansen et al. (2011). This approach is usually used when the number of models in the set is high. The range statistics, T_R, are given by,

$$T_R = \max_{z_1, z_2 \in M_0} \left|t_{z_1 z_2}\right| = \max_{z_1, z_2 \in M_0} \frac{\left|\overline{d_{z_1 z_2}}\right|}{\sqrt{\widehat{var\left(\overline{d_{z_1 z_2}}\right)}}} \qquad (B.18)$$

Finally, the MCS procedure assigns a p-value, $\hat{p}_z$, to each model z in the initial set M_0. The resulting optimal set of models is denoted by $\widehat{M}^*_{1-\alpha}$ if and only if $\hat{p}_z \geq \alpha$.

B.4.2 The Modified Diebold-Mariano test

The Modified Diebold-Mariano test conducts a pairwise comparison of the loss function derived from each model and is computed as

$$MDM = \sqrt{\frac{h-1}{\frac{1}{h}\sum_{t=1}^{h}(d_t-\bar{d})^2}}\,\bar{d} \qquad (B.19)$$

where $d_t = g(e_{t,1}) - g(e_{t,2})$, being $g(e_{t,z})$, $z = 1, 2$ the loss function for each of the two models, which we specify as the Mean Square Error (MSE) and QLIKE, h is the number of out-of-sample forecasts, and $\bar{d}$ is the average of the difference. In this paper, we follow Sévi (2014) and obtain the MDM test results with a simple Wald test. The expected value of $\bar{d}$ is zero. Also, we replace the asymptotic standard deviation in equation E2 with the HAC long-run variance from Newey and West (1987). We set the lag equal to $(h_{max})^{0.25} = 2$, which is the common practice suggested in Greene (2002).

Table B.1 presents the MDM to distinguish the forecasting performance between iterative and direct forecasting methods. If MDM>0, the forecasting error of iterative method is higher than the error of direct method, and vice versa. Results suggest that for the linear models' group, there are no statistically significant differences between the iterative and direct forecasting methods. Some statistically significant differences are however identified in hybrid and nonlinear models, especially in ANN-HAR and RF, which suggest one-step ahead misspecification of these models and favor the direct over the iterative method.

We further use the MDM test to distinguish the forecasting performance between hybrid models and their linear components. The results are reported in Tables E.2A and E.2B. If MDM>0, the forecasting error of hybrid ANN models is higher than the error of their linear components, and vice versa. As we can see in Tables E.2A and E.2B, most significant differences are negative, which means the forecasting error of hybrid ANN models is significantly lower than their linear components. Also, the significant differences mostly occur when $h = 1$. The MDM test fails to distinguish between the two models when $h = 5$ and 22. This is consistent with the results in

Tables 3.4A and 3.4B, which fail to differentiate forecasting accuracy between hybrid and linear results with the MCS test.

B.5 The dynamics of nonlinear components in the wheat market

As we can see in Figure B.1, the nonlinear components across the three forecasting horizons all reach maximum levels in 2013 and decrease afterward. The reason behind the increase in $\widehat{IV}_{i,t}^{j,NL}$ in 2013 is the wheat price boom between 2013 and 2014. The existing linear dynamics failed to explain the price spikes in the wheat market, which boosted the nonlinear component. More importantly, except for small scattered periods such as the first half of 2014 for $h = 22$, the nonlinear components have the shape of a step function which suggests the nonlinear component is constant, with its specific value being adjusted each time the model is re-estimated. A constant nonlinear component is indicative that the data are not characterized by nonlinear dynamics. Conclusions for other markets are similar and not presented here to preserve space.

The heatmaps in Figure B.2 indicate that the kernel weights for linear components are much larger than the weights for nonlinear components in the wheat market. Unlike the results in Table 6, we do not multiply the kernel weights by 100 in Figure B.2. While the linear kernels are mostly between $1e^{-1}$ and $1e^{-2}$, most nonlinear kernel weights are well below $1e^{-4}$ for most of the forecasting window. The results are similar to Audrino and Knaus (2016) with penalized regression models. The heatmap of linear components indicates that the kernel weights for lagged IVs with $m = 1$ to 5 are higher than lagged IVs with $m > 5$. Further, the decay pattern does not exist among the nonlinear kernel values, which is consistent with our findings in Table 6. We find short periods with relatively significant nonlinear patterns, such as the slots around January 1st, 2014 when investors began to bet on extreme weather conditions reversing the downward trend in the wheat price, a bet that was later proved wrong. For other periods, the ANN-IV model assigns

near-zero weights to all lagged IVs when fitting the nonlinear components. The near-zero kernel values are achieved via the L1 kernel regularization described in Appendix B.

