Inference in Direct Multi-Step and Long Horizon Forecasting Regressions

Asad Dossani*

December 16, 2022

Abstract

This paper proposes and evaluates a new method for inference in direct multi-step and long horizon forecasting regressions. The residuals from both direct multi-step and long horizon forecasting regressions are serially correlated, and can be expressed as a vector moving average (VMA) process of the one step ahead forecast residuals. The proposed estimator imposes the VMA structure on the serially correlated residuals to estimate the covariance matrix of the OLS estimates of direct multi-step and long horizon forecasting regressions. The parameters governing the VMA process are estimated using OLS regressions. A simulation study indicates that the proposed estimator is substantially more accurate and efficient, relative to existing methods that do not impose any structure on the autocorrelation process of the residuals. The paper presents two empirical applications: the first uses the variance risk premium to forecast currency returns, and the second uses inflation to forecast the equity premium.

JEL Codes: C5

Keywords: Autocorrelation, Return Predictability, Vector Moving Average

*Colorado State University, asad.dossani@colostate.edu. Thank you to participants of the Colorado State seminar series, for valuable comments and suggestions. All errors are my own.

1 Introduction

This paper proposes and evaluates a new method for inference in direct multi-step and long horizon forecasting regressions, using the heteroskedasticity and autocorrelation consistent (HAC) covariance estimator in West (1997). I present two empirical applications to illustrate the relevance of the approach: the first uses the variance risk premium to forecast currency returns, and the second uses inflation to forecast the equity premium.

A direct multi-step forecast is a regression of a vector of variables observed at a future horizon on variables observed in the current time period. A long horizon forecast is a regression of the cumulative sum of a vector of variables observed through a future horizon on variables observed in the current period. The West (1997) covariance matrix estimator is applicable when the regression residuals follow a vector moving average (VMA) process of a known order. The estimator, which is parametric, is $\sqrt{T}$ consistent, and asymptotically more efficient than nonparametric estimators, such as the commonly used estimator in Newey *et al.* (1987). The estimator requires estimates of the moving average coefficients of the residuals.

Under the assumption that the data have a vector autoregression (VAR) representation, the residuals from both direct multi-step and long horizon forecasting regressions are serially correlated, and can be expressed as a VMA of the VAR residuals. The proposed estimator imposes the VMA structure on the serially correlated residuals to estimate the covariance matrix of the OLS estimates of direct multi-step and long horizon forecasting regressions. The parameters governing the VMA process are estimated using OLS regressions. This results in substantially more accurate and efficient estimates of the covariance matrix, relative to existing methods that do not impose any structure on the autocorrelation process of the residuals. A simulation study comparing the proposed estimator with the most commonly used Newey *et al.* (1987) HAC estimator illustrates the benefits of the approach. The proposed estimator has a lower root mean squared error and size tests closer to its nominal values.

Direct multi-step and long horizon forecasting regressions have a variety of applications in empirical macroeconomics and finance. Common applications of direct multi-step forecasts are local projections in empirical macroeconomics, as per Jordà (2005). In the forecasting literature, McCracken & McGillicuddy (2019) show that conditional direct multi-step forecasts outperform iterated vector autoregression based approaches at short horizons, with substantial improvement during recent data. In such applications, it is common practice to compute Newey *et al.* (1987) HAC standard errors; see Ramey (2016) for details. Lazarus *et al.* (2018) provide recommendations for inference in such contexts. In related literature Montiel Olea & Plagborg-Møller (2021) propose lag augmented local projections that do not require correcting standard errors for serial correlation. Lusompa (2021) derives the autocorrelation process of local projection residuals, and proposes using generalized least squares to estimate the regression parameters.

Common applications of long horizon forecasts are tests of return predictability in empirical finance. For example, Cochrane & Piazzesi (2005) forecast annual bond returns using monthly data. Common approaches to estimating the covariance matrix include Newey *et al.* (1987) and Hansen & Hodrick (1980) standard errors. Alternatives are proposed in Valkanov (2003), Hodrick (1992), Richardson & Smith (1991), and Hjalmarsson (2011) for specific tests and cases. More generally, Wei & Wright (2013) propose reverse regressions that regress one period returns onto sums of predictor variables. Britten-Jones *et al.* (2011) propose a similar estimator that regresses one period returns onto a transformed matrix of predictor variables. Both these approaches improve inference by reducing the autocorrelation in the error term.

This paper makes the following contributions, relative to the above literature. First, in both direct multi-step and long horizon forcasts, I explicitly model the autocorrelation process of the residuals to estimate the covariance matrix. This results in substantial improved efficiency and accuracy, relative to conventional HAC estimation that imposes no structure on the autocorrelation process of the residuals. Second, I present a unified framework for inference in both direct multi-step and long horizon forecasts, by taking advantage of the fact that error terms in a long horizon forecast are equal to the sum of the error terms from direct multi-step forecasts.

The remainder of this paper is organized as follows: Section 2 presents the theory behind the proposed estimator. This includes deriving the autocorrelation process of the residuals and the covariance matrix of the OLS estimates. Section 3 shows how to estimate the covariance matrix. Section 4 presents the results of a simulation that evaluates the accuracy and efficiency of the proposed estimator. Section 5 presents two empirical examples of the proposed estimator using actual data, and section 6 concludes.

2 Theory

2.1 Autocorrelation Process

Let $\boldsymbol{y}_t$ denote an $(n \,\mathrm{x}\, 1)$ vector of variables observed at time t, with $t = 1, \ldots, T$. $\boldsymbol{a}^{(d,0)}$ is an $(n \,\mathrm{x}\, 1)$ vector and $(\boldsymbol{A}_1^{(d,0)}, \ldots, \boldsymbol{A}_p^{(d,0)})$ are $(n \,\mathrm{x}\, n)$ coefficient matrices. I assume that $\boldsymbol{y}_t$ has a VAR(p) representation, and the $(n \,\mathrm{x}\, 1)$ vector of residuals $\boldsymbol{\epsilon}_t$ are a martingale difference sequence. This allows for conditionally heteroskedastic innovations.

$$\boldsymbol{y}_t = \boldsymbol{a}^{(d,0)} + \boldsymbol{A}_1^{(d,0)} \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(d,0)} \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t \tag{1}$$

$$\mathrm{E}(\boldsymbol{\epsilon}_t | \boldsymbol{\epsilon}_{t-1}, \ldots) = 0$$

Suppose $h \geq 1$. Direct multi-step forecasts are OLS regressions of $\boldsymbol{y}_{t+h}$ on $(1, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p})$. Long horizon forecasts are OLS regressions of $\sum_{s=0}^{h} \boldsymbol{y}_{t+s}$ on $(1, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p})$.

$$\boldsymbol{y}_{t+h} = \boldsymbol{a}^{(d,h)} + \boldsymbol{A}_1^{(d,h)} \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(d,h)} \boldsymbol{y}_{t-p} + \boldsymbol{e}_{t+h}^{(d,h)} \tag{2}$$

$$\sum_{s=0}^{h} \boldsymbol{y}_{t+s} = \boldsymbol{a}^{(l,h)} + \boldsymbol{A}_1^{(l,h)} \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(l,h)} \boldsymbol{y}_{t-p} + \boldsymbol{e}_{t+h}^{(l,h)} \tag{3}$$

The error terms $\boldsymbol{e}_{t+h}^{(d,h)}$ and $\boldsymbol{e}_{t+h}^{(l,h)}$ are serially correlated, and can be expressed as a VMA of the VAR residuals $(\boldsymbol{\epsilon}_{t+h}, \dots, \boldsymbol{\epsilon}_t)$. A detailed proof of this result can be found in Lusompa (2021). Consider the direct multi-step forecast for $h = 1$. I iterate equation 1 one period ahead, and substitute for $\boldsymbol{y}_t$.

$$\boldsymbol{y}_{t+1} = \boldsymbol{a}^{(d,0)} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{y}_t + \cdots + \boldsymbol{A}_p^{(d,0)}\boldsymbol{y}_{t-p+1} + \boldsymbol{\epsilon}_{t+1} \tag{4}$$

$$\begin{aligned}\boldsymbol{y}_{t+1} = {} & \boldsymbol{a}^{(d,0)} + \boldsymbol{A}_1^{(d,0)}(\boldsymbol{a}^{(d,0)} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(d,0)}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t) \\ & + \cdots + \boldsymbol{A}_p^{(d,0)}\boldsymbol{y}_{t-p+1} + \boldsymbol{\epsilon}_{t+1}\end{aligned} \tag{5}$$

$$\boldsymbol{y}_{t+1} = \boldsymbol{a}^{(d,1)} + \boldsymbol{A}_1^{(d,1)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(d,1)}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_{t+1} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t \tag{6}$$

$$\boldsymbol{e}_{t+1}^{(d,1)} = \boldsymbol{\epsilon}_{t+1} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t \tag{7}$$

