# Factor-based Quantile Forecasting with Textual Data

Luiz Renato Lima
Department of Economics, The University of Tennessee, Knoxville.

Lucas Lúcio Godeiro
UFERSA

Jie Wei *
School of Economics, Huazhong University of Science and Technology, Wuhan, China

January 24, 2024

**Abstract**

Quantile forecasting has become an important research topic in econometrics as policy makers and investors are increasingly interested to focus more on downside (upside) risks rather than learning about the average or most likely outcome. Simultaneously, practitioners have largely used textual data to construct new predictors of financial and economic variables. This paper develops a novel methodology for out-of-sample quantile forecasting with textual data. It relies on elastic net quantile regression to incorporate "attention" into the quantile forecasting model, allowing for the dictionary to vary across time and over the quantiles of the conditional distribution. We next extract common factors from the refined specific dictionary, explore them for quantile forecasting, and establish new asymptotic theory for forecasting. We apply this approach to study U.S. dollar appreciation (depreciation) events against the Canadian dollar. The proposed approach yields density, interval and probability forecasts that outperform the ones from the random-walk and other conditional quantile models by economically meaningful margins.

*Keywords:* Textual data ; quantile regression; factor model; density forecast; portfolio analysis.

*JEL Classification codes:* C53, C55, E37, E47.

*Corresponding author (Jie Wei): jiewei.econ@gmail.com

# 1 Introduction

As mentioned by Adrian et al. (2019), point forecasts ignore risks around the central forecast and, as such, are not useful to predict tail risks related to the occurrence of rare events. Quantile forecasting has arisen as a potential solution to this problem as documented by a large and increasing literature that includes empirical applications in areas as diverse as profitability forecasting (Tian et al., 2021), economic growth Adrian et al. (2019), fiscal policy and debt ceiling (Lima et al., 2008); risk management (Gaglianone et al., 2012a); density forecasting (Gaglianone and Lima, 2012, 2014); international trade agreements Figueiredo and Lima (2020), among many other applications in economics and finance. More recently, practitioners have started using textual data to construct new predictors of financial and economic variables. For instance, Ellingsen et al. (2022) show that news data are useful to predict the U.S. GDP, consumption and investment growth as they contain information not captured by hard economic indicators. These predictors are computed by using textual information which includes unstructured data from news published in newspapers, reports, etc. Most of the analyses using textual data rely on fixed (ad-hoc) dictionaries. However, as demonstrated by Lima and Godeiro (2023), the assumption of a fixed dictionary may not be appropriate as new words are introduced over time and vocabulary usage in published texts may differ across media. Indeed, even if the vocabulary were the same across media texts, the predictive power of some words might differ, say, in times of recession and economic expansion. This problem is exacerbated as one uses textual data to do quantile forecasting since the optimal dictionary is now assumed to be constant not only over time but also across quantiles of the conditional distribution.

This paper proposes addressing this problem by using a quantile factor model. In the literature of Latent Semantic Analysis (LSA), a factor model for textual data is interpreted as belonging to a Singular Value Decomposition (SVD) - see Hendry (2012). However, traditional factor estimation requires the model to be dense, but textual data are mostly sparse since not all words have predictive content at a certain quantile. Although it is conventional to standardize data prior to estimation, the lack of dense textual data has refrained practitioners from using factor models for text analysis. There have been increasing concerns looking into the inconsistency of forecasting based on factors from the sparse environment, also known as weak factor models (Onatski, 2012; Chao et al., 2022; Giglio et al., 2023). In the machine learning literature, "attention" is a technique that enhances some parts of the input data while diminishing other parts. As "attention" targets the most predictive words before computing the common factors, it avoids sparsity and makes textual data suitable to principal component estimation.[1]

In this paper, we apply the above approach to the problem of quantile forecasting. As in Bai and Ng (2008) - but focusing now on quantile forecasting with textual data - we can be better off throwing away those words without predictive content. In order to do so, the most predictive words have to be selected not only at different forecast origins but also across different quantiles of the conditional distribution. We perform such a selection by minimizing the elastic-net penalized Huber loss regression in high dimensional settings. Such a minimization process is conducted through the semismooth Newton coordinate descent (SNCD) algorithm, recently proposed by Yi and Huang (2017). Unlike existing coordinate descent type algorithms, the SNCD updates a regression coefficient and its corresponding subgradient simultaneously in each iteration. It combines the strengths of the coordinate descent and the semismooth Newton algorithm, and effectively solves the computational challenges posed by high dimensionality and nonsmoothness.[2]

The methodology developed in this paper has three steps:[3] Firstly, we transform words into numerical values (time series) without imposing any restriction on which terms can be included. This implies we are not using a fixed (ad-hoc) dictionary and, consequently, the obtained numerical representation is sparse and high dimensional; the SNCD algorithm is used in the second step to select the most predictive time series (words) at each forecast origin and across different quantiles of the conditional distribution - words are selected in terms of their predictive content at different quantiles and time periods; in the the third step, one estimates common factors for the dense set of selected terms and use them to predict conditional quantiles. Factor estimation is done through principal component analysis (PCA) and the quantile factor

[1]Latent Dirichlet Allocation (LDA) models are largely used in the forecasting literature, but they are computationally intensive and updating the training sample recursively is costly. LDA models are sensitive to the choice of the number of topics and also subject to the look-ahead bias. Empirical applications of LDA have also overlooked the role played by n-grams (or collocations). See additional comments on LDA models in Lima and Godeiro (2023)

[2]Yi and Huang (2017) also establishes the convergence properties of the algorithm and present an adaptive version of the “strong rule”for screening predictors to gain extra efficiency.

[3]Details on the proposed methodology and the out-of-sample forecasting exercise are presented in Section 2.

analysis (QFA) estimator recently introduced by Chen et al. (2021). This step is an extension of difussion index (DI) model of Stock and Watson (1989) towards quantile forecasting, but we additionally establish a new theoretical foundation and show under what regularity conditions and convergence rates the DI model approach can be successfully extended to the problem of quantile forecasting.

We apply the proposed method to the problem of exchange rate predictability. Models of exchange rate determination assume asset prices are set in efficient markets, but limits to arbitrage and rationally limited attention challenge this assumption as information in news texts are not absorbed fully by market prices instantaneously. Even if news does not have predictive content for forecasting the mean, it may still have predictive content at the quantiles of the exchange rate distribution. If this occurs, then news can be useful to predict left and right tail outcomes of exchange rates, that is, they are useful to predict depreciation (appreciation) risks. As exemplified in Rossi (2013), multinational corporations consider depreciation or appreciation risks when deciding whether or not to hedge future payables and receivables which are in foreign currencies. Moreover, cash flows of all international transactions are affected by expectations about depreciation (appreciation) of the foreign currency. Decisions on short-term investment and capital budgeting are also affected by depreciation risks since banks are looking for the ideal currency for borrowing money as well allocating the one which maximizes return on investment. Furthermore, Central Banks balance sheets respond to expectations on depreciation or appreciation of foreign currencies. These examples suggest downside and upside risks affect many decisions undertaken by firms, governments and individuals.

Our primary evaluation will be based on forecasts of the US-CAN exchange rate at different quantiles of the conditional distribution. The extant literature on exchange-rate forecasting (e.g., Filippou et al. 2021 and references therein) has mainly focused on point (conditional mean) forecasting (rather than quantile forecasting) and either did not study the US-CAN exchange rate or were not able to predict it better than what is obtained with a random walk model.[4] Thus, testing the methodologies on the US-CAN exchange rate puts the proposed method to a non-trivial challenge. Textual data have been used

[4] Amat et al. (2018) found a very small forecasting gain over the random walk model.

to construct new predictors for out-of-sample forecasting of financial variables (see Lima and Godeiro (2023) and references therein), but not much has been done in terms of using textual data to predict exchange-rate returns. A noticeable exception is Narayan et al. (2021) that utilizes a fixed dictionary proposed by Loughran and McDonald (2011) to construct indexes of positive and negative news that are then used as predictors of exchange rate returns. The usage of fixed (ad-hoc) dictionaries is hard do justify since words predictability may change across quantiles and over time. [5] The forecasting ability of fixed dictionaries is investigated in the empirical section of this paper.

Our empirical results confirm that it is hard to improve point forecasting of the US-CAN beyond the one established by the random walk model. However, we find that predictability of the US-CAN exchange rate is stronger at the low and high-end quantiles of the conditional distribution. The proposed text-based quantile model generates one-month-ahead quantile, interval, density and directional forecasts (i.e., predicted probability of depreciation or appreciation) of the US-CAN exchange rate that consistently outperform the corresponding forecasts from the benchmark random walk (unconditional quantile) model and from several other conditional quantile models by economically meaningful margins. In particular, the predicted probability of dollar depreciation (appreciation) is related to the downside (upside) entropy of the US-CAN exchange rate, meaning it carries information about downside (upside) risks. We identify words associated to such downside and upside risks. We show that terms related to economic recessions increase the downside (depreciation) risk of the U.S. dollar whereas terms related to geopolitical events and market volatility increase the upside (appreciation) risk, suggesting that the U.S. dollar is seen by investors as a "safe heaven" in times of geopolitical conflicts and high market volatility.

The organization of this paper is as follows. Section 2 introduces the quantile forecasting problem. Section 3 describes the proposed methodology with a new asymptotic theory for the diffusion index model established in Section 4. Section 5 explains our out-of-sample forecasting exercise with the main results presented in section 6. Section 7 concludes.

[5] Loughran and McDonald (2011) acknowledge the limitation of the fixed-dictionary approach by saying that word usage is not identical across media, especially in samples that are highly dimensional with respect to terms.

# 2 The Forecasting Problem

We want to predict the one-month ahead return to the US-CAN exchange rate, $\Delta s_{t+1}$, based on the textual information set available up to month $t$, $\Omega_t$. Our monthly textual data are collected from FACTIVA database for New York Times ($NYT$) and Wall Street Journal ($WSJ$) from January 1980 to July 2022. An advantage of our textual database is that news published at the NYT and WSJ impact a large audience not only in the United States but also in Canada. Data on US-CAN exchange rate were collected from the FRED-Saint Louis database for the 1980.1-2022.7 time period. This time period corresponds to the interval in that textual data are fully available. It also covers a long-span of economic episodes, such as: "dot.com bubble", "mortgage bubble", "European-bond crisis", "COVID-19-induced recession" and "Ukraine-war crisis". We let $s_t$ denote the logarithm of the US-CAN exchange rate measured as the number of units of Canadian dollar per one U.S. dollar. In such a background, $\Delta s_{t+1} = s_{t+1} - s_t < 0$ refers to a depreciation of the U.S. dollar. We divide the total sample of $T = R + P$ observations into in-sample and out-of-sample portions. We will obtain in total of $P$ one-month-ahead predictions based on recursive in-sample estimation.

Let $\Omega_t = (\Omega_{1t}, ..., \Omega_{Nt})'$ be an $N \times 1$ vector, where $\Omega_{it}$ measures the number of times (frequency) a given word "$i$" appears on documents published at time $t$.[6] The equation that utilizes information $\Omega_t$ is given by:

$$\Delta s_{t+1} \quad = \quad \alpha(\tau) + \Gamma(\tau)' \Omega_t + \varepsilon_{t+1}(\tau) \tag{1}$$

where $\varepsilon_{t+1}(\tau)$ satisfies the quantile restriction $Q_\tau(\varepsilon_{t+1}(\tau) | \Omega_t) = 0$. This implies that the quantile forecasting equation is

$$Q_\tau(\Delta s_{t+1} | \Omega_t) \quad = \quad \alpha(\tau) + \Gamma(\tau)' \Omega_t \tag{2}$$

[6]Texts are preprocessed before any word counting. In the next section we explain how text data enter our database.

where $\alpha(\tau)$ is an intercept and $\Gamma(\tau)$ is an $N \times 1$ vector of slope coefficients at some $\tau \in (0,1)$. We generate quantile forecasts recursively by estimating quantile regressions (Koenker and Basset, 1978) of $\Delta s_{\upsilon+1}$ on the $N \times 1$ observed predictors (words) $\Omega_\upsilon$ for $\upsilon = 1, ..., t$ and $t \in \{R, ..., T-1\}$. Under this recursive scheme, parameters are estimated at each new forecast origin by using the largest possible observation window. Specifically, at each forecast origin the parameter estimates depend explicitly on all information from 1 to $t$ where $t \in \{R, ..., T-1\}$. Thus, a quantile forecast that makes use of all available predictors is $\widehat{\Delta} s_{t+1} = \widehat{\alpha}_t(\tau) + \widehat{\Gamma}_t(\tau)\Omega_t$, where $\widehat{\alpha}_t(\tau)$ and $\widehat{\Gamma}_t(\tau)$ are recursive quantile regression (QR) estimates of $\alpha(\tau)$ and $\Gamma(\tau)$.[7]

Estimation of $\alpha_t(\tau)$ and $\Gamma_t(\tau)$ in model (2) is forbidden when $N \gg T$. In order to solve the high-dimensionality problem that affects the performance of the quantile forecasting model (2), one could extend the DI approach by Stock and Watson (2002) towards quantile regression. However, consideration of text data naturally leads to the issue of what predictors should be used to form the diffusion indices. Indeed, it may be the case that not all terms in $\Omega_t$ are important to predict the $\tau$th conditional quantile of $\Delta s_{t+1}$, i.e., some elements of $\Gamma_t(\tau)$ are indeed equal to zero, leading to the notorious weak factor model and the consequent issues in consistent estimation. In this paper, we find that the DI approach combined with text data can improve out-of-sample quantile forecasts only if one can select the most predictive textual information at forecast origin $t$ and quantile $\tau$ before computing the common factors. In what follows, we describe each step of the proposed methodology in detail.

# 3 Methodology

In this section, we describe the 3-step approach used to make quantile forecasting with textual data. Step 1 is common to the literature of natural language processing, whereas step 2 is used to identify the most predictive dictionary across time and over different quantiles of the conditional distribution. Step 3 extends the DI model of Stock and Watson (2002) towards quantile forecasting. We derive asymptotic results that are new to the literature of factor models and also important to validate our forecasting approach.

[7]Notice that $\widehat{\alpha}_t(\tau)$ and $\widehat{\Gamma}_t(\tau)$ change as we change the forecast origin $t \in \{R, ..., T-1\}$.

## 3.1 Textual Data

In this section, we explain how textual data enter the database. In order to transform news into numerical data we first save all downloaded articles into what the literature calls "a corpus", which is a collection of texts, i.e., a set of articles published in the $NYT$ and $WSJ$.

