ABSTRACT

Title of dissertation:	ESSAYS ON ECONOMETRICS AND MACRO-FINANCE
	Mao, Zi-Ying
	Doctor of Philosophy, 2019
Dissertation directed by:	Prof. Steven Heston, Department of Finance
	Prof. Guido Kuersteiner, Department of Economics

This dissertation consists of three chapters on empirical macro-finance and the associated econometric methods.

In the first chapter, I develop a semiparametric single-index method for estimating multivariate jump-diffusion processes to model federal funds futures. I find that high-frequency changes in federal funds futures around FOMC announcements, which are the predominant measure of monetary policy shocks in the asset pricing and macroeconomic literature, are strongly forecastable by the estimated models, suggesting that they are not truly exogeneous. In contrast, the unexpected changes in federal funds futures on FOMC announcement days constructed from the semiparametric method are unforecastable by construction, and are strongly correlated with, but not the same as such high-frequency changes around FOMC announcements, suggesting that they are a better measure of monetary shocks.

In the second chapter, I study the predictability of bond yields. I find that federal funds futures, a proxy of monetary shocks, exhibit strong forecasting power on bond yields conditional on information contained in the cross section of the yield curve. Such additional return-forecasting information is effectively summarized by a single factor, and is

not captured by unspanned macro factors. By focusing on the return-forecastability of trading strategies that take opposite positions at two different tenors by equal amount and unwind these positions one-day later, I bypass common econometric issues arising from the overlapping nature of bond excess returns.

In the third chapter, I study macro factors in the risk premia of G10 currencies. Motivated by the finding from a structural model with minimalistic assumptions that the predictability of currency risk premium arises from the differences in the market prices of risks between the home and foreign countries, I tackle this problem by identifying return-forecasting macro factors for the G10 currencies. Based on dynamic factor analysis on a large panel of macro variables, it is found that common macro factors possess strong forecasting power on the risk premia of G10 currencies, especially at longer maturities. The single most important factor loads heavily on activities in the US housing market and bond yields, which exhibits uniform and nonlinear forecasting power across all currencies and at a variety of maturities. The strong in-sample forecasting power preserves out-of-sample.

Essays on Econometrics and Macro-Finance

by

Mao, Zi-Ying

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Steven Heston, Co-Chair
Professor Guido Kuersteiner, Co-Chair
Professor Alberto Rossi
Professor Luminita Stevens
Professor Joshua Linn

Dedication

To my parents.

Acknowledgments

I am still looking for words that express my gratitude to my advisors, Steven Heston and Guido Kuersteiner. Without their blend of guidance, high expectations, patience and encouragement, this dissertation would not have been possible. I am deeply grateful for Gurdip Bakshi for his advice and support in the early stage of my research career. I am also thankful for my comittee members, Joshua Linn, Alberto Rossi, and Luminita Stevens, for their invaluable time and interest.

For the first chapter of this dissertation, I thank Refet Gürkaynak for generously sharing data with me. This dissertation has also been benefited greatly from the constructive comments and suggestions made by Hidehiko Ichimura, Michal Kolesár, Albert "Pete" Kyle, Mikkel Plagborg-Moller, Benoit Perron, Shrihari Santosh, and Neil Shephard.

Contents

List of Tables

List of Figures

Chapter 1: Monetary Shocks and Market Expectations: A Semiparametric Identification Approach

I Introduction

Identifying monetary shocks, defined as unexpected news about future monetary policy ex ante conditional on currently available information, is a fundamental measurement issue underlying all empirical research on monetary policy, as well as on the effects of macroeconomic news. Specifically, a class of monetary shocks resulting from the Federal Reserve's monetary policy actions, which is referred to as monetary policy shocks, has been drawing most of the attention in economic research. Pioneered by Cook and Hahn (1989), Kuttner (2001), Cochrane and Piazzesi (2002), and Gürkaynak, Sack, and Swanson (2005), among others, the predominant approach to identifying monetary policy shocks in asset pricing and monetary economics is to measure them as the changes in interest rate future prices or Treasury yields around a narrow window bracketing each Federal Open Market Committee (FOMC) announcement, where the window length is commonly taken as 30-minute long. Their reasoning is that since the prices of interest rate futures and bonds are forward-looking and are very sensitive to monetary policy (e.g. Angrist, Jordà, and Kuersteiner, 2018), the arrivals of any information about future monetary policy through FOMC announcements should be quickly priced in these securities. By taking the window length narrow enough, it is hoped that any confounding information, such as the FOMC's responses to the financial markets before the announcements are made, and any variations in the future prices or Treasury yields caused by factors unrelated to FOMC announcements can be eliminated. It is argued that this identification strategy is superior to the more conventional approach that measures monetary policy shocks as the structural shocks to the policy rate in a structural VAR (e.g. Christiano, Eichenbaum, and Evans, 1999) in that it is model-free, and is able to capture news about unconventional monetary policy unrelated to the short-term policy rate.

There is, however, no established empirical evidence for the exogeneity of these high-frequency changes around FOMC announcements. While Nakamura and Steinsson (2018) find some evidence that such high-frequency changes in a 30-minute window bracketing each FOMC announcement are less contaminated by noise than in the one-day window, it does not eliminate the possibility of simultaneity. There may be new information even in such a narrow window that moves both the Federal Reserve's decision and the Treasury or futures market so that such high-frequency changes in bond yields or interest rate futures are not merely causal effects of unexpected monetary news. If, in addition, there is some predictability in this information then there should be a significant correlation between high-frequency movements and lagged predictors. I examine this possibility and propose an alternative measure of monetary shocks, which nests monetary policy shocks, based on an alternative identification strategy.

I model the federal funds future (FFF) price at each horizon as a jump-diffusion process governed by five state variables capturing the cross section of the yield curve, yield curve volatility and real-time real business conditions, and take the unexpected daily changes of FFF conditional on the values of the state variables on the previous days as a new measure of monetary shocks at that specific horizon. I achieve this by extending the econometrics literature on kernel-based estimation of continuous-time processes to multivariate jump-diffusion cases. Pioneered by Bandi and Phillips (2003), Bandi and Nguyen (2003), and Bandi and Moloche (2018), this approach uses local constant kernel to fit each infinitesimal moment of the continuous-time process under consideration as a function of some state variables. Estimates of the parameters governing the process are in turn derived as functions of these moments. While the literature has developed the asymptotic theories for scalar diffusion, scalar jump-diffusion and multivariate diffusion processes, there is no existing theory for multivariate jump-diffusion processes specific to this approch. Given that I take five state variables to capture the dynamics of FFFs, such a theory is needed. Moreover, given that my sample spans only a short period of time with about 3,500 data points at daily frequency, naively conducting multivariate kernel regressions of the infinitesimal moments would run into the curse of dimensionality. I thus adapt the single-index specification of Ichimura (1993) to this problem by assuming that each infinitesimal moment is a function of some linear combination of the state variables, which reduces the dimension from five to one, and derive the corresponding asymptotic theory.

I look at the forecastability of the changes in the first six FFFs in a 30-minute window bracketing each FOMC announcement. For each FOMC announcement day, I calculate the drift and the expected value of the jump of each FFF price conditional on the state variables on the day before. If these high-frequency changes were truly shocks, they should not be predictable by the sum of these two quantities. It turns out, however, they are strongly forecastable in both a statistical and an economic sense with adjusted R^2s as high as 13.16%. This is strong evidence that such high-frequency changes are not exogeneous and thus are invalid measure of monetary policy shocks. Moreover, neither the estimated single index nor the state variables exhibit significant forecasting power in the linear regression setting, suggesting that the forecastable components in these high-frequency changes are related to the state variables in some highly nonlinear fashion.

By contrast, the alternative measures proposed in this research are unpredictable, since by construction they have conditional expectations equal to zero. They are highly correlated with the conventional high-frequency measures of monetary policy shocks on FOMC days but also with significant amount of differences in their information contents in that the signal-to-noise ratios of the predictive regressions can be as high as 0.16.

The remainder of this chapter is organized as follows. Section 2 lays out the semi-parametric single-index method for estimating multivariate jump-diffusion processes, and discusses how it is applied to identifying monetary shocks from FFFs. The proofs of the asymptotic properties, however, are left to Appendix A. Section 3 describes the data, and presents the main results along with an array of robustness checks on their sensitivity to the choice of sample period and bandwidth selection. Finally, Section 4 concludes and discusses directions for future research.

II Semiparametric Identification of Monetary Shocks

The n-th FFF contract matures at the end of the last trading day n months ahead. The underlying fundamental is the averaged daily effective federal funds rate over the month in which the contract matures. In particular, the fundamental of the first FFF contract is the averaged daily effective federal funds rate within the current month. The first 36 FFF contracts are available for trading. Denote the n-th FFF price on day t as $f_t^{(n)}$, the effective federal funds rate as r_t, the first (last) trading day of the n-th month as $T_{0t}^{(n)}$ ($T_{1t}^{(n)}$), the number of trading days as $m_t^{(n)} \equiv T_{1t}^{(n)} - T_{0t}^{(n)} + 1$, and the underlying risk-neutral expectation as $E_t^* [\cdot]$. Consider a portfolio of a sequence of consecutive of FFFs, $f_t^{(n_1)}$, ...,

$f_t^{(n_K)}$ with $n_k = n_{k-1} + 1$ and $2 \leq k \leq K$, whose portfolio weight on the n_k-th contract is proportional to the number of trading days in the n_k-th month. Standard no-arbitrage arguments imply that the price of this portfolio, denoted as $f_t^{(n_1, n_K)}$, is

$$f_t^{(n_1, n_K)} = \frac{\sum_{k=1}^{K} m_t^{(n_k)} f_t^{(n_k)}}{\sum_{k=1}^{K} m_t^{(n_k)}} = E_t^* \left[\frac{\sum_{s=T_{0t}^{(n_1)}}^{T_{1t}^{(n_K)}} r_s}{T_{1t}^{(n_K)} - T_{0t}^{(n_1)} + 1} \right] \tag{1.1}$$

which is nothing but the market's risk-adjusted expectation of the average effective federal funds rate from the n_1-th month to the n_K-th month.

Denote some generic FFF contract or portfolio as $f_t = f(X_t)$. I model its time-series dynamics as a function of some underlying state variables. It is widely documented in the literature that the cross-sectional variations of the US Treasury yield curve are almost entirely spanned by its first three principal components, which are referred to as the level (L_t), slope (S_t) and curvature (C_t) factors (e.g. Litterman and Scheinkman, 1991). It is also documented that the volatilities of bond yields are driven by some hidden variable indepdendent of the level, slope and curvature factors. While this variable does not span the cross section of the yield curve, it is priced in fixed income options (e.g. Collin-Dufresne and Goldsten, 2002). I take TYVIX index ($TYVIX_t$), the 30-day implied volatility of the 10-year Treasury future contract published daily by CBOE, as a proxy of such unobserved stochastic volatility factor. Furthermore, to capture other aspects of real business conditions not incorporated in the current yield curve and its derivatives, but that may affect the Federal Reserve's monetary policy decisions, I include the Aruoba-Diebold-Scotti (2009) business condition index (ADS_t) as an additional state variable. I denote the vector of these five state variables collectively as $X_t \equiv (L_t \;\; S_t \;\; C_t \;\; TYVIX_t \;\; ADS_t)'$.

I assume that $f(X_t)$ follows some jump-diffusion process denoted as

$$d\log f(X_t) \quad = \quad \mu(X_t)\,dt + \sigma(X_t)\,dW_t + J_t(X_t)\,dN_t \tag{1.2}$$

where W_t is a Brownian motion, N_t is a Poisson process with jump intensity $\lambda(X_t)$, and the jump size $J_t(X_t)$ is independent of dN_t conditional on the state variables X_t.[1] The unexpected components of $df(X_t)$ and $f(X_t)$ are thus

$$d\bar{Z}_t \equiv df(X_t) - E_t[df(X_t)] = \left(e^{J_t(X_t)} - 1\right) d\bar{N}_t + \sigma(X_t)\,dW_t \tag{1.3}$$

$$\bar{Z}_t \equiv \int_{-\infty}^{t} d\bar{Z}_s = \int_{-\infty}^{t} \left(e^{J_s(X_s)} - 1\right) d\bar{N}_s + \int_{-\infty}^{t} \sigma(X_s)\,dW_s \tag{1.4}$$

respectively, where $d\bar{N}_t \equiv dN_t - \lambda(X_t)\,ds$ is the compensated Poisson process. Assumption 1 in Appendix A implies that the process $\bar{Z}_t$ is null recurrent, and $\bar{Z}_t < \infty$ a.s. for $\forall t$. Hence, in the following I shall abuse the notation slightly by denoting $\bar{Z}_t = \int_1^t \left(e^{J_s(X_s)} - 1\right) d\bar{N}_s + \int_1^t \sigma(X_t)\,dW_t$ instead. $d\bar{Z}_t$ is the new measure of monetary surprises I shall refer to, which is unforecastable by construction since $E[d\bar{Z}_t \mid X_t] = 0$. Correspondingly $\bar{Z}_t$ can be interpreted as the cumulative information inflow.

Suppose the data are observed at evenly-spaced discrete times $1 = t_1 \leq \ldots \leq t_N = T$, where N is the number of observations, and denote the length of the time interval between two consecutive observations as $\Delta \equiv (T-1)/N$. Under some regularity conditions as outlined in Appendix A, it can be shown that the infinitesimal conditional moments of

[1] The functional form (2) ensures that the instantaneous jump in $df(X_t)$, which is $\left(e^{J_t} - 1\right) dN_t$ by Ito's lemma, cannot be smaller than -1 even if J_t has unbounded support. I.e. a negative instantaneous jump can at most drag $f(X_{t+dt})$ to zero. This is an important constraint to impose since $f(X_t)$ cannot be negative.

$f(X_t)$ satisfy (Gikhman and Skorohod, 1972, pp. 68-69; Johannes, 2004)

$$\operatorname*{plim}_{\Delta\downarrow 0} E\left(\frac{\left[\log f_{t+\Delta}(X_t) - \log f_t(X_t)\right]^1}{\Delta}|X_t\right) = \mu(X_t) + \lambda(X_t) E_t\left[J_t(X_t)\right] \tag{1.5}$$

$$\operatorname*{plim}_{\Delta\downarrow 0} E\left(\frac{\left[\log f_{t+\Delta}(X_t) - \log f_t(X_t)\right]^2}{\Delta}|X_t\right) = \sigma^2(X_t) + \lambda(X_t) E_t\left[J_t(X_t)^2\right] \tag{1.6}$$

$$\operatorname*{plim}_{\Delta\downarrow 0} E\left(\frac{\left[\log f_{t+\Delta}(X_t) - \log f_t(X_t)\right]^k}{\Delta}|X_t\right) = \lambda(X_t) E_t\left[J_t(X_t)^k\right] \tag{1.7}$$

for all $k \geq 3$. If some consistent estimates of the infinitesimal conditional moments can be obtained, and the conditional distribution of $J_t(X_t)$ can be pinned down by finitely many of such moments, then it is possible to identify all model parameters of the stochastic process (1.2) from equations (1.5)-(1.7).