The coefficient matrix on $\boldsymbol{\epsilon}_t$ is the same as the coefficient matrix on $\boldsymbol{y}_{t-1}$ in the VAR. The coefficient matrices on $(1, \boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_{t-p})$ are now indexed by the superscript $(d, 1)$. Next, consider the long horizon forecast for $h = 1$. I sum equations 1 and 6 to derive an expression for $\boldsymbol{e}_{t+1}^{(l,1)}$. Let $\boldsymbol{I}_n$ denote the identity matrix of order n.

$$\begin{aligned}\boldsymbol{y}_{t+1} + \boldsymbol{y}_t = {} & (\boldsymbol{a}^{(d,1)} + \boldsymbol{A}_1^{(d,1)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(d,1)}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_{t+1} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t) \\ & + (\boldsymbol{a}^{(d,0)} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(d,0)}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t)\end{aligned} \tag{8}$$

$$\sum_{s=0}^{1} \boldsymbol{y}_{t+s} = \boldsymbol{a}^{(l,1)} + \boldsymbol{A}_1^{(l,1)}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p^{(l,1)}\boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_{t+1} + (\boldsymbol{I}_n + \boldsymbol{A}_1^{(d,0)})\boldsymbol{\epsilon}_t \tag{9}$$

$$\boldsymbol{e}_{t+1}^{(l,1)} = \boldsymbol{\epsilon}_{t+1} + (\boldsymbol{I}_n + \boldsymbol{A}_1^{(d,0)})\boldsymbol{\epsilon}_t \tag{10}$$

$$\boldsymbol{e}_{t+1}^{(l,1)} = \boldsymbol{\epsilon}_{t+1} + \boldsymbol{B}_1^{(l,0)}\boldsymbol{\epsilon}_t \tag{11}$$

$$\boldsymbol{B}_1^{(l,0)} \equiv \boldsymbol{I}_n + \boldsymbol{A}_1^{(d,0)} \tag{12}$$

The coefficient matrix on $\boldsymbol{\epsilon}_t$ is a linear transformation of the coefficient matrix on $\boldsymbol{y}_{t-1}$ in the VAR. The coefficient matrices on $(1, \boldsymbol{y}_{t-1}, \dots, \boldsymbol{y}_{t-p})$ are now indexed by the superscript

$(h, 1)$. I derive expressions for $\boldsymbol{e}_{t+h}^{(d,h)}$ and $\boldsymbol{e}_{t+h}^{(l,h)}$ through repeated iteration and substitution.

$$\boldsymbol{e}_{t+h}^{(d,h)} = \boldsymbol{\epsilon}_{t+h} + \sum_{j=0}^{h-1} \boldsymbol{A}_1^{(d,j)} \boldsymbol{\epsilon}_{t+h-1-j} \tag{13}$$

$$\boldsymbol{e}_{t+h}^{(l,h)} = \boldsymbol{\epsilon}_{t+h} + \sum_{j=0}^{h-1} \boldsymbol{B}_1^{(l,j)} \boldsymbol{\epsilon}_{t+h-1-j} \tag{14}$$

$$\boldsymbol{B}_1^{(l,j)} \equiv \boldsymbol{I}_n + \sum_{k=0}^{j} \boldsymbol{A}_1^{(d,k)}$$

For the horizon h direct multi-step and long horizon forecasting regressions, the coefficient matrices on the lagged VAR residuals are linear combinations of the coefficient matrices from the VAR and the 1 through $(h-1)$ direct multi-step regressions.

2.2 Covariance Matrix of OLS Estimates

I derive the covariance matrix of the OLS estimates as a function of the VAR residuals. West (1997) derives the covariance matrix for the general case where the residuals follow a VMA process. I apply that derivation to the case of direct multi-step and long horizon forecasting regressions. For convenience, I rewrite the direct multi-step and long horizon forecasting regressions so that the regression coefficients can be stacked into a vector.

$$\boldsymbol{y}_{t+h} = \boldsymbol{X}_{t-1}' \boldsymbol{\beta}^{(d,h)} + \boldsymbol{e}_{t+h}^{(d,h)} \tag{15}$$

$$\sum_{s=0}^{h} \boldsymbol{y}_{t+s} = \boldsymbol{X}_{t-1}' \boldsymbol{\beta}^{(l,h)} + \boldsymbol{e}_{t+h}^{(l,h)} \tag{16}$$

$$\boldsymbol{X}_{t-1} \equiv \boldsymbol{I}_n \otimes \begin{bmatrix} 1 & \boldsymbol{y}_{t-1}' & \dots & \boldsymbol{y}_{t-p}' \end{bmatrix}' \tag{17}$$

$$\boldsymbol{\beta}^{(d,h)} \equiv \text{vec} \begin{bmatrix} \boldsymbol{a}^{(d,h)} & \boldsymbol{A}_1^{(d,h)} & \dots & \boldsymbol{A}_p^{(d,h)} \end{bmatrix}' \tag{18}$$

$$\boldsymbol{\beta}^{(l,h)} \equiv \text{vec} \begin{bmatrix} \boldsymbol{a}^{(l,h)} & \boldsymbol{A}_1^{(l,h)} & \dots & \boldsymbol{A}_p^{(l,h)} \end{bmatrix}' \tag{19}$$

If the VAR is restricted, one or more of the regression coefficients are equal to zero. In this case, the corresponding row is removed from the $\boldsymbol{X}_{t-1}$ matrix, and corresponding parameter removed from the $\boldsymbol{\beta}^{(d,h)}$ and $\boldsymbol{\beta}^{(l,h)}$ vectors. Let $\hat{\boldsymbol{\beta}}^{(d,h)}$ and $\hat{\boldsymbol{\beta}}^{(l,h)}$ denote the OLS estimates of $\boldsymbol{\beta}^{(d,h)}$ and $\boldsymbol{\beta}^{(l,h)}$, and $\boldsymbol{V}^{(d,h)}$ and $\boldsymbol{V}^{(l,h)}$ denote the corresponding covariance matrices, respectively. These are given by:

$$\sqrt{T-p-h}\,(\hat{\boldsymbol{\beta}}^{(d,h)} - \boldsymbol{\beta}^{(d,h)}) \overset{a}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{V}^{(d,h)}) \tag{20}$$

$$\sqrt{T-p-h}\,(\hat{\boldsymbol{\beta}}^{(l,h)} - \boldsymbol{\beta}^{(l,h)}) \overset{a}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{V}^{(l,h)}) \tag{21}$$

I derive the expression for $\boldsymbol{V}^{(d,1)}$ as a function of the VAR residuals.

$$\boldsymbol{V}^{(d,1)} = \Big(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}]\Big)^{-1} \boldsymbol{S}^{(d,1)} \Big(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}]\Big)^{-1} \tag{22}$$

$$\boldsymbol{S}^{(d,1)} \equiv \boldsymbol{\Gamma}_0^{(d,1)} + \boldsymbol{\Gamma}_1^{(d,1)} + \boldsymbol{\Gamma}_1^{(d,1)'} \tag{23}$$

$$\begin{aligned}
\boldsymbol{\Gamma}_0^{(d,1)} &\equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+1}^{(d,1)}\boldsymbol{e}_{t+1}^{(d,1)'}\boldsymbol{X}'_{t-1}] \\
&= \mathrm{E}[\boldsymbol{X}_{t-1}(\boldsymbol{\epsilon}_{t+1} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t)(\boldsymbol{\epsilon}_{t+1} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t)'\boldsymbol{X}'_{t-1}] \\
&= \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}\boldsymbol{X}'_{t-1}] + \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t\boldsymbol{A}_1^{(d,0)'}\boldsymbol{X}'_{t-1}] \\
&= \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}\boldsymbol{X}'_{t-1}] + \mathrm{E}[\boldsymbol{X}_{t}\boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}\boldsymbol{A}_1^{(d,0)'}\boldsymbol{X}'_{t}]
\end{aligned} \tag{24}$$

$$\begin{aligned}
\boldsymbol{\Gamma}_1^{(1)} &\equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+1}^{(d,1)}\boldsymbol{e}_{t}^{(d,1)'}\boldsymbol{X}'_{t-2}] \\
&= \mathrm{E}[\boldsymbol{X}_{t-1}(\boldsymbol{\epsilon}_{t+1} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t)(\boldsymbol{\epsilon}_{t} + \boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_{t-1})'\boldsymbol{X}'_{t-2}] \\
&= \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t\boldsymbol{X}'_{t-2}] \\
&= \mathrm{E}[\boldsymbol{X}_{t}\boldsymbol{A}_1^{(d,0)}\boldsymbol{\epsilon}_{t+1}\boldsymbol{\epsilon}'_{t+1}\boldsymbol{X}'_{t-1}]
\end{aligned} \tag{25}$$

$$\boldsymbol{S}^{(d,1)} = \mathrm{E}[\boldsymbol{d}_{t+1}^{(d,1)}\boldsymbol{d}_{t+1}^{(d,1)'}] \tag{26}$$

$$\boldsymbol{d}_{t+1}^{(d,1)} \equiv (\boldsymbol{X}_{t-1} + \boldsymbol{X}_t\boldsymbol{A}_1^{(d,0)})\boldsymbol{\epsilon}_{t+1} \tag{27}$$