Prior to performing any word counting, we preprocess the raw texts in several steps with the purpose to reduce the vocabulary to a set of words that are mostly meaningful. For instance, we remove stopwords (e.g. the, but, for), hyphens, apostrophes, numbers, and words that only contain two characters. We also reduce words to their stem such that, for example, words as "Economy" and "Economics" are identified as "Econom". Additionally, we identify collocations or sequence of words that have a specific meaning. For example, "economic growth", "military spending", "trade deficit" correspond to a single economic concept but it is composed of two separate words. We create a single term for two-word sequences (bigrams) by using the algorithm developed by Straka and Straková (2017). As documented by Lima and Godeiro (2023), the utilization of collocations (bigrams) can improve out-of-sample forecasting substantially.[8]

Finally, as in Hansen et al. (2017), we rank the remaining terms (words and bigrams) by using their term frequency-inverse document frequency (tf-idf), which is a measure that punishes terms that are both rare and too frequent, and drop all terms ranking 100 or lower. Indeed, too frequent words are always used in documents regardless of the occurrence of important economic events and, for this reason, do not contain relevant predictive information. Rare words are mostly associated to "typos" which are not correlated with important economic events either. In order to preprocess the raw texts, we consider all articles published up to the first forecast origin. This guarantees that we are using a long period of time to reduce the original bag of words to a meaningful set of words and bigrams (i.e., a preprocessed dictionary).[9]

Given the above preprocessed dictionary, we then compute the frequency of its terms (words and bigrams) on articles published at month $t$. We save this information in a $N \times 1$ vector $\Omega_t = (\Omega_{1t}, ...\Omega_{Nt})'$,

[8] Hereafter, we use "term" to denote both words or combination of words (collocations).

[9] This is essentially the preprocessing approach of Lima and Godeiro (2023). Results are robust to using an alternative (but computationally costly) scheme in which the preprocessing of raw texts includes more recent articles - beyond the ones available up to the first forecast origin.

where each element shows the frequency that a given term $i = 1, ..., N$ (listed in the preprocessed dictionary) appears on texts published at time $t$. Notice that $N$ can be very large here since we have not used any specific dictionary to pre-process the data. In order to make these time series data suitable to principal components estimation, we standardize the data (demean and divide by the simple standard deviation) before estimating the model. In order to avoid the look-ahead bias,[10] these transformations are carried out by using time-series observations of $\Omega_t$ available up to each forecast origin $t \in \{R, ..., T-1\}$.

Thus, the above procedure transforms news from $NYT$ and $WSJ$ into an observed $N \times T$ panel of "meaningful" data, where the cross-sectional dimension, $N$, is larger than the number of time periods. How to use this information is not immediately obvious because estimates of linear quantile regressions is unfeasible when $N \gg T$, that is, estimation quickly runs into the degrees of freedom problem because of the high dimension of $\Omega_t$.[11] Moreover, the data set $\Omega_t$ is sparse and, therefore, not suitable for common factor estimation. Hence, dimension reduction techniques that retain the most predictive words have to be employed.

## 3.2 Incorporating attention into the model

In Boivin and Ng (2006), it is found that expanding the sample size simply by adding data with little information about the factor components does not necessarily improve forecasts when the data are too noisy. Furthermore, Giglio et al. (2023) find that prediction based on data with extremely sparse structure even produces biased predictions. We will remedy this issue by incorporating attention into the model, which is similar in spirit to the first step of screening employed by Giglio et al. (2024). In neural networks, attention is a technique that mimics cognitive attention. The idea is to enhance some parts of the input data while diminishing other parts. Learning which part of the data is more important than another depends on the context or the target variable being predicted. We incorporate attention into the forecasting model by targeting terms (time series) in $\Omega_t$ that are important to predict the quantiles of the exchange-rate return

[10] A looking-ahead bias occurs when the entire sample is used to compute a model parameter that is subsequently used to make out-of-sample predictions on a future that the parameter estimates has extracted information from.

[11] In the first forecast origin of the empirical section, i.e. 1989.12, there are $N = 4,250$ parameters to estimate and only 120 time series observations.

($\Delta s_{t+1}$) distribution. In doing so, we allow the dictionaries to vary not only over time, but also across the quantiles of the conditional distribution. Our approach relies on recursive estimation of elastic-net quantile regressions. In general, given a linear regression model $\Delta s_{t+1} = \alpha + \Gamma'\Omega_t + \varepsilon_{t+1}$, the coefficient vector $\Gamma = (\Gamma_1, ..., \Gamma_N)'$ is estimated by minimizing the following *elastic net* (*EN*) objective function:

$$\min_{\Gamma} \sum_{t} \mathcal{L}(\Delta s_{t+1} - \alpha - \Gamma'\Omega_t) + \lambda_{1\tau} \left\|\Gamma\right\|_1 + \lambda_{2\tau} \left\|\Gamma\right\|_2^2 \tag{3}$$

where $\mathcal{L}$ is a generic loss function whereas $\left\|\Gamma\right\|_1 = \sum_{i=1}^{N} |\Gamma_i|$ and $\left\|\Gamma\right\|_2^2 = \sum_{i=1}^{N} \Gamma_i^2$. Thus, the "Elastic Net" penalty is a combination of the *LASSO* and *Ridge* penalty. If the two tuning parameters $\lambda_{1\tau}$ and $\lambda_{2\tau}$ satisfy the relation $\frac{\lambda_{1\tau}}{\lambda_{1\tau}+\lambda_{2\tau}} > 0$, then the *EN* penalty is strictly convex, which enforces highly-correlated predictors to have similar coefficients. As a result, *EN* can capture all significant predictors even if they are highly correlated.

Lima and Godeiro (2023)'s choice for $\mathcal{L}$ was the squared loss $\mathcal{L}(t) = t^2/2$, corresponding to the least squares regression in classical regression literature. Although the squared loss is analytically simple, it is not suitable for selecting words at different quantiles. Indeed, some words may have predictive power at the low-end quantiles, but not at the high-end ones and vice versa. In order to select words for each specific quantile $\tau \in (0, 1)$, we replace the loss $\mathcal{L}(t) = t^2/2$ by the quantile loss (Koenker and Basset, 1978), that is:

$$\mathcal{L}(t) = \rho_\tau(t) = t(\tau - I(t < 0)),\ t \in R \tag{4}$$

where $0 < \tau < 1$. This is a generalization of the absolute loss for median with $\tau = 0.5$. Indeed, by choosing different $\tau$'s, one can identify the most predictive words across different quantiles by just solving problem given by (3) with loss given by (4) at different values of $\tau$'s. Thus, elastic-net quantile regression provides a powerful technique for exploring textual data heterogeneity in addition to be robust to outliers. An important output of this step is a possibly dense data set,, i.e., an $n \times 1$ vector ($n < N$) $\Omega_t^\tau = (\Omega_{1t}^\tau, ..., \Omega_{nt}^\tau)'$, being a subset of $\Omega_t$, of targeted words at quantile $\tau$ and forecast origin $t$.

An important aspect of our methodology is that the identification of the most predictive dictionary $\Omega_t^\tau$

is adapted to the problem of quantile forecasting since the word $\Omega_{it}^{\tau}$, $i = 1, ..., n$ is selected only if it helps predict the $\tau$'th quantile of $\Delta s_{t+1}$ at forecast origin $t$.[12] For example, if we change the dependent variable from exchange rate return to inflation rate, then the optimal dictionary would also change because we would be now selecting the most predictive words for the inflation rate.

This paper utilizes the novel semismooth Newton coordinate descent (SNCD) algorithm recently proposed by Yi and Huang (2017) for computing solution paths of the elastic-net quantile regression. This algorithm combines the coordinate descent method with the semismooth Newton algorithm (SNA) for solving nonsmooth equations. Yi and Huang (2017) show that such an algorithm is highly efficient and scalable in high dimensional settings and its implementation can be easily done by using the R-package "hqreg" publicly available on CRAN.[13] In the empirical section, we employ cross-validations for dependent data as in Elliott and Timmermann (2013) to tune the parameters $\lambda_{1\tau}$ and $\lambda_{2\tau}$, which is essential to avoid overfitting. Specically, we combine the `cv.hqreg` function, which allows for cross-validation in Elastic-Net penalized Huber loss regression, with the cross validation method for dependent (contiguous) data developed by the Elliot and Timmermann (2013). This is applied to all quantile forecasting models that rely on regularized methods.[14]

The above process is repeated recursively, that is, one could estimate the elastic net quantile regression for $\tau \in (0, 1)$ by regressing $\Delta s_{\upsilon+1}$ on the predictors $\Omega_{\upsilon}$ for $\upsilon = 1, ...., t$ and $t \in \{R, ..., T-1\}$.[15] The notation $t \in \{(R, ..., T-1)\}$ indicates that elastic net quantile regression is estimated using observations up to a specific updating period. However, in order to minimize computational costs and avoid the look-ahead bias when updating the dictionary, we initially select the most predictive dictionary at the first forecast origin $R$ and then updated it at later periods by using an updating rule based on the Sahm recession indicator.[16] Such an indicator signals the start of a recession when the three-month moving av-

[12] Subsection 3.2 will provide details as to how such a word selection is performed. Also note that $n$ is allowed to go to infinity with $N$.

[13] (http://cloud.rproject.org/package=hqreg). Convergence of the SNCD iterations to a global minimizer is also established by Yi and Huang (2017).

[14] In an attempt to disseminate the methodology and render the results reproducible, we provide R codes that replicate all results presented in this paper.

[15] We end at $T-1$ because we need to use observation $T$ to evaluate forecasts made at $T-1$.

[16] Sahm index is available here: https://fred.stlouisfed.org/series/SAHMREALTIME

erage of the national unemployment rate rises by 0.50 percentage points or more relative to its low during the previous 12 months. Thus, we update the dictionary at time $t$ and quantile $\tau$ if the Sahm indicator has gone above (below) 0.5 at time $t-1$. For example, if the indicator went above 0.5 in 2001.06, then we optimize (3) with loss (4) in the forecast origin 2001.7 to obtain the most predictive dictionary available at quantile $\tau \in (0,1)$ and forecast origin 2001.7. This dictionary stays the same until 2002.11 when the indicator goes below 0.5. By following this rule, the empirical section of this paper updates the dictionary in these months: 1990.2, 1991.1, 1992.12, 2001.7, 2002.12, 2008.6, 2010.7, 2020.5 and 2021.5. This updating rule attempts to mimic a real-time forecasting exercise faced by professional forecasters in that she (he) decides to update the dictionary when the Sahm indicator signals either the start or the end of a recession.[17]

### 3.3 The factor-based forecasting

We have used elastic-net quantile regression to select the most predictive terms at specific quantiles. These terms are separate but correlated observables, which means that one can exploit their intrinsic factor structure to further extract concentrated information. Specifically, it is expected that

$$\Omega_{it}^{\tau} = \lambda_{i,\tau}' F_{t,\tau} + e_{it}, \tag{5}$$

where $\lambda_{i,\tau}$ is the exposure (loading) of word $i$ to the topic (factor) $F_{t,\tau}$, both being $\ell \times 1$, and $e_{it}$ is the idiosyncratic frequency of word $i$ at time $t$. Since factors are loaded with quantile-dependent information, $\Omega_t^{\tau}$, the notation used here for loadings and factors (i.e., $\lambda_{i,\tau}$ and $F_{t,\tau}$) stress such a quantile dependence.

The factor-based approach to one-month-ahead quantile forecasting quantile is defined as

$$Q_{\tau}\left[\Delta s_{t+1} \mid \underline{\Omega}_t^{\tau}\right] = \alpha(\tau) + \beta(\tau)' F_{t,\tau}, \tag{6}$$

where $\underline{\Omega}_t^{\tau} = (\Omega_1^{\tau\prime}, \Omega_2^{\tau\prime}, \ldots, \Omega_t^{\tau\prime})'$. Equations (5) and (6) thus constitute the "diffusion index" framework

[17] In the empirical section of this paper, we will set $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$.

of Stock and Watson (2002) without lags in $F_{t,\tau}$ and adapted to the problem of quantile forecasting with targeted text data.

Given that $F_{t,\tau}$ is latent, we estimate it from the least squares (LS) problem:

$$\min_{\{\lambda_{i,\tau}, F_{s,\tau}\}} \frac{1}{nt} \sum_{i=1}^{n} \sum_{s=1}^{t} \left(\Omega_{is}^{\tau} - \lambda_{i,\tau}' F_{s,\tau}\right)^2, \tag{7}$$

subject to the identification restrictions imposed by Bai (2003). Intuitively, the estimated factors are linear combinations of each element of the $n \times 1$ vector $\Omega_t^{\tau} = (\Omega_{1t}^{\tau}, ..., \Omega_{nt}^{\tau})'$ where the linear combination is chosen optimally to minimize the sum of squared residuals as shown in equation (7). This makes clear that the $\ell$ estimated factors are only loaded with the most predictive textual information available at quantile $\tau$ and forecast origin $t$ ($\Omega_t^{\tau}$). This is an important aspect of the methodology proposed in this paper as it will be essential to improve out-of-sample quantile forecasts. Finally, for each forecast origin $t$ and quantile $\tau$, we apply the method developed by Ahn and Horenstein (2013) to determine the value of $\ell$.

The resulting factor estimator, denoted as $\widetilde{F}_{t,\tau}$, is an estimator for $HF_{t,\tau}$, with the rotation matrix $H$ defined in Bai (2003). Due to the presence of the matrix $H$, let us rewrite (6) as

$$Q_{\tau}\left[\Delta s_{t+1} \mid \underline{\Omega}_t^{\tau}\right] = \alpha(\tau) + \left(H^{-1}\beta(\tau)\right)' HF_{t,\tau} \equiv \delta(\tau)'\bar{Z}_{t,\tau}, \tag{8}$$

where $\delta(\tau) = \left(\alpha(\tau), (H^{-1}\beta(\tau))'\right)'$, and

$$\bar{Z}_{t,\tau} = \left(1, F_{t,\tau}'\right)'.$$

To estimate $Q_{\tau}\left[\Delta s_{t+1} \mid \underline{\Omega}_t^{\tau}\right]$, we will first replace $\bar{Z}_{t,\tau}$ by the feasible estimator

$$\widehat{Z}_{t,\tau} \equiv \left(1, \widetilde{F}_{t,\tau}'\right)',$$

and then estimate the parameter $\delta(\tau)$ by quantile regression to obtain $\widehat{\delta}(\tau) = \left(\widehat{\alpha}(\tau), \widehat{\beta}(\tau)'\right)'$. That is,

$$\widehat{\delta}(\tau) = \arg\min_{\delta} \sum_{s=1}^{t} \rho_\tau \left(\Delta s_{s+1} - \delta' \widehat{Z}_{s,\tau}\right),$$

with the check loss function defined previously in (4). Then our estimator of target is

$$\widehat{Q}_\tau \left(\Delta s_{t+1} \mid \underline{\Omega}_t^\tau\right) = \widehat{\delta}(\tau)' \widehat{Z}_{t,\tau}. \tag{9}$$

The out-of-sample forecasting is also implemented in the same recursive scheme as introduced previously.

# 4 Asymptotic Theory

We will show the asymptotic distributions of estimated coefficient $\widehat{\delta}(\tau)$ as well as the quantile forecast $\widehat{Q}_\tau \left(\Delta s_{t+1} \mid \underline{\Omega}_t^\tau\right)$. To this end, some regularity conditions are needed for each $\tau$. Hereafter, $M$ denotes a bounded positive constant.

**Assumption 1** $\{F_{s,\tau}, e_{is}^\tau\}_{s\geq 1}$ *is stationary.*

**Assumption 2** *The conditional distribution* $F\left(x \mid \underline{\Omega}_s^\tau\right) = P\left(\Delta s_{s+1} \leq x \mid \underline{\Omega}_s^\tau\right)$ *admits a density function* $f\left(x \mid \underline{\Omega}_s^\tau\right)$ *being bounded and continuous in* $x$.