To this end, I assume that $J_t(X_t)$ is independent conditional on X_t and follows a normal distribution denoted as[2]

$$J_t(X_t) \sim N\left(\mu_J(X_t), \sigma_J(X_t)^2\right) \tag{1.8}$$

[2]It is possible to make this model more general by assuming that J_t is in the exponential family, which nests a wide range of commonly seen distributions such as the normal, beta, Poisson, gamma, Bernoulli, and Wishart distributions, for which the asymptotic theory in Appendix A can be easily extended. However, as shall be seen in Section 3, the normal assumption already captures the dynamics of FFFs very well.

The model parameters can then be identified by the first, second, third, fourth and sixth moments, namely

$$\operatorname*{plim}_{\Delta \downarrow 0} E\left(\frac{[\log f_{t+\Delta} - \log f_t]^1}{\Delta} \mid X_t \right) = \mu_t + \lambda_t \mu_{Jt} \tag{1.9}$$

$$\operatorname*{plim}_{\Delta \downarrow 0} E\left(\frac{[\log f_{t+\Delta} - \log f_t]^2}{\Delta} \mid X_t \right) = \sigma_t^2 + \lambda_t \left[\mu_{Jt}^2 + \sigma_{Jt}^2 \right] \tag{1.10}$$

$$\operatorname*{plim}_{\Delta \downarrow 0} E\left(\frac{[\log f_{t+\Delta} - \log f_t]^3}{\Delta} \mid X_t \right) = \lambda_t \left[\mu_{Jt}^3 + 3\mu_{Jt}\sigma_{Jt}^2 \right] \tag{1.11}$$

$$\operatorname*{plim}_{\Delta \downarrow 0} E\left(\frac{[\log f_{t+\Delta} - \log f_t]^4}{\Delta} \mid X_t \right) = \lambda_t \left[\mu_{Jt}^4 + 6\mu_{Jt}^2\sigma_{Jt}^2 + 3\sigma_{Jt}^4 \right] \tag{1.12}$$

$$\operatorname*{plim}_{\Delta \downarrow 0} E\left(\frac{[\log f_{t+\Delta} - \log f_t]^5}{\Delta} \mid X_t \right) = \lambda_t \left[\mu_{Jt}^5 + 10\mu_{Jt}^3\sigma_{Jt}^2 + 15\mu_{Jt}\sigma_{Jt}^4 \right] \tag{1.13}$$

$$\operatorname*{plim}_{\Delta \downarrow 0} E\left(\frac{[\log f_{t+\Delta} - \log f_t]^6}{\Delta} \mid X_t \right) = \lambda_t \left[\mu_{Jt}^6 + 15\mu_{Jt}^4\sigma_{Jt}^2 + 45\mu_{Jt}^2\sigma_{Jt}^4 + 15\sigma_{Jt}^6 \right] \tag{1.14}$$

where for notational simplicity I write $\mu_t \equiv \mu(X_t)$, etc. For scalar diffusion and jump-diffusion processes with a single state variable, Bandi and Phillips (2003), and Bandi and Nguyen (2003), respectively, propose estimating the infinitesimal conditional moments using local constant kernels. Bandi and Moloche (2018) subsequently extend this approach to the estimation of multivariate diffusion processes. There is, however, no existing theory on the asymptotic properties of this method for multivariate jump-diffusion processes. Moreover, due to limited data availability in that $TYVIX$ became available only since 2003, even if such asymptotic theory existed, applying the multivariate local contant kernel to this problem would nonetheless run into the curse of dimensionality. For these reasons, I propose a new approach in which the infinitesimal conditional moments are determined by the state variables via some index structure à la Ichimura (1993).

Specifically, I assume the following single index structure for the j-th infinitesimal moments for each $j = 1, 2, \ldots$

$$\begin{aligned} & \operatorname*{plim}_{\Delta\downarrow 0} E_t \left[\frac{\left[\log f_{t+\Delta}\left(X_t\right) - \log f_t\left(X_t\right)\right]^j}{\Delta} \mid X_t \right] \\ = & \operatorname*{plim}_{\Delta\downarrow 0} E_t \left[\frac{\left[\log f_{t+\Delta}\left(X_t\right) - \log f_t\left(X_t\right)\right]^j}{\Delta} \mid X_t\beta_0 \right] \qquad (1.15) \\ \equiv & g_j\left(X_t\beta_0\right) \qquad (1.16) \end{aligned}$$

where the index parameter β_0 is a column vector and is the same for all the six moments. Following Ichimura (1993), I estimate the infinitesimal conditional moment $g_j\left(X_t\beta\right)$ using the local constant kernel estimator

$$\hat{g}_j\left(X_t\beta\right) = \frac{\frac{1}{T}\sum_{t'=1}^{T} \mathbb{I}\left(X_{t'}\beta \in A_{jn}\right) K\left(\frac{X_{t'}\beta - X_t\beta}{h_j}\right)\left(\log f_{t'+\Delta} - \log f_{t'}\right)^j}{\frac{\Delta}{T}\sum_{t'=1}^{T} \mathbb{I}\left(X_{t'}\beta \in A_{jn}\right) K\left(\frac{X_{t'}\beta - X_t\beta}{h_j}\right)} \qquad (1.17)$$

$$A_{jn} = \left\{X_s\beta \,|\, \left| \left(X_s - X_{s'}\right)\beta \right| \le 2h_n,\ \exists X_{s'} \in A\right\} \qquad (1.18)$$

where $K(\cdot)$ is taken as the Gaussian kernel, $h_j = h_j(N, T)$ is the bandwidth parameter that depends on both the time span of the sample T and the number of observations N, $(A, \mathcal{A}) \subset \left(\mathbb{R}^5, \mathcal{B}^5\right)$ is the state space of the stochastic process X_t, and $\mathbb{I}(\cdot)$ denotes the indicator function. The term $\mathbb{I}\left(X_{t'}\beta \in A_{jn}\right)$ is merely used to trim out observations whose values inside the summation sign in the denominator of equation (1.17) are too small, and has no effect on the asymptotic properties of the estimator. Following Ichimura (1993), to avoid numerical problems in which $\sum_{t'=1}^{T} \mathbb{I}\left(X_{t'}\beta \in A_{jn}\right) K\left(\frac{X_{t'}\beta - X_t\beta}{h_j}\right) \simeq 0$, in such situations I set $\hat{g}_j\left(X_t\beta\right)$ equal to either $\max_s\left(\log f_{s+\Delta} - \log f_s\right)^j$ or $\min_s\left(\log f_{s+\Delta} - \log f_s\right)^j$, depending on which one $\left(\log f_{t+\Delta} - \log f_t\right)^j$ is closer to. The true value of the index parameter is estimated by nonlinear least square. I.e.

$$\hat{\beta} \equiv \underset{\beta}{arg\,min} \sum_{j=1}^{6} \frac{1}{T} \sum_{t=1}^{T} \left[(\log f_{t+\Delta} - \log f_t)^j - \hat{g}_j \left(X_t \beta \right) \right]^2 \tag{1.19}$$

The infinitesimal conditional moment estimator $\hat{g}_j \left(X_t \beta \right)$ has many local minima as it is a superposition of single peaked kernel functions. I thus first solve the problem using simulated annealing with 10,000 iterations, which is a global optimization method that converges to the global minimum given arbitrary initial value but at a very slow rate (e.g., Brémaud, 2013, Section 7.8). I then take the solution as the initial value for some local optimization routine to get a more accurate estimate.

It is shown in Appendix A that the model parameters μ_t, σ_t^2, etc. can be identified from the six moment conditions (1.9)-(1.14) as functions of the index $X_t \beta$. Denote $\hat{\mu}_t$, $\hat{\sigma}_t^2$, etc. as the estimators of μ_t, σ_t^2, etc. obtained by inverting equations (1.9)-(1.14) with the kernel estimators of the infinitesimal conditional moments (1.17)-(1.18) on the LHS'. Note that there are two forecastable components of df_t conditional on X_t, namely the instantaneous drift and the expected value of the jump component. By Ito's lemma, the sum of these values is

$$E_t \left[df_t \right] = f_t \left(\mu_t + \frac{1}{2} \sigma_t^2 \right) dt + \left(\exp \left[\mu_{Jt} + \frac{\sigma_{Jt}^2}{2} \right] - 1 \right) \lambda_t dt \tag{1.20}$$

and can be estimated by

$$\hat{E}_t \left[\Delta f_t \right] \equiv f_t \left(\hat{\mu}_t + \frac{1}{2} \hat{\sigma}_t^2 \right) \Delta + \left(\exp \left[\hat{\mu}_{Jt} + \frac{\hat{\sigma}_{Jt}^2}{2} \right] - 1 \right) \hat{\lambda}_t \Delta \tag{1.21}$$

The estimator of monetary surprise (3) is then taken as

$$\Delta\hat{\hat{Z}}_t \equiv \Delta f_t - \hat{E}_t\left[\Delta f_t\right] \tag{1.22}$$

$$= \Delta f_t - \left(f_t\left(\hat{\mu}_t + \frac{1}{2}\hat{\sigma}_t^2\right) + \left(\exp\left[\hat{\mu}_{Jt} + \frac{\hat{\sigma}_{Jt}^2}{2}\right] - 1\right)\hat{\lambda}_t\right)\Delta \tag{1.23}$$

In Appendix A, I derive the asymptotic theory of the index estimator $\hat{\beta}$ and the infinitesimal conditional moment estimators $\hat{g}_j\left(X_t\beta\right)$. The main result is the following:

Theorem. Under Assumptions 1-5 in Appendix A, and conditions on the limiting behaviors of the bandwidth h_j in Theorem 3 in Appendix A, $\hat{\beta} \xrightarrow{p} \beta_0$ and $\hat{g}_j\left(X_t\hat{\beta}\right) \xrightarrow{p} g_j\left(X_t\beta_0\right)$ as $T \to \infty$, $N \to \infty$ and $\Delta \equiv \left(T-1\right)/N \to 0$, for $j = 1, \ldots, 6$.

Given the consistency of the index and the infinitesimal moment estimators, the model parameters on the RHS' of equations (1.9)-(1.14), and so the estimates of $\hat{E}_t\left[\Delta f_t\right]$ and $\Delta\hat{\hat{Z}}_t$, are solved consistently. In Appendix B, I outline a modification of the randomized procedure devised by Bandi, Corradi, and Moloche (2009) to select the appropriate bandwidths that ensure consistency. Since $\hat{E}_t\left[\Delta f_t\right]$ and $\Delta\hat{\hat{Z}}_t$, and the yield curve factors, L_t, S_t and C_t, are all generated regressors, which shall introduce additional uncertainty in subsequent regression analyses, I conduct all statistical inferences under a bootstrap procedure detailed in Appendix C.

III Data and Results

I use the daily Treasury yield dataset assembled by Gürkaynak, Sack, and Wright (2007), which is available on the Federal Reserve Bank of New York website. They fit the yield curve using outstanding US Treasury notes and bonds that are to mature in at least three months. The dataset includes yields with maturities ranging from 1 year to 30 years with 1-year increment, as well as fitted parameters for the Svensson (1994) method of yield curve interpolation. Using these parameters, I additionally calculate yields with maturities ranging from 3 months to 3 years with 3-month increment. I estimate the level (L_t), slope (S_t) and curvature (C_t) factors by principal component analysis on this extended panel of bond yields. The daily TYVIX index is retrieved from the CBOE website, which is available since the beginning of 2003. The data on FFF prices are retrieved from Bloomberg. Finally, the ADS index is retrieved from the Federal Reserve Bank of Philadelphia website. I thus estimate the semiparametric jump-diffusion models of the first six FFF prices using daily data starting from the beginning of 2003. The entire dataset ends on May 18, 2018. The high-frequency monetary shocks dataset, which Refet Gürkaynak graciously shared with me, includes changes in the first six FFF prices within a 30-minute window bracketing each FOMC announcement from February 1994 to October 2015.

Table 1.1 lists the AR(1) coefficient and the p-value of Phillips-Perron (1988) unit root test for the first difference of each log-FFF and for each state variable. While the first differences of log-FFFs are stationary, the state variables are highly persistent and Phillips-Perron tests fail to reject the unit root hypotheses for the level, slope and ADS state variables. Nonetheless, the semiparametric procedure as outlined in Section 2 does not yield spurious correlation between the independent and dependent variables commonly encountered in linear regression settings. Intuitively, this is because the state variables X_t enter the kernel estimator (1.17) as a first difference, $X_{t'}\beta - X_t\beta$, making the RHS stationary regardless of whether X_t is cointegrated. In fact, Assumption 2 in Appendix A, the only assumption on X_t to ensure the consistency of the index parameter estimator $\hat{\beta}$ and the moment estimators $\hat{g}_j(\cdot)$, allows X_t to be nonstationary.

Figure 1.1 plots the daily changes Δf_t along with the unexpected component $\Delta\hat{\hat{Z}}_t$. As is seen, most of the daily variations in FFFs are driven by the residuals conditional on all information available on the previous days, suggesting that FFFs are a potentially rich source of monetary news. It is also noted that FFFs of higher horizons are more volatile and are governed more by the arrivals of unexpected information, which is intuitive as the level of uncertainty increases with the underlying investment horizon. Across all the first six horizons, FFFs are much more tranquil in the zero-lower-bound (ZLB) period, which is expected as that period was dominated by unconvetional monetary policy influencing market expectations of the long term well beyond the first six months. To guard against such a potential structural break, I shall repeat the same empirical analyses in the following using only the subsample that spans the pre-ZLB period.

To assess model fits, I regress each infinitesimal moment $[\log f_{t+\Delta} - \log f_t]^j / \Delta$ on its fitted values $\hat{g}_{jt}$, and report the in-sample R^2 in Table 1.2. The model fits the data reasonably well in that the R^2s for the third to the sixth moments, which are critical to the estimation of model parameters governing the jump components, λ_t, μ_{Jt} and σ^2_{Jt}, range from 12.33% to 53.58% across all horizons except for the fifth moment at the 1-month horizon. The R^2s for the second moment range from 20.89% to 44.15%, and those for the first moment range from 9.98% to 19.76%. To assess the robustness of the main results in the following to the choice of bandwidths, I shall repeat the same empirical exercises using bandwidths that are five times of the optimal ones chosen by the procedure in Appendix B, as well as taking the bandwidth for each horizon as the median value of the optimal bandwidths across all six horizons.