Note that the error structure for $\boldsymbol{e}_{t+1}^{(l,1)}$ is the same as it is for $\boldsymbol{e}_{t+1}^{(d,1)}$. Thus, by the same logic, I derive the expression for $\boldsymbol{V}^{(l,1)}$ as a function of the VAR residuals.

$$\boldsymbol{V}^{(l,1)} = \left(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}]\right)^{-1}\boldsymbol{S}^{(l,1)}\left(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}]\right)^{-1} \tag{28}$$

$$\boldsymbol{S}^{(l,1)} \equiv \boldsymbol{\Gamma}_0^{(l,1)} + \boldsymbol{\Gamma}_1^{(l,1)} + \boldsymbol{\Gamma}_1^{(l,1)'} \tag{29}$$

$$\boldsymbol{\Gamma}_0^{(l,1)} \equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+1}^{(l,1)}\boldsymbol{e}_{t+1}^{(l,1)'}\boldsymbol{X}'_{t-1}] \tag{30}$$

$$\boldsymbol{\Gamma}_1^{(l,1)} \equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+1}^{(l,1)}\boldsymbol{e}_{t}^{(l,1)'}\boldsymbol{X}'_{t-2}] \tag{31}$$

$$\boldsymbol{S}^{(l,1)} = \mathrm{E}[\boldsymbol{d}_{t+1}^{(l,1)}\boldsymbol{d}_{t+1}^{(l,1)'}] \tag{32}$$

$$\boldsymbol{d}_{t+1}^{(l,1)} \equiv (\boldsymbol{X}_{t-1} + \boldsymbol{X}_t\boldsymbol{B}_1^{(l,0)})\boldsymbol{\epsilon}_{t+1} \tag{33}$$

Next, I derive the expression for $\boldsymbol{V}^{(d,h)}$ as a function of the VAR residuals.

$$\boldsymbol{V}^{(d,h)} = \left(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}]\right)^{-1}\boldsymbol{S}^{(d,h)}\left(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}'_{t-1}]\right)^{-1} \tag{34}$$

$$\boldsymbol{S}^{(d,h)} \equiv \boldsymbol{\Gamma}_0^{(d,h)} + \sum_{j=1}^{h}\left(\boldsymbol{\Gamma}_j^{(d,h)} + \boldsymbol{\Gamma}_j^{(d,h)'}\right) \tag{35}$$

$$\boldsymbol{\Gamma}_0^{(d,h)} \equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+h}^{(d,h)}\boldsymbol{e}_{t+h}^{(d,h)'}\boldsymbol{X}'_{t-1}] \tag{36}$$

$$\boldsymbol{\Gamma}_j^{(d,h)} \equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+h}^{(d,h)}\boldsymbol{e}_{t+h-j}^{(d,h)'}\boldsymbol{X}'_{t-1-j}] \tag{37}$$

$$\boldsymbol{S}^{(d,h)} = \mathrm{E}[\boldsymbol{d}_{t+h}^{(d,h)}\boldsymbol{d}_{t+h}^{(d,h)'}] \tag{38}$$

$$\boldsymbol{d}_{t+h}^{(d,h)} \equiv \left(\boldsymbol{X}_{t-1} + \sum_{j=0}^{h-1}\boldsymbol{X}_{t+j}\boldsymbol{A}_1^{(d,j)}\right)\boldsymbol{\epsilon}_{t+h} \tag{39}$$

Last, I derive the expression for $\boldsymbol{V}^{(l,h)}$ as a function of the VAR residuals.

$$\boldsymbol{V}^{(l,h)} = \left(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}']\right)^{-1}\boldsymbol{S}^{(l,h)}\left(\mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}']\right)^{-1} \tag{40}$$

$$\boldsymbol{S}^{(l,h)} \equiv \boldsymbol{\Gamma}_0^{(l,h)} + \sum_{j=1}^{h}\left(\boldsymbol{\Gamma}_j^{(l,h)} + \boldsymbol{\Gamma}_j^{(l,h)'}\right) \tag{41}$$

$$\boldsymbol{\Gamma}_0^{(l,h)} \equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+h}^{(l,h)}\boldsymbol{e}_{t+h}^{(l,h)'}\boldsymbol{X}_{t-1}'] \tag{42}$$

$$\boldsymbol{\Gamma}_j^{(l,h)} \equiv \mathrm{E}[\boldsymbol{X}_{t-1}\boldsymbol{e}_{t+h}^{(l,h)}\boldsymbol{e}_{t+h-j}^{(l,h)'}\boldsymbol{X}_{t-1-j}'] \tag{43}$$

$$\boldsymbol{S}^{(l,h)} = \mathrm{E}[\boldsymbol{d}_{t+h}^{(l,h)}\boldsymbol{d}_{t+h}^{(l,h)'}] \tag{44}$$

$$\boldsymbol{d}_{t+h}^{(l,h)} \equiv \left(\boldsymbol{X}_{t-1} + \sum_{j=0}^{h-1}\boldsymbol{X}_{t+j}\boldsymbol{B}_1^{(l,j)}\right)\boldsymbol{\epsilon}_{t+h} \tag{45}$$

The covariance matrices $\boldsymbol{V}^{(d,h)}$ and $\boldsymbol{V}^{(l,h)}$ are functions of the observed data $\boldsymbol{X}_{t-1}$, the VAR residuals $\boldsymbol{\epsilon}_{t+h}$, and the coefficient matrices on $\boldsymbol{y}_{t-1}$ from the VAR and the 1 through $(h-1)$ direct multi-step regressions, $\boldsymbol{A}_1^{(d,j)}$ and $\boldsymbol{B}_1^{(l,j)}$.

3 Estimation

I now show that the parameters of the VMA process can be estimated by OLS. This then facilitates estimation of the covariance matrices $\boldsymbol{V}^{(d,h)}$ and $\boldsymbol{V}^{(l,h)}$, as per the method in West (1997). Suppose $h = 1$. Let $\hat{\boldsymbol{A}}_1^{(d,0)}$ denote the OLS estimates of $\boldsymbol{A}_1^{(d,0)}$ from the VAR, and $\hat{\boldsymbol{\epsilon}}_{t+1}$ the residuals from the VAR. The estimated covariance matrix $\hat{\boldsymbol{V}}^{(d,1)}$ is given by:

$$\hat{\boldsymbol{V}}^{(d,1)} = \left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1}\hat{\boldsymbol{S}}^{(d,1)}\left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1} \tag{46}$$

$$\hat{\boldsymbol{S}}^{(d,1)} \equiv (T-p-1)^{-1}\sum_{t=p+1}^{T-1}\hat{\boldsymbol{d}}_{t+1}^{(d,1)}\hat{\boldsymbol{d}}_{t+1}^{(d,1)'} \tag{47}$$

$$\hat{\boldsymbol{d}}_{t+1}^{(d,1)} \equiv (\boldsymbol{X}_{t-1} + \boldsymbol{X}_t\hat{\boldsymbol{A}}_1^{(d,0)})\hat{\boldsymbol{\epsilon}}_{t+1} \tag{48}$$