**Assumption 3** *(a)* $E\left\|F_s\right\|^4 \leq M$ *and* $\frac{1}{t}\sum_{s=1}^{t} F_s F_s' \xrightarrow{p} \Sigma_F$ *for a* $\ell \times \ell$ *positive definite (non-random) matrix* $\Sigma_F$.

*(b) The loading* $\lambda_i$ *is either deterministic such that* $\|\lambda_i\| \leq M$, *or it is stochastic such that* $E\|\lambda_i\|^4 \leq M$. *In either case,* $\frac{1}{n}\sum_{i=1}^{n} \lambda_i \lambda_i' \xrightarrow{p} \Sigma_\Lambda$ *as* $n \to \infty$ *for some* $\ell \times \ell$ *positive definite non-random matrix* $\Sigma_\Lambda$.

**Assumption 4** *(a)* $E\left(e_{is}\right) = 0, E\left|e_{is}\right|^8 \leq M$*;*

*(b)* $E\left(e_{is}e_{js'}\right) = \sigma_{ij,rs}, |\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ *for all* $(s, s')$ *and* $|\sigma_{ij,ss'}| \leq \tau_{ss'}$ *for all* $(i, j)$ *such that*

$$\frac{1}{n}\sum_{i,j=1}^{n} \bar{\sigma}_{ij} \leq M, \frac{1}{T}\sum_{s,s'=1}^{t} \tau_{ss'} \leq M, \text{ and } \frac{1}{nt}\sum_{i,j,s,s'=1}^{n,t} |\sigma_{ij,ss'}| \leq M$$

*(c) For every* $(s, s')$, $E\left|n^{-1/2}\sum_{i=1}^{n}\left[e_{is}e_{is'}-E\left(e_{is}e_{is'}\right)\right]\right|^4 \leq M$.

*(d) For each* $t$, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i e_{it} \xrightarrow{d} N\left(0,\Gamma_t\right)$, *where* $\Gamma_t = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(\lambda_i\lambda_j{}'e_{it}e_{jt}\right)$.

**Assumption 5** $Q_\tau\left[\Delta s_{t+1} \mid \underline{\Omega}_t^\tau\right] = \alpha(\tau)+\beta(\tau)'F_{t,\tau}$, *where* $\underline{\Omega}_t^\tau = \left(\Omega_1^{\tau\prime}, \Omega_2^{\tau\prime}, \ldots, \Omega_t^{\tau\prime}\right)'$.

**Theorem 1** *Suppose Assumptions 1-5 hold. If* $\frac{\sqrt{t}}{n}\to 0$, *then*

*(a) for the estimated coefficient* $\hat{\delta}(\tau)$,

$$\sqrt{t}(\hat{\delta}(\tau)-\delta(\tau)) \xrightarrow{d} N\left(0, \tau(1-\tau)A(\tau)^{-1}E\left(\bar{Z}_{s,\tau}\bar{Z}_{s,\tau}'\right)A(\tau)^{-1}\right),$$

*where* $A(\tau) = E\left[f\left(Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) \mid \underline{\Omega}_s^\tau\right)\bar{Z}_{s,\tau}\bar{Z}_{s,\tau}'\right]$;

*(b) for the estimated forecast*

$$\widehat{Q}_\tau\left(\Delta s_{t+1} \mid \underline{\Omega}_t^\tau\right) - Q_\tau\left(\Delta s_{t+1} \mid \underline{\Omega}_t^\tau\right) \xrightarrow{d} N(0, \Xi_{t,\tau}).$$

*where* $\Xi_{t,\tau} = \frac{1}{t}\widehat{Z}_{t,\tau}'Avar\left(\hat{\delta}(\tau)\right)\widehat{Z}_{t,\tau} + \frac{1}{n}\hat{\beta}(\tau)'Avar\left(\widetilde{F}_{t,\tau}\right)\hat{\beta}(\tau)$.

**Remark 1** *Theorem 1 reveals several interesting linkage of our asymptotic theory to that established by Bai and Ng (2006) for the classical DI forecasts. First, the asymptotic distribution of* $\hat{\delta}(\tau)$ *is the same as if the latent factors were observed, implying that the estimation error from PCA is negligible. Second, the quantile forecast* $\widehat{Q}_\tau\left(\Delta s_{t+1} \mid \underline{\Omega}_t^\tau\right)$ *convergences to the truth at the rate of* $O_p(n^{-1/2}+t^{-1/2})$, *whose distribution relies on those of* $\hat{\delta}(\tau)$ *and* $\widetilde{F}_{t,\tau}$. *Both of these findings coincide with results from Bai and Ng (2006).*

**Remark 2** *In practice, we can estimate* $Avar\left(\hat{\delta}(\tau)\right)$ *by*

$$\widehat{Avar\left(\hat{\delta}(\tau)\right)} = \widehat{A}(\tau)^{-1}\widehat{D}(\tau)\widehat{A}(\tau)^{-1},$$

*where* $\widehat{D}(\tau) = \frac{\tau(1-\tau)}{t}\sum_{s=1}^{t}\widehat{Z}_{s,\tau}\widehat{Z}_{s,\tau}'$, *and* $\widehat{A}(\tau) = (th)^{-1}\sum_{s=1}^{t}K\left(\frac{\Delta s_{s+1}-\widehat{Z}_{s,\tau}'\hat{\delta}(\tau)}{h}\right)\widehat{Z}_{s,\tau}\widehat{Z}_{s,\tau}'$ *with* $K(\cdot)$ *being a kernel function and* $h$ *being the bandwidth. The consistency of* $Avar\left(\hat{\delta}(\tau)\right)$ *follows Ma et al.*

*(2021) together with the specifications of $K(\cdot)$ and $h$. Moreover, we can estimate $Avar\left(\widetilde{F}_{t,\tau}\right)$ consistently by*

$$\widehat{Avar}\left(\widetilde{F}_{t,\tau}\right) = \widetilde{V}^{-1}\widetilde{\Gamma}_{t,\tau}\widetilde{V}^{-1}.$$

*Here $\widetilde{V}$ is the $r \times r$ diagonal matrix consisting of the $r$ largest singular values of $\underline{\Omega}_t^{\tau}$, and $\widetilde{\Gamma}_{t,\tau}$ is proposed by $(5c)$ by Bai and Ng (2006) to allow for cross sectional dependence of $e_{it}$.*

# 5 Forecast Evaluation

This section investigates if selected textual data can add predictability at different quantiles of the US-CAN return distribution. It is known that factors obtained by PCA using LS criteria are sensitive to outliers and heavy-tailed errors. For robustness, we consider the Quantile Factor Analysis (QFA) estimator recently developed by Chen et al. (2021). Indeed, the QFA for $\tau = 0.5$ can be interpreted as a least absolute deviation ($LAD$) estimator of common factors which is more robust than PCA in the presence of outliers and heavy-tails distribution. Hence, at $\tau = 0.5$, the QFA can be defined as

$$\min_{\{\lambda_{i,\tau}, F_{s,\tau}\}} \frac{1}{nt} \sum_{i=1}^{n} \sum_{s=1}^{t} \left| \Omega_{i,s}^{\tau} - \lambda_{i,\tau}' F_{s,\tau} \right| \tag{10}$$

In order to estimate common factors via QFA, Chen et al. (2021) proposed a computational algorithm called "iterative quantile regression" ($IQR$), which can effectively locate the stationary points of the object function (10). Chen et al. (2021) developed a method, called rank minimization through which one can select the correct number of factors under the QFA approach. They also establish that the QFA factor estimator is consistent with a similar asymptotically first order expansion to that from the PCA estimator, implying that it is possible to derive the corresponding asymptotic theory for the QFA factor based quantile forecast, following the analysis conducted in Section 4, which we leave for the future research.

We evaluate different conditional quantile models in terms of their out-of-sample forecasting performance. Such an evaluation relies on measures of accuracy relative to the benchmark unconditional quantile model. It is important to keep in mind that each model evaluated in this section adopts the same recursive estimation scheme towards the end of the sample; makes direct one-step forecasts of exchange-rate returns; uses same number of observations to estimate quantile models for $\tau \in (0.05, 0.25, 0.50, 0.75, 0.95)$ and same number of out-of-sample observations to evaluate their respective quantile forecasts.[18] They differ in terms of predictors included in the model (economic versus textual-based predictors), the method used to estimate the common factors, and the method used to construct the density forecast (Epanechnikov Kernel versus skewed-t distribution). In what follows, we describe each quantile model.[19]

(1) **Quantile forecasting using words selected by PCA**. This is our preferred model. It uses elastic-net quantile regression to target the most predictive words at quantile $\tau$ and forecast origin $t$. Then, common factors of targeted textual information are estimated by PCA. Finally, the linear quantile regression (9) is used to obtain (out-of-sample) quantile forecasts. This is done for $\tau = 0.05, 0.25, 0.50, 0.75, 0.95$. This is a dense quantile forecasting model since all selected terms (times series) have a role to play. We label this model as $\widehat{\Delta} s_{t+1|\Omega_t^\tau}$;

(2) **Quantile forecasting using words selected by QFA**. Same as model (1), but common factors of selected words are computed by using QFA at the median. We label this model as $\widehat{\Delta} s^{QFA}_{t+1|\Omega_t^\tau}$;

(3) **Quantile forecasting using full (sparse) textual information set**. We do not use elastic-net quantile regression to select the most predictive words. Instead, we follow Chen et al. (2021) in that $\tau$-dependent common factors are computed using QFA. Then, a linear quantile regression as in (9) is used to obtain (out-of-sample) quantile forecasts by using the factors from QFA. Factor models may not work well with sparse data, so we do not expect a good out-of-sample forecasting performance from this model. We label this model as $\widehat{\Delta} s_{t+1|\Omega_t}$;

(4) **Quantile forecasting using economic indicators**. Same as (3), but instead of textual data, we use

[18] Our out-of-sample period starts at 1990.1 and ends in 2022.7.

[19] For the first three factor-based quantile models below, we further reduce the used factors by only keeping those significant at $5\%$ level in empirical study, although the resulting difference is negligible.

a database of 105 time series of economic indicators (FRED-MD) listed in McCracken and Ng (2016) web-page.[20] We label this model as $\widehat{\Delta} s_{t+1|M_t}$;

(5) **Quantile forecasting using targeted economic indicators**. Same as (1) and (2), respectively. However, instead of textual data, we use elastic net quantile regression to target the economic indicators in the FRED-MD database. We label this model as $\widehat{\Delta} s_{t+1|M_t^\tau}$ and $\widehat{\Delta}^{QFA} s_{t+1|M_t^\tau}$, respectively .

(6) **Quantile forecasting using a fixed (ad-hoc) dictionary**. Same as (1), but we replace the full "bag of words" $\Omega_t$ by the set of 4,150 words from the (fixed) dictionary of Loughran and McDonald (2011). Everything else is the same as in (1). We label this quantile forecast as $\widehat{\Delta} s_{t+1|LM_t^\tau}$.

Following Li et al. (2015), Engel et al. (2019), and others, we use the random walk (RW) model as the benchmark to which all other models are compared to when assessing forecasting performance. This is standard practice within the exchange-rate forecasting literature. The quantile equivalent of RW model is the $\tau$'th unconditional quantile computed using all the observations available up to the forecast origin $t$. The $\tau$'th unconditional quantile is equivalent to a linear quantile regression with an intercept only. Taking the unconditional quantile model as a benchmark also facilitates the comparison of the proposed text-data based quantile forecasts to other conditional quantile forecasts. We label forecasts based on the unconditional quantile model as $\overline{\Delta} s_{t+1}$.

**Remark 3** *A more rigorous forecasting evaluation of various quantile models with the benchmark would be based on a statistical tests of equal accuracy for nested models. Tests of this type under mean forecasting have been proposed by McCracken (2007) and Gonçalves et al. (2017) with or without estimated factors, among others. However, the extension to quantile forecasting is challenging mainly due to two issues. First, the object function is not differentiable so that the standard analysis such as in McCracken (2007) is not applicable. Second, the score function associated quantile regression comes with an additional (conditional) density component which is an unknown nonparametric function to be estimated, and this implies that the normalization employed in McCracken (2007, for the test statistic OOS-$F$) or*

[20] See appendix for a complete description of this database.

*Gonçalves (2017, for the test statistic $MSE\text{-}F_{\tilde{f}}$) may not be simply a constant any more. A possible solution to the tackle this problem is to study the loss differential by directional derivative of the quantile loss used in Chen et al. (2024), or to consider bootstrapping the test following Tchatoka and Haque (2023).*

To gauge the goodness-of-fit of linear quantile regressions, Koenker and Machado (1999) proposed a Pseudo R-squared statistic. In this paper, we follow Narayan et al. (2021) and compute the out-of-sample Pseudo R-squared, which compares the out-of-sample performance of the unconditional quantile benchmark ($\overline{\Delta}s_{t+1,\tau}$) to conditional quantile forecasts $\widehat{\Delta}s^i_{t+1,\tau}$, where

$$\widehat{\Delta}s^i_{t+1,\tau} \in (\widehat{\Delta}s_{t+1|\Omega^\tau_t}; \widehat{\Delta}^{QFA}s_{t+1|\Omega^\tau_t}; \widehat{\Delta}s_{t+1|\Omega_t}; \widehat{\Delta}s_{t+1|M_t}, \widehat{\Delta}s_{t+1|M^\tau_t}; \widehat{\Delta}^{QFA}s_{t+1|M^\tau_t}, \widehat{\Delta}s_{t+1|LM^\tau_t}) \tag{11}$$

Thus, we define the Pseudo $R^2_{OOS}(\tau)$ as:

$$R^2_{OOS}(\tau) = 1 - \frac{\sum\limits_{t=R}^{T-1} \rho_\tau(\Delta s_{t+1} - \widehat{\Delta}s^i_{t+1,\tau})}{\sum\limits_{t=R}^{T-1} \rho_\tau(\Delta s_{t+1} - \overline{\Delta}s_{t+1,\tau})} \tag{12}$$

where $\rho_\tau(\cdot)$ is the $\tau$'th quantile loss function, $\widehat{\Delta}s^i_{t+1,\tau}$ is the $\tau$'th quantile forecast obtained using conditional quantile model $i = 1, ..., 7$, respectively, whereas $\overline{\Delta}s_{t+1,\tau}$ in the denominator is the corresponding unconditional quantile forecast.[21] A positive $R^2_{OOS}(\tau)$ means the alternative model has lower loss $\rho_\tau(\cdot)$ $i$ and, therefore, it outperforms the unconditional quantile benchmark at quantile $\tau$. Additionally, we will plot the cumulative sum of loss differentials to pin down the relative performance of each conditional quantile model.[22]

Point forecasts is a measure of central tendency of the exchange rate return, but policy makers and investors are concerned about depreciation and appreciation risks of exchange rates, especially during episodes of unexpected shocks. In this paper, we follow the literature, i.e., Adrian et al. (2019),

[21] Recall that the forecast $\widehat{\Delta}s^i_{t+1}$ in the numerator and the unconditional quantile forecast $\overline{\Delta}s_{t+1}$ in the denominator are predicting the same $\tau$'th quantile of the exchange-rate return.