III.I Main Results

As is discussed, as long as X_t includes all state variables underlying the FFF market, the identified unexpected daily change $\Delta\hat{\hat{Z}}_t^{(n)}$ is unforecastable by construction. On the other hand, the very claim made by the literature that high-frequency changes of FFFs in a 30-minute window bracketing each FOMC announcement do not contain endogeneous information has not been statistically justified. If these high-frequency changes are truly exogeneous, they should not be predictable by any information available before the FOMC announcements are made, including information contained in the state variables X_t. If not, then the confounding information in these high-frequency changes contributes to the discrepency between the information contents of the conventional high-frequency measures of monetary policy shocks and those of the new measure $\Delta\hat{\hat{Z}}_t^{(n)}$.

For each of the first six FFFs, the expected daily change $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$ as in equation (1.21) is a natural candidate for predicting the high-frequency change around FOMC announcement at the same horizon n on day t, $\Delta FOMC_t^{(n)}$. For each horizon n, I run predictive regression of $\Delta FOMC_t^{(n)}$ on $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$, and report the results in Panel A of Table 1.3. The slope coefficients are all positive and are significant at the 5% level or above at all but the 2-month and the 3-month horizons. This is strong evidence that the so-called monetary policy shocks identified as high-frequency changes in FFFs around FOMC announcements are predictable, and thus are invalid measure of monetary policy shocks. Moreover, the signal-to-noise ratios, defined as the variance of the fitted values of the regression divided by the variance of the residuals, range from 0.06 to 0.16 across all but the 2-month and the 3-month horizons, suggesting that while there is significant amount of overlap between the information contents of such high-frequency changes and the new measure of monetary shocks (1.23), there is also vast difference between them.

For comparison, I run predictive regressions of high-frequency changes in FFFs around FOMC announcements, $\Delta FOMC_t^{(n)}$, on the values of the state variables one-day before, X_{t-1}. The results are summarized in Panel B of Table 1.3. In contrast to $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$, which is nothing but some highly nonlinear function of X_{t-1}, X_{t-1} itself does not exhibit any joint significance in the linear setting as is evident from the Wald tests. I also run predictive regressions of such high-frequency changes in FFFs on the forecastable components $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$ conditional on the state variables X_{t-1}. As is seen in Panel C, $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$ remain significant at the 5% level or above at all but the 2-month and the 3-month horizons. While one explanation of these phenomena is that the forecastable component of high-frequency changes at, say, the n-month horizon $\Delta FOMC_t^{(n)}$ is related to the state variables X_{t-1} in some highly nonlinear way not captured by the linear combination of X_{t-1} found via ordinary least square, another potential explanation is that $\Delta FOMC_t^{(n)}$ is indeed proportional to some linear combination of X_{t-1} that is better captured in finite sample by the single index $X_{t-1}\hat{\beta}$ estimated from the problem (1.19). To rule out this second possibility, I run predictive regressions of $\Delta FOMC_t^{(n)}$ on the estimated single index $X_{t-1}\hat{\beta}$, whose results are summarized in Panel D. The slope coefficients are more than three orders of magnitude smaller than those in Panel A, and are insignificant at the 10% level across all horizons. Thus, the semiparametric model developed in this research is indeed necessary to capture the highly nonlinear forecastable components in $\Delta FOMC_t^{(n)}$. Such high degree of nonlinearity is perhaps also why the forecastability of the so called high-frequency monetary policy shocks has not been discovered for such a long time in the literature.

Finally, as a robustness check that the new measure of monetary shocks proposed in this research is indeed unforecastable, I run predictive regression of the unexpected daily changes in FFF as identified by the semiparametric jump-diffusion model, $\Delta\hat{\bar{Z}}_t^{(n)}$, on the set of state variables X_t for each horizon $n = 1, \ldots, 6$. Panel E of Table 1.3 summarizes the results for the version in which the regressions are run over both FOMC and non-FOMC days, and Panel F summarizes the results over FOMC announcement days only. In both samples, the F-statistics for the joint significance of the state variables are insignificant at the 10% level across all horizons. While the new measure of monetary shocks, $d\bar{Z}_t^{(n)} \equiv df^{(n)}(X_t) - E_t\left[df^{(n)}(X_t)\right]$, is unpredictable by definition, the resulting estimates may nonetheless be forecastable if the semiparametric jump-diffusion model of $f^{(n)}(X_t)$ as described in Section 2 is misspecified. A special situation of such violation is that $d\bar{Z}_t^{(n)}$ is not orthogonal to the $\mathcal{L}^2$ space spanned by the state vector X_t. As is verified here, this is not the case.

III.II Robustness Checks

As is seen in Figure 1.1, FFFs exhibit little variability in the ZLB period, raising the concern that the semiparametric model may behave very differently during this period. To ensure robustness of the results reported in Section 3.1 to such pontential structural break, I estimate the semiparametric model and then repeat the same regression analyses using only the pre-ZLB sample period, where the bandwidths are taken the same as the optimal values selected for the whole sample. The results are summarized in Table 1.4. The forecastable component of daily changes at horizon n, $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$, exhibits significant forecasting power on high-frequency changes in FFF around FOMC announcements, $\Delta FOMC_t^{(n)}$, at the 1% level at the 2-month, 3-month, 4-month and 6-month horizons, and at the 5% level at the 5-month horizon. Such forecasting power is also economically significant in that the adjusted R^2s range from 11.40% to 47.55% across all but the 1-month horizon. The high signal-to-noise ratios, which range from 0.15 to 0.94 across all but the 1-month horizon, indicate that the forecastable component in $\Delta FOMC_t^{(n)}$ still counts for large amount of discrepency between the information contents of $\Delta FOMC_t^{(n)}$ and those of $\Delta\hat{\hat{Z}}_t^{(n)}$. The same pattern remains when the state variables X_{t-1} are included as control variables, and neither X_{t-1} nor the estimated single index $X_{t-1}\hat{\beta}$ exhibit significant forecasting power at the 10% level. Finally, the new measure of monetary shocks, $\Delta\hat{\hat{Z}}_t^{(n)}$,remain unforecastable by the state variables one period both through the entire sample period and on FOMC announcement days only. These are all consistent with the main results in Section 3.1.

To guard against potential overfitting of the semiparametric models, and to investigate the sensitivity of the main results to the choice of bandwidth, I repeat the same exercises using bandwidths five times of the optimal ones as selected by the procedure outlined in Appendix B. The results are summarized in Table 1.5. The high-frequency changes around FOMC announcements, $\Delta FOMC_t^{(n)}$, remain significantly forecastable by the expected daily change, $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$, at the 5% level or above across all horizons both with and without the state variables X_{t-1} being included as control variables. On the other hand, neither the state variables X_{t-1} nor the index $X_{t-1}\hat{\beta}$ exibit significant forecasting power at the 5% level. Finally, the new measure of monetary shocks $\Delta\hat{\hat{Z}}_t^{(n)}$ remains unforecastable by the state variables X_{t-1}.

Additionally, I repeat the same exercises using the median bandwidth across all horizons, whose results are summarized in Table 1.6. $\Delta FOMC_t^{(n)}$ is still strongly forecastable by the expected daily change $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$ at the 1% significance level for $n = 1, 4, 5, 6$;

and if controlled for the state variables X_{t-1}, it is forecastable by $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$ at the 5% significance level or above for $n = 1, 2, 3, 4, 5$. The estimated index $X_{t-1}\hat{\beta}$ exhibits no forecasting power at the 5% significance level across all horizons. Finally, the state variables still exhibit no forecasting power on the new measure of monetary shocks $\Delta\hat{\hat{Z}}_t^{(n)}$ both in the entire sample and only on FOMC announcement days.

As is discussed earlier, the semiparametric model developed in this research does not generate spurious correlation when the independent and/or the dependent variables are nonstationary. In fact, if some of the independent variables are cointegrated, and deviations from such equilibrum relationship are correlated with the (stationary) dependent variable, the conventional prescription for spurious regressions by taking first differences of the independent variables is inappropriate as it will mistakenly wipe out this correlation. To demonstrate this, I reestimate the semiparametric jump-diffusion model taking the first differences of the state variables used previously, ΔX_t, as the new set of state variables, and then repeat the same empirical analyses of the information contents of the conventional high-frequency measure of monetary shocks as well as the new measure (1.23). The results are summarized in Table 1.7. At the 10% significance level, $\hat{E}_{t-1}\left[\Delta f_{t-1}^{(n)}\right]$ no longer possesses any forecasting power on $\Delta FOMC_t^{(n)}$ with and without conditioning on ΔX_t across all six horizons. Finally, ΔX_t do not exhibit significant forecasting power on the new measure of monetary shocks $\Delta\hat{\hat{Z}}_t^{(n)}$ at the 10% level.

IV Conclusions

I demonstrate that the predominant measure of monetary policy shocks in the academic literature for the past twenty years, which are changes in FFF prices in a narrow window bracketing each FOMC announcement, are in fact forecastable and thus invalid.

In contrast, the alternative measures of monetary shocks proposed in this research are unforecastable by construction. To estimate the shocks, I extend the method of kernel-based estimation of continuous-time processes to multivariate jump-diffusion processes with single-index structure, and derive the asymptotic theory.

A follow-up study may reestimate the model using intraday data, which shall be able to disentangle shocks triggered by different events on the same day. Moreover, given that some previous research on the effects of monetary policy shocks looks at high-frequency price changes in the Treasury market instead of in the futures market around FOMC announcements (e.g. Hanson and Stein, 2015), a follow-up study may repeat the same exercises conducted in this research on high-frequency Treasury data. Given that the main focus of this research is on the validity of the conventional high-frequency identification strategy of monetary shocks, and that in this research I do not have access to high-frequency data on Treasury yields, I do not investigate this question.

Given that future rates and forward rates are highly correlated, it is very likely that the unforecastable component in forward rate covering the period, say, $(T_0,\, T_1)$

$$
\begin{aligned}
F_t^{(T_0,\,T_1)} &\equiv P_t^{(T_0)}/P_t^{(T_1)} && (1.24)\\
&= E_t^*\left[\exp\left(-\int_{T_0}^{T_1} r_s ds\right)\right] && (1.25)
\end{aligned}
$$

capture monetary surprises specific to the horizon $(T_0,\, T_1)$. Since yield curve data are available at a much wider range of maturities than FFFs and eurodollar futures, this can be very useful to research on the effects of shocks indicative of monetary policy in the long term.

Finally, FFF prices are not really the physical expectations of future policy rates. Piazzesi and Swanson (2008) illustrate that the risk premium term, i.e. the difference between the physical expectation and FFF price, can be measured by macroeconomic and financial variables, and that future prices corrected for such risk adjustments have stronger forecasting power on monetary policy rates. Subsequent research may look into how to incorporate such risk adjustments in the semiparametric model developed in this research.

Incorporating such risk adjustments may also shed light on the source of the forecastable components in high-frequency changes of FFFs around FOMC announcements. One possibility is that such a high-frequency change is partly a change in the risk premium and partly a change in the physical expectation of future monetary policy rates. While the arrival of monetary news that triggers the change in expectation can be exogeneous, the change in risk aversion may well be predictable. In fact, a recent literature on FOMC cycle has found that large portions of the excess returns on US equities (e.g. Lucca and Morench, 2015; Cieslak, Morse and Vissing-Jorgensen, 2019) and currencies (Mueller, Tahbaz-Salehi, and Vedolin, 2017) are earned within periods before the scheduled FOMC announcements. Similar phenomena may also be present in the FFF market.

Appendix A: Asymptotic Theory of Continuous-Time Single-Index Estimators

Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space endowed with some filtration $\{\mathcal{F}_t\}_{t\geq 0}$, where P is the physical probability measure. The vector of state variables $\{X_t\}_{t\geq 1}$ is a family of measurable mappings from the probability space to its state space $(A, \mathcal{B}(A))$, where $A \subseteq \mathbb{R}^5$ and $\mathcal{B}(A)$ denotes the Borel σ-algebra of A, and is adapted to $\{\mathcal{F}_t\}_{t\geq 0}$. I label the start of time as $t = 0$ and the time of the first observation as $t = 1$.

I start by imposing the following standard condition on $\log f(X_t\beta)$ defined in equation (1.2), which guarantees the existence and uniqueness of a càdlàg strong solution (e.g. Gikhman and Skorohod, 1972, Chapter 2; Bandi and Nguyen, 2003):

Assumption 1. The process $\log f(X_t\beta)$ whose time-series dynamics is given by equation (1.2) satisfies:

1. $\mu(\cdot)$, $\sigma(\cdot)$ and $\lambda(\cdot)$ are twice continuously differentiable, and satisfy the local Lipschitz condition. I.e. for evey compact subset S of the range D of the process, there exists a constant C_1 such that

$$|\mu(x) - \mu(z)| + |\sigma(x) - \sigma(z)| + \lambda(x) E[J(x) - J(z)] \leq C_1 |x - z| \tag{1.26}$$

for $\forall x, z \in S$. They also satisfy the growth condition that there exists a constant C_2 such that

$$|\mu(x)| + |\sigma(x)| + \lambda(x) E|J| \leq C_2 (1 + |x|), \quad \forall x \in S \tag{1.27}$$

2. For any $\alpha > 2$, there exists a constant C_3 such that

$$\lambda(x) \int_{\mathbb{R}} |y|^{\alpha} N_{\mu,\sigma^2}(dy) \leq C_3 \{1 + |x|^{\alpha}\}, \quad \forall x \in D \tag{1.28}$$

where $N_{\mu,\sigma^2}(\cdot)$ is the distribution function of $N(\mu, \sigma^2)$.

3. $\lambda(\cdot) \geq 0$ and $\sigma^2(\cdot) > 0$ on D.

Under Assumption 1, it can be shown using Ito's lemma and the dominated convergence theorem that the infinitesimal conditional moments on the LHS' of equations (1.9)-(1.14) converges to the RHS (Gikhman and Skorohod, 1972, pp. 68-69; Johannes, 2004).

I impose the same condition on the state variables X_t.

Assumption 2. The vector of state variables X_t is a null recurrent multivariate jump-diffusion process with each element satisfying the same set of conditions as in Assumption 1. This implies that there exists a σ-finite measure $\Phi(\cdot)$ on $(A, \mathcal{B}(A))$ such that

$$\Phi(D) = \int_A P(X_t \in D) \Phi(dx), \quad \forall D \in \mathcal{B}(A) \tag{1.29}$$

Moreover, $\sum_{0<s\leq t} ||X_s - X_{s-}||_2 < \infty$ a.s. for $\forall t > 0$.