Similarly, the estimated covariance matrix $\hat{\boldsymbol{V}}^{(l,1)}$ is given by:

$$\hat{\boldsymbol{V}}^{(l,1)} = \left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1}\hat{\boldsymbol{S}}^{(l,1)}\left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1} \tag{49}$$

$$\hat{\boldsymbol{S}}^{(l,1)} \equiv (T-p-1)^{-1}\sum_{t=p+1}^{T-1}\hat{\boldsymbol{d}}_{t+1}^{(l,1)}\hat{\boldsymbol{d}}_{t+1}^{(l,1)\prime} \tag{50}$$

$$\hat{\boldsymbol{d}}_{t+1}^{(l,1)} \equiv (\boldsymbol{X}_{t-1} + \boldsymbol{X}_t\hat{\boldsymbol{B}}_1^{(l,0)})\hat{\boldsymbol{\epsilon}}_{t+1} \tag{51}$$

$$\hat{\boldsymbol{B}}_1^{(l,0)} \equiv \boldsymbol{I}_n + \hat{\boldsymbol{A}}_1^{(d,0)} \tag{52}$$

Thus, to estimate the covariance matrices $\hat{\boldsymbol{V}}^{(d,1)}$ and $\hat{\boldsymbol{V}}^{(l,1)}$, we use the OLS estimates from the VAR. Let $(\hat{\boldsymbol{A}}_1^{(d,1)}, \ldots, \hat{\boldsymbol{A}}_1^{(d,h-1)})$ denote the OLS estimates of $(\boldsymbol{A}_1^{(d,1)}, \ldots, \boldsymbol{A}_1^{(d,h-1)})$ from the 1 through $(h-1)$ direct multi-step regressions. The estimated covariance matrix $\hat{\boldsymbol{V}}^{(d,h)}$ is given by:

$$\hat{\boldsymbol{V}}^{(d,h)} = \left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1}\hat{\boldsymbol{S}}^{(d,h)}\left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1} \tag{53}$$

$$\hat{\boldsymbol{S}}^{(d,h)} \equiv (T-p-h)^{-1}\sum_{t=p+1}^{T-h}\hat{\boldsymbol{d}}_{t+h}^{(d,h)}\hat{\boldsymbol{d}}_{t+h}^{(d,h)\prime} \tag{54}$$

$$\hat{\boldsymbol{d}}_{t+h}^{(d,h)} \equiv \left(\boldsymbol{X}_{t-1} + \sum_{j=0}^{h-1}\boldsymbol{X}_{t+j}\hat{\boldsymbol{A}}_1^{(d,j)}\right)\hat{\boldsymbol{\epsilon}}_{t+h} \tag{55}$$

Similarly, the estimated covariance matrix $\hat{\boldsymbol{V}}^{(l,h)}$ is given by:

$$\hat{\boldsymbol{V}}^{(l,h)} = \left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1}\hat{\boldsymbol{S}}^{(l,h)}\left((T-p)^{-1}\sum_{t=p+1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1} \tag{56}$$

$$\hat{\boldsymbol{S}}^{(l,h)} \equiv (T-p-h)^{-1}\sum_{t=p+1}^{T-h}\hat{\boldsymbol{d}}_{t+h}^{(l,h)}\hat{\boldsymbol{d}}_{t+h}^{(l,h)\prime} \tag{57}$$

$$\hat{\boldsymbol{d}}_{t+h}^{(l,h)} \equiv \left(\boldsymbol{X}_{t-1} + \sum_{j=0}^{h-1}\boldsymbol{X}_{t+j}\hat{\boldsymbol{B}}_1^{(l,j)}\right)\hat{\boldsymbol{\epsilon}}_{t+h} \tag{58}$$

$$\hat{\boldsymbol{B}}_1^{(l,j)} \equiv \boldsymbol{I}_n + \sum_{k=0}^{j}\hat{\boldsymbol{A}}_1^{(d,k)} \tag{59}$$

Thus, to estimate the covariance matrices $\hat{\boldsymbol{V}}^{(d,h)}$ and $\hat{\boldsymbol{V}}^{(l,h)}$, we use the OLS estimates from the VAR and the 1 through $(h-1)$ direct multi-step regressions.

4 Simulation

I conduct a simulation using a VAR(1) model to assess the performance of the proposed estimator's standard errors, relative to the Newey *et al.* (1987) estimator's standard errors, henceforth denoted NW. The data generating process (DGP) is given by the following:

$$y_{1,t} = a_1 + A_{1,11}y_{1,t-1} + \epsilon_{1,t} \tag{60}$$

$$y_{2,t} = a_2 + A_{1,21}y_{1,t-1} + \epsilon_{2,t} \tag{61}$$

$$\begin{bmatrix}\epsilon_{1,t}\\ \epsilon_{2,t}\end{bmatrix} \sim \mathrm{N}\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right] \tag{62}$$

The first variable y_1 follows an AR(1) process. The covariance matrix is a function only of $\epsilon_{1,t}$, since y_2 has no effect on y_1. The multivariate part of the DGP is intended to capture a return predictability regression, where y_1 is a persistent predictor and y_2 is an asset return.

In the simulation, $a_1 = a_2 = 0$, $A_{1,11} = 0.8$, $A_{1,21} = 0.2$, and $\rho_{12} \equiv \frac{\sigma_{12}}{\sigma_1\sigma_2} = -0.2$. σ_1

and σ_2 are set so that y_1 and y_2 have unit variance. Then $A_{1,21}$ is also the correlation coefficient between y_1 and y_2. This implies an R^2 in the return predictability regression of $(0.2)^2 = 0.04$. I run two sets of simulations, and the total number of simulations for each case is $M = 100,000$. Let H denote the farthest horizon, so that $h = 1, \ldots, H$. In the first simulation, $T = 1260$ and $H = 20$. This is equivalent to observing five years of daily data, and forecasting one month ahead. In the second simulation, $T = 240$ and $H = 5$. This is equivalent to observing twenty years of monthly data, and forecasting six months ahead.

I assess the performance of the proposed estimator's standard errors for direct multi-step and long horizon forecasts of both y_1 and y_2.

$$y_{1,t+h} = a_1^{(d,h)} + A_{1,11}^{(d,h)} y_{1,t-1} + e_{1,t+h}^{(d,h)} \tag{63}$$

$$\sum_{s=0}^{h} y_{1,t+s} = a_1^{(l,h)} + A_{1,11}^{(l,h)} y_{1,t-1} + e_{1,t+h}^{(l,h)} \tag{64}$$

$$y_{2,t+h} = a_2^{(d,h)} + A_{1,21}^{(d,h)} y_{1,t-1} + e_{2,t+h}^{(d,h)} \tag{65}$$

$$\sum_{s=0}^{h} y_{2,t+s} = a_2^{(l,h)} + A_{1,21}^{(l,h)} y_{1,t-1} + e_{2,t+h}^{(l,h)} \tag{66}$$

The parameters of interest are $A_{1,11}^{(d,h)}$, $A_{1,11}^{(l,h)}$, $A_{1,21}^{(d,h)}$ and $A_{1,21}^{(l,h)}$. For each simulation, the proposed standard error (P) and the NW standard error are recorded. The NW standard error has a lag length set to $(5+h)$.[1] The standard errors are compared to the true standard error. I report the root mean squared error (RMSE), and the 90%, 95%, and 99% coverage probabilities. These are defined as follows.

Let se denote the true or theoretical standard error, and $\hat{se}_m$ denote the estimated standard error from simulation m, where $m = 1, \ldots, M$. Note that for ease of interpretation and comparability across horizons and coefficients, the RMSE is divided by the true standard

[1]The results are not sensitive to setting the lag length equal to h.

error. It is given by:

$$\text{RMSE} = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{se}_m - se)^2}}{se} \tag{67}$$

Let b denote the true or theoretical value of the regression parameter and $\hat{b}_m$ the OLS estimate from simulation m. Let α denote the significance level, so that $(1-\alpha)$ is the coverage probability. Let $t_{crit(\alpha/2)}$ denote the critical value from the student-t distribution of a two tailed test with significance level α. The $(1-\alpha)$ coverage probability is given by:

$$CP_{(1-\alpha),m} = \begin{cases} 1 & \text{if} \quad \hat{b}_m - \hat{se}_m t_{crit(\alpha/2)} < b < \hat{b}_m + \hat{se}_m t_{crit(\alpha/2)} \\ 0 & \text{otherwise} \end{cases} \tag{68}$$

$$CP_{(1-\alpha)} = \frac{1}{M}\sum_{m=1}^{M} CP_{(1-\alpha),m} \tag{69}$$

Table 1 and 2 report results for each first simulation by horizon. Figures 1 plots the RMSE for both simulations by horizon. Table 3 presents a summary of the simulation results. The RMSE column is the average ratio of the RMSE of the proposed estimator to the NW estimator, across horizons. The CP columns are the average ratios of the absolute difference between the coverage probability and the nominal level of the proposed estimator to the NW estimator, across horizons. Ratios less than one indicate an outperformance of the proposed estimator, and ratios greater than one indicate outperformance of the NW estimator.