B.6 Tables and Figures

Table B.1: Hyperparameters For Selected Models

Model	Hyperparameters	Grid Resolution
LASSO	λ: $[10^{-6}, \dots, 10^{-1}]$	100
Ridge	λ: $[10^{-6}, \dots, 10^{-1}]$	100
RF	$mtry$ (P'): $[1, \dots, 22]$ $ntree$ (B): $[10, \dots, 400]$	5
SVR	λ: $[2^{-4}, \dots, 2^{4}]$ γ: $[2^{-5}, \dots, 2^{3}]$	10
XGB	max depth (B): $[1, \dots, 10]$ min child weight (S):$[1, \dots, 10]$ η: $[0.1, \dots, 1]$	5

Note: The table describe the range of each hyperparameters and the resolution of grid search. During the grid search process, the range is divided equally by the corresponding grid resolution. The values on the grids are used in the cross-validation process. Grid resolutions and ranges of hyperparameter are chosen to keep the balance between completeness and computational efficiency.

Table B.2: The Modified Diebold-Mariano test: Iterative vs. Direct

Model	h	Corn		Soybean		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
HAR	5	0.8075	0.7710	0.7903	0.6787	0.0649	0.0590	0.0597	0.0720	0.9705	0.9997
HAR-PCA	5	0.9940	0.9084	0.7641	0.6834	0.1174	0.0916	0.1515	0.1513	0.9407	0.9334
LASSO	5	0.9807	0.6906	0.6459	0.7208	0.9008	0.7930	0.7932	0.8054	0.7302	0.8017
Ridge	5	0.7138	0.6993	0.3393	0.3370	0.2248	0.2024	0.7969	0.8677	0.0851	0.0538
ANN-IV	5	0.5942	0.6720	0.2309	0.1991	0.2518	0.3554	0.2955	0.2415	0.3668	0.1813
ANN-CumSum	5	0.9371	0.7376	0.3534	0.2562	0.0126*	0.0273*	0.1051	0.0778	0.9515	0.7257
ANN-HAR	5	**0.0000****	**0.0000****	**0.0000****	**0.0000****	0.0141*	0.0198*	**0.0012****	**0.0029****	0.1409	0.8073
RF	5	**0.0002****	**0.0014****	0.3039	0.4564	0.0262*	0.0273*	0.5028	0.4757	0.3340	0.5276
SVR	5	0.8515	0.2839	0.9466	0.4291	0.8139	0.7854	0.0877	0.0740	0.1038	0.0852
XGB	5	0.8702	0.8070	0.1264	0.2170	0.0841	0.1211	0.3876	0.4846	0.0222*	0.0267*
HAR	22	0.8567	0.9404	0.2348	0.2283	0.3166	0.4724	0.8247	0.7274	0.2172	0.5748
HAR-PCA	22	0.5288	0.7546	0.1828	0.1326	0.1260	0.2298	0.6373	0.7816	0.1551	0.5438
LASSO	22	0.3239	0.4832	0.5474	0.5496	0.1211	0.1736	0.2247	0.3303	0.1409	0.5725
Ridge	22	0.6186	0.8364	0.4000	0.3614	0.0989	0.1133	0.8703	0.9389	0.6441	0.9304
ANN-IV	22	0.0637	0.0938	0.2699	0.2499	0.1793	0.1497	0.5281	0.5367	0.9836	0.7662
ANN-CumSum	22	0.1157	0.1373	0.1532	0.1184	0.5239	0.5585	0.5753	0.7327	0.2740	0.1701
ANN-HAR	22	**0.0000****	**0.0000****	0.0252*	0.1134	0.0788	0.0874	0.2627	0.1995	**0.0023****	0.0139*
RF	22	**0.0000****	**0.0000****	0.0582	0.0542	0.0392*	0.0302*	0.2576	0.2020	**0.0028****	0.0137*
SVR	22	0.4610	0.2754	0.5217	0.2544	0.2159	0.1097	0.0616	0.0307*	0.1378	0.0894
XGB	22	0.2882	0.4697	0.3620	0.4972	**0.0068****	**0.0066****	0.4058	0.3813	0.1513	0.1799

Note: Each cell represents the MDM test of the difference between iterative and direct forecasting methods; if MDM > 0, then the forecasting error of the iterative method is higher than the direct method, and vice versa; if the difference is significant at the 5% level, the cell is denoted by one asterisk; if the difference is significant at the 1% level, the cell is marked by two asterisks and bold fonts; the Model column contains the names of each model; h represents the number of out-of-sample forecasting steps; the MSE and QLIKE columns represent MDM test based on corresponding loss functions.