$\Phi(\cdot)$ is referred to as the invariant measure of X_t. For Lévy processes, $\Phi(\cdot)$ is absolutely continuous with resepect to Lebesgue measure, and I denote $\phi(x) \equiv \Phi(dx)/dx$. The existence of an invariant measure $\Phi(\cdot)$ leads to the following result:

Lemma 1 (Quotient Limit Theorem). Given Assumption 2, for any Borel measurable functions $f(\cdot)$ and $g(\cdot)$ that are integrable with respect to $\Phi(dx)$

$$\frac{\int_1^T f(X_s)\, ds}{\int_1^T g(X_s)\, ds} \overset{a.s.}{\to} \frac{\int_A f(x)\, \Phi(dx)}{\int_A g(x)\, \Phi(dx)} \tag{1.30}$$

as $T \to \infty$, provided $\int_A g(x)\, \Phi(dx) > 0$.

Proof. See Revuz and Yor (2013, Theorem 3.12, Chapter 10). #

Let $[X\beta]_t^c$ denote the quadratic variation of the continuous component of the index $X_t\beta$. A useful tool for the consistency proof below is the notion of local time

$$L_{X\beta}(t, a) \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_1^t \mathbb{I}(X_s\beta \in [a, a+\epsilon))\, d[X\beta]_s^c \tag{1.31}$$

which, intuitively speaking, measures the amout of time the index $X_t\beta$ spends around a in information unit. The scaled local time is defined as

$$\bar{L}_{X\beta}(t, a) \equiv \frac{L_{X\beta}(t, a)}{\sigma^2_{X\beta}(a)} \tag{1.32}$$

where $\sigma^2_{X\beta}(\cdot)$ denotes the diffusion term of $X\beta$. The local time has the following properties:

Lemma 2. Given Assumption 2:

1. $L_{X\beta}(t, a)$ is continuous in t, and a.s. càdlàg in a.

2. Let $g(\cdot)$ be a bounded Borel measurable function. Then

$$\int_1^t g\left(X_{s-}\beta\right) d\left[X\beta\right]_s^c \overset{a.s.}{=} \int_A L_{X\beta}\left(t, a\right) g\left(a\right) da \tag{1.33}$$

Proof. See Protter (1995, Theorem 56 and Corollary 1, Chapter 4). #

I make the following standard assumption on the kernel function (e.g. Bandi and Nguyen, 2003; Ichimura, 1993):

Assumption 3. The kernel function $K(\cdot) : \mathbb{R} \to \mathbb{R}^{+}$ is twice continuously differentiable, bounded and symmetric around 0 such that

$$\int_{\mathbb{R}} K(s)\, ds = 1 \tag{1.34}$$

$$\int_{\mathbb{R}} \left[K(s)\right]^2 ds < \infty \tag{1.35}$$

$$\int_{\mathbb{R}} s^2 K(s)\, ds < \infty \tag{1.36}$$

$$\int_{\mathbb{R}} \left|\frac{\partial K(s)}{\partial s}\right| ds < \infty \tag{1.37}$$

and its second derivative satisfies a Lipschitz condition.

In the following, I first show that the functional estimators of the infnitesimal moments are consistent as long as $\hat{\beta} = \beta_0 + o_p(1)$. I then show that $\hat{\beta}$ is indeed consistent. They together imply that the functional estimators of the model parameters on the RHS' of equations (1.9)-(1.14) are consistent. The following preliminary results are useful for proving the first step, Theorem 1:

Lemma 3.

1. Let C be a time change that is a.s. finite and Y be a continuous $\mathcal{F}_t$-local martingale. If Y is C-continuous, i.e. Y is continuous in $[C_{t-}, C_t]$ for any t, then Y_C is a continuous $\mathcal{F}_{C_t}$-local martingale and $[Y_C]_t = [Y]_{C_t}$, where $[Y]_t$ denotes the quadratic variation of Y.

2. (Dambis-Dubins-Schwarz). Let M^n be a sequence of continuous local martingales such that a.s. $M_0^n = 0$ and $[M^n]_\infty = \infty$. Define time change $T_t^n \equiv \inf\{s \mid [M^n]_s > t\}$, then $B_t^n \equiv M_{T_t}^n$ is a Brownian motion and $M_t^n = B_{[M^n]_t}^n$. Moreover, B^n converges in distribution to a Brownian motion.

Proof. See Revuz and Yor (2013, Proposition 1.5, Chapter 5) for the first part, and Revuz and Yor (2013, Theorem 1.6, Chapter 5, and Theorem 2.3, Chapter 8) for the second part. #

Lemma 4 (Chow's Law of Large Number for MDS). Let $\{Y_i\}_{i\geq 1}$ be a martingale difference sequence (MDS). If $\underset{i}{sup}E\left(Y_i^2\right) \leq K < \infty$ for some constant K, then

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{p} 0 \tag{1.38}$$

Proof. See Stout (1974, Theorem 3.3.8). #

Lemma 5 (Central Limit Theorem for MDS). Let $\{Y_{ni}\}_{i=1}^{n}$ be a MDS, and $\mathcal{F}_{ni}$ be the σ-algebra generated by Y_{ni}. Suppose that: (1) $\max_i \mid Y_{ni} \mid \xrightarrow{p} 0$; (2) $\sum_i Y_{ni}^2 \xrightarrow{p} \eta^2$, where η^2 is some positive constant; and (3) $E\left[\max_i Y_{ni}^2\right]$ is bounded in n. Then

$$\sum_i Y_{ni} \xrightarrow{d} N\left(0, \eta^2\right) \tag{1.39}$$

Proof. See Eagleson (1975), and Hall and Heyde (1980, Theorem 3.2, Chapter 3). #

Bandi and Nguyen (2003) propose the following local constant kernel estimator for scaled local time, which I state without proof:

Proposition 1. Let Y be a scalar jump-diffusion process satisfying the same set of conditions as in Assumption 1, which is observed at evenly-spaced discrete times $1 = t_1 \leq \ldots \leq t_N = T$. Let $K(\cdot)$ be some kernel function satisfying Assumption 3, and the bandwidth h be such that $(1/h)(\Delta \log(1/\Delta))^{1/2} = o_p(1)$, where $\Delta \equiv (T-1)/N$. Then

$$\hat{\bar{L}}_Y(T, a) \equiv \frac{\Delta}{h}\sum_j K\left(\frac{Y_j - a}{h}\right) \xrightarrow{a.s.} \bar{L}_Y(T, a) \tag{1.40}$$

Proof: See Bandi and Nguyen (2003, Theorem 1). #

Theorem 1. Suppose Assumptions 1-3 hold. As $N \to \infty$, $T \to \infty$ and $\Delta \equiv T/N \to 0$, if the bandwidth h_j is such that $\left(L_{X\hat{\beta}_j}\left(T, X\hat{\beta}\right)/h_j\right)(\Delta \log(1/\Delta))^{1/2}$, $\left(L_{\log f}(T, \log f)/h_j\right)(\Delta \log(1/\Delta))^{1/2}$

and $1/\left(h_1 \bar{L}_{X\hat{\beta}}\left(T,\, X\hat{\beta}\right)\right)^{-1/2}$ converge in probability to 0, then

$$\hat{g}_j\left(X_t\hat{\beta}\right) \xrightarrow{p} g_j\left(X_t\beta_0\right) \tag{1.41}$$

for $j = 1, \ldots, 6$, provided that $\hat{\beta} = \beta_0 + o_p(1)$.

Proof. Denote $a \equiv X_t\beta_0$ and $\hat{a} \equiv X_t\hat{\beta}$. Write

$$\begin{aligned}
\hat{g}_1(\hat{a}) &= \frac{\frac{1}{h_1}\sum_{t'=1,\,t'\neq t}^{T} \mathbb{I}\left(X_{t'} \in A_n\right) K\left(\frac{X_{t'}\hat{\beta}-\hat{a}}{h_1}\right)\left(\log f_{t'+\Delta} - \log f_{t'}\right)}{\frac{\Delta}{h_1}\sum_{t''=1}^{T} \mathbb{I}\left(X_{t''}\hat{\beta} \in A_n\right) K\left(\frac{X_{t''}\hat{\beta}-\hat{a}_1}{h_1}\right)} &(1.42)\\
&= \sum_{t'=1,\,t'\neq t}^{T} \frac{\frac{1}{h_1}\mathbb{I}\left(X_{t'}\hat{\beta} \in A_n\right) K\left(\frac{X_{t'}\hat{\beta}-\hat{a}}{h_1}\right)\int_{t'}^{t'+\Delta}\left[\mu\left(X_s\right) + \mu_J\left(X_s\right)\lambda\left(X_s\right)\right] ds}{\frac{\Delta}{h_1}\sum_{t''=1}^{T} \mathbb{I}\left(X_{t''}\hat{\beta} \in A_n\right) K\left(\frac{X_{t''}\hat{\beta}-\hat{a}_1}{h_1}\right)}\\
&\quad + \sum_{t'=1,\,t'\neq t}^{T} \frac{\frac{1}{h_1}\mathbb{I}\left(X_{t'}\hat{\beta} \in A_n\right) K\left(\frac{X_{t'}\hat{\beta}-\hat{a}}{h_1}\right)\int_{t'}^{t'+\Delta}\sigma\left(X_s\right) dW_s}{\frac{\Delta}{h_1}\sum_{t''=1}^{T} \mathbb{I}\left(X_{t''\hat{\beta}} \in A_n\right) K\left(\frac{X_{t''}\hat{\beta}-\hat{a}}{h_1}\right)}\\
&\quad + \sum_{t'=1,\,t'\neq t}^{T} \frac{\frac{1}{h_1}\mathbb{I}\left(X_{t'}\hat{\beta} \in A_n\right) K\left(\frac{X_{t'}\hat{\beta}-\hat{a}_1}{h_1}\right)\int_{t'}^{t'+\Delta} d\bar{Z}_s}{\frac{\Delta}{h_1}\sum_{t''=1}^{T} \mathbb{I}\left(X_{t''}\hat{\beta} \in A_n\right) K\left(\frac{X_{t''}\hat{\beta}-\hat{a}}{h_1}\right)} &(1.43)\\
&\equiv \sum_{t'=1,\,t'\neq t}^{T} k_{11,\,t'+\Delta}(\hat{a}) + \sum_{t'=1,\,t'\neq t}^{T} k_{12,\,t'+\Delta}(\hat{a}) + \sum_{t'=1,\,t'\neq t}^{T} k_{13,\,t'+\Delta}(\hat{a}) &(1.44)\\
&\equiv K_{11}(\hat{a}) + K_{12}(\hat{a}) + K_{13}(\hat{a}) &(1.45)
\end{aligned}$$

where $\bar{Z}_t \equiv \int_0^t J_s\left(X_s\right) dN_s - \int_0^t \mu_J\left(X_s\right)\lambda\left(X_s\right) ds$ denotes the compensated Poisson jump process. I follow similar steps as in Bandi and Nguyen (2003, Theorem 2) to show convergence of $K_{11}(\hat{a})$. Denote $K^*\left(X_s\hat{\beta}\right) \equiv \mathbb{I}\left(X_s\hat{\beta} \in A_n\right) K\left(\frac{X_s\hat{\beta}-\hat{a}}{h_1}\right)$ and $\mu^*\left(X_s\right) \equiv \mu\left(X_s\right) +$

$\mu_J(X_s)\,\lambda(X_s)$. Note that by the mean value theorem

$$
\begin{aligned}
&|\sum_{t'=1,\,t'\neq t}^{T} K^*\left(X_{t'}\hat{\beta}\right)\int_{t'}^{t'+\Delta}\mu^*(X_s)\,ds-\int_1^T K^*\left(X_s\hat{\beta}\right)\mu^*(X_s)\,ds| \\
=&|\sum_{t'=1}^{T}\int_{t'}^{t'+\Delta}\left[K^*\left(X_{t'}\hat{\beta}\right)-K^*\left(X_s\hat{\beta}\right)\right]\mu^*(X_s)\,ds| \qquad (1.46)\\
\leq&\sum_{t'=1}^{T}\int_{t'}^{t'+\Delta}|\frac{dK^*}{dx}\left(\tilde{X}_s\hat{\beta}\right)\mu^*(X_s)||X_s\hat{\beta}-X_{t'}\hat{\beta}|\,ds \qquad (1.47)\\
\leq&\left(\sum_{t'=1}^{T}\int_{t'}^{t'+\Delta}|\frac{dK^*}{dx}\left(\tilde{X}_s\hat{\beta}\right)\mu^*(X_s)|\,ds\right)\\
&\qquad\times\left(\max_{t'}\sup_{t'\leq s\leq t'+\Delta}|X_s-X_{t'}|\right)\|\hat{\beta}\|_1 \qquad (1.48)
\end{aligned}
$$

for some $\tilde{X}_s\hat{\beta}$ between $X_{t'}\hat{\beta}$ and $X_s\hat{\beta}$. Note that the first term on the RHS satisfies

$$
\begin{aligned}
&\sum_{t'=1}^{T}\int_{t'}^{t'+\Delta}|\frac{dK^*}{dx}\left(\tilde{X}_s\hat{\beta}\right)\mu^*(X_s)|\,ds\\
\leq&\int_1^T|\frac{dK}{dx}\left(\frac{X_s\hat{\beta}-\hat{a}}{h_1}+o_p(1)\right)g_1\left(X_s\hat{\beta}\right)|\,ds \qquad (1.49)\\
=&\int_{-\infty}^{\infty}|\frac{dK}{dx}\left(q+o_p(1)\right)g_1(q)|\,\bar{L}_{X\hat{\beta}_1}(T,\,qh_1+\hat{a})\,dq \qquad (1.50)\\
\leq&C_5O_{a.s.}\left(\bar{L}_{X\hat{\beta}_1}(T,\,\hat{a})\right) \qquad (1.51)
\end{aligned}
$$

for some constant C_5, where the second step follows from Lemma 2. Moreover, Lévy's modulus of continuity implies that the second term

$$
\left(\max_{t'}\sup_{t'\leq s\leq t'+\Delta}|X_{is}-X_{it'}|\right)=O_{a.s.}\left(\sqrt{\Delta\log(1/\Delta)}\right) \qquad (1.52)
$$

for each element i of X_t. Combining equations (1.48), (1.51) and (1.52) yields

$$\begin{aligned}
&\frac{1}{h_1}\sum_{t'=1,\, t'\neq t}^{T} K^*\left(X_{t'}\hat{\beta}\right)\int_{t'}^{t'+\Delta}\mu^*\left(X_s\right)ds \\
=&\frac{1}{h_1}\int_1^T K^*\left(X_s\hat{\beta}\right)\mu^*\left(X_s\right)ds + O_p\left(\bar{L}_{X\hat{\beta}}\left(T,\,\hat{a}\right)\sqrt{\Delta\log\left(1/\Delta\right)}/h\right)\parallel\hat{\beta}\parallel_1
\end{aligned} \tag{1.53}$$