First, consider the results for $A_{1,11}^{(d,h)}$. The proposed estimator has an RMSE ratio of 0.547 in the first simulation and 0.426 in second simulation. The 95% coverage probability ratio is 0.158 in the first simulation, and 0.403 in the second simulation. For $A_{1,11}^{(l,h)}$, the proposed estimator has an RMSE ratio of 0.476 in the first simulation and 0.410 in second simulation. The 95% coverage probability ratio is 0.325 in the first simulation, and 0.374 in the second simulation. For both parameters, the coverage probabilities are closer to their nominal levels in the first simulation, when the sample size is larger.

Next, consider the results for $A_{1,21}^{(d,h)}$. The proposed estimator has an RMSE ratio of 0.636 in the first simulation and 0.765 in second simulation. The 95% coverage probability ratio is 0.327 in the first simulation, and 0.084 in the second simulation. These regressions show the smallest gain from using the proposed standard erros, partly because these regressions have less autocorrelation in the errors, compared to the others. For $A_{1,21}^{(l,h)}$, the proposed estimator has an RMSE ratio of 0.384 in the first simulation and 0.557 in second simulation. The 95% coverage probability ratio is 0.055 in the first simulation, and 0.047 in the second simulation. For both parameters, the coverage probabilities of the proposed estimator are approximately equal to the nominal level in both simulations. The gain in coverage probabilities are strongest for the long horizon return predictability regressions.

Overall, the simulation results indicate a substantial improvement in efficiency and accuracy of the proposed estimator, relative to the NW estimator. This result holds for both direct multi-step and long horizon forecasts, univariate and multivariate forecasts, large and small sample sizes.

5 Empirical Examples

5.1 Variance Risk Premium and Currency Returns

I present two empirical examples that mirror the model specification and samples sizes in the simulation. In the first example, I use the euro currency variance risk premium to predict the euro currency futures return. The variance risk premium measures the cost of increasing against an increase in variance. Bollerslev *et al.* (2011) show that it is is a good proxy for overall market risk aversion. It has been shown to have predictive power over currency returns in certain contexts. Della Corte *et al.* (2016) show that currency volatility risk premia predict exchange rate returns. Currencies with cheap volatility insurance outperform currencies with expensive volatility insurance. Londono & Zhou (2017) find that currency variance risk premia predict future changes in exchange rates. Menkhoff *et al.* (2012) find a strong link between currency variance risk premia and returns from currency carry trades.

I use data on the euro currency futures and options contracts traded on the Chicago Mercantile Exchange, sampled from December 2016 until December 2021[2]. The sample size is $T = 1260$. The farthest forecast horizon is $H = 20$ (one month). The euro currency futures contract is the quantity of US dollars per euro. Using the options and futures data, I compute daily values of the variance risk premium, denoted by VP_t. Next, I compute daily log returns from holding the futures contracts, denote by r_t. The variance risk premium requires estimation of implied variance and expected variance. I estimate implied variance using model free implied variance, as per the method in Andersen & Bondarenko (2007). I estimate expected variance using using a long horizon mixed data sampling (MIDAS) volatility forecasting regression, as per the methods in Ghysels *et al.* (2006) and Ghysels *et al.* (2019). For further detail on the construction of currency variance risk premia, please see Dossani (2021). I estimate the following direct and long horizon forecasting regressions, which follow the same structure as the model used in the simulation.

$$VP_{t+h} = a_1^{(d,h)} + A_{1,11}^{(d,h)} VP_{t-1} + e_{1,t+h}^{(d,h)} \tag{70}$$

$$\sum_{s=0}^{h} VP_{t+s} = a_1^{(l,h)} + A_{1,11}^{(l,h)} VP_{t-1} + e_{1,t+h}^{(l,h)} \tag{71}$$

$$r_{t+h} = a_2^{(d,h)} + A_{1,21}^{(d,h)} VP_{t-1} + e_{2,t+h}^{(d,h)} \tag{72}$$

$$\sum_{s=0}^{h} r_{t+s} = a_2^{(l,h)} + A_{1,21}^{(l,h)} VP_{t-1} + e_{2,t+h}^{(l,h)} \tag{73}$$

The parameters of interest are $A_{1,11}^{(d,h)}$, $A_{1,11}^{(l,h)}$, $A_{1,21}^{(d,h)}$ and $A_{1,21}^{(l,h)}$. The variable VP_t is standardized so that coefficients $A_{1,21}^{(d,h)}$ and $A_{1,21}^{(l,h)}$ can be interpreted as the impact of a one standard deviation increase in the variance risk premium. Table 4 reports results from the regressions. For each regression, I report the coefficient, the proposed standard error (P), the NW standard error with lag length set to $(5 + h)$, and the associated p-values testing

[2]The futures data are from IQ Feed and the options data are from Quandl.

the null hypothesis that the coefficient is equal to zero. The top half of figure 2 presents the results in the form of the regression coefficient by horizon alongside the 95% confidence intervals using both the proposed and NW standard errors.

Across most coefficients and horizons, the proposed standard errors are larger than the NW standard errors. Associated p-values for the proposed estimator are larger, and confidence intervals are wider. This mirrors the simulation results, whereby coverage probabilities for the proposed estimator are larger and closer to their nominal values. The NW standard errors tend to over-reject more often than the proposed estimator in the simulation, and this patterns holds in the empirical example. The magnitudes are economically meaningful. Consider the coefficient $A_{1,21}^{(l,h)}$ for $h = 20$. This coefficient measures the predictive power of the variance risk premium for the subsequent one month return. This horizon is particularly meaningful because the variance risk premium is computed at the one month horizon, and is thus an estimate of the subsequent one month risk premium. The coefficient estimate suggests that a one standard deviation increase in the variance risk premium is associated with a 0.365% subsequent depreciation of the euro, relative to the US dollar. Economically, this is consistent with an increase in risk aversion leading to a decline in the value of the euro. The proposed standard error is approximately three times larger than the NW standard error. The p-value is 0.045 using the proposed standard error, and 0.000 using the NW standard error. Using the proposed standard errors suggests that the evidence of predictability is weaker, significant at the 5% but not 1% level. For a substantial number of coefficients reported in the table, the proposed standard errors would not reject the null hypothesis of a zero coefficient, whereas the NW standard errors would reject.

5.2 Inflation and the Equity Premium

In the second empirical example, I use the lagged rate of inflation to predict the equity premium. A variety of studies, such as Campbell & Vuolteenaho (2004), Fama & Schwert (1977), and Fama (1981), document a negative empirical relationship between the inflation rate and the equity premium. These studies use the lagged rate of inflation, as the CPI is reported around 2-3 weeks after the month ends.

The infation rate is computed as the log change in the consumer price index. The data come from FRED[3]. The equity premium is the excess return on the market, defined as the value weighted return of all CRSP firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ. The data come from the Ken French Data Library[4]. Inflation rates and the equity premium are sampled monthly. The data go from December 2001 until December 2021, and the sample size is $T = 240$. The farthest forecast horizon is $H = 5$ (six months). Let Inf_t denote the lagged inflation rate, and r_t denote the equity premium. I estimate the following direct and long horizon forecasting regressions, which follow the same structure as the model used in the simulation.

$$Inf_{t+h} = a_1^{(d,h)} + A_{1,11}^{(d,h)} Inf_{t-1} + e_{1,t+h}^{(d,h)} \tag{74}$$

$$\sum_{s=0}^{h} Inf_{t+s} = a_1^{(l,h)} + A_{1,11}^{(l,h)} Inf_{t-1} + e_{1,t+h}^{(l,h)} \tag{75}$$

$$r_{t+h} = a_2^{(d,h)} + A_{1,21}^{(d,h)} Inf_{t-1} + e_{2,t+h}^{(d,h)} \tag{76}$$

$$\sum_{s=0}^{h} r_{t+s} = a_2^{(l,h)} + A_{1,21}^{(l,h)} Inf_{t-1} + e_{2,t+h}^{(l,h)} \tag{77}$$

The parameters of interest are $A_{1,11}^{(d,h)}$, $A_{1,11}^{(l,h)}$, $A_{1,21}^{(d,h)}$ and $A_{1,21}^{(l,h)}$. Table 5 reports results from the regressions. For each regression, I report the coefficient, the proposed standard error (P), the NW standard error with lag length set to $(5 + h)$, and the associated p-values testing the null hypothesis that the coefficient is equal to zero. The bottom half of figure 2 presents the results in the form of the regression coefficient by horizon alongside the 95% confidence intervals using both the proposed and NW standard errors.