Table B.3A: The Modified Diebold-Mariano Test for ANN Models: Hybrid vs. Linear (Iterative forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
ANN-IV[1]	1	-0.0008	**-0.0027****	-0.0003	**-0.0011****	-0.0006	**-0.0018****	-0.0003	**-0.0004****	-0.0001	**-0.0007****
ANN-CumSum	1	-0.0006	**-0.0022****	-0.0005	**-0.0012****	-0.0005	**-0.0017****	-0.0004*	**-0.0005****	-0.0001	**-0.0008****
ANN-HAR	1	-0.0007	**-0.0021****	-0.0005	**-0.0012****	**-0.0017****	**-0.0021****	-0.0018	-0.0013	-0.0078	-0.0038*
ANN-IV	5	0.0001	-0.0007	-0.0001	-0.0003	0.0008	-0.0002	-0.0001	-0.0003	0.0006	-0.0003
ANN-CumSum	5	0.0003	-0.0005	-0.0002	-0.0004	0.0010	-0.0001	-0.0001	-0.0003	0.0008	-0.0003
ANN-HAR	5	0.0006	-0.0004	-0.0001	-0.0005	-0.0005	-0.0008*	-0.0014	-0.0010	-0.0075	-0.0034*
ANN-IV	22	0.0029	0.0008	0.0006	0.0000	0.0046	0.0018	0.0005	-0.0001	0.0028	0.0006
ANN-CumSum	22	0.0030	0.0009	0.0007	-0.0000	0.0049	0.0019	0.0006	-0.0001	0.0033	0.0008
ANN-HAR	22	**0.0053****	0.0019*	0.0013	0.0003	0.0029	0.0010	-0.0006	-0.0007	-0.0062	-0.0026

Note: Each cell represents the MDM test of the difference between the forecasting error of hybrid ANN models and their linear components; if MDM> 0, then the forecasting error of hybrid predictions is higher than linear components, and vice versa; if the difference is significant at the 5% level, the cell is denoted by one asterisk; if the difference is significant at the 1% level, the cell is marked by two asterisks and bold fonts; the Model column contains the names of each model; h represents the number of out-of-sample forecasting steps; the MSE and QLIKE columns represent MDM test based on corresponding loss functions.

Table B.3B: The Modified Diebold-Mariano Test for ANN Models: Hybrid vs. Linear (Direct forecasting)

Model	h	Corn		Soybeans		Wheat		Crude Oil		S&P 500 E-mini	
		MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE	QLIKE
ANN-IV	1	-0.0003	**-0.0018****	-0.0005	**-0.0012****	-0.0005	**-0.0017****	-0.0003	**-0.0004****	-0.0000	**-0.0007****
ANN-CumSum	1	-0.0005	**-0.0021****	-0.0004	**-0.0012****	-0.0004	**-0.0017****	-0.0004*	**-0.0005****	-0.0005	**-0.0010****
ANN-HAR	1	-0.0007	**-0.0020****	-0.0005	**-0.0011****	-0.0010	**-0.0019****	**-0.0006****	**-0.0006****	-0.0021	-0.0015
ANN-IV	5	-0.0005	-0.0012	-0.0011	-0.0011*	-0.0007	-0.0012	-0.0018*	**-0.0014****	0.0002	-0.0013
ANN-CumSum	5	-0.0009	-0.0014	-0.0017*	**-0.0014****	-0.0009	-0.0014	-0.0012	-0.0011*	0.0006	-0.0011
ANN-HAR	5	-0.0008	-0.0013	-0.0013	-0.0012*	-0.0018	-0.0017*	-0.0091	-0.0041	-0.0008	-0.0017*
ANN-IV	22	-0.0041	-0.0036	-0.0036	-0.0028	-0.0001	-0.0012	-0.0071	-0.0053	0.0029	-0.0026
ANN-CumSum	22	-0.0035	-0.0033	-0.0050	-0.0037	-0.0036	-0.0032	-0.0060	-0.0048	0.0034	-0.0023
ANN-HAR	22	-0.0047	-0.0033	-0.0056	-0.0040	-0.0089	-0.0056*	-0.0071	-0.0055	0.0024	-0.0027