[22] Loss differentials are defined as $D_\tau = [\rho_\tau(\Delta s_{t+1} - \overline{\Delta}s_{t+1,\tau}) - \rho_\tau(\Delta s_{t+1} - \widehat{\Delta}s^i_{t+1,\tau})]$. Notice this plot varies across quantiles as well.

Gaglianone and Lima (2012), Gaglianone and Lima (2014), among others, in that the available information is used to construct density forecasts. Such density forecasts will then be used to forecast the probability of depreciation (appreciation) of the US-CAN exchange rate, which is important to formulate directional trading strategies involving such currencies. Our density forecasts will be obtained by using the skewed-t distribution approach by Adrian et al. (2019) and the Epanechnikov kernel approach by Gaglianone and Lima (2012). We will compare the performance of these methods by using the log predictive density score (LPDS). Our choice of recursive forecasts is based on the fact that quantile regressions require large samples, especially when estimation occurs at extreme values of $\tau$, such as $\tau = 0.05, 0.95$. For this reason, we are left with few density evaluation tests that works with recursive forecasting. Following the literature on density forecast evaluation, we apply the density test of Berkowitz Berkowitz (2001) for one-step ahead forecasts. We quantify upside and downside entropy of future exchange rate return as the "extra" probability mass that the conditional density assigns to extreme right and left tail outcomes, relative to the probability of these outcomes under the unconditional density (Adrian et al., 2019, page 1276). We will use such entropy to demonstrate how our predicted probabilities of depreciation (appreciation) of the U.S. dollar are associated with downside (upside) risks.

We also evaluate the performance of our density forecasts in terms of the cumulative wealth of a portfolio that utilizes the predicted probability of depreciation (appreciation) of the U.S. dollar. We consider a Canadian firm that import most of its inputs from the U.S. and, for this reason, its cash flow is highly affected by fluctuations of the U.S. dollar. Accordingly, changes in the US-CAN exchange rate, and in particular a strengthening of the U.S. dollar, will negatively affect the Company's profit and gross margins. There is a risk that the company will have to adjust local currency product pricing due to competitive pressures if the U.S. dollar appreciates significantly. We assume this company hedges against exchange rate risk by using a simple managed portfolio, which consist of a single risky asset (US-CAN exchange rate) and a risk free asset (Canadian t-bond). We use our density forecasts to predict the probability of U.S. dollar depreciation at $t+1$ based on information available up to time $t$, i.e., $p_{t+1,t} = prob[(\Delta s_{t+1} < 0)|\Im_t]$, where $\Im_t$ is a measurable information set at time $t$.[23] Let $I_{t+1,t}$ be an indicator function that equals 1 if $p_{t+1,t} > 0.5$ and zero otherwise. We assume the optimal portfolio allocation is fully invested on the risk-free asset (Canadian t-bonds) at the end of time $t$ whenever an investor predicts a probability of depreciation of the U.S. dollar above 50%, that is $I_{t+1,t} = 1$ or equivalently $p_{t+1,t} > 0.5$, and fully invest on the U.S. dollar at the end of time $t$ whenever the probability of appreciation of the US dollar is above or equal to 50%, that is, when $I_{t+1,t} = 0$ or equivalently $p_{t+1,t} \leq 0.5$. As shown by de de Castro and Galvao (2022), this trading strategy is expected in quantile preference settings where the presence of a risk-free asset produces an all-or-nothing optimal allocation in that investors avoid downside risks by investing their wealth on the risk-free asset when they expect a depreciation of the U.S. dollar (negative return rate) and invest their wealth on the U.S. dollar when they expect an appreciation of the U.S. dollar.

We evaluate the above portfolio in terms of its cumulative performance, that is, we will compute the cumulative wealth of an investor who makes the above investment decisions and faces a transaction cost of $c$ dollars. Such a transaction cost aims to penalize investors who re-balance their portfolio too often. We want to learn whether different information sets used to calculate the directional forecasts $I_{t+1,t}$ will ultimately affect the investor's cumulative wealth, and by how much. In particular, we are interested to learn whether an investor who uses text-based directional forecasts will be wealthier than those who do not use it.

# 6 Empirical Results

Table A in the online appendix describes the US-CAN exchange rate return for the the out-of-sample period. It has an annualized mean of about 0.30%, autocorrelation close to zero, and its distribution is fairly symmetric with excess kurtosis, which is consistent with a high annualized volatility of 7.07%. Thus, an important question is whether such excess kurtosis could be explained by economic or textual predictors and if one could use such a predictability to improve investment decisions. We recall that $\Delta s_{t+1} < 0$ indicates a U.S. dollar depreciation and that a Canadian firm whose cash flow is sensitive to changes in

[23] The probability of appreciation is $1 - p_{t+1,t}$

the U.S dollar would be interested to predict $\Delta s_{t+1}$ in order to hedge itself against exchange rate risks. However, as recently shown by Filippou et al. (2021), even a large set of economic variables does not improve predictability of the US-CAN exchange rate at the conditional mean (i.e., point forecasting). This paper wonders if (targeted) textual data increases predictability of the US-CAN exchange rate and if such a predictability is more concentrated in the tails of the conditional distribution.

Table 1 displays the results for the pseudo-R2 as developed by Koenker and Machado (1999). One can see that the method proposed in this paper ($\widehat{\Delta} s_{t+1|\Omega_t^\tau}$ and $\widehat{\Delta}^{QFA} s_{t+1|\Omega_t^\tau}$) indicates the existence of predictability of the US-CAN exchange-rate return in the tails of the distribution, specially at the left tail. This is particularly useful for investors concerned about avoiding downside risks.[24] Although using LASSO to select the most predictive words improves forecasting accuracy relative to the benchmark, it does not improve over the corresponding model that uses elastic net. This happens because LASSO is bounded by the sample size, $T$, meaning that important conditioning information may be thrown away even if it is contains a predictive content. Moreover, LASSO is not robust to group effects, which is particularly important to our analysis because the occurrence of many words, such as inflation and unemployment, may be highly correlated.[25] This explains the relative good performance of elastic net as it not only selects the most predictive words, but also guarantees that equally predictive words will not be thrown away. In the forecasting literature, a short list of papers that successfully employed elastic net includes Bai and Ng (2008), Li et al. (2015), Li (2015), Lima et al. (2020) and Lima and Godeiro (2023). We also notice that models based on selected economic variables ($\widehat{\Delta} s_{t+1|M_t^\tau}$ and $\widehat{\Delta}^{QFA} s_{t+1|M_t^\tau}$ ) as well as the model based on a fixed dictionary ($\widehat{\Delta} s_{t+1|LM_t^\tau}$) perform no better than the benchmark unconditional quantile model.

[24]Even an apparently small degree of return predictability (low out-of-sample pseudo $R^2_{OOS}(\tau)$ can translate into substantial financial gains to investors - see Campbell and Thompson (2008).

[25]When predictors are highly correlated, LASSO tends to select only one variable from the group and does not care which one - Bai and Ng (2008, page 307).

Table 1: Out-of-sample Pseudo $R^2_{OOS}(\tau)$

| Model/Quantile | 05 | 25 | 50 | 75 | 95 |
|---|---|---|---|---|---|
| $\widehat{\Delta s}_{t+1\vert\Omega_t^\tau}$ | 3.78 | 3.01 | 1.77 | 2.40 | 1.84 |
| $\widehat{\Delta s}_{t+1\vert\Omega_t^\tau(LASSO)}$ | 1.79 | 1.42 | 0.74 | 1.17 | 0.99 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\vert\Omega_t^\tau}$ | 3.61 | 2.45 | 1.97 | 2.29 | 1.57 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\vert\Omega_t^\tau(LASSO)}$ | 1.73 | 1.24 | 0.90 | 1.15 | 0.60 |
| $\widehat{\Delta s}_{t+1\vert\Omega_t}$ | 3.20 | 0.87 | 0.07 | 0.22 | 0.97 |
| $\widehat{\Delta s}_{t+1\vert LM_t^\tau}$ | -0.05 | 2.04 | 0.09 | 0.12 | -0.92 |
| $\widehat{\Delta s}_{t+1\vert LM_t^\tau(LASSO)}$ | -0.92 | 0.75 | -0.24 | -0.70 | -1.37 |
| $\widehat{\Delta s}_{t+1\vert M_t^\tau}$ | 0.34 | -0.85 | -0.59 | -1.98 | -1.25 |
| $\widehat{\Delta s}_{t+1\vert M_t^\tau(LASSO)}$ | -0.45 | -1.12 | -0.99 | -2.32 | -1.65 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\vert M_t^\tau}$ | -0.76 | -0.76 | -0.41 | -0.43 | -1.77 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\vert M_t^\tau(LASSO)}$ | -0.96 | -1.22 | -0.71 | -0.62 | -2.28 |

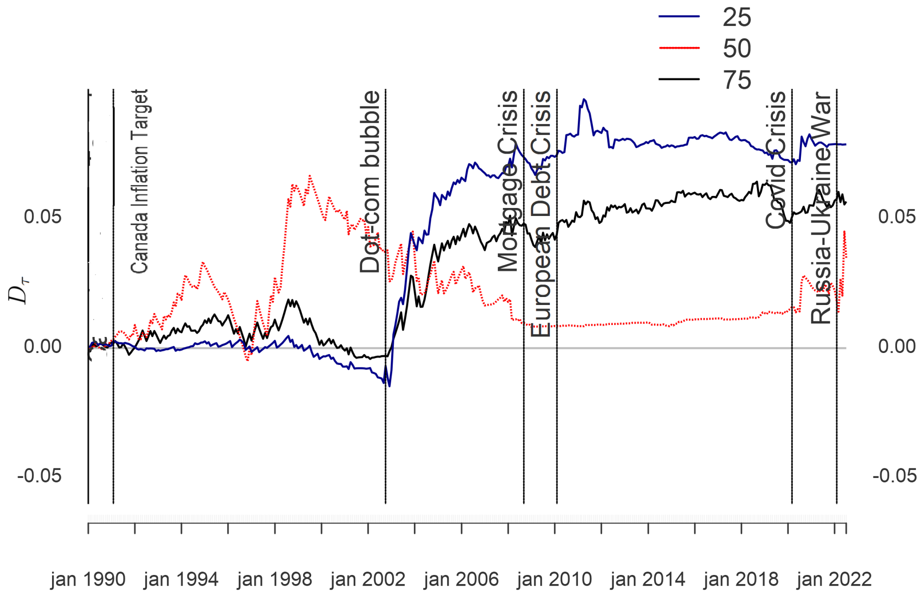
The table reports Koenker and Machado (1999) $Pseudo-R^2$ statistics in percentage value for out-of-sample forecasts of monthly log exchange rate changes ($\widehat{\Delta} s_{t+1}|\Im_t$), where $\Im_t$ is one of the information sets listed on the the table. The out-of-sample pseudo-$R^2$ statistic measures the proportional reduction in quantile loss $\rho(\tau)$ of the conditional model vis-á-vis the unconditional model.

Figures 1 presents time series plots of the cumulative difference between the prediction loss of the historical (unconditional) quantile forecast and the prediction loss of the quantile forecasts based on (selected) textual data. We show results for quantiles $\tau = 0.25, 0.50, 0.75$ in the out-of-sample period. This graphical analysis informs the cumulative performance of a given forecasting model compared to the benchmark model over time and across selected quantiles. When the curve in each panel increases, the conditional model outperforms the benchmark, while the opposite holds when the curve decreases. Moreover, if the curve is higher at the end of the period, the conditional model has a lower loss $\rho_\tau(\cdot)$ than the benchmark over the whole out-of-sample period. Figure 1 shows that before the blast of the "dot.com" bubble, our text-based quantile forecasting model outperforms the historical (unconditional) one at the median ($\tau = 0.50$) but not so much at the tails of the distribution ($\tau = 0.25, 0.75$), implying that (selected) textual information was useful to compute point forecasts, but not very relevant to predict events in the tails of the US-CAN return distribution. This situation changes substantially after 2002, as the text-based quantile model starts finding more (less) predictability at the tails (median) of the US-CAN return distribution. Indeed, as shown in Figure 1, tail predictability with (selected) textual data is particularly strong during periods of market turmoil, such as the mortgage crisis, European-bond crisis, COVID-19-

induced recession, and the Russia-Ukraine war. This result suggests that selected textual information could be useful to predict probabilities of depreciation (appreciation) of the US dollar, helping investors to formulate directional trading strategies.

Figures A in the online appendix shows what happens when we do not select the most predictive terms over time and across quantiles. Indeed, one can see that the quantile forecasting model based on sparse (non-targeted) textual data hardly outperforms the benchmark (unconditional) quantile model for $\tau = 0.25, 0.50, 0.75$. This result confirms that incorporating attention into the quantile forecasting model is essential to improve the out-of-sample forecasting performance of the text-based quantile model. Figures B and C in the online appendix show what happens when we replace targeted textual data with targeted economic data. Remember that the database $M_t$ contains information about 105 time series of economic indicators listed in McCracken and Ng (2016) - see appendix for a complete description of this dataset - and the quantile forecasts $\widehat{\Delta} s_{t+1|M_t^\tau}$ rely on factors loaded with selected economic indicators over time and across quantiles. These figures confirm the findings by Filippou et al. (2021) in that it is hard to predict the US-CAN return by using economic indicators. Additionally, our results suggest that economic indicators also fail to predict the quantiles of the US-CAN return distribution. The median forecast ($\tau = 0.5$) provides a point (central) forecast for exchange rate return, but policy makers as well as investors are also concerned about downside and upside risks associated to future values of exchange-rate returns during episodes of unexpected shocks. Our preliminary results point out to the existence of predictive content of targeted textual data, especially in the tails of the distribution and during periods of economic turmoil. These results provide the first empirical evidence in favor of the predictive content of targeted textual data. In what follows, we provide more statistical and economic evidence to further support the methodology developed in this paper.

Figure 1: Cumulative Sum of Loss Differentials: $\widehat{\Delta}s_{t+1|\Omega_t^\tau}$ versus $\overline{\Delta}s_{t+1,\tau}$




This figure shows the cumulative of loss differentials $D_\tau = [\rho_\tau(\Delta s_{t+1} - \overline{\Delta}s_{t+1,\tau})] - [\rho_\tau(\Delta s_{t+1} - \widehat{\Delta}s_{t+1|\Omega_t^\tau})]$. Notice this plot varies across quantiles as well. The blue, red, and black curves show plots when $\tau = 0.25, 0.50, 0.75$, respectively. An increasing curve means the conditional model ($\widehat{\Delta}s_{t+1|\Omega_t^\tau}$) outperforms the unconditional one ($\overline{\Delta}s_{t+1,\tau}$).