Across most coefficients and horizons for the long horizon regressions, the proposed stan-

[3]https://fred.stlouisfed.org/series/CPIAUCSL

[4]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

dard errors are larger than the NW standard errors. Associated p-values for the proposed estimator are larger, and confidence intervals are wider. This mirrors the simulation results, whereby coverage probabilities for the proposed estimator are larger and closer to their nominal values. For the direct multi-step regressions, the proposed standard errors are sometimes larger the the NW standard errors, and occasionally smaller. As compared wit the first empirical example, the inflation rate is less persistent than the variance risk premium, and this reduces the gap between the standard errors in the direct multi-step forecasts. The pattern of the p-values is similar to the first empirical example. There are a substantial number of coefficients reported in the table where the proposed standard errors would not reject the null hypothesis of a zero coefficient, whereas the NW standard errors would reject.

6 Conclusion

This paper proposes and evaluates a new method for inference in direct multi-step and long horizon forecasting regressions, using the HAC covariance estimator in West (1997), and presents two relevant empirical applications. The residuals from both direct multi-step and long horizon forecasting regressions are serially correlated, and can be expressed as a VMA of the VAR residuals. The proposed estimator imposes the VMA structure on the serially correlated residuals to estimate the covariance matrix of the OLS estimates of direct multi-step and long horizon forecasting regressions.

The computation of the proposed estimator is straightforward, as it is a function of OLS parameter estimates. A simulation study indicates substantial improvement in both the efficiency and accuracy of the proposed estimator, relative to the NW estimator. A potential future extension of this paper is an application to Bayesian estimation of direct multi-step and long horizon forecasting regressions. For example, Lusompa (2021) and Miranda-Agrippino *et al.* (2021) propose methods for Bayesian local projections. Applications to long horizon regression forecasts could improve the accuracy of return predictability regressions.

References

Andersen, Torben, & Bondarenko, Oleg. 2007. Construction and Interpretation of Model-Free Implied Volatility. *Volatility as an Asset Class*, 10, 141–181.

Bollerslev, Tim, Gibson, Michael, & Zhou, Hao. 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of econometrics*, **160**(1), 235–245.

Britten-Jones, Mark, Neuberger, Anthony, & Nolte, Ingmar. 2011. Improved inference in regression with overlapping observations. *Journal of Business Finance & Accounting*, **38**(5-6), 657–683.

Campbell, John Y, & Vuolteenaho, Tuomo. 2004. Inflation illusion and stock prices. *American Economic Review*, **94**(2), 19–23.

Cochrane, John H, & Piazzesi, Monika. 2005. Bond risk premia. *American economic review*, **95**(1), 138–160.

Della Corte, Pasquale, Ramadorai, Tarun, & Sarno, Lucio. 2016. Volatility risk premia and exchange rate predictability. *Journal of Financial Economics*, **120**(1), 21–40.

Dossani, Asad. 2021. Monetary Policy and Currency Variance Risk Premia. *Available at SSRN 3300244*.

Fama, Eugene F. 1981. Stock returns, real activity, inflation, and money. *The American economic review*, **71**(4), 545–565.

Fama, Eugene F, & Schwert, G William. 1977. Asset returns and inflation. *Journal of financial economics*, **5**(2), 115–146.

Ghysels, Eric, Santa-Clara, Pedro, & Valkanov, Rossen. 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, **131**(1-2), 59–95.

Ghysels, Eric, Plazzi, Alberto, Valkanov, Rossen, Rubia, Antonio, & Dossani, Asad. 2019. Direct Versus Iterated Multiperiod Volatility Forecasts. *Annual Review of Financial Economics*, **11**, 173–195.

Hansen, Lars Peter, & Hodrick, Robert J. 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of political economy*, **88**(5), 829–853.

Hjalmarsson, Erik. 2011. New methods for inference in long-horizon regressions. *Journal of Financial and Quantitative Analysis*, **46**(3), 815–839.

Hodrick, Robert J. 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies*, **5**(3), 357–386.

Jordà, Òscar. 2005. Estimation and inference of impulse responses by local projections. *American economic review*, **95**(1), 161–182.

Lazarus, Eben, Lewis, Daniel J, Stock, James H, & Watson, Mark W. 2018. HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics*, **36**(4), 541–559.

Londono, Juan M, & Zhou, Hao. 2017. Variance risk premiums and the forward premium puzzle. *Journal of Financial Economics*, **124**(2), 415–440.

Lusompa, Amaze. 2021. Local projections, autocorrelation, and efficiency. *Federal Reserve Bank of Kansas City Working Paper.*

McCracken, Michael W, & McGillicuddy, Joseph T. 2019. An empirical investigation of direct and iterated multistep conditional forecasts. *Journal of Applied Econometrics*, **34**(2), 181–204.

Menkhoff, Lukas, Sarno, Lucio, Schmeling, Maik, & Schrimpf, Andreas. 2012. Carry trades and global foreign exchange volatility. *The Journal of Finance*, **67**(2), 681–718.

Miranda-Agrippino, Silvia, Ricco, Giovanni, *et al.* 2021. *Bayesian Local Projections.* University of Warwick, Department of Economics.

Montiel Olea, José Luis, & Plagborg-Møller, Mikkel. 2021. Local projection inference is simpler and more robust than you think. *Econometrica*, **89**(4), 1789–1823.

Newey, Whitney K, West, Kenneth D, *et al.* 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**(3), 703–708.

Ramey, Valerie A. 2016. Macroeconomic shocks and their propagation. *Handbook of macroeconomics*, **2**, 71–162.

Richardson, Matthew, & Smith, Tom. 1991. Tests of financial models in the presence of overlapping observations. *The Review of Financial Studies*, **4**(2), 227–254.

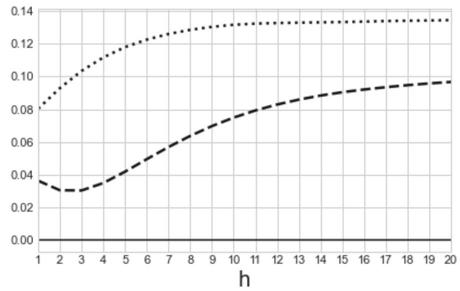
Valkanov, Rossen. 2003. Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics*, **68**(2), 201–232.

Wei, Min, & Wright, Jonathan H. 2013. Reverse regressions and long-horizon forecasting. *Journal of Applied Econometrics*, **28**(3), 353–371.

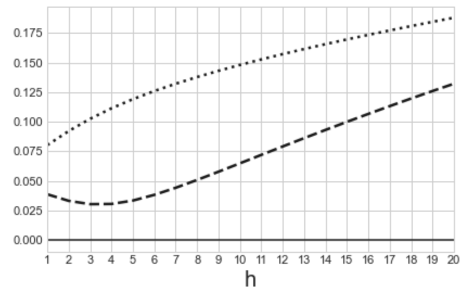
West, Kenneth D. 1997. Another heteroskedasticity-and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics*, **76**(1-2), 171–191.