Note: Each cell represents the difference between the forecasting error of hybrid ANN models and their linear components; if the difference > 0, then the forecasting error of hybrid predictions is higher than linear components, and vice versa; if the difference is significant at the 5% level, the cell is denoted by one asterisk; if the difference is significant at the 1% level, the cell is marked by two asterisks and bold fonts; the Model column contains names of each hybrid ANN models; h represents the number of out-of-sample forecasting steps; the MSE and QLIKE columns represent differences of corresponding loss functions.

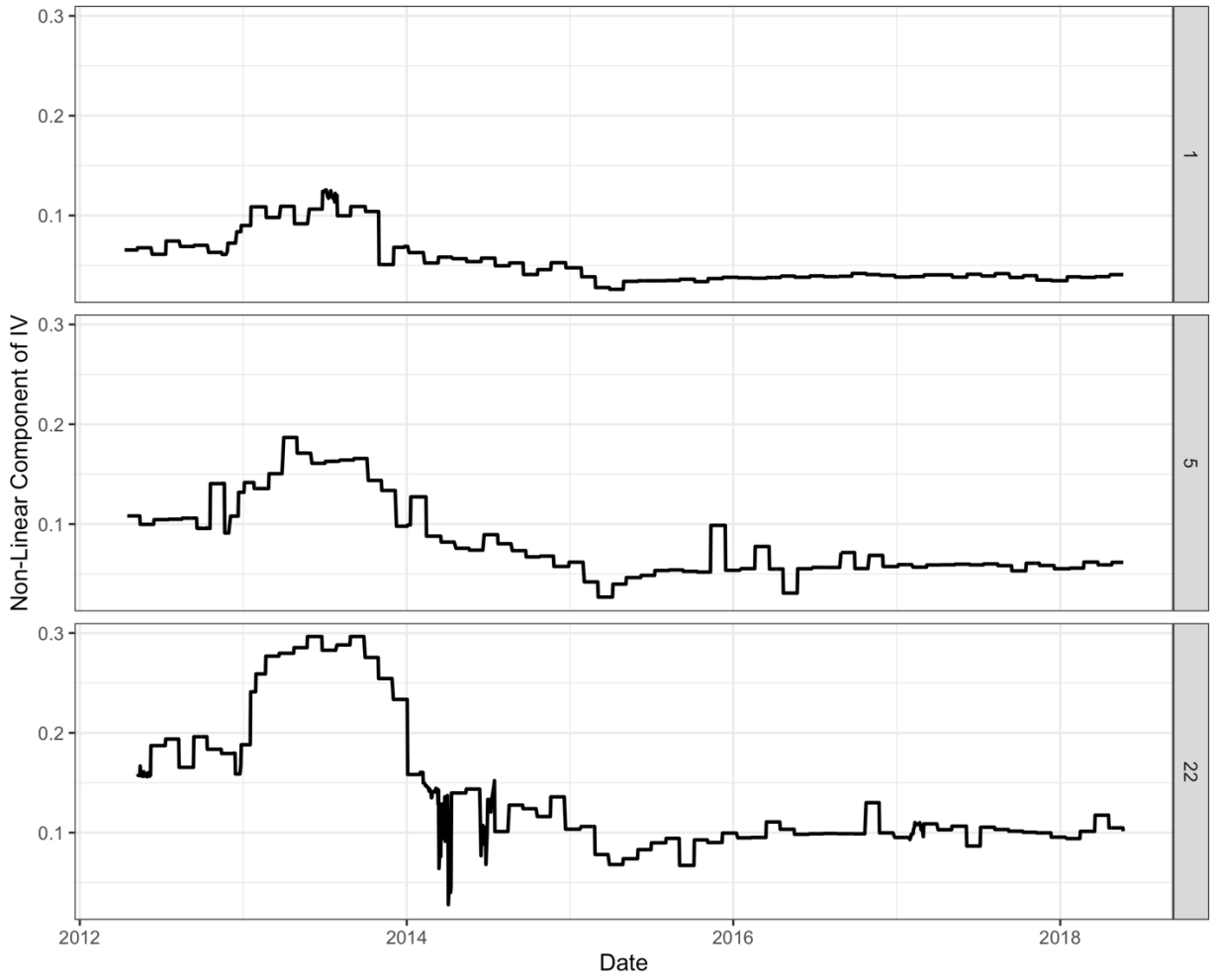

Figure B.1. Non-linear component of $\widehat{IV}_t$ in the wheat futures market with the ANN-IV Model under direct forecasting method