## 6.1 Density, interval, and probability forecasts

We next use the quantile forecasting models to construct density forecasts. This is done by interpolating conditional quantiles through the skewed-t student distribution approach suggested by Adrian et al. (2019) as well as the Epanechnikov Kernel approach suggested by Gaglianone et al. (2012b) and Gaglianone and Lima (2014). For the former, we follow Adrian et al. (2019) and use quantile forecasts at $\tau = (0.05, 0.25, 0.75, 0.95)$. For the Epanechnikov Kernel, we additionally include the conditional median ($\tau = 0.5$). We keep the number of quantiles small to be fair with the skewed-t distribution method that relies on a few quantiles. Of course more quantiles could be used in both methods at the expense of higher computational costs.[26] Table 2 reports the results of the log-predictive score and their respective p-values. Two main results stand out. First, the density forecast constructed by using the proposed text-based quantile model $\widehat{\Delta}s_{t+1|\Omega_t^\tau}$ has the highest LPDS (the higher, the better) among all densities constructed by using other quantile forecasting models (i.e., models that rely on targeted and non-targeted economic

[26] As shown by Adrian et al. (2019), one could use more quantiles and allow the parameters of the skewed t-distribution to be over-identified.

indicators, non-targeted textual data, fixed dictionary and the unconditional quantile model). Second, for the proposed quantile model $\widehat{\Delta}s_{t+1|\Omega_t^\tau}$, the densities computed by using either the skewed-t distribution or the Epanechnikov Kernel seem to have similar LPDS, but the latter slightly outperforms the former. If we rank the models in terms of the p-value of Berkowitz (2001) and the LPDS scores, then one can see that both rankings agree with each other. Indeed, densities forecasts constructed from the text-based quantile models outperforms those based on economic data, and the unconditional quantile model yields the worst density forecast. Among the models that uses text data, the ones that rely on targeted textual information are the best, and PCA works very well with targeted textual data.

Table 2 also shows that the method used to select the most predictive words does affect the out-of-sample forecasting performance of the proposed method. Our results show that using LASSO to select the words does increase forecasting performance of the DI model, but greater improvement is obtained when elastic net is used. Indeed, although $LASSO$ is successful at variable selection, in the particular case where the number of words is larger than the sample size, $LASSO$ selects at most $T$ words before it saturates, therefore excluding large portions of the conditioning information set and potentially reducing the accuracy of the forecasts. Table 2 suggests that incorporating attention into the text-based quantile forecasting model ends up improving density forecasts. As expected, the model based on sparse (non-targeted) textual data is outperformed by the ones based on dense (targeted) textual data. This occurs because factor models require data to be dense, and this is met after transforming sparse textual data set into a dense one by incorporating "attention" into the model. Our conclusion from Table 2 is that (out-of-sample) density forecasts of the US-CAN exchange rate return are mainly driven by targeted textual data with model $\widehat{\Delta}s_{t+1|\Omega_t^\tau}$ yielding the best results regardless whether skewed-t student distribution or Epanechnikov Kernel is used to estimate the conditional density.

The quantile forecasts can be used to make interval forecasts. We use quantile forecasts for $\tau = 0.025; 0.975$ to obtain a $95\%$-coverage interval. The width of the interval, which is basically the difference between the quantile forecast at $\tau = 0.975$ and the one at $\tau = 0.025$, can be used as measure of volatility forecast. We expected a wider interval in times of market turmoil as opposed to a narrow interval in times

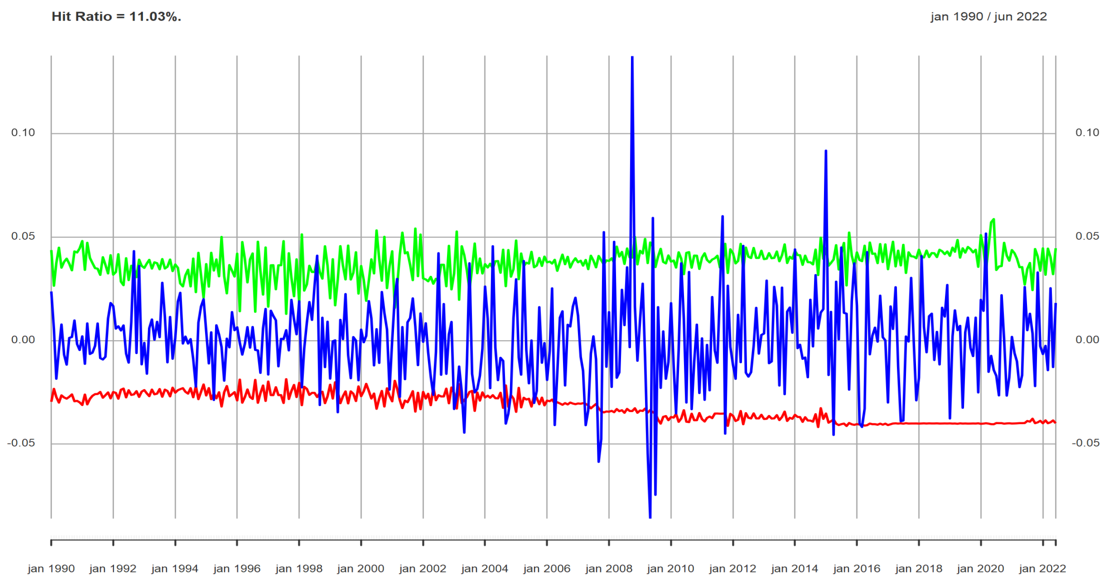
Table 2: Log predictive density score (LPDS) and Berkowitz (2001) density test

| | Panel A: Epanechnikov | | Panel B: Skewed t-student | |
|---|---|---|---|---|
| Model | LPDS | P-value | LPDS | P-value |
| $\widehat{\Delta s}_{t+1\mid\Omega_t^{\tau}}$ | 2.832 | 0.27 | 2.801 | 0.22 |
| $\widehat{\Delta s}_{t+1\mid\Omega_t^{\tau}(LASSO)}$ | 2.672 | 0.12 | 2.664 | 0.11 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\mid\Omega_t^{\tau}}$ | 2.689 | 0.16 | 2.675 | 0.14 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\mid\Omega_t^{\tau}(LASSO)}$ | 2.669 | 0.10 | 2.665 | 0.09 |
| $\widehat{\Delta s}_{t+1\mid\Omega_t}$ | 2.661 | 0.13 | 2.652 | 0.12 |
| $\widehat{\Delta s}_{t+1\mid LM_t^{\tau}}$ | 2.632 | 0.07 | 2.626 | 0.05 |
| $\widehat{\Delta s}_{t+1\mid LM_t^{\tau}(LASSO)}$ | 2.600 | 0.06 | 2.596 | 0.05 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\mid M_t^{\tau}}$ | 2.832 | 0.02 | 2.676 | 0.02 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\mid M_t^{\tau}(LASSO)}$ | 2.782 | 0.03 | 2.976 | 0.04 |
| $\widehat{\Delta s}_{t+1\mid M_t^{\tau}}$ | 2.457 | 0.01 | 2.430 | 0.00 |
| $\widehat{\Delta s}_{t+1\mid M_t^{\tau}(LASSO)}$ | 2.427 | 0.01 | 2.401 | 0.00 |
| $\overline{\Delta s}_{t+1}$ | 2.417 | 0.00 | 2.382 | 0.00 |

Table entries show log predictive density scores (a higher score implies a better model) and the p-values for the Berkowitz (2001) density test, respectively. Panel A shows results for densities estimated by the Epanechnikov Kernel whereas Panel B shows results for densities estimated by the skewed t-distribution method.

of market normality. Figure 2 shows an out-of-sample interval forecast obtained by using the random walk (unconditional quantile) model. The blue line corresponds to actual observations of $\Delta s_{t+1}$ whereas the green and red lines correspond to out-of-sample quantile forecasts for $\tau = 0.975$ and $\tau = 0.025$, respectively. It is well known that the random-walk model performs well in terms of out-of-sample point (conditional mean) forecasting. However, as shown in Figure 1, most of the predictability of $\Delta s_{t+1}$ is away from the center and more concentrated in the tails of the conditional distribution. For this reason, we investigate the performance of the random walk model in terms of out-of-sample interval forecasts. Figure 2 shows that the random model fails to anticipate increases in market volatility during periods of market turmoils. Indeed, the interval forecast from such a model does not vary much over time, leading to about 11% of violations (hits) during the out-of-sample period, which is twice as much as the nominal level of 5%.

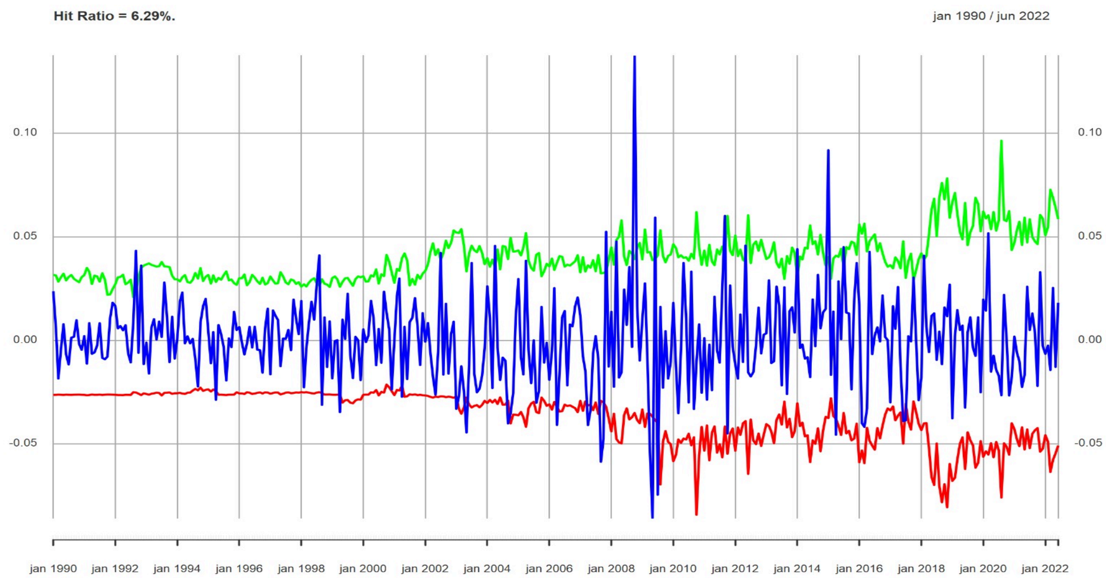
Figure 2: 95%-coverage interval forecast from the unconditional (RW) model ($\overline{\Delta} s_{t+1}$)




This figure shows a 95%-coverage interval forecast from the RW model ($\overline{\Delta} s_{t+1}$). The green line corresponds to quantiles forecasts with $\tau = 0.975$ whereas the red line corresponds to quantile forecasts with $\tau = 0.025$. The blue line shows actual observations of exchange-rate return ($\Delta S_t$) during the 1990.1-2022.7 out of sample period. The percentage of hits (11.03%) is well above 5% nominal level.

Figure 3 shows that selected textual information across time and quantiles ($\Omega_t^\tau$ ) can improve interval forecasts. It strikes the fact that, unlike the RW model, the interval forecast obtained with target textual information are quite responsive to changes in the market. The width of the interval (which is a proxy for volatility) widens in periods of market turmoil through through increases in both upside and downside risks. This is why the interval forecasts from the text-based model $\widehat{\Delta s}_{t+1|\Omega_t^\tau}$ yields a number of violations (hits) of about 6%, which is very close to the nominal level of 5%. In other words, by outperforming the RW model in the tails of the conditional distribution, the text-based model $\widehat{\Delta s}_{t+1|\Omega_t^\tau}$ can deliver more accurate interval forecasts.

Figure 3: 95%-coverage interval forecast from the conditional model $\widehat{\Delta s}_{t+1|\Omega_t^\tau}$.

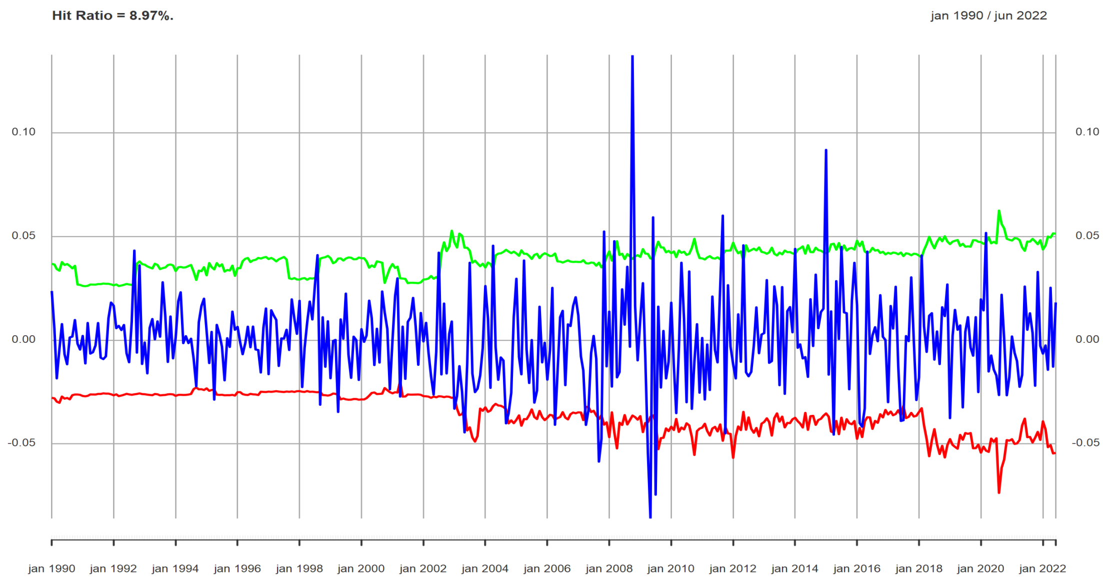



This figure shows a 95%-coverage interval forecast from the text-based model ($\widehat{\Delta s}_{t+1|\Omega_t^\tau}$). The green line corresponds to quantiles forecasts with $\tau = 0.975$ whereas the red line corresponds to quantile forecasts with $\tau = 0.025$. The blue line shows actual observations of exchange-rate return ($\Delta S_t$) during the 1990.1-2022.7 out of sample period. The percentage of hits (6.3%) is close to the nominal level of 5%.