Similarly, the denominator of $K_{11}\left(\hat{a}\right)$

$$\begin{aligned}
&\frac{\Delta}{h_1}\sum_{t'=1,\, t'\neq t}^{T} K^*\left(X_{t'}\hat{\beta}\right) \\
=&\frac{1}{h_1}\int_1^T K^*\left(X_s\hat{\beta}\right)ds + O_p\left(\bar{L}_{X\hat{\beta}}\left(T,\,\hat{a}\right)\sqrt{\Delta\log\left(1/\Delta\right)}/h\right)\parallel\hat{\beta}\parallel_1
\end{aligned} \tag{1.54}$$

Thus

$$K_{11}\left(\hat{a}\right) \overset{p}{\to} \frac{\int_1^\infty K^*\left(X_s\beta_0\right)g_1\left(X_s\beta_0\right)ds}{\int_1^\infty K^*\left(X_s\beta_0\right)ds} \tag{1.55}$$

$$\overset{a.s.}{=} \frac{\int_A K\left(q\right)g_1\left(qh_1+a\right)\phi\left(qh_1+a\right)dq}{\int_A K\left(q\right)\phi\left(qh_1+a\right)dq} \tag{1.56}$$

$$= g_1\left(a\right) \tag{1.57}$$

I use Lemma 3 to derive the asymptotic property of $K_{12}\left(\hat{a}\right)$. Denote the nominator of K_{12} as $\tilde{k}_{12,\,T} \equiv \frac{1}{h_1}\sum_{t'=1}^{T}\mathbb{I}\left(X_{t'}\hat{\beta}\in A_n\right)K\left(\frac{X_{t'}\hat{\beta}-\hat{a}}{h_1}\right)\int_{t'}^{t'+\Delta}\sigma\left(X_s\right)dW_s$ and define time change $C_T \equiv \inf\left\{s \mid \left[\tilde{k}_{12}\right]_s > T\right\}$. Since C_T is continuous and a.s. increasing, $\tilde{k}_{12,\,T}$ is C-continuous. Therefore, Lemma 3, Part 1 implies that $M_T \equiv \tilde{k}_{12,\,C_T}$ is a continuous local martingale. Since $\left[\tilde{k}_{12}\right]_\infty = \infty$, C_T is a.s. finite and $\left[M\right]_\infty = \left[\tilde{k}_{12}\right]_{C_\infty} = \infty$. Define another time change $D_T \equiv \inf\left\{s \mid \left[M\right]_s > T\right\}$, then by Lemma 3, Part 2, $B_T \equiv M_{D_T}$ is a Brownian motion that coverges in distribution to a Brownian motion BM_T. Thus $\tilde{k}_{12,\,T} = B_{V_{12,\,T}} \overset{d}{\to} BM_{\underset{N}{lim}V_{12,\,T}}$, where $V_{12} \equiv [M]_{[\tilde{k}_{12}]} = [\tilde{k}_{12}]_{C_{[\tilde{k}_{12}]}} = [\tilde{k}_{12}]$. This together with Proposition 1 implies

$$\hat{\bar{L}}_{X\hat{\beta}_1}(T, \hat{a}) \frac{K_{12}(\hat{a})}{\sqrt{\mathrm{Var}(\tilde{k}_{12})}} \xrightarrow{d} N(0, 1) \tag{1.58}$$

I.e.

$$K_{12}(\hat{a}) = O_p\left(\frac{\sqrt{\mathrm{Var}(\tilde{k}_{12})}}{\bar{L}_{X\hat{\beta}}(T, \hat{a})}\right) = O_p\left(\frac{1}{\sqrt{h_1 \bar{L}_{X\hat{\beta}}(T, \hat{a})}}\right) \tag{1.59}$$

To show the convergence of $K_{13}(\hat{a})$, define $\tilde{k}_{13,t'} \equiv \frac{1}{\sqrt{h_1}} \mathbb{I}\left(X_{t'}\hat{\beta} \in A_n\right) K\left(\frac{X_{t'}\hat{\beta}-\hat{a}}{h_1}\right)$ $\times \int_{t'}^{t'+\Delta} d\bar{Z}_s$ and $\tilde{V}_{13} \equiv \mathrm{Var}\left(\tilde{k}_{13,t'}\right)$. Then both $\tilde{k}^{STD}_{13,t'} \equiv \tilde{k}_{13,t'}/\sqrt{N^2\tilde{V}_{13}}$ and $\left(\tilde{k}^{STD}_{13,t'}\right)^2 - 1$ are MDS'. Levy-Khintchine formula implies that $\left(\tilde{k}^{STD}_{13,t'}\right)^2 = O_p\left(\sqrt{\Delta}\right)$. Thus, there exists some $K \geq O(\Delta)$ such that $\underset{t'}{sup} Var\left[\left(\tilde{k}^{STD}_{13,t'}\right)^2 - 1\right] \leq K$. I.e. $\Sigma_{t'}\left(\tilde{k}^{STD}_{13,t'}\right)^2 \xrightarrow{p} 1 \equiv \eta^2$ according to Lemma 4. Moreover, $\underset{t'}{max} \mid \tilde{k}^{STD}_{13,t'} \mid \xrightarrow{p} 0$ and $E\left[\underset{t'}{max}\left(\tilde{k}^{STD}_{13,t'}\right)^2\right]$ is bounded. Lemma 5 together with Proposition 1 thus implies

$$K_{13}(\hat{a}) = O_p\left(\frac{1}{\sqrt{h_1 \bar{L}_{X\hat{\beta}}(T, \hat{a})}}\right) \tag{1.60}$$

In summary, $\hat{g}_1(\hat{a}) \xrightarrow{p} \mu(a) + \mu_J \lambda(a)$.

I now turn to the second infinitesimal moment $\hat{g}_2(\hat{a})$. Following Bandi and Nguyen (2003, Theorem 2), I use Ito's lemma to write

$$\begin{aligned}
&\left[\log f_{t+\Delta} - \log f_t\right]^2 \\
=&\left(\log f_{t+\Delta}\right)^2 - \left(\log f_t\right)^2 - 2\log f_t\left[\log f_{t+\Delta} - \log f_t\right] &(1.61)\\
=&2\int_t^{t+\Delta}\left(\log f_s - \log f_t\right)\mu^*\left(X_s\right)ds + 2\int_t^{t+\Delta}\left(\log f_s - \log f_t\right)\sigma\left(X_s\right)dW_s \\
&-2\int_t^{t+\Delta}\log f_t\left(J_s\left(X_s\right)dN_s - \mu_J\lambda\left(X_s\right)ds\right) \\
&+\int_t^{t+\Delta}\sigma^2\left(X_s\right)ds + \int_t^{t+\Delta}\left[\mu_J\left(X_s\right)^2 + \sigma_J\left(X_s\right)^2\right]\lambda\left(X_s\right)ds \\
&+\int_t^{t+\Delta}\left[\left(\log f_s + J_s\right)^2 - \left(\log f_s\right)^2\right]\left[J_s\left(X_s\right)dN_s - \mu_J\left(X_s\right)\lambda\left(X_s\right)ds\right] &(1.62)
\end{aligned}$$

Thus

$$\begin{aligned}
\hat{g}_2(\hat{a}) =& \frac{\frac{1}{h_2}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)2\int_t^{t+\Delta}\left(\log f_s - \log f_t\right)\mu_s^* ds}{\frac{\Delta}{h_1}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)} \\
&+\frac{\frac{1}{h_2}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)2\int_t^{t+\Delta}\left(\log f_s - \log f_t\right)\sigma_s dW_s}{\frac{\Delta}{h_1}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)} \\
&-\frac{\frac{1}{h_2}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)2\int_t^{t+\Delta}\log f_t J_s d\bar{N}_s}{\frac{\Delta}{h_1}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)} \\
&+\frac{\frac{1}{h_2}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)\int_t^{t+\Delta}\left[\sigma_s^2 + \left[\mu_{Js}^2 + \sigma_{Js}^2\right]\lambda_s\right]ds}{\frac{\Delta}{h_1}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)} \\
&+\frac{\frac{1}{h_2}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)\int_t^{t+\Delta}\left[\left(\log f_s + J_s\right)^2 - \left(\log f_s\right)^2\right]J_s d\bar{N}_s}{\frac{\Delta}{h_1}\sum_{t'=1}^T K^*\left(X_{t'}\hat{\beta}\right)} &(1.63)\\
\equiv& K_{21}\left(\hat{a}_2\right) + K_{22}\left(\hat{a}_2\right) + K_{23}\left(\hat{a}_2\right) + K_{24}\left(\hat{a}_2\right) + K_{25}\left(\hat{a}_2\right) &(1.64)
\end{aligned}$$

$K_{22}(\hat{a})$, $K_{23}(\hat{a})$ and $K_{25}(\hat{a})$ converge to zero following the same reasoning for showing $K_{12}(\hat{a})$ and $K_{13}(\hat{a})$ converge to zero. Moreover, $K_{24}(\hat{a}) \xrightarrow{p} \sigma^2(a)\left[\mu_J\left(X_s\right)^2 + \sigma_J\left(X_s\right)^2\right]\lambda(a)$ following the same reasoning for proving $K_{11}(\hat{a}) \xrightarrow{p} g_1(a)$. As for $K_{21}(\hat{a})$, the same arguments for proving $K_{11}(\hat{a}) \xrightarrow{p} g_1(a)$ leads to

$$\begin{aligned} K_{21}(\hat{a}) &\leq O_p\left(\bar{L}_{\log f}(T, \log f_t)\sqrt{\Delta \log(1/\Delta)}/h_2\right) \\ &\quad \times \frac{\frac{1}{h_2}\sum_{t'=1,\, t'\neq t}^{T} K^*\left(X_{t'}\hat{\beta}\right) 2\int_t^{t+\Delta} \mu^*(X_s)\, ds}{\frac{\Delta}{h_1}\sum_{t'=1}^{T} K^*\left(X_{t'}\hat{\beta}\right)} \qquad (1.65) \\ &= O_p\left(\bar{L}_{\log f}(T, \log f)\sqrt{\Delta \log(1/\Delta)}/h_2\right)\left(2\mu^*(\hat{a}) + o_p(1)\right) \qquad (1.66) \\ &\xrightarrow{p} 0 \qquad (1.67) \end{aligned}$$

In summary, $\hat{g}_2(\hat{a}) \xrightarrow{p} g_2(a)$.

The consistency of the other moments follows from similar arguments. #

I now proceed to proving that the nonlinear least square estimator $\hat{\beta}$ is indeed consistent. To start with, I show that the true value of the index parameter β_0 is identified up to a multiplicative constant. Similar to the serially uncorrelated case studied by Ichimura (1993), identification of the true value β_0 is based on the observation that on the contour line $X_t\beta_0 = a$, where a is some constant, variations in the residual

$$\epsilon_{jt}(\beta_0) \equiv (\log f_{t+\Delta} - \log f_t)^j - g_j(X_t\beta_0) \qquad (1.68)$$

result only from innovations independent of the state variables X_t, whereas for some other $\beta \neq \beta_0$ variations in $\epsilon_{jt}(\beta)$ along the contour line $X_t\beta = a$ may also result from variations in $X_t\beta_0$, since the true index $X_t\beta_0$ is no longer necessarily a constant. Finding the true value

β_0 thus amounts to minimizing the joint variability

$$\mathcal{J}(\beta) \equiv \sum_{j=1,2,3,4,6} E\left\{\left[\operatorname*{plim}_{\Delta\downarrow 0}\left[\frac{\Delta \log f_{t+\Delta}(X_t)}{\Delta}\right]^j - \operatorname*{plim}_{\Delta\downarrow 0}\left(E\left[\frac{\Delta \log f_{t+\Delta}(X_t)}{\Delta}\right]^j \mid X_t\beta\right)\right]^2\right\} \tag{1.69}$$

which motivates the estimator $\hat{\beta}$ defined by equation (1.19). To establish conditions for the uniqueness of such β_0 that minimizes $\mathcal{J}(\beta)$, I follow essentially the same arguments as in Ichimura (1993, Theorem 4.1):

Assumption 4. The unknown function $g_j(\cdot)$ is differentiable and not a.e. constant on the support of $x\beta_0$, where $j = 1, \ldots, 6$.

Proposition 2. Under Assumptions 1, 2 and 4, the index parameter β_0 is unique up to a multiplicative constant.

Proof. Suppose both β_0 and β_1 minimize the objective function $\mathcal{J}(\beta)$. Then

$$\mathcal{J}(\beta_0) = \mathcal{J}(\beta_1) \tag{1.70}$$

$$= \mathcal{J}(\beta_0) + \sum_{j=1}^{6} E\left\{\left[g_j(X_t\beta_0) - E\left[g_j(X_t\beta_0) \mid X_t\beta_1\right]\right]^2\right\} \tag{1.71}$$

This implies that

$$g_j(X_t\beta_0) \overset{a.e.}{=} E\left[g_j(X_t\beta_0) \mid X_t\beta_1\right] \tag{1.72}$$

for $j = 1, \ldots, 6$. Fix $a = X_t\beta_1$ and denote $\gamma_m \equiv \beta_{0,m} - \beta_{0,1}\beta_{1,m}$, where $\beta_{0,m}$ ($\beta_{1,m}$) is the m-th element of β_0 (β_1). The above equation can be rewritten as

$$g_j\left(\beta_{0,1}r_j + \gamma_2 X_{2t} + \cdots \gamma_5 X_{5t}\right) \overset{a.e.}{=} E\left[g_j\left(X_t\beta_0\right) \mid a\right] = g_j\left(a\right) \tag{1.73}$$

Taking partial derivatives with respect to the m-th state variable, X_{mt}, gives

$$g_j'\left(X_t\beta_0\right)\gamma_m \overset{a.e.}{=} 0, \quad \text{for } m = 2, 3, 4, 5 \tag{1.74}$$

Since g_j is not a constant with strictly positive measure, this implies $\gamma_2 = \cdots = \gamma_5$. I.e. $\beta_0 = c\beta_1$ for some constant c. #

Since the index parameter is identified only up to a multiplicative constant, I follow the convention in the literature by taking $\beta_{0,1} = 1$.