Notes: The x-axis represents the dates; the y-axis represents the absolute level of non-linear components; 1, 5, and 22 represents different steps (h).

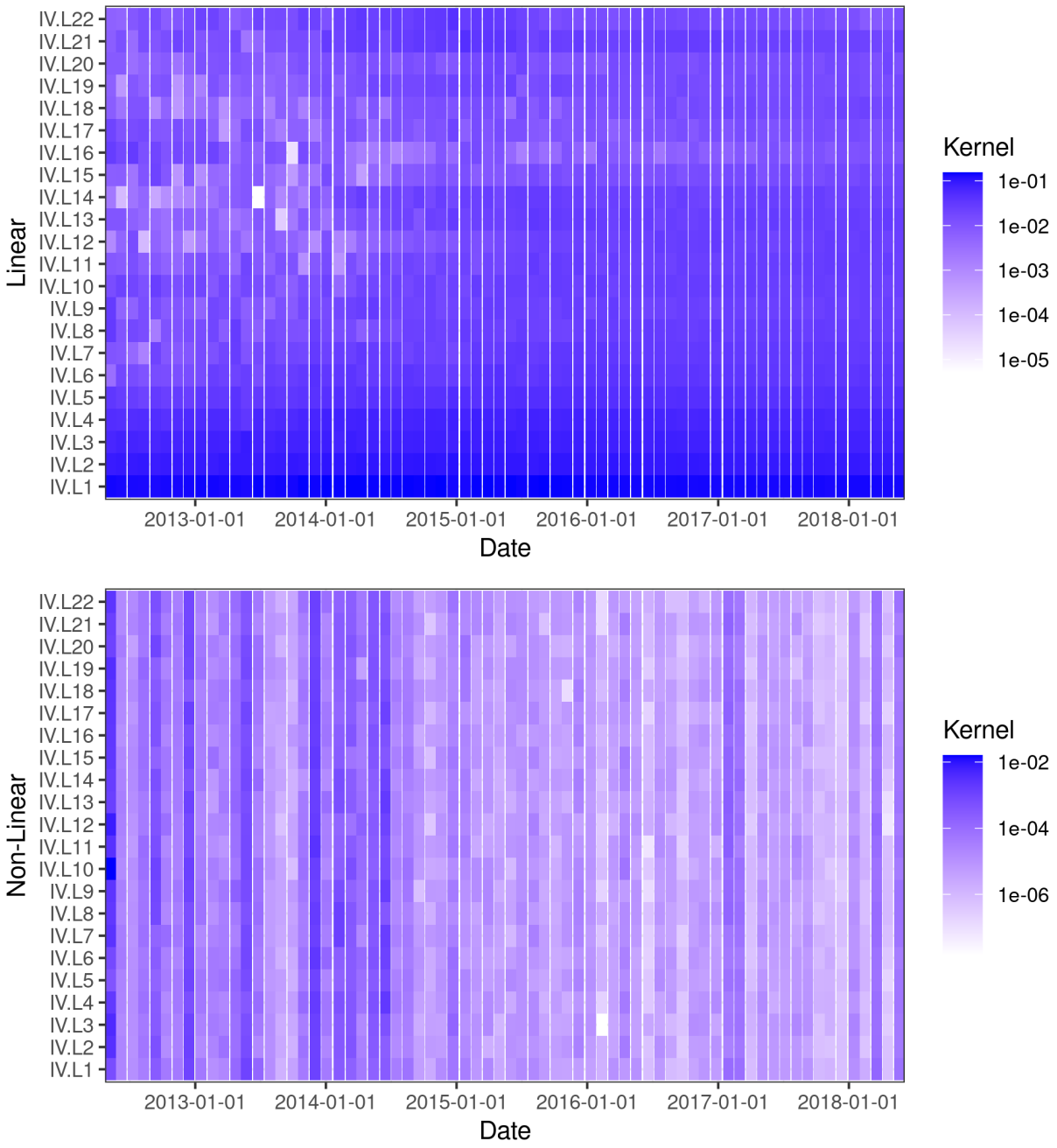

Figure B.2. Heatmap of average ANN-IV kernel values for each input lag of linear and nonlinear components in the wheat market (direct forecasting method + $h = 22$)