Figure 1: Simulation Results: RMSE

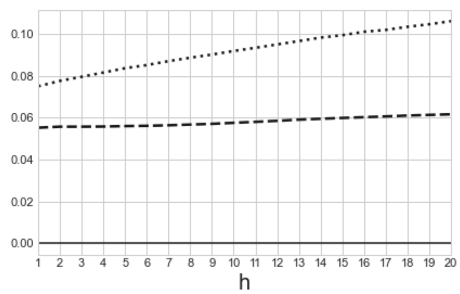
(a) $A^{(d,h)}_{1,11}$, $T = 1260$, $H = 20$



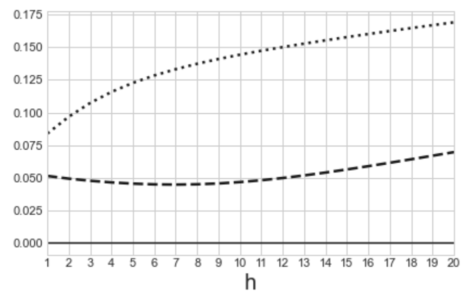
(b) $A^{(l,h)}_{1,11}$, $T = 1260$, $H = 20$



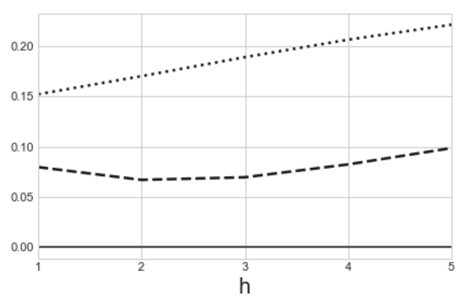
(c) $A^{(d,h)}_{1,21}$, $T = 1260$, $H = 20$



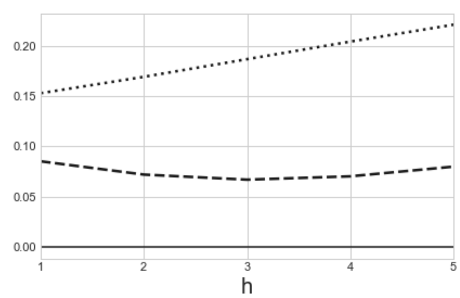
(d) $A^{(l,h)}_{1,21}$, $T = 1260$, $H = 20$



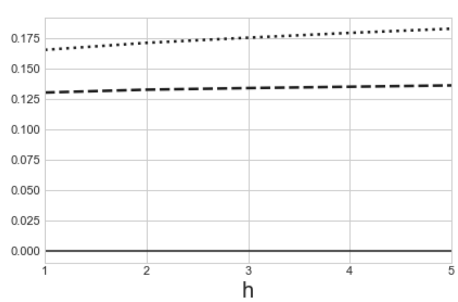
(e) $A^{(d,h)}_{1,11}$, $T = 240$, $H = 5$



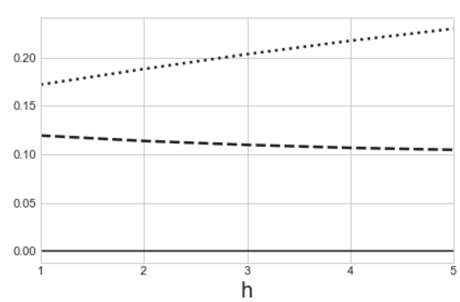
(f) $A^{(l,h)}_{1,11}$, $T = 240$, $H = 5$



(g) $A^{(d,h)}_{1,21}$, $T = 240$, $H = 5$



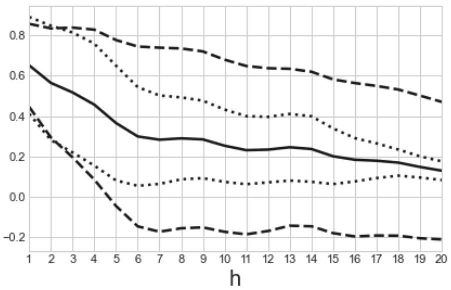
(h) $A^{(l,h)}_{1,21}$, $T = 240$, $H = 5$



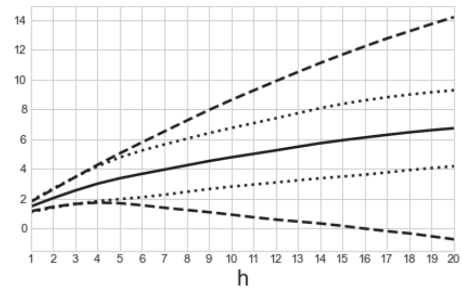
These figures plot the root mean squared error (RMSE) from the simulation results by horizon. The dashed line is the proposed estimator and the dotted line is the NW estimator.

Figure 2: Empirical Examples

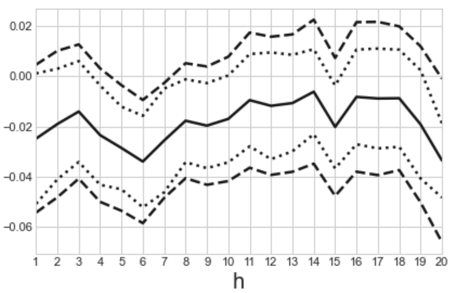
(a) $A_{1,11}^{(d,h)}$: Variance Risk Premium



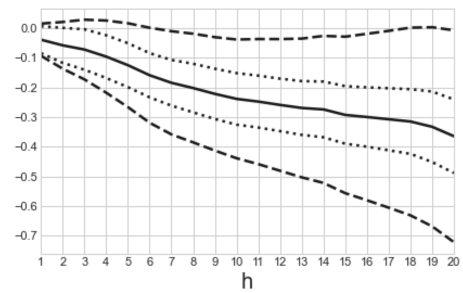
(b) $A_{1,11}^{(l,h)}$: Variance Risk Premium



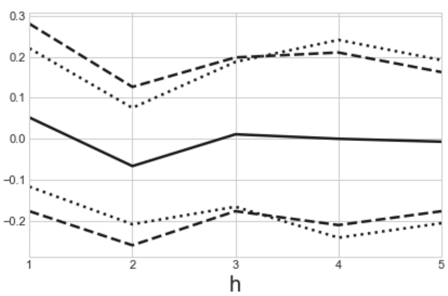
(c) $A_{1,21}^{(d,h)}$: Variance Risk Premium



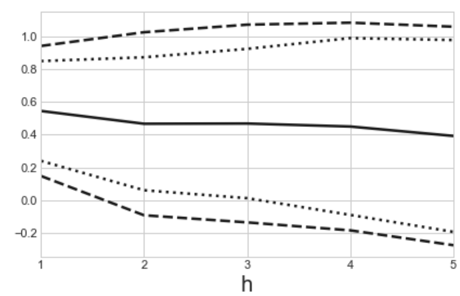
(d) $A_{1,21}^{(l,h)}$: Variance Risk Premium



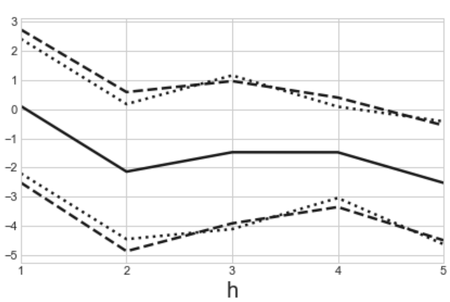
(e) $A_{1,11}^{(d,h)}$: Inflation



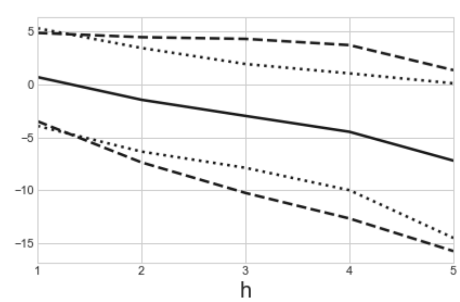
(f) $A_{1,11}^{(l,h)}$: Inflation



(g) $A_{1,21}^{(d,h)}$: Inflation



(h) $A_{1,21}^{(l,h)}$: Inflation



These figures report results for the two empirical examples (Variance Risk Premium and Inflation). Each figure plots the coefficient estimates (solid line) by horizon, alonside the 95% confidence intervals using the proposed estimator (dashed line) and the NW estimator (dotted line).

Table 1: **Simulation Results:** $T = 1260$ & $H = 20$

		RMSE		90% CP		95% CP		99% CP	
	h	P	NW	P	NW	P	NW	P	NW
$A_{1,11}^{(d,h)}$	1	0.036	0.080	0.894	0.877	0.945	0.933	0.988	0.983
	5	0.042	0.118	0.892	0.858	0.943	0.919	0.985	0.976
	10	0.075	0.132	0.894	0.858	0.944	0.918	0.985	0.976
	15	0.090	0.133	0.900	0.862	0.949	0.922	0.988	0.978
	20	0.097	0.134	0.905	0.866	0.954	0.925	0.992	0.979
$A_{1,11}^{(l,h)}$	1	0.039	0.080	0.894	0.879	0.945	0.934	0.988	0.983
	5	0.033	0.119	0.891	0.857	0.942	0.916	0.986	0.975
	10	0.065	0.148	0.887	0.844	0.936	0.906	0.982	0.968
	15	0.100	0.169	0.880	0.837	0.931	0.898	0.976	0.962
	20	0.132	0.188	0.875	0.831	0.924	0.892	0.971	0.956
$A_{1,21}^{(d,h)}$	1	0.055	0.075	0.899	0.897	0.949	0.946	0.990	0.988
	5	0.056	0.084	0.901	0.895	0.952	0.946	0.990	0.988
	10	0.058	0.092	0.902	0.891	0.952	0.944	0.991	0.987
	15	0.060	0.099	0.903	0.888	0.953	0.941	0.991	0.986
	20	0.062	0.106	0.906	0.890	0.954	0.941	0.992	0.986
$A_{1,21}^{(l,h)}$	1	0.052	0.084	0.900	0.884	0.949	0.937	0.990	0.985
	5	0.046	0.123	0.900	0.864	0.951	0.924	0.990	0.979
	10	0.047	0.144	0.901	0.857	0.952	0.918	0.991	0.977
	15	0.056	0.158	0.902	0.854	0.952	0.917	0.991	0.975
	20	0.070	0.169	0.903	0.853	0.953	0.915	0.991	0.975

This table reports the simulation results by horizon for $T = 1260$ and $H = 20$, and 100,000 simulations. Results are reported for the proposed estimator (P), and the NW estimator with lag length equal to $(5 + h)$. The results include root mean squared error (RMSE), 90% coverage probability (90% CP), 95% coverage probability (95% CP), and 99% coverage probability (99% CP).