Figures 4 and 5 show interval forecasts from models that rely on a fixed dictionary ($\widehat{\Delta s}_{t+1|LM_t^\tau}$) and economic indicators ($\widehat{\Delta s}_{t+1|M_t^\tau}$), respectively. It can be noted that the upper and lower bounds of such interval forecasts do not respond satisfactory to changes in the market volatility. This implies in a violation rate (hits) above the 5% level. In conclusion, the model based on target textual information makes interval forecasts with almost the right coverage, outperforming not only the random walk (unconditional) model but other conditional models that rely on economic indicators or a fixed dictionary.[27]

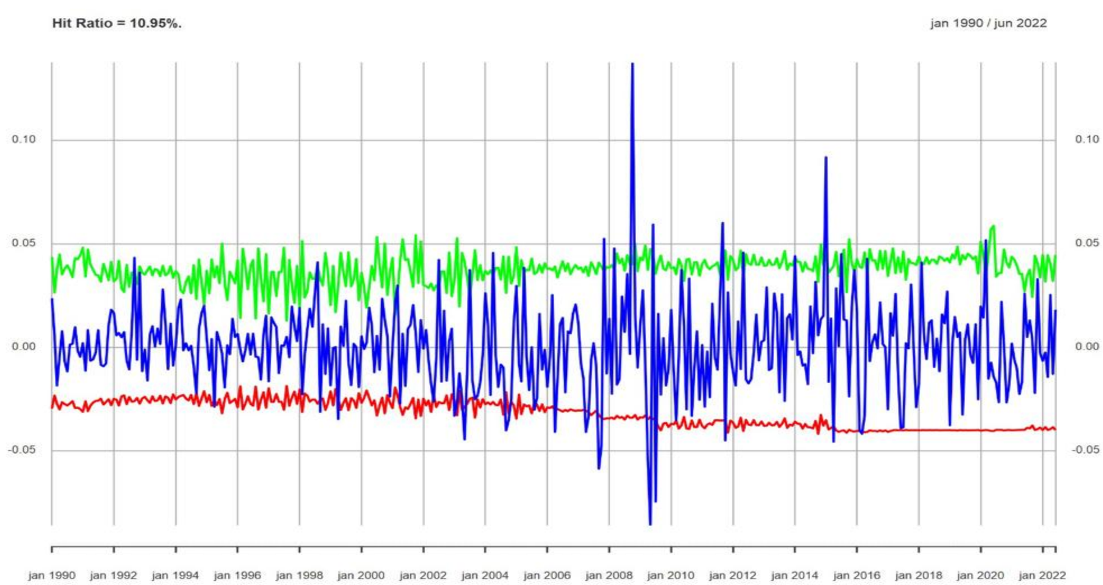
[27] The text-based model $\widehat{\Delta s}_{t+1|\Omega_t^\tau}$ also outperforms the model based on untargeted textual data $\widehat{\Delta s}_{t+1|\Omega_t}$.

Figure 4: 95%-coverage interval forecast from the conditional model $\widehat{\Delta s}_{t+1|LM_t^{\tau}}$




This figure shows a 95%-coverage interval forecast from the from the text-based model $\widehat{\Delta s}_{t+1|LM_t^{\tau}}$. The green line corresponds to quantiles forecasts with $\tau = 0.975$ whereas the red line corresponds to quantile forecasts with $\tau = 0.025$. The blue line shows actual observations of exchange-rate return ($\Delta S_t$) during the 1990.1-2022.7 out of sample period. The percentage of hits (9%) is above the 5% nominal level.

Figure 5: 95%-coverage interval forecast from the conditional model $\widehat{\Delta s}_{t+1|M_t^{\tau}}$




This figure shows a 95%-coverage interval forecast from the text-based model $\widehat{\Delta s}_{t+1|M_t^{\tau}}$. The green line corresponds to quantiles forecasts with $\tau = 0.975$ whereas the red line corresponds to quantile forecasts with $\tau = 0.025$. The blue line shows actual observations of exchange-rate return ($\Delta S_t$) during the 1990.1-2022.7 out of sample period. The percentage of hits (11.00%) is well above 5% nominal level.

The above graphs can be used to obtain time series of violations or hits, that is, a binary variable $H_t$ which equals 1 when the actual out-of-sample observations of dollar return $\Delta s_{t+1}$ is outside the interval forecast (i.e, above the $97.5\%$ quantile forecast or below the $2.5\%$ quantile forecast), and zero otherwise. Then, we can use $H_t$ and the vast literature on backtesting of value-at-risk models to test the null hypothesis of correct specification of the interval forecasting model. This paper considers the Kupiec (1995), Christoffesen (1998), and Engle and Maganelli (2004) backtests, which are all based on (out-of-sample) hit observations $H_t$. Results are shown on Table 3 where one can see that interval forecasts based on selected textual information, that is $\widehat{\Delta s}_{t+1|\Omega_t^\tau}$, $\widehat{\Delta s}_{t+1|\Omega_t^\tau(LASSO)}$, and $\widehat{\Delta s}^{(QFA)}_{t+1|\Omega_t^\tau}$ pass all the backtests at the 5% level (p-value larger than 0.05). On the other hand, interval forecasts from the random-walk model ($\overline{\Delta s}_{t+1}$), or from conditional quantile models that rely either on selected economic information ($\widehat{\Delta s}_{t+1|M_t^\tau}$, $\widehat{\Delta s}_{t+1|M_t^\tau(LASSO)}$ ) or on selected words from the fixed ($LM$ ) dictionary ( $\widehat{\Delta s}_{t+1|LM_t^\tau}$, $\widehat{\Delta s}_{t+1|LM_t^\tau(LASSO)}$) fail all the backtests at the 5% level (p-value smaller than 0.05). These results add to the ones shown in Figure 1 in that the predictability of the random-walk model lies at the central tendency whereas targeted textual data display strong predictive content in the tails of the conditional distribution of dollar return.

Table 3: 5%-level Backests

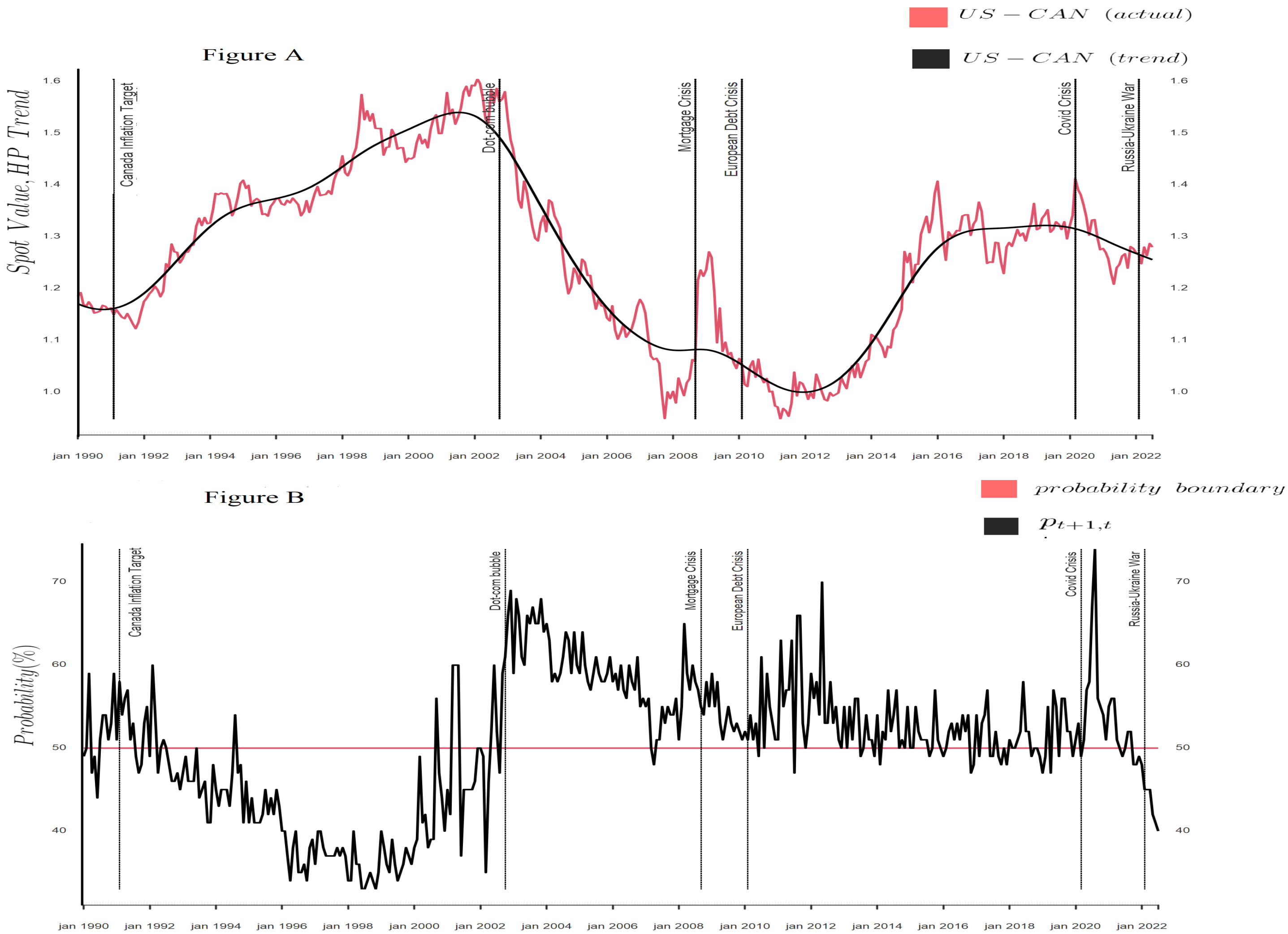
| Model | Kupiec test | Kupiec p-value | Christoffesen test | Christoffesen p-value | Engle and Manganelli test | Engle and Manganelli p-value |
|---|---|---|---|---|---|---|
| $\widehat{\Delta s}_{t+1\|\Omega_t^\tau}$ | 0,51 | 0,47 | 2,29 | 0,31 | 10,24 | 0,17 |
| $\widehat{\Delta s}_{t+1\|\Omega_t^\tau(LASSO)}$ | 0,69 | 0,32 | 2,58 | 0,25 | 11,55 | 0,12 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\|\Omega_t^\tau}$ | 0,58 | 0,38 | 2,42 | 0,27 | 10,85 | 0,14 |
| $\widehat{\Delta s}^{(QFA)}_{t+1\|\Omega_t^\tau(LASSO)}$ | 0,79 | 0,09 | 2,93 | 0,08 | 13,15 | 0,06 |
| $\widehat{\Delta s}_{t+1\|\Omega_t}$ | 28,11 | 0,01 | 24,97 | 0,01 | 59,32 | 0,00 |
| $\widehat{\Delta s}_{t+1\|M_t^\tau}$ | 27,81 | 0,02 | 21,97 | 0,03 | 53,32 | 0,00 |
| $\widehat{\Delta s}_{t+1\|M_t^\tau(LASSO)}$ | 28,99 | 0,01 | 22,37 | 0,02 | 56,41 | 0,01 |
| $\widehat{\Delta s}_{t+1\|M_t^\tau}$ | 27,81 | 0,02 | 21,97 | 0,03 | 53,32 | 0,00 |
| $\widehat{\Delta s}_{t+1\|M_t^\tau}$ | 27,81 | 0,02 | 21,97 | 0,03 | 53,32 | 0,00 |
| $\widehat{\Delta s}_{t+1\|LM_t^\tau}$ | 92,71 | 0,00 | 99,69 | 0,00 | 54,74 | 0,00 |
| $\widehat{\Delta s}_{t+1\|LM_t^\tau(LASSO)}$ | 93,51 | 0,01 | 100,88 | 0,02 | 55,84 | 0,00 |
| $\overline{\Delta s}_{t+1}$ | 62,37 | 0,01 | 39,33 | 0,01 | 61,78 | 0,00 |

The tables shows several backtests: (i) The statistical tests of Kupiec (1995), Christoffesen (1998) and Engle and Manganelli (2004). A p-value lower than 0.05 indicates rejection of the null hypothesis of correct specification of the interval forecasting model

## 6.2 Probability Forecasts

The existence of predictability in the tails can be used to predict depreciation risks, i.e., $p_{t+1,t}$, which can then be used by investors to formulate directional trading strategies such as selling US dollar and buy the risk-free t-bond whenever such probability is above a certain threshold. We now use the density forecast to compute the probability of the U.S. dollar depreciation, i.e., $p_{t+1,t} = prob[(\Delta s_{t+1} < 0)|\Im_t]$. Recall we have in mind a Canadian firm that utilizes inputs imported from the U.S. and, for this reason, its cash flow is highly affected by fluctuations of the U.S. dollar. Thus, the goal is to employ the approach developed in this paper to predict $p_{t+1,t}$ and use it to make trading decisions for the Canadian firm at the end of time $t$. In order to motivate our exercise, Figure 6 (part A) shows the value of the US-CAN spot exchange rate ($S_t$) during the out-of-sample period (1990.1-2022.7) as well as its trend estimated by the Hodrick-Prescott filter. A good directional forecast is the one that maximizes the portfolio value in the long run, correctly predicting the directional changes implied by the trend. We list five economic events that triggers depreciation of the US dollar with respect to the Canadian Dollar: the introduction of the inflation target policy in Canada; the "dot.com" bubble in the U.S.; the U.S. mortgage crisis; the European bond crisis and; the COVID-19-induced recession. We also list an economic event (Russia-Ukraine war) that triggers appreciation of the US Dollar. Figure 6 shows that $S_t$ falls shortly ($\Delta s_{t+1} < 0$) during the implementations of the inflation-target policy in Canada but soon starts a long period of appreciation ($\Delta s_{t+1} > 0$) until the blast of the "dot.com" bubble in 2002. At that point, the US dollar starts a long period of depreciation that ultimately reverses to a very short period of appreciation in 2007. Such an appreciation momentum is interrupted by the blast of the subprime bubble (mortgage crisis) at the end of 2008. By the end of 2011,the US dollar gain a new momentum with a steady appreciation that is shortly interrupted by the COVID-19-induced recession. Finally, the beginning of the Russia-Ukraine war leads the U.S. dollar to period of appreciation. Given this sequence of economic events, a Canadian company concerned about downside risks might be interested to predict the directional changes implied by the trend as shown in Figure 6 (part A) and use it to make informed financial decisions.

Figure 6: US-CAN Exchange Rate: spot value and Hodrick-Prescott trend and Probability of depreciation from conditional density with information set $\Omega_t^\tau$




The figure A shows the US-CAN spot exchange rate and its trend estimated by using a Hodrick Prescott (HP) filter with tuning parameter $\lambda = 14,400$.The figure B shows the Probability of depreciation from conditional density. The black line represents the probability of depreciation, i.e. $p_{t+1,t} = prob[(\Delta s_{t+1} < 0)|\Im_t]$, calculated from the conditional density with information set $\Im_t = \Omega_t^\tau$ and estimated by using the Epanechnikov Kernel. The red line represents the 50% probability boundary.

To save space, our predicted probability of depreciation will be computed from the unconditional density forecasts as well as from conditional density ones with information sets $\Im_t = (\Omega_t^\tau, M_t^\tau)$. We only considered the Epanechnikov kernel but the results from densities computed by using the skewed-t student distribution are similar and available under request. Figure 6 (part B) shows the predicted probability of depreciation obtained when $\Im_t = \Omega_t^\tau$ and a red horizontal line representing a 50% probability boundary. One can see that our model is able to predict the depreciation of the US dollar quite well. Indeed, we have

$p_{t+1,t} > 50\%$ during the implementation of the inflation-target policy in Canada. Then, the probability of depreciation remains below 50% during several months until it goes above 50% again in the end of 2002 with the blast of the "dot.com" bubble. Then, the probability of deprecation starts going down slowly towards 50% but increases again because of the European debt crisis in 2010, and remains most of the time above 50% until the outbreak of the COVID pandemic in early 2020 when it goes above 70%. Finally, the probability of depreciation falls below 50% with the beginning of the Russia-Ukraine war.