Notes: The x-axis represents the dates; the y-axis represents different lagged IV_{t-s}, where $s = 1, \ldots, 22$; The colored tiles represent the mean values of absolute kernel coefficients of each lagged IV_{t-s} across different neurons in the first layer of the ANN-IV model; the top figure depicts the kernel coefficients from the linear components; the bottom figure depicts the kernel coefficients from the non-linear components; each figure has its own scales.

B.7 References

Audrino, F., and S.D. Knaus. 2016. "Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics." *Econometric Reviews* 35(8–10):1485–1521. Available at: http://dx.doi.org/10.1080/07474938.2015.1092801.

Bandi, F.M., and J.R. Russell. 2006. "Separating microstructure noise from volatility." *Journal of Financial Economics* 79(3):655–692. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304405X05001534.

Christensen, K., R.C.A. Oomen, and M. Podolskij. 2014. "Fact or friction: Jumps at ultra high frequency." *Journal of Financial Economics* 114(3):576–599. Available at: http://dx.doi.org/10.1016/j.jfineco.2014.07.007.

Corsi, F. 2008. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7(2):174–196. Available at: https://academic.oup.com/jfec/article-lookup/doi/10.1093/jjfinec/nbp001.

Couleau, A., T. Serra, and P. Garcia. 2020. "Are Corn Futures Prices Getting 'Jumpy'?" *American Journal of Agricultural Economics* 102(2):569–588. Available at: https://onlinelibrary.wiley.com/doi/10.1002/ajae.12030.

Fleming, J., C. Kirby, and B. Ostdiek. 2003. "The economic value of volatility timing using 'realized' volatility." *Journal of Financial Economics* 67(3):473–509. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0304405X02002593.

Glorot, X., and Y. Bengio. 2010. "Understanding the difficulty of training deep feedforward neural networks." In *The 13th International Conference on Artificial Intelligence and Statistics*. pp. 249–256.

Gonçalves, S., and H. White. 2005. "Bootstrap Standard Error Estimates for Linear Regression."

Journal of the American Statistical Association 100(471):970–979. Available at: http://www.tandfonline.com/doi/abs/10.1198/016214504000002087.

Greene, H.W. 2002. *Econometric Analysis* 5th ed. Prentice Hall PTR.

Hansen, P.R. 2015. "A martingale decomposition of discrete Markov chains." *Economics Letters* 133:14–18. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0165176515001895.

Hansen, P.R., G. Horel, A. Lunde, and I. Archakov. 2015. "A Markov Chain Estimator of Multivariate Volatility from High Frequency Data." In *The Fascination of Probability, Statistics and their Applications*. pp. 361–394.

Hansen, P.R., and A. Lunde. 2006. "Realized Variance and Market Microstructure Noise." *Journal of Business & Economic Statistics* 24(2):127–161. Available at: http://www.tandfonline.com/doi/abs/10.1198/073500106000000071.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification."

Ioffe, S., and C. Szegedy. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift."

Jacod, J., Y. Li, P.A. Mykland, M. Podolskij, and M. Vetter. 2009. "Microstructure noise in the continuous case: The pre-averaging approach." *Stochastic Processes and their Applications* 119(7):2249–2276. Available at: http://dx.doi.org/10.1016/j.spa.2008.11.004.

Kingma, D.P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization."

Newey, W.K., and K.D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55(3):703.

Oomen, R.C.A. 2006. "Comment." *Journal of Business & Economic Statistics* 24(2):195–202. Available at: http://www.tandfonline.com/doi/abs/10.1198/073500106000000125.

Rossi, B., J. Stock, and M. Wolf. 2011. “The Model Confidence Set.” *Econometrica* 79(2):453–497. Available at: http://doi.wiley.com/10.3982/ECTA5771.

Sévi, B. 2014. “Forecasting the volatility of crude oil futures using intraday data.” *European Journal of Operational Research* 235(3):643–659.