Given a sequence of positive numbers $\left\{M_{jN}\right\}$, denote $a \equiv x\beta$ and

$$\begin{aligned}
A_t^{(j)}(x, \beta) &\equiv \frac{1}{Nh_j}\sum_{t' \neq t} \mathbb{I}\left(X_{t'}\beta \in A_n\right) K\left(\frac{X_{t'}\beta - a}{h_j}\right) \\
&\qquad \times \left(\log f_{t'+\Delta} - \log f_{t'}\right)^j \qquad (1.75)\\
B_t^{(j)}(x, \beta) &\equiv \frac{1}{Nh_j}\sum_{t' \neq t} \mathbb{I}\left(X_{t'}\beta \in A_n\right) K\left(\frac{X_{t'}\beta - a}{h_j}\right) \qquad (1.76)\\
\tilde{l}_{t+\Delta}^{(j)}(x, \beta) &\equiv \mathbb{I}\left(X_t\beta \in A_n\right) K\left(\frac{X_t\beta - a}{h}\right)\left(\log f_{t+\Delta} - \log f_t\right)^j \\
&\qquad \times \mathbb{I}\left(\left(\log f_{t+\Delta} - \log f_t\right) \notin \left[-M_{jN}, M_{jN}\right]\right) \qquad (1.77)\\
l_{t+\Delta}^{(j)}(x, \beta) &\equiv \mathbb{I}\left(X_t\beta \in A_n\right) K\left(\frac{X_t\beta - a}{h_j}\right)\left(\log f_{t+\Delta} - \log f_t\right)^j \\
&\qquad \times \mathbb{I}\left(\left(\log f_{t+\Delta} - \log f_t\right) \in \left[-M_{jN}, M_{jN}\right]\right) \qquad (1.78)
\end{aligned}$$

and the probability limits of $A_t^{(j)}$ and $B_t^{(j)}$ as $\bar{A}_t^{(j)}$ and $\bar{B}_t^{(j)}$, respectively. Note that

$$\begin{aligned}
& \sup_{(x,\beta)\in A\times B} \mid A_t^{(j)} - \bar{A}_t^{(j)} \mid \\
\leq \quad & \sup_{(x,\beta)\in A\times B} \mid \frac{1}{Nh_j} \sum_{t'\neq t} \left(l_{t'+\Delta}^{(j)}(x,\beta) - E_{t'}\left[l_{t'+\Delta}^{(j)}(x,\beta) \right] \right) \mid \\
& \quad + \sup_{(x,\beta)\in A\times B} \mid \frac{1}{Nh_j} \sum_{t'\neq t} \left(\tilde{l}_{t'+\Delta}^{(j)}(x,\beta) - E_{t'}\left[\tilde{l}_{t'+\Delta}^{(j)}(x,\beta) \right] \right) \mid \\
& \quad + \sup_{(x,\beta)\in A\times B} \mid \sum_{t'\neq t} E_{t'} \left(l_{t'+\Delta}^{(j)}(x,\beta) + \tilde{l}_{t'+\Delta}^{(j)}(x,\beta) \right) - \bar{A}_t^{(j)} \mid
\end{aligned} \tag{1.79}$$

Similar to Ichimura (1993), the goal here is to find an appropriate sequence $\{M_{jN}\}$ such that each term on the RHS of the above inequality converges in probability to 0. I first show that the third term on the RHS of inequality (1.79) converges to zero.

Proposition 3. Let $\{Y_t\}_{t\geq 1}$ be a discrete-time homogeneous first-order Markov process with transition density f_Y, and $F : \mathbb{R} \to \mathbb{R}$ be a function. If Ff is twice continuously differentiable, the second derivative satisfies a Lipschitz condition, the kernel function $K(\cdot)$ satisfies Assumption 3, and y is an interior point of the support of Y, then for $h > 0$ and $h \to 0$ as $n \to \infty$,

$$\mid E_{t-1}\left[\frac{F(Y_t)}{h} K\left(\frac{y - Y_t}{h} \right) \right] - F(y) f_Y(y) \mid \xrightarrow{p} 0 \tag{1.80}$$

provided that $E_{t-1}\left[F(Y_t) K\left(\frac{y-Y_t}{h} \right) \right]$ exists.

Proof. Denote $\phi(x) \equiv F(x) f_Y(x)$. Note that

$$
\begin{aligned}
& \mid \int_{\mathbb{R}} \frac{F(x)}{h} K\left(\frac{y-x}{h}\right) f_Y(x)\, dx - F(y) f_Y(y) \mid \\
= & \mid \int_{\mathbb{R}} \phi(y-hs) K(s)\, ds - \phi(y) \mid & (1.81) \\
= & \mid \int_{\mathbb{R}} \left[\phi(y) - hs\phi'(\bar{y})\right] K(s)\, ds - \phi(y) \mid + o(1) & (1.82) \\
\rightarrow & 0 & (1.83)
\end{aligned}
$$

for some $\bar{y}$ between y and $y - hs$. #

Proposition 4 gives the stochastic order of the second term in inequality (1.79).

Assumption 5. Let B denote the parameter space of β. $A \subseteq \mathbb{R}^5$ and $B \subseteq \mathbb{R}^5$ are compact. Moreover, the true value β_0 is in the interior of B.

Proposition 4. Suppose Assumptions 1, 2, 3 and 5 hold, then

$$\sup_{(x,\beta)\in A\times B} \mid \frac{1}{Nh_j} \sum_{t' \neq t} \left(\tilde{l}_{t'}^{(j)}(x,\beta) - E_{t'-1}\left[\tilde{l}_{t'}^{(j)}(x,\beta)\right]\right) \mid = O_p\left(\left[M_{jN}^{m-1} h_j\right]^{-1}\right) \tag{1.84}$$

Proof. Note that

$$P\left(\sup_{(x,\beta)\in A\times B}\mid\frac{1}{Nh_j}\sum_{t'=1}^{T}\left(\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)-E_{t'}\left[\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)\right]\right)\mid>\epsilon_n\right)$$

$$\leq P\left(\sup_{(x,\beta)\in A\times B}\sum_{t'=1}^{T}\mid\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)-E_{t'}\left[\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)\right]\mid>Nh_j\epsilon_n\right) \tag{1.85}$$

$$\leq\frac{E\left(\sup_{(x,\beta)\in A\times B}\sum_{t'=1}^{T}\mid\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)-E_{t'}\left[l^{\tilde{(j)}}{}_{t'+\Delta}(x,\beta)\right]\mid\right)}{Nh_j\epsilon_n} \tag{1.86}$$

Since the kernel function $K(\cdot)$ is bounded, there exists some large constant C_6 such that

$$E\left(\sup_{(x,\beta)\in A\times B}\sum_{t'=1}^{T}\mid\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)\mid\right)$$

$$\leq C_6\sum_{t'=1}^{T}E\left[\mid\log f_{t'+\Delta}-\log f_{t'}\mid^j\mathbb{I}\left((\log f_{t'+\Delta}-\log f_{t'})\notin\left[-M_{jN},M_{jN}\right]\right)\right] \tag{1.87}$$

Moreover, Hölder's and Chebyshev's inequalities imply that

$$\sum_{t'=1}^{T}E\left[\mid\log f_{t'+\Delta}-\log f_{t'}\mid^j\mathbb{I}\left((\log f_{t'+\Delta}-\log f_{t'})\notin\left[-M_{jN},M_{jN}\right]\right)\right]$$

$$\leq\sum_{t'=1}^{T}E\left[\left(\mid(\log f_{t'+\Delta}-\log f_{t'})^j\mid^m\right)^{1/m}\right. \tag{1.88}$$

$$\left.\times\ P\left(\mathbb{I}\left((\log f_{t'+\Delta}-\log f_{t'})\notin\left[-M_{jN},M_{jN}\right]\right)\right)^{1-1/m}\right] \tag{1.89}$$

$$\leq\sum_{t'=1}^{T}E\left(\mid(\log f_{t'+\Delta}-\log f_{t'})^j\mid^m\right)^{1/m}M_{jN}^{1-m} \tag{1.90}$$

where the moment $E\left(\mid\log f_{t+\Delta}-\log f_t\mid^{jm}\right)$ exists for $\forall j,m$ since both $\log f_t$ and X_t satisfy the growth condition specified in Assumptions 1-2. I.e.

$$E\left(\sup_{(x,\beta)\in A\times B}\sum_{t'=1}^{T}\mid\tilde{l}_{t'+\Delta}^{(j)}(x,\beta)\mid\right)\leq\frac{C_6E\left(\mid(\log f_{t+\Delta}-\log f_t)^j\mid^m\right)^{1/m}}{M_{jN}^{m-1}h_j\epsilon_n} \tag{1.91}$$

Similarly,

$$E\left(\sup_{(x,\beta)\in A\times B}\sum_{t'=1}^{T}E_{t'}\mid\tilde{l}_{t'+\Delta}^{(j)}\left(x,\,\beta\right)\mid\right)\leq\frac{C_6 E\left(\mid\left(\log f_{t+\Delta}-\log f_t\right)^j\mid^m\right)^{1/m}}{M_{jN}^{m-1}h_j\epsilon_n}\tag{1.92}$$

provided that the constant C_6 is taken large enough.

Hence, inequalities (1.85), (1.90) and (1.91) together imply

$$P\left(\sup_{(x,\beta)\in A\times B}\mid\frac{1}{Nh_j}\sum_{t'=1}^{T}\left(\tilde{l}_{t'+\Delta}^{(j)}\left(x,\,\beta\right)-E_{t'}\left[\tilde{l}_{t'+\Delta}^{(j)}\left(x,\,\beta\right)\right]\right)\mid>\epsilon_n\right)$$

$$\leq\frac{2C_6\sum_{t'=1}^{T}E\left(\mid\left(\log f_{t'+\Delta}-\log f_{t'}\right)^j\mid^m\right)^{1/m}}{M_{jN}^{m-1}Nh_j\epsilon_n}\tag{1.93}$$

$$\leq\frac{2C_6 E\left(\mid\left(\log f_{t+\Delta}-\log f_t\right)^j\mid^m\right)^{1/m}}{M_{jN}^{m-1}h_j\epsilon_n}\tag{1.94}$$

$$\leq O_p\left(\left[M_{jN}^{m-1}h_j\right]^{-1}\right)\tag{1.95}$$

by triangular and Jensen's inequalities. #

To prove the convergence of the first term on the RHS of inequality (1.79), I resort to Freedman's (1975) inequality for MDS.

Lemma 6 (Freedman's Inequality). Suppose $\{Y_n\}_{n\geq1}$ is a sequence of random variables adapted to a filtration $\{\mathcal{F}_n\}_{n\geq0}$ such that $E\left(Y_n\mid\mathcal{F}_{n-1}\right)=0$. Let τ be a stopping time, and K a positive real number. Suppose $P\left(\mid Y_i\mid\leq K\text{ for }i\leq\tau\right)=1$. Denote $S_n\equiv\sum_{i=1}^{n}Y_i$ and $V_n\equiv\sum_{i=1}^{n}\mathrm{Var}\left(Y_i\mid\mathcal{F}_{i-1}\right)$. Then for all positive real numbers s and v

$$P\left(S_n\geq s\text{ and }V_n\leq v\text{ for some }n\leq\tau\right)\leq\left[\left(\frac{v}{Ks+v}\right)^{Ks+v}\exp\left(Ks\right)\right]^{1/K^2}\tag{1.96}$$

Proof. See Freedman (1975, Proposition 2.1). #

Proposition 5. Suppose Assumptions 1-4 hold. If $h_j = o\left(\frac{M_{jN}}{N}\right)$ and $NM_{jN}^{-1} \to \infty$, then

$$P\left(\frac{1}{Nh_j}\sup_{A\times B} \mid \sum_{t'\neq t}\left(l_{t'+\Delta}^{(j)}(x,\beta) - E_{t'}\left[l_{t'+\Delta}^{(j)}(x,\beta)\right]\right) \mid \geq \epsilon_n\right) \to 0 \tag{1.97}$$

Proof. First note that $\mid l_{t'}^{(j)}(x,\beta) - E_{t'}\left[l_{t'}^{(j)}(x,\beta)\right] \mid \leq C_7 M_{jN}^j$ for some constant C_7, since the kernel function is bounded and the indicator function $\mathbb{I}\left((\log f_{t+\Delta} - \log f_t) \in \left[-M_{jN}, M_{jN}\right]\right)$ implies that $\mid \log f_{t+\Delta} - \log f_t \mid$ cannot be larger than M_{jN}. Popovicio's (1935) inequality then implies that

$$\mathrm{Var}\left(l_{t'+\Delta}^{(j)}(x,\beta) \mid \mathcal{F}_{t'}\right) \leq C_7^2 M_{jN}^j \tag{1.98}$$

Thus

$$V_N \equiv \sum_{t'\neq t}\mathrm{Var}\left(l_{t'+\Delta}^{(j)}(x,\beta) \mid \mathcal{F}_{t'}\right) \leq C_7^2 N M_{jN}^j \tag{1.99}$$

Take the stopping time $\tau \equiv \infty$, $K \equiv C_7 M_{jn}^j$, $s \equiv C_7^{-1} N h_j \epsilon_n$, and $v \equiv C_7^2 N M_{jN}^j$. Freedman's inequality implies that

$$\begin{aligned}
&P\left(\frac{1}{Nh_j}\sup_{A\times B} \mid \sum_{t'\neq t}\left(l_{t'+\Delta}^{(j)}(x,\beta) - E_{t'}\left[l_{t'+\Delta}^{(j)}(x,\beta)\right]\right) \mid \geq \epsilon_n\right) \\
&\leq P\left(\mid \sum_{t'\neq t}\left(l_{t'+\Delta}^{(j)}(x,\beta) - E_{t'}\left[l_{t'+\Delta}^{(j)}(x,\beta)\right]\right) \mid \geq C_7^{-1} N h_j \epsilon_n\right) && (1.100) \\
&\leq 2\left(\frac{Nh_j\epsilon_n}{C_7^2 N} + 1\right)^{-\frac{Nh_j\epsilon_n/C_7^2+N}{M_{jN}^j}} \exp\left(\frac{Nh_j\epsilon_n}{C_7^2 M_{jN}^j}\right) && (1.101) \\
&\to 0 && (1.102)
\end{aligned}$$

if $h_j = o\left(\frac{M_{jN}^{j}}{N}\right)$ and $NM_{jN}^{-j} \to \infty$. #

The following uniform convergence result follows directly from Propositions 3-5.

Proposition 6. Under Assumptions 1-5, if $Nh_j \to \infty$, then for $\forall \epsilon > 0$

$$P\left\{\sup_{(x,\beta)\in A\times B} \mid \hat{g}_j\left(x\beta\right) - g_j\left(x\beta\right) \mid > \epsilon\right\} \to 0 \tag{1.103}$$

as $N \to \infty$.