Thus, our model seems to predict well the depreciation (appreciation) of the U.S. dollar with respect to the Canadian dollar. We now evaluate the performance of the unconditional quantile model, that is, the model represented by a quantile regressions with an intercept only. We also use the Epanechnikov Kernel and the red horizontal line still represents the 50% probability cutoff. If we continue using $p_{t+1.t} > 50\%$ to predict depreciation of U.S. dollar, then one can easily see in Figure D of the online appendix that the unconditional quantile model fails to detect the depreciation of the U.S. dollar provoked by the blast of the "dot.com" bubble. This failure remains for many months after the blast of the bubble, suggesting that the unconditional model takes too long to predict directional changes in the US-CAN exchange rate. It wrongly predicts a prolonged period of depreciation between 2010 and 2020 besides missing the appreciation of the U.S. dollar caused by the Russia-Ukraine war. We also predicted the probability of U.S. dollar depreciation based on selected economic indicators (i.e., model $\widehat{\Delta} s_{t+1|M_t^\tau}$). The consequence of using such a model to predict probabilities of depreciation is shown in Figure E of the online appendix, where one can see $p_{t+1.t}$ changing values too often from time to time. In practice, this would lead an investor to incur a high transaction cost since she would need to re-balance the portfolio many times. Table 4 reports the p-value of the Mincer-Zarnowitz regression test for these 3 models. One can see that we fail to reject the null hypothesis even at the 10% level (p-value=0.52) only when targeted textual data is used to compute $p_{t+1,t}$. This suggests that (target) textual data can be useful to compute the expected directional change $E[y_{t+1}|\Im_t]$. The same does not happen when one uses either economic or the unconditional quantile model to estimate $p_{t+1,t}$ (p-values less than 0.10). The directional forecast also anticipates the trend of depreciation of the U.S. dollar with respect to the Canadian Dollar. Indeed, the second column of Table 4 indicates the directional forecast $I_{t+1,t} = 1$ can correctly predict about 78% of the depreciation implied by the trend shown in Figure 6, whereas the economic-based and unconditional directional forecast only correctly predict about 47% and 45%, respectively. For an investor concerned about downside risks, predicting the trend of depreciation is essential for the formulation of directional trading strategies, and our results suggest that forecasts based on targeted textual data is important to this end.

Table 4: Mincer-Zarnowitz and changes correctly predicted

| Model | P-value Wald Test | Proportion of change predicted(%) |
|---|---|---|
| $\widehat{\Delta} s_{t+1\|\Omega_t^\tau}$ | 0.52 | 77.71 |
| $\widehat{\Delta} s_{t+1\|M_t^\tau}$ | 0.09 | 46.98 |
| $\overline{\Delta} s_{t+1}$ | 0.01 | 44.75 |

This table shows the p-value of the null hypothesis $H_0 : \theta_0 = 0$ and $\theta_1 = 1$ in Equation $y_{t+1} = \theta_0 + \theta_1 I_{t+1,t} + \eta_{t+1}$, where $y_{t+1}$ equals 1 if the HP trend implies a depreciation and zero otherwise. The last column shows the percentage of directional changes implied by the HP trend correctly predicted by the directional forecast $I_{t+1,t}$.

An important question is whether the quantile model ($\widehat{\Delta} s_{t+1|\Omega_t^\tau}$) is indeed able to identify the depreciation (appreciation) risk of the U.S. dollar. To shed light on this, we show the relationship between the predicted probability of the U.S. dollar depreciation ($p_{t+1,t}$) - computed by using model $\widehat{\Delta} s_{t+1|\Omega_t^\tau}$ - and the downside and upside risks, here represented by the downside and upside entropy measures $\pounds_t^D$, $\pounds_t^{UP}$, respectively [28]. The triple ($p_{t+1,t}$, $\pounds_t^D$, $\pounds_t^{UP}$) is calculated for each out-of-sample observation and relies on the Epanechnikov estimates of the density forecasts. Figure H in the online appendix shows a scatter plot of the 391 out-of-sample observations of $p_{t+1,t}$, $\pounds_t^D$, and $\pounds_t^{UP}$. The $Y$-axis shows the entropy (downside and upside) whereas the $X$-axis shows the probability of depreciation. Red dots correspond to the depreciation risk (downside entropy) whereas the blue dots correspond to the appreciation risk (upside entropy). Conditional on the same $p_{t+1,t}$, one sees the distribution of downside and upside entropy - i.e., months showing the same $p_{t+1,t}$ can display different entropy. A striking result coming from that figure is that months for which the probability of depreciation $p_{t+1,t}$ is larger than 50% also experienced more downside entropy than upside one, whereas the opposite occurs at months for which $p_{t+1,t} < 50\%$. This result suggests the larger the predicted probability of depreciation from the text-based quantile model,

[28]See online appendix for an explanation on the computation of $\pounds_t^D$, $\pounds_t^{UP}$

the larger the downside entropy, implying that such forecasts carry information about downside risks of the of the US-CAN exchange rate. Hence, investment strategies designed to minimize downside risks could potentially benefit of such forecasts. Figures J and K in the online appendix show the same pattern does not occur when the triple $(p_{t+1,t}, \pounds_t^D, \pounds_t^{UP})$ is computed with ((non-selected) textual data and a fixed dictionary, meaning that "attention" and a varying dictionary are important to establish this link between the predicted probability of depreciation and the downside (upside) entropy.

We now turn to our portfolio analysis. We assume that an investor starts with 100 dollars and face a fixed transaction cost ($c$) of \$0.30 and \$0.50, respectively, to re-balance her portfolio. Figures F and G of the online appendix indicate that if in January 1990 an investor had begun with \$100 and started using the above described investment strategy, then his/her portfolio would be worth about \$240 (\$220) in July 2022 if they had used the information set $\Im_t = \Omega_t^\tau$ to predict the probabilities of depreciation and had payed a 30 (50) cents transaction cost, respectively. If she had used the unconditional density forecast, then her portfolio would be worth slightly more than \$100 in July 2022. regardless the transaction cost. This happens because the probability of depreciation computed from the unconditional density forecast does not vary much over time and, for this reason, the portfolio is re-balanced a few times.[29] Finally, if the investor had used the information set $\Im_t = M_t^\tau$ to predict the probabilities of depreciation, then her portfolio would be worth about\$130 (\$30) in July 2022 if she had payed a 30 (50) cents transaction cost, respectively. This happens because the predicted probability of depreciation based on economic indicators changes very often, incurring a high transaction cost for the investor. Indeed, the portfolio performance worsens as we increase the transaction cost.

Table B in the online appendix indicates that the set of most predictive terms (dictionary) is quite heterogeneous, changing not only across time but also across different quantiles of the distribution (i.e., left tail versus right tail). This is important since the depreciation and appreciation risks are related to outcomes occurring at the left and right tail of the distribution respectively. For instance, the most im-

[29]Indeed, Figure 4 in the appendix shows a flat segment on the green line between 2010 and 2020 reflecting the fact that the unconditional model predicts $p > 0.5$ exactly when the t-bond rate is close to zero percent, wrongly inducing the investor to make a bad financial decision.

portant terms affecting the appreciation risk during the Russia-Ukraine war (2022.2) are those related to geopolitical or military events (i.e.,"soviet union" and "military spending") and market volatility, suggesting that the U.S. dollar is seen by investors as a "safe heaven" in times of geopolitical conflicts and high market volatility. Terms specific to trade, exchange rate and energy (i.e., "exchange market", "trade deficit", "energy price", and "short position") are also commonly listed, perhaps because they reflect the mechanisms through which geopolitical conflicts affect the economic system.

The terms "market collapse", "price drop", "market turmoil" are commonly cited during events related to appreciation and depreciation of the U.S. dollar, suggesting that such terms reflect the existence of economic crises in general. However, other terms such as "weakness dollar", "dollar decline", "fall dollar", "dollar drop", "stabilize dollar", and "dollar rally" are exclusively found during periods of depreciation of the U.S. dollar, implying that such terms are not commonly cited during periods of appreciation of the U.S. dollar. Interestingly, the term "economic growth" is particularly important since it appears quite frequently during months in that the U.S dollar depreciates. Additionally, other terms related to economic recession such as "consumer spending", "investor confidence", "unemployment rate", "economic recovery", "soft landing", "stock fall", "economic activity" also contribute to increase the depreciation risk of the U.S. dollar. Importantly, these terms do not seem to affect the appreciation risk much. Other terms that also contribute to increase the depreciation risk - but do not affect the appreciation risk -are exactly those related to interest rates (i.e.,"bond price", "fed easing", "home loan", "budget deficit", "raise tax", "lower rate"); stock market (i.e., "stock market", "corporate profit", "stock buyback", "stock price", "equity market", "stock rise") and; oil market (i.e., petroleum exporter, "oil market", "oil price").

In sum, we found that depreciation risks of the U.S. dollar are mostly driven by terms related to financial markets and economic recessions, whereas terms related to geopolitical conflicts and market volatility seem to drive more the appreciation risk, suggesting that the U.S. dollar may be seen by investors as a "safe heaven" in times of war and high market volatility. We also found terms that affect both depreciation and appreciation risks, perhaps because they signal the existence of economic crises in general. This analysis illustrates the benefits of using recent advances in machine learning to identify the most predictive dictionary over time and across different quantiles of the exchange rate distribution.

# 7 Conclusion

Words matter, but attention is important! We introduce "attention" in the quantile factor model to study the predictive content of textual data. The proposed methodology delivers not only the most predictive dictionary available up to a forecast origin, but also the most predictive dictionary at different quantiles of the distribution. Our results indicate that targeted textual data improve predictability of the US-CAN exchange rate over the random walk benchmark, whereas target economic data do not. We found that such a predictability is concentrated in the tails of the distribution, especially after the burst of the "dot.com" bubble. Text-based quantile models yields directional forecasts that are strongly related to the downside entropy of the exchange rate, meaning they carry information about downside risks of currency investments. We consider a simple managed portfolio to illustrate the economic value of our directional forecasts and found that the model based on targeted textual data yields accurate predictions of depreciation of the U.S. dollar with respect to the Canadian dollar, leading to a larger cumulative wealth.

Words that signal the existence of economic recessions are shown to increase the depreciation risk of the U.S. dollar, whereas words related to geopolitical events and market volatility are shown to increase the appreciation risk, suggesting that the U.S. dollar is seen by investors as a "safe heaven" in times of geopolitical conflicts and high market volatility. The approach developed in this paper may be useful to many other applications in finance and economics, where the ability to predict directional changes (such as exchange rate depreciation) is essential for decision making. Future research could consider more performant machine learning methods, such as neural network and also allow for the possibility that words are exposure to topics (factors) dynamically. In this regard, a dynamic factor model would be a natural choice.

**Proof of Theorem 1.**

Proof of (a). Let $u = \sqrt{t}(d - \delta(\tau))$ and $\widehat{u}(\tau) = \sqrt{t}(\widehat{\delta}(\tau) - \delta(\tau))$. Then

$$\widehat{u}(\tau) = \arg\min_u \sum_{s=1}^{t} \rho_\tau \left( \Delta s_{s+1} - \widehat{Z}'_{s,\tau} \left( \delta(\tau) + \frac{u}{\sqrt{t}} \right) \right) \tag{13}$$

$$= \arg\min_u \sum_{s=1}^{t} \left[ \rho_\tau \left( \Delta s_{s+1} - \widehat{Z}'_{s,\tau} \left( \delta(\tau) + \frac{u}{\sqrt{t}} \right) \right) - \rho_\tau \left( \Delta s_{s+1} - \widehat{Z}'_{s,\tau}\delta(\tau) \right) \right]. \tag{14}$$

By the Knight's (1998) identity, we have

$$\rho_\tau(\mu - \pi) - \rho_\tau(\mu) = -\pi(\tau - 1\{\mu \leqslant 0\}) + \int_0^\pi (1\{\mu \leqslant s\} - 1\{\mu \leqslant 0\})ds.$$

So, it follows that

$$\begin{aligned}
&\sum_{s=1}^{t} \left[ \rho_\tau \left( \Delta s_{s+1} - \widehat{Z}'_{s,\tau} \left( \delta(\tau) + \frac{u}{\sqrt{t}} \right) \right) - \rho_\tau \left( \Delta s_{s+1} - \widehat{Z}'_{s,\tau}\delta(\tau) \right) \right] \\
&= \sum_{s=1}^{t} [ - \frac{\widehat{Z}'_{s,\tau} u}{\sqrt{t}} \left( \tau - 1 \left\{ \Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) \right\} \right) \\
&\qquad + \int_0^{\frac{\widehat{Z}'_{s,\tau} u}{\sqrt{t}}} 1 \left\{ \Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) + v \right\} - 1 \left\{ \Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) \right\} dv] \\
&= -\widehat{W}_t(\tau)'u + \widehat{A}_t(u, \tau),
\end{aligned}$$

where

$$\widehat{W}_t(\tau) = \sum_{s=1}^{t} \frac{1}{\sqrt{t}} \left[ \widehat{Z}'_{s,\tau} \left( \tau - 1 \left\{ \Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) \right\} \right) \right],$$

and

$$\widehat{A}_t(u, \tau) = \sum_{s=1}^{t} \int_0^{\frac{\widehat{Z}'_{s,\tau} u}{\sqrt{t}}} 1 \left\{ \Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) + v \right\} - 1 \left\{ \Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) \right\} dv.$$

Let $g_t(u, \tau) = -\widehat{W}_t(\tau)'u + \widehat{A}_t(u, \tau)$, and we note that $g_t(u, \tau)$ is a convex function in $u$ for each $\tau$.

Then we only need to prove (i) $\widehat{A}_t(u,\tau) \xrightarrow{p} \frac{u'A(\tau)u}{2}$ for some positive definite $A(\tau)$, and (ii) $\widehat{W}_t(\tau) \xrightarrow{p} W(\tau)$, where $W(\tau)$ is normally distributed. Then we would be able to claim, by the Convexity Lemma (Kato, 2009), that

$$\widehat{u}(\tau) = A(\tau)^{-1}W(\tau) + R(\tau), \text{ where } R(\tau) = o_p(1),$$

and hence obtain the asymptotic distribution of $\widehat{u}(\tau) = \sqrt{t}(\widehat{\delta}(\tau) - \delta(\tau))$.