Table 2: Simulation Results: $T = 240$ **&** $H = 5$

		RMSE		90% CP		95% CP		99% CP	
	h	P	NW	P	NW	P	NW	P	NW
$A_{1,11}^{(d,h)}$	1	0.080	0.152	0.877	0.840	0.931	0.902	0.982	0.967
	2	0.067	0.170	0.871	0.823	0.926	0.887	0.978	0.958
	3	0.069	0.189	0.866	0.810	0.921	0.877	0.973	0.950
	4	0.082	0.206	0.863	0.801	0.917	0.869	0.970	0.945
	5	0.098	0.221	0.860	0.794	0.913	0.862	0.968	0.941
$A_{1,11}^{(l,h)}$	1	0.085	0.153	0.877	0.840	0.933	0.903	0.983	0.967
	2	0.072	0.169	0.872	0.823	0.927	0.888	0.980	0.957
	3	0.067	0.187	0.867	0.809	0.923	0.875	0.976	0.949
	4	0.070	0.204	0.862	0.797	0.918	0.864	0.971	0.941
	5	0.080	0.221	0.857	0.786	0.912	0.853	0.967	0.932
$A_{1,21}^{(d,h)}$	1	0.131	0.166	0.897	0.879	0.948	0.933	0.989	0.983
	2	0.133	0.172	0.898	0.875	0.949	0.930	0.990	0.981
	3	0.134	0.176	0.899	0.874	0.951	0.929	0.991	0.981
	4	0.135	0.180	0.901	0.872	0.952	0.928	0.991	0.979
	5	0.136	0.183	0.903	0.869	0.953	0.925	0.992	0.978
$A_{1,21}^{(l,h)}$	1	0.119	0.172	0.895	0.859	0.946	0.919	0.989	0.976
	2	0.114	0.188	0.896	0.847	0.947	0.909	0.989	0.971
	3	0.110	0.204	0.897	0.838	0.949	0.900	0.989	0.966
	4	0.107	0.217	0.899	0.831	0.950	0.894	0.990	0.963
	5	0.105	0.230	0.901	0.825	0.951	0.889	0.991	0.959

This table reports the simulation results by horizon for $T = 240$ and $H = 5$, and 100,000 simulations. Results are reported for the proposed estimator (P), and the NW estimator with lag length equal to $(5 + h)$. The results include root mean squared error (RMSE), 90% coverage probability (90% CP), 95% coverage probability (95% CP), and 99% coverage probability (99% CP).

Table 3: Simulation Results Summary

	$T = 1260$ & $H = 20$			
	RMSE	90% CP	95% CP	99% CP
$A_{1,11}^{(d,h)}$	0.547	0.131	0.158	0.233
$A_{1,11}^{(l,h)}$	0.476	0.272	0.325	0.394
$A_{1,21}^{(d,h)}$	0.636	0.274	0.327	0.240
$A_{1,21}^{(l,h)}$	0.384	0.041	0.055	0.051
	$T = 240$ & $H = 5$			
	RMSE	90% CP	95% CP	99% CP
$A_{1,11}^{(d,h)}$	0.426	0.376	0.403	0.406
$A_{1,11}^{(l,h)}$	0.410	0.372	0.374	0.349
$A_{1,21}^{(d,h)}$	0.765	0.077	0.084	0.095
$A_{1,21}^{(l,h)}$	0.557	0.054	0.047	0.045

This table reports the summary of the simulation results. The RMSE column is the average ratio of the RMSE of the proposed estimator to the NW estimator, across horizons. The CP columns are the average ratios of the absolute difference between the coverage probability and the nominal level of the proposed estimator to the NW estimator, across horizons. Ratios less than one indicate an outperformance of the proposed estimator, and ratios greater than one indicate outperformance of the NW estimator.

Table 4: Empirical Example: Variance Risk Premium and Currency Returns

	h	Coef	P-se	NW-se	P-pv	NW-pv
$A_{1,11}^{(d,h)}$	1	0.651	0.105	0.122	0.000	0.000
	5	0.365	0.209	0.145	0.081	0.012
	10	0.253	0.217	0.091	0.244	0.006
	15	0.201	0.194	0.070	0.300	0.004
	20	0.130	0.173	0.024	0.454	0.000
$A_{1,11}^{(l,h)}$	1	1.470	0.161	0.190	0.000	0.000
	5	3.373	0.854	0.714	0.000	0.000
	10	4.782	1.963	1.005	0.015	0.000
	15	5.932	2.936	1.246	0.044	0.000
	20	6.738	3.806	1.303	0.077	0.000
$A_{1,21}^{(d,h)}$	1	−0.025	0.015	0.013	0.098	0.061
	5	−0.029	0.013	0.008	0.025	0.001
	10	−0.017	0.013	0.009	0.179	0.054
	15	−0.020	0.014	0.008	0.150	0.017
	20	−0.033	0.017	0.008	0.044	0.000
$A_{1,21}^{(l,h)}$	1	−0.039	0.028	0.023	0.156	0.089
	5	−0.125	0.072	0.038	0.083	0.001
	10	−0.239	0.102	0.044	0.019	0.000
	15	−0.293	0.134	0.049	0.030	0.000
	20	−0.365	0.182	0.063	0.045	0.000

This table report results for the empirical example using data on dollar euro futures returns and variance risk premia. For each parameter and horizon, the table reports the coefficient estimate (Coef), the proposed standard error (P-se), the NW standard error (NW-se), the p-value using the proposed standard errors (P-pv), and the p-value using NW standard errors (NW-pv). The p-value tests the hypothesis that the coefficient is equal to zero.

Table 5: Empirical Example: Inflation and Equity Risk Premium

	h	Coef	P-se	NW-se	P-pv	NW-pv
	1	0.052	0.116	0.086	0.655	0.545
	2	−0.067	0.098	0.072	0.497	0.355
$A_{1,11}^{(d,h)}$	3	0.011	0.095	0.090	0.910	0.904
	4	−0.000	0.107	0.123	0.999	0.999
	5	−0.007	0.086	0.101	0.934	0.944
	1	0.544	0.201	0.154	0.007	0.001
	2	0.466	0.283	0.206	0.101	0.024
$A_{1,11}^{(l,h)}$	3	0.467	0.306	0.231	0.128	0.044
	4	0.449	0.321	0.274	0.164	0.102
	5	0.392	0.338	0.297	0.248	0.188
	1	0.103	1.330	1.169	0.938	0.930
	1	0.103	1.330	1.169	0.938	0.930
	2	−2.140	1.383	1.174	0.123	0.070
$A_{1,21}^{(d,h)}$	3	−1.476	1.237	1.336	0.234	0.270
	4	−1.478	0.951	0.794	0.122	0.064
	5	−2.523	0.998	1.067	0.012	0.019
	1	0.665	2.112	2.335	0.753	0.776
	2	−1.473	2.993	2.476	0.623	0.552
$A_{1,21}^{(l,h)}$	3	−2.991	3.682	2.480	0.417	0.229
	4	−4.487	4.144	2.792	0.280	0.109
	5	−7.203	4.325	3.695	0.097	0.052

This table report results for the empirical example using data on inflation and the equity risk preium. For each parameter and horizon, the table reports the coefficient estimate (Coef), the proposed standard error (P-se), the NW standard error (NW-se), the p-value using the proposed standard errors (P-pv), and the p-value using NW standard errors (NW-pv). The p-value tests the hypothesis that the coefficient is equal to zero.