Proof. To show $A_t^{(j)} \xrightarrow{p} \bar{A}_t^{(j)}$ uniformly using Propositions 3-5, it is sufficient to choose some appropriate m and M_{jN} such that $h_j = O\left(\frac{M_{jN}}{N}\right)$, $NM_{jN}^{-j} \to \infty$, and $\left(M_{jN}^{j}\right)^{m-1} h_j \to \infty$. Given $Nh_j \to \infty$, the choice of $m = 2$ and $M_{jN}^{j} = O_p\left(\sqrt{N}\right)$ shall do the job. Since $B_t^{(j)}$ is a special case of $A_t^{(j)}$ with $(\log f_{t'+\Delta} - \log f_{t'})^j$ replaced by 1, the same arguments imply that $B_t^{(j)} \xrightarrow{p} \bar{B}_t^{(j)}$ uniformly under the same conditions. Thus $\hat{g}_j = A_t^{(j)}/B_t^{(j)} \xrightarrow{p} \bar{A}_t^{(j)}/\bar{B}_t^{(j)}$ uniformly. #

I am now ready to state the consistency of $\hat{\beta}$. The proof follows the standard outline used in the literature (e.g. Ichimura, 1993, Theorem 5.1; Newey and McFadden, 1994).

Theorem 2. Suppose Assumptions 1-5 hold. As $N \to \infty$, $T \to \infty$ and $\Delta \equiv (T-1)/N \to 0$, if the bandwidth h_j is such that $\left(L_{X\hat{\beta}}\left(T,\, X\hat{\beta}\right)/h_j\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2}$, $\left(L_{\log f}\left(T,\, \log f\right)/h_j\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2}$ and $1/\left(h_1 \bar{L}_{X\hat{\beta}}\left(T,\, X\hat{\beta}\right)\right)^{-1/2}$ converge in probability to 0, and $Nh \to \infty$, then the index estimator $\hat{\beta}$ is consistent.

Proof. Denote

$$\hat{J}_{jN}(\beta) \equiv \frac{1}{T}\sum_{t=1}^{T}\left[(\log f_{t'+\Delta} - \log f_{t'})^j - \hat{g}_j(X_t\beta)\right]^2 \tag{1.104}$$

$$J_{jN}(\beta) \equiv \frac{1}{T}\sum_{t=1}^{T}\left[(\log f_{t'+\Delta} - \log f_{t'})^j - g_j(X_t\beta)\right]^2 \tag{1.105}$$

$$J_j(\beta) \equiv \frac{1}{T}\sum_{t=1}^{T}E\left[(\log f_{t'+\Delta} - \log f_{t'})^j - g_j(X_t\beta)\right]^2 \tag{1.106}$$

and recall that the true value of the index parameter β is denoted as β_0.

By definition of the nonlinear least square estimator $\hat{\beta}$,

$$P\left(\hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}(\beta_0)\right) = 1, \quad \forall j = 1, \ldots, 6 \tag{1.107}$$

On the other hand, for any open neighborhood of β_0, $U(\beta_0)$,

$$\begin{aligned}
&P\left(\hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}(\beta_0)\right) \\
=&P\left(\hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}(\beta_0) \text{ and } \hat{\beta} \in U(\beta_0)\right) \\
&\quad + P\left(\hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}(\beta_0) \text{ and } \hat{\beta} \in B \setminus U(\beta_0)\right) && (1.108)\\
\leq&P\left(\hat{\beta} \in U(\beta_0)\right) + P\left(\underset{\hat{\beta}\in B\setminus U(\beta_0)}{inf} \hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}(\beta_0)\right) && (1.109)
\end{aligned}$$

Thus

$$P\left(\underset{\hat{\beta}\notin U(\beta_0)}{inf} \hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}(\beta_0)\right) \to 0, \quad \forall j = 1, \ldots, 6 \tag{1.110}$$

implies consistency.

Note that

$$
\begin{aligned}
& P\left(\inf_{\hat{\beta}\in B\backslash U(\beta_0)} \hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}\left(\beta_0\right)\right) \\
\leq & P\left(\inf_{\hat{\beta}\in B\backslash U(\beta_0)} \left[\hat{J}_{jN}\left(\hat{\beta}\right) - J_{jN}\left(\hat{\beta}\right)\right] + \inf_{\hat{\beta}\in B\backslash U(\beta_0)} \left[J_{jN}\left(\hat{\beta}\right) - J_j\left(\hat{\beta}\right)\right]\right. \\
& \left. + \inf_{\hat{\beta}\in B\backslash U(\beta_0)} J_j\left(\hat{\beta}_j\right) + J_j\left(\beta_0\right) \leq \hat{J}_{jN}\left(\beta\right) + J_j\left(\beta_0\right)\right) \qquad (1.111) \\
\leq & P\left(\sup_{\hat{\beta}\in B\backslash U(\beta_0)} \mid \hat{J}_{jN}\left(\hat{\beta}\right) - J_{jN}\left(\hat{\beta}\right) \mid + \sup_{\hat{\beta}\in B\backslash U(\beta_0)} \mid J_{jN}\left(\hat{\beta}\right) - J_j\left(\hat{\beta}\right) \mid\right. \\
& \left. + \mid \hat{J}_{jN}\left(\beta_0\right) - J_j\left(\beta_0\right) \mid \geq \inf_{\hat{\beta}\in B\backslash U(\beta_0)} J_j\left(\hat{\beta}\right) - J_j\left(\beta_0\right)\right) \qquad (1.112)
\end{aligned}
$$

Proposition 1 implies that there exists $\epsilon > 0$ such that

$$\inf_{\hat{\beta}\in B\backslash U(\beta_0)} J_j\left(\hat{\beta}\right) - J_j\left(\beta_0\right) > \epsilon \qquad (1.113)$$

Proposition 4 implies that

$$\sup_{\hat{\beta}\in B\backslash U(\beta_0)} \mid \hat{J}_{jN}\left(\hat{\beta}\right) - J_{jN}\left(\hat{\beta}\right) \mid \xrightarrow{p} 0 \qquad (1.114)$$

Moreover,

$$J_{jN}\left(\hat{\beta}\right) - J_j\left(\hat{\beta}\right) \xrightarrow{p} 0 \qquad (1.115)$$

following the same arguments as in the proof of Theorem 1 except that here $\hat{\beta}$ is taken as a fixed constant instead of $\xrightarrow{p} \beta_0$. Since B is compact and $J_{jN}\left(\cdot\right)$ is continuous

$$\sup_{\hat{\beta}\in B\backslash U(\beta)} \mid J_{jN}\left(\hat{\beta}\right) - J_j\left(\hat{\beta}\right) \mid \xrightarrow{p} 0 \qquad (1.116)$$

Thus, $P\left(\underset{\hat{\beta}\in B\setminus U(\beta)}{inf} \hat{J}_{jN}\left(\hat{\beta}\right) \leq \hat{J}_{jN}\left(\beta_0\right)\right) \to 0$. #

Finally, Theorems 1-2 together immediately imply the consistency of the functional estimates of the infinitesimal conditional moments:

Theorem 3. Suppose Assumptions 1-5 hold. As $N \to \infty$, $T \to \infty$ and $\Delta \equiv (T-1)/N \to 0$, if the bandwidth h_j is such that $\left(L_{X\hat{\beta}}\left(T, X\hat{\beta}\right)/h_j\right)(\Delta \log(1/\Delta))^{1/2}$, $\left(L_{\log f}\left(T, \log f\right)/h_j\right)(\Delta \log(1/\Delta))^{1/2}$ and $1/\left(h_j \bar{L}_{X\hat{\beta}}\left(T, X\hat{\beta}\right)\right)^{-1/2}$ converge in probability to 0, and $Nh_j \to \infty$, then

$$\hat{g}_j\left(X_t\hat{\beta}\right) \xrightarrow{p} g_j\left(X_t\beta_0\right) \tag{1.117}$$

for j = 1, ..., 6. #

It remains to identify the model parameters in equations (1.9)-(1.14), $\mu(\cdot)$, $\sigma^2(\cdot)$, $\lambda(\cdot)$, $\mu_J(\cdot)$ and $\sigma_J^2(\cdot)$, from the first six moment conditions.[3] For convenience, I omit the time subscripts in the following. Denote the ratio between the i-th and the j-th infinitesimal

[3]Mathematica codes are available from the author upon request.

moments as $g_{i|j} \equiv g_i/g_j$, and the n-th smallest real roots of

$$r_{3|4}(x) \equiv 375g_{4|5}^4 - 3600g_{4|5}^3 x + 8088g_{4|5}^2 x^2 - 6912g_{4|5}x^3 + 2048x^4 \tag{1.118}$$

$$\begin{aligned} r_{5|6}(x) \equiv{} & 256g_{3|4}^5 g_{4|5}^2 + \left(-1920g_{3|4}^4 g_{4|5}^2 + 800g_{3|4}^3 g_{4|5}^3\right) x \\ & + \left(384g_{3|4}^4 g_{4|5} + 4080g_{3|4}^3 g_{4|5}^2 - 3000g_{3|4}^2 g_{4|5}^3 + 375g_{3|4}g_{4|5}^4\right) x^2 \\ & + \left(-1248g_{3|4}^3 g_{4|5} - 2120g_{3|4}^2 g_{4|5}^2 + 2400g_{3|4}g_{4|5}^3 - 375g_{4|5}^4\right) x^3 \\ & + \left(48g_{3|4}^3 + 720g_{3|4}^2 g_{4|5} - 150g_{3|4}g_{4|5}^2 - 250g_{4|5}^3\right) x^4 \end{aligned} \tag{1.119}$$

$$r_{\mu_J}(x) \equiv 3 - 18g_{4|5}x + \left(24g_{3|4}g_{4|5} + 15g_{4|5}^2\right) x^2 - 40g_{3|4}g_{4|5}^2 x^3 + 16g_{3|4}^2 g_{4|5}^2 x^4 \tag{1.120}$$

as $r_{3|4}^{(n)}$, $r_{5|6}^{(n)}$ and $r_{\mu_J}^{(n)}$, respectively. The solution of $\mu_J(\cdot)$ is given by

$$\mu_J = \begin{cases} \mu_{J1}, & \text{if } g_{4|5} < 0 \\ \mu_{J2}, & \text{if } g_{4|5} > 0 \end{cases} \tag{1.121}$$

where

$$\mu_{J1} \equiv \begin{cases} r_{\mu_J}^{(1)}, & \text{if } g_{3|4} \leq g_{4|5} \text{ and } g_{5|6} = r_{5|6}^{(1)} \\ r_{\mu_J}^{(2)}, & \text{if } g_{3|4} \leq g_{4|5} \text{ and } g_{5|6} = r_{5|6}^{(2)}, \text{ or } g_{4|5} \leq g_{3|4} < r_{3|4}^{(2)} \\ r_{\mu_J}^{(4)}, & \text{otherwise} \end{cases} \tag{1.122}$$

$$\mu_{J2} \equiv \begin{cases} r_{\mu_J}^{(1)}, & \text{otherwise} \\ r_{\mu_J}^{(2)}, & \text{if } g_{4|5} \leq g_{3|4} < r_{3|4}^{(2)} \text{ and } g_{5|6} = r_{5|6}^{(2)} \end{cases} \tag{1.123}$$

Given $\mu_J(\cdot)$, $\sigma_J^2(\cdot)$ can be solved as

$$\sigma_J^2 = \frac{3\mu_J^2 - 5g_{4|5}\mu_J^3 + \sqrt{2\left(3\mu_J^4 - 6g_{4|5}\mu_J^5 + 5g_{4|5}^2\mu_J^6\right)}}{3\left(-1 + 5g_{4|5}\mu_J\right)} \tag{1.124}$$

and then $\lambda(\cdot)$ as

$$\lambda = \frac{g_3}{\mu_J^3 + 3\mu_J\sigma_J^2} \tag{1.125}$$

Finally, $\mu(\cdot)$ and $\sigma^2(\cdot)$ can be solved as

$$\mu = g_1 - \lambda\mu_J \tag{1.126}$$

$$\sigma^2 = g_2 - \lambda\left(\mu_J^2 + \sigma_J^2\right) \tag{1.127}$$

Appendix B: A Randomized Bandwidth Selection Procedure

Set $h_1 = \cdots = h_6 = h$. In line with the requirement that $Nh \to \infty$, I consider bandwidths within the range $H \equiv \left\{\frac{5c+1}{N} \mid c = 0, 1, \ldots, 60\right\}$. Theorem 3 imposes three asymptotic conditions on h, namely

$$\left(L_{X\hat{\beta}}\left(T, X\hat{\beta}\right)/h\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2} \xrightarrow{p} 0 \tag{1.128}$$

$$hL_{X\hat{\beta}}\left(T, X\hat{\beta}\right) \xrightarrow{p} \infty \tag{1.129}$$

$$\left(L_{\log f}\left(T, \log f\right)/h\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2} \xrightarrow{p} 0 \tag{1.130}$$

Following Bandi, Corradi, and Moloche (2009), to test if each candidate bandwidth $h \in H$ satisfies these conditions I consider the following test statistic:

$$V_{R,h} \equiv \min\left\{V_{1,R,h},\ V_{2,R,h},\ V_{f,R,h}\right\} \tag{1.131}$$

where

$$V_{i,R,h} = \int_U \left[\frac{2}{\sqrt{R}}\sum_{r=1}^{R}\left(\mathbb{I}\left\{v_{i,h,r} \leq u\right\} - \frac{1}{2}\right)\right]^2 \pi\left(u\right) du \tag{1.132}$$

for $i = 1, 2$, $\pi\left(u\right)$ is some weight function satisfying $\int_U \pi\left(u\right) du = 1$ and $\pi\left(u\right) \geq 0$, and

$$v_{1,h,r} = \left(\exp\left[\left(\int h^{(1+\epsilon)}\hat{\hat{L}}_{X\hat{\beta}}\left(T, a\right) da\right)^{-1}\right]\right)^{1/2}\eta_{j1,r} \tag{1.133}$$

$$v_{2,h,r} = \left(\exp\left[\int \frac{\hat{\hat{L}}_{X\hat{\beta}}\left(T, a\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2}}{h^{(1+\epsilon)}} da\right]\right)^{1/2}\eta_{j2,r} \tag{1.134}$$

$$v_{f,h,r} = \left(\exp\left[\int \frac{\hat{\hat{L}}_{\log f}\left(T, a\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2}}{h^{(1+\epsilon)}} da\right]\right)^{1/2}\eta_{f,r} \tag{1.135}$$

with $\left(\eta_{1,r},\ \eta_{2,r},\ \eta_{f,r}\right) \overset{i.i.d.}{\sim} N\left(0,\ I_3\right)$ for $r = 1, \ldots, R$, and $\epsilon > 0$ is some arbitrarily small number. The index and functional estimates here are calculated with h being the bandwidth.