**Step 1.** We prove (i). Taking expectation on $\widehat{A}_t(u,\tau)$, we get

$$\begin{aligned}
E\left(\widehat{A}_t(u,\tau)\right) &= tE\int_0^{\frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}}} \{F[Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) + \left(\widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau}\right)'\delta(\tau) + v \mid \underline{\Omega}_s^\tau] \\
&- F[Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) + \left(\widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau}\right)'\delta(\tau)] \mid \underline{\Omega}_s^\tau\}dv \\
&= tE\left\{f\left[Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) + \left(\widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau}\right)'\delta(\tau) + \widetilde{v} \mid \underline{\Omega}_s^\tau\right]\int_0^{\frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}}} sds\right\} \\
&= tE\left\{f\left[Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) + \left(\widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau}\right)'\delta(\tau) + \widetilde{v} \mid \underline{\Omega}_s^\tau\right]\frac{u'\widehat{Z}_{s,\tau}\hat{Z}'_{s,\tau}u}{2t}\right\},
\end{aligned}$$

where $\widetilde{v}$ is between 0 and $\frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}}$. Given that $\widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau} = \left(0, \left(\widetilde{F}_{s,\tau} - HF_{s,\tau}\right)'\right)'$, the well established properties for $\widetilde{F}_{s,\tau} - HF_{s,\tau}$ by Bai (2003), and the fact that $\widetilde{v} \xrightarrow{p} 0$, we have

$$E\left(\widehat{A}_t(u,\tau)\right) = \frac{1}{2}u'E\left[f\left(Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) \mid \underline{\Omega}_s^\tau\right)\bar{Z}_{s,\tau}\bar{Z}'_{s,\tau}\right]u + o(1).$$

Hence,

$$A(\tau) = E\left[f\left(Q_\tau\left(\Delta s_{s+1} \mid \underline{\Omega}_s^\tau\right) \mid \underline{\Omega}_s^\tau\right)\bar{Z}_{s,\tau}\bar{Z}'_{s,\tau}\right].$$

To complete Step 1, we also have to show that

$$\widehat{A}_t(u,\tau) - E\left(\widehat{A}_t(u,\tau)\right) \xrightarrow{p} 0.$$

Given that

$$\widehat{A}_t(u,\tau) = \sum_{s=1}^{t} \int_0^{\frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}}} 1\left\{\Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) + v\right\} - 1\left\{\Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau)\right\} dv$$
$$= \sum_{s=1}^{t} \int_0^{1} \frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}} \left[1\left\{\Delta s_{s+1} \leq \hat{Z}'_{s,\tau}\delta(\tau) + v\frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}}\right\} - 1\left\{\Delta s_{s+1} \leq \widehat{Z}'_{s,\tau}\delta(\tau)\right\}\right] dv$$
$$= \frac{1}{t}\sum_{s=1}^{t} \sqrt{t}\widehat{Z}'_{s,\tau}u \int_0^{1} \left[1\left\{\Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau) + v\frac{\widehat{Z}'_{s,\tau}u}{\sqrt{t}}\right\} - 1\left\{\Delta s_{s+1} \leqslant \widehat{Z}'_{s,\tau}\delta(\tau)\right\}\right] dv,$$

let us introduce the so defined $A_t(u,\tau)$ as

$$A_t(u,\tau) = \frac{1}{t}\sum_{s=1}^{t} \sqrt{t}Z'_{s,\tau}u \int_0^{1} \left[1\left\{\Delta s_{s+1} \leqslant Z'_{s,\tau}\delta(\tau) + v\frac{Z'_{s,\tau}u}{\sqrt{t}}\right\} - 1\left\{\Delta s_{s+1} \leqslant Z'_{s,\tau}\delta(t)\right)\right] dv.$$

Further, notice that by the law of iterated expectation, we have

$$A_t(u,\tau) = \frac{1}{t}\sum_{s=1}^{t} \sqrt{t}u'Z_{s,\tau} \int_0^{1} \left[F\left(Z'_{s,\tau}\delta(\tau) + v\frac{Z'_{s,\tau}u}{\sqrt{t}} \mid \underline{\Omega}_s^{\tau}\right) - F\left(Z'_{s,\tau}\delta(\tau) \mid \underline{\Omega}_s^{\tau}\right)\right] dv + o_p(1)$$
$$= \frac{1}{t}\sum_{s=1}^{t} \sqrt{t}u'Z_{s,\tau} \quad f\left(Z'_{s,\tau}\delta(\tau) \mid \underline{\Omega}_s^{\tau}\right) \quad \frac{Z'_{s,\tau}u}{\sqrt{t}} \int_0^{1} v dv + op(1)$$
$$= \frac{1}{2t}\sum_{s=1}^{t} u'Z_{s,\tau}Z'_{s,\tau}uf\left(Z'_{s,\tau}\delta(\tau) \mid \underline{\Omega}_s^{\tau}\right) + op(1).$$

Similarly,

$$\widehat{A}_t(u,\tau) = \frac{1}{2t}\sum_{s=1}^{t} u'\widehat{Z}_{s,\tau}\hat{Z}'_{s,\tau}uf\left(\widehat{Z}'_{s,\tau}\delta(\tau) \mid \underline{\Omega}_s^{\tau}\right) + o_p(1)$$
$$= \frac{1}{2t}\sum_{s=1}^{t} u'\hat{Z}_{s,\tau}\hat{Z}'_{s,\tau}uf\left(Z'_{s,\tau}\delta(\tau) \mid \underline{\Omega}_s^{\tau}\right) + o_p(1).$$

Our main task is to show $\mathcal{L} = o_p(1)$, where

$$\mathcal{L} = \frac{1}{t}\sum_{s=1}^{t} u'\hat{Z}_{s,\tau}\hat{Z}'_{s,\tau}u - \frac{1}{t}\sum_{s=1}^{t} u'\bar{Z}_{s,\tau}\bar{Z}'_{s,\tau}u.$$

Notice that

$$\mathcal{L}=\frac{2}{t}\sum_{s=1}^{t}u'\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)\bar{Z}_{s,\tau}'u+\frac{1}{t}\sum_{s=1}^{t}u'\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)'u.$$

Given that $\frac{1}{t}\sum_{s=1}^{t}\left\|\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right\|^{2}=o_p(1)$ by Lemma A1 (i) of Bai and Ng (2006), and that $\frac{1}{T}\sum_{s=1}^{t}\left\|\bar{Z}_{s,\tau}\bar{Z}_{s,\tau}'\right\|^{2}=O_p(1)$ implied by Assumption 3(a), it follows that $\mathcal{L}=o_p(1)$ as desired. Meanwhile, as $A_t(u,\tau)\xrightarrow{p}\frac{1}{2}u'E\left[f\left(Q_\tau\left(\Delta s_{s+1}\mid\underline{\Omega}_s^{\tau}\right)\mid\underline{\Omega}_s^{\tau}\right)\bar{Z}_{s,\tau}\bar{Z}_{s,\tau}'\right]u=E\left(\widehat{A}_t(u,\tau)\right)$, implied by Assumptions 2 and 3(a), the proof is complete.

**Step 2.** We let

$$W(\tau)=\frac{1}{\sqrt{t}}\sum_{s=1}^{t}\bar{Z}_{s,\tau}\left(\tau-1\left\{\Delta s_{s+1}\leqslant\bar{Z}_{s,\tau}'\delta(\tau)\right\}\right).$$

Then it follows that

$$\begin{aligned}
&\widehat{W}(\tau)-W(\tau)=\frac{1}{\sqrt{t}}\sum_{s=1}^{t}\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)\left(\tau-1\left\{\Delta s_{s+1}\leqslant\bar{Z}_{s,\tau}'\delta(\tau)\right\}\right)\\
&-\frac{1}{\sqrt{t}}\sum_{s=1}^{t}\bar{Z}_{s,\tau}\left[1\left\{\Delta s_{s+1}\leqslant\hat{Z}_{s,\tau}'\delta(\tau)\right\}-1\left\{\Delta s_{s+1}\leqslant\bar{Z}_{s,\tau}'\delta(\tau)\right\}\right]\\
&-\frac{1}{\sqrt{t}}\sum_{s=1}^{t}\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)\left[1\left\{\Delta s_{s+1}\leqslant\hat{Z}_{s,\tau}'\delta(\tau)\right\}-1\left\{\Delta s_{s+1}\leqslant\bar{Z}_{s,\tau}'\delta(\tau)\right\}\right]\\
&=I-II-III.
\end{aligned}$$

We can show $I=o_p(1)$ by Lemma A1 (iv) of Bai and Ng (2006).

For $II$, note that

$$II=\frac{1}{\sqrt{t}}\sum_{s=1}^{t}\bar{Z}_{s,\tau}f\left[Q_\tau\left(\Delta s_{s+1}\mid\underline{\Omega}_s^{\tau}\right)\mid\underline{\Omega}_s^{\tau}\right]\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)'\delta(\tau)+o_p(1).$$

By Lemma Lemma A1 (ii) of Bai and Ng (2006), $\frac{1}{t}\sum_{s=1}^{t}\left(\widehat{Z}_{s,\tau}-\bar{Z}_{s,\tau}\right)\bar{Z}_{s,\tau}'=O_p\left(\frac{1}{n}+\frac{1}{t}\right)$. So $II=o_p(1)$ given $\frac{\sqrt{t}}{n}\to 0$.

For $III$,

$$\begin{aligned} III &= \frac{1}{\sqrt{t}} \sum_{s=1}^{t} \left( \widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau} \right) \left( \widehat{Z}_{s,\tau} - \bar{Z}_{s,\tau} \right)' f \left( Z'_{s,\tau} \delta(\tau) \mid \Omega_s^\tau \right) \delta(\tau) + o_p(1) \\ &= o_p(1) + o_p(1), \end{aligned}$$

by Lemma A1 (i) of Bai and Ng (2006), given $\frac{\sqrt{t}}{n} \to 0$.

So step 2 is also proved with $W(\tau) \xrightarrow{d} N\left(0, \tau(1-\tau) E\left(\bar{Z}_{s,\tau} \bar{Z}'_{s,\tau}\right)\right)$.

Putting everything together, we get

$$\widehat{u}(\tau) = \sqrt{t}(\hat{\delta}(\tau) - \delta(\tau)) \xrightarrow{p} A(\tau)^{-1} W(\tau) \xrightarrow{d} N\left(0, \tau(1-\tau) A(\tau)^{-1} E\left(\bar{Z}_{s,\tau} \bar{Z}'_{s,\tau}\right) A(\tau)^{-1}\right).$$

Proof of (b). Given that

$$\begin{aligned} &\widehat{Q}_\tau \left( \Delta s_{t+1} \mid \underline{\Omega}_t^\tau \right) - Q_\tau \left( \Delta s_{t+1} \mid \underline{\Omega}_t^\tau \right) \\ =& \frac{1}{\sqrt{t}} \widehat{Z}'_{t,\tau} \sqrt{t}(\hat{\delta}(\tau) - \delta(\tau)) + \frac{1}{\sqrt{n}} \beta'(\tau) H^{-1} \sqrt{n} \left( \widetilde{F}_{t,\tau} - H F_{t,\tau} \right), \end{aligned}$$

it then follows that

$$\widehat{Q}_\tau \left( \Delta s_{t+1} \mid \underline{\Omega}_t^\tau \right) - Q_\tau \left( \Delta s_{t+1} \mid \underline{\Omega}_t^\tau \right) \xrightarrow{d} N(0, \Xi_{t,\tau}).$$

where $\Xi_{t,\tau} = \frac{1}{t} \widehat{Z}'_{t,\tau} Avar\left(\hat{\delta}(\tau)\right) \widehat{Z}_{t,\tau} + \frac{1}{n} \hat{\beta}(\tau)' Avar\left(\widetilde{F}_{t,\tau}\right) \hat{\beta}(\tau)$.

# References

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–89.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Amat, C., Michalski, T., and Stoltz, G. (2018). Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance*, 88:1–24.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Chao, J., Liu, Y., and Swanson, N. (2022). Consistent factor estimation and forecasting in factor-augmented var models. *Available at SSRN 4010249*.

Chen, L., Dolado, J. J., and Gonzalo, J. (2021). Quantile factor models. *Econometrica*, 89(2):875–910.

Chen, Z., Cheng, V. X., and Liu, X. (2024). Hypothesis testing on high dimensional quantile regression. *Journal of Econometrics*, 238(1):105525.

de Castro, L. and Galvao, A. F. (2022). Static and dynamic quantile preferences. *Economic Theory*, 73(2):747–779.

Ellingsen, J., Larsen, V. H., and Thorsrud, L. A. (2022). News media versus fred-md for macroeconomic forecasting. *Journal of Applied Econometrics*, 37(1):63–81.

Elliott, G. and Timmermann, A. (2013). *Handbook of economic forecasting*. Elsevier.

Engel, C., Lee, D., Liu, C., Liu, C., and Wu, S. P. Y. (2019). The uncovered interest parity puzzle, exchange rate forecasting, and taylor rules. *Journal of International Money and Finance*, 95:317–331.

Figueiredo, E. and Lima, L. R. (2020). Do economic integration agreements affect trade predictability? a group effect analysis. *Canadian Journal of Economics/Revue canadienne d'économique*, 53(2):637–664.

Filippou, I., Rapach, D., Taylor, M. P., and Zhou, G. (2021). Exchange rate prediction with machine learning and a smart carry portfolio. *Available at SSRN 3455713*.

Gaglianone, W. P. and Lima, L. R. (2012). Constructing density forecasts from quantile regressions. *Journal of Money, Credit and Banking*, 44(8):1589–1607.

Gaglianone, W. P. and Lima, L. R. (2014). Constructing optimal density forecasts from point forecast combinations. *Journal of Applied Econometrics*, 29(5):736–757.

Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2012a). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*.

Gaglianone, W. P., Lima, L. R., Linton, O., and Smith, D. R. (2012b). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics*.

Giglio, S., Xiu, D., and Zhang, D. (2023). Prediction when factors are weak. *Available at SSRN 4397485*.

Giglio, S., Xiu, D., and Zhang, D. (2024). Test assets and weak factors. *The Journal of Finance*, page Forthcoming.

Gonçalves, S., McCracken, M. W., and Perron, B. (2017). Tests of equal accuracy for nested models with estimated factors. *Journal of Econometrics*, 198(2):231–252.

Hansen, S., McMahon, M., and Prat, A. (2017). Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics*, 133(2):801–870.

Hendry, S. (2012). Central bank communication or the media's interpretation: What moves markets? Technical report, Bank of Canada Working Paper.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100(8):1816–1829.

Koenker, R. and Basset, G. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, 73(363):618–622.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310.

Li, J., Tsiakas, I., and Wang, W. (2015). Predicting exchange rates out of sample: Can economic fundamentals beat the random walk? *Journal of Financial Econometrics*, 13(2):293–341.

Lima, L. R., Gaglianone, W. P., and Sampaio, R. M. (2008). Debt ceiling and fiscal sustainability in brazil: a quantile autoregression approach. *Journal of Development Economics*, 86(2):313–335.

Lima, L. R. and Godeiro, L. (2023). Equity premium prediction: Attention is all you need. *Journal of Applied Econometrics*, page Forthcoming.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Ma, S., Linton, O., and Gao, J. (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics*, 222(1):295–323.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Narayan, P. K., Bannigidadmath, D., and Narayan, S. (2021). How much does economic news influence bilateral exchange rates? *Journal of International Money and Finance*, 115:102410.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258.

Rossi, B. (2013). Exchange rate predictability. *Journal of economic literature*, 51(4):1063–1119.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, 4:351–394.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Straka, M. and Straková, J. (2017). Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.

Tchatoka, F. D. and Haque, Q. (2023). On bootstrapping tests of equal forecast accuracy for nested models. *Journal of Forecasting*, 42(7):1844–1864.

Tian, H., Yim, A., and Newton, D. P. (2021). Tail-heaviness, asymmetry, and profitability forecasting by quantile regression. *Management Science*, 67(8):5209–5233.

Yi, C. and Huang, J. (2017). Semismooth newton coordinate descent algorithm for elastic-net penalized huber loss regression and quantile regression. *Journal of Computational and Graphical Statistics*, 26(3):547–557.