The idea of the test is as follows. Suppose the condition $\left(L_{X\hat{\beta}}\left(T,\ X\hat{\beta}\right)/h\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2} \xrightarrow{p} 0$ is violated, then $\left(L_{X\hat{\beta}}\left(T,\ X\hat{\beta}\right)/h^{(1+\epsilon)}\right)\left(\Delta \log\left(1/\Delta\right)\right)^{1/2} \xrightarrow{p} \infty$ since $h \to 0$. This implies that $v_{2,h,r}$ diverges to ∞ with probability $1/2$ and to $-\infty$ with probability $1/2$. I.e. the random variable $\mathbb{I}\left\{v_{i,h,r} \leq u\right\} - \frac{1}{2}$ follows a Bernoulli distribution, and so $V_{i,R,h} \xrightarrow{d} \chi^2_1$. If instead $\left(L_{X\hat{\beta}}\left(T,\ X\hat{\beta}\right)/h\right) \times \left(\Delta \log\left(1/\Delta\right)\right)^{1/2} \xrightarrow{p} 0$ is satisfied, then $\frac{2}{\sqrt{R}}\sum_{r=1}^{R}\left(\mathbb{I}\left\{v_{ji,h,r} \leq u\right\} - \frac{1}{2}\right)$ diverges to infinity. The same reasoning also applies to the other two conditions. They together imply that $V_{R,h} \xrightarrow{d} \chi^2_3$ if any of the three asymptotic conditions is violated, and diverges otherwise.

I take $\pi\left(u\right)$ as the standard normal density, $U = [-2.5,\ 2.5]$, $\epsilon = 10^{-5}$, and $R = 100{,}000$. I estimate the integral terms in $v_{1,h,r}$, $v_{2,h,r}$ and $v_{f,h,r}$ using trapzoidal rule with 1,000,000 grid points over the range of the sample, and those in $V_{i,R,h}$ and $V_{i,R,h}$ with 1,000 grid. Following Bandi, Corradi, and Moloche (2009), among all $h \in H$ that satisfy these asymptotic conditions, I pick the one that minimizes the Kolmogorov-Smirnov distance

$$d_{KS} \equiv \sup_x \mid F_h\left(x\right) - N\left(x\right) \mid \tag{1.136}$$

where $N\left(x\right)$ is the distribution function of $N\left(0,\ 1\right)$, and $F_h\left(x\right)$ is the empirical distribution function of

$$\hat{\epsilon}_{t+\Delta} \equiv \frac{\Delta \log f_t - \hat{\mu}_t}{\sqrt{\Delta}\hat{\sigma}_t} \tag{1.137}$$

which is asymptotically distributed as $N(0, 1)$ if all these three conditions on h are satisfied.[4]

[4] To avoid numerical problem in which the estimated standard deviation $\hat{\sigma}^2(X_t) \simeq 0$, I add a small value 10^{-50} to $\hat{\sigma}^2(X_t)$ in such occasions.

Appendix C: Bootstrap Procedure

Denote the residuals of some generic regression model estimated under the null hypothesis as $\hat{w}_t$. For the i-th bond yield in the dataset, y_{it}, denote its estimated loadings on $PC_t \equiv (L_t \;\; S_t \;\; C_t)'$ as $\hat{\lambda}_i$, and its residuals as $\hat{e}_{it} = y_{it} - \hat{\lambda}_i PC_t$. I start by fitting an AR(1) model $\hat{e}_{it} = \hat{\rho}_i \hat{e}_{it-1} + \hat{u}_{it}$ for each i, and stack the residuals $\hat{u}_{it}$ across i as well as $\hat{w}_t$ into a vector $\hat{u}_t$.

I conduct resampling in the following way. If the regression model is conducted on both FOMC and non-FOMC days, I simply stack $\hat{w}_t$ with $\hat{u}_t$ into $\hat{v}_t \equiv (\hat{w}_t, \hat{u}_t)$. Denote $\hat{V}$ as the matrix whose t-th row is $\hat{v}_t$. I conduct independent sampling with replacement from the rows of $\hat{V}$. If instead the regression model is conducted on FOMC days only, I define a new sequence of residuals $\tilde{w}_t$, which is equal to $\hat{w}_t$ if t is an FOMC day, and is labeled as missing value otherwise, and stack $\tilde{w}_t$ with $\hat{u}_t$ into $\tilde{v}_t \equiv (\tilde{w}_t, \hat{u}_t)$. Denote $\tilde{V}$ as the matrix whose t-th row is $\tilde{v}_t$, $\tilde{V}_{FOMC}$ as the submatrix of $\tilde{V}$ keeping only rows correponding to FOMC days, and $\tilde{V}_{non-FOMC}$ as the submatrix keeping only rows correponding to non-FOMC days. For each $t = 1, \ldots, T$, I conduct independent sampling with replacement from the rows of $\tilde{V}_{FOMC}$ if t is an FOMC day, and from the rows of $\tilde{V}_{non-FOMC}$ otherwise.

I use the bootstrapped sample of $\hat{u}_t$, denoted as $\hat{u}_t^*$, to form a bootstrapped sample of the residuals $\hat{e}_{it}^* \equiv \hat{\rho}_i \hat{e}_{it-1}^* + \hat{u}_{it}^*$, which are in turn used to form a bootstrapped sample of each individual yield $\hat{y}_{it}^* \equiv \hat{\lambda}_i PC_t + \hat{e}_{it}^*$. I conduct principal component analysis on the bootstrapped yields to get the first three principal components PC_t^*. I then take $X_t^* \equiv (PC_t^{*\prime} \;\; TYVIX_t \;\; ADS_t)'$ as the bootstrapped state vector to re-estimate the semiparametric jump-diffusion models.[5]

[5] To reduce the excessive computing resources needed, I do not re-estimate the bandwidth for each bootstrapped sample using simulated annealing. Instead I run a local optimization routine to choose the bandwidth by taking the optimal bandwidth found in-sample as the initial value.

I construct bootstrapped sample of the dependent variable of the regression under consideration by adding the bootstrapped sample of $\hat{w}_t$, denoted as $\hat{w}_t^*$, to the fitted values estimated under the null hypothesis. Finally, I run the regression using the bootstrapped independent and dependent variables. The bootstrapped p-value for some t-statistic, t, is taken as the fraction of the same statistics in the bootstrapped samples, denoted as t^*, satisfying $\mid t^* \mid > \mid t \mid$. Similarly, the p-value for some F-statistic, F, is taken as the proportion in the bootstrapped samples satisfying $F^* > F$.

I generate 1,000 bootstrapped samples for statistical inferences.

Tables and Figures

Table 1.1: Summary Statistics

	AR(1) Coefficient	Phillips-Perron p-Value
Panel A: Daily Changes of log-Federal Funds Futures		
1-Month Ahead	0.0630	< 0.01
2-Month Ahead	0.0141	< 0.01
3-Month Ahead	0.0105	< 0.01
4-Month Ahead	0.0075	< 0.01
5-Month Ahead	0.0209	< 0.01
6-Month Ahead	0.0173	< 0.01
Panel B: State Variables		
Level	0.9991	0.9221
Slope	0.9988	0.6065
Curvature	0.9939	0.0475
TYVIX	0.9879	< 0.01
Business Condition	0.9994	0.4193

Notes: AR(1) coefficients and p-values of Phillips-Perron (1988) unit root tests.

Table 1.2: In-Sample R^2s of Fitted log-Federal Funds Futures

	1 Month Ahead	2 Months Ahead	3 Months Ahead	4 Months Ahead	5 Months Ahead	6 Months Ahead
Moment #1	19.76%	13.27%	11.37%	11.46%	10.22%	9.98%
Moment #2	26.13%	44.15%	43.11%	40.31%	29.90%	20.89%
Moment #3	23.81%	50.98%	51.68%	51.07%	41.89%	22.07%
Moment #4	20.77%	53.17%	53.32%	53.13%	48.21%	22.91%
Moment #5	0.91%	53.07%	53.29%	53.07%	46.62%	12.33%
Moment #6	15.40%	53.64%	53.58%	53.52%	52.29%	21.61%

Notes: Goodness-of-fit of the single-index estimators for the infinitesimal moments of federal funds futures.

Table 1.3: Information Contents and Endogeneity of High-Frequency Monetary Shocks

	1 Month Ahead	2 Months Ahead	3 Months Ahead	4 Months Ahead	5 Months Ahead	6 Months Ahead
Panel A: Predictability of HF Monetary Shocks by Expected Daily Changes						
Expected Changes in FFF	0.36**	0.25	0.29*	0.63***	0.82***	0.64***
	(p =0.01)	(p =0.10)	(p =0.08)	(p =0.00)	(p =0.00)	(p =0.00)
Adjusted R^2	5.20%	1.28%	2.29%	10.04%	13.16%	8.88%
Signal-to-Noise Ratio	0.06	0.02	0.03	0.12	0.16	0.11
Panel B: Predictability of HF Monetary Shocks by State Variables						
State Variables	(1.24,	(1.12,	(0.97,	(1.08,	(1.16,	(1.22,
	p =0.29)	p =0.34)	p =0.43)	p =0.34)	p =0.28)	p =0.25)
Adjusted R^2	1.10%	0.53%	-0.13%	0.35%	0.74%	0.98%
Signal-to-Noise Ratio	0.06	0.05	0.05	0.05	0.06	0.06
Panel C: Predictability of HF Monetary Shocks by Expected Daily Changes, Conditional on State Variables						
Expected Changes in FFF	0.39**	0.29*	0.29*	0.65**	0.80***	0.60**
	(p =0.03)	(p =0.07)	(p =0.09)	(p =0.01)	(p =0.00)	(p =0.02)
Adjusted R^2	6.60%	2.38%	2.01%	10.39%	12.81%	8.55%
Signal-to-Noise Ratio	0.13	0.08	0.08	0.18	0.21	0.16
Panel D: Predictability of HF Monetary Shocks by Index of State Variables						
Expected Changes in FFF	0.00	0.00	0.00	0.00	0.00	0.00
	(p =0.50)	(p =0.31)	(p =0.75)	(p =0.86)	(p =0.77)	(p =0.74)
Adjusted R^2	-0.33%	-0.84%	-0.84%	-0.90%	-0.85%	-0.84%
Signal-to-Noise Ratio	0.01	0.00	0.00	0.00	0.00	0.00
Panel E: Predictability of Unexpected Changes in FFF, All Days						
State Variables	(4.09,	(15.47,	(16.47,	(16.02,	(12.08,	(9.70,
	p =0.46)	p =0.46)	p =0.47)	p =0.46	p =0.46)	p =0.48)
Adjusted R^2	0.40%	1.86%	1.98%	1.92%	1.43%	1.12%
Signal-to-Noise Ratio	0.01	0.02	0.02	0.02	0.02	0.01
Panel F: Predictability of Unexpected Changes in FFF, FOMC Days						
State Variables	(0.68,	(2.71,	(2.40,	(2.16,	(2.00,	(2.19,
	p =0.50)	p =0.49)	p =0.48)	p =0.46)	p =0.44)	p =0.44)
Adjusted R^2	-1.50%	7.32%	6.10%	5.09%	4.44%	5.21%
Signal-to-Noise Ratio	0.03	0.13	0.12	0.10	0.10	0.11

Notes: Panel A – predictive regressions of high-frequency (HF) changes in federal funds futures (FFFs) within a 30-minute window bracketing each FOMC announcement on expected daily changes in FFFs on FOMC announcement days, where the slope coefficients and the p-values are reported. Panel B – predictive regressions of HF changes on values of the state variables the days before, where the F-statistics along with the p-values are reported in brackets. Panel D – Predictive regressions of HF changes on estimated index of state variables, where the slope coefficients are reported. Panel E – predictive regressions of unexpected daily changes in FFF on the state variables over entire sample, where the F-statistics are reported in brackets. Panel F – same as in Panel E except the regressions are conducted over FOMC announcement days only. All significance levels are calculated from the boostrap procedure outlined in Appendix C.

"***": Significance at 1% level; "**": Significance at 5% level; "*": Significance at 10% level.

Table 1.4: Robustness Check - Pre-Zero-Lower-Bound Sample

	1 Month Ahead	2 Months Ahead	3 Months Ahead	4 Months Ahead	5 Months Ahead	6 Months Ahead
Panel A: Predictability of HF Monetary Shocks by Expected Daily Changes						
Expected Changes in FFF	0.24	0.89***	0.97***	0.67***	0.56**	0.72***
	(p =0.35)	(p =0.00)	(p =0.00)	(p =0.00)	(p =0.01)	(p =0.00)
Adjusted R^2	-0.34%	19.35%	47.55%	14.30%	11.40%	16.81%
Signal-to-Noise Ratio	0.02	0.26	0.94	0.19	0.15	0.23
Panel B: Predictability of HF Monetary Shocks by State Variables						
Expected Changes in FFF	(0.74,	(1.06,	(1.39,	(1.34,	(1.12,	(0.96,
	p =0.64)	p =0.41)	p =0.25)	p =0.28)	p =0.37)	p =0.46)
Adjusted R^2	-2.47%	0.59%	3.52%	3.13%	1.13%	-0.40%
Signal-to-Noise Ratio	0.08	0.11	0.14	0.14	0.12	0.10
Panel C: Predictability of HF Monetary Shocks by Expected Daily Changes, Conditional on State Variables						
Expected Changes in FFF	0.54	0.92***	0.91***	0.67**	0.54**	0.70***
	(p =0.12)	(p =0.00)	(p =0.00)	(p =0.01)	(p =0.03)	(p =0.00)
Adjusted R^2	2.04%	21.84%	46.09%	17.02%	11.01%	14.83%
Signal-to-Noise Ratio	0.15	0.44	1.09	0.36	0.27	0.32
Panel D: Predictability of HF Monetary Shocks by Index of State Variables						
Expected Changes in FFF	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(p =0.79)	(p =0.39)	(p =0.55)	(p =0.45)	(p =0.36)	(p =0.48)
Adjusted R^2	-1.80%	-0.45%	-1.19%	-0.01	-0.16%	-0.84%
Signal-to-Noise Ratio	0.00	0.01	0.01	0.01	0.02	0.01
Panel E: Predictability of Unexpected Changes in FFF, All Days						
State Variables	(1.42,	(1.61,	(0.60,	(0.48,	(0.26,	(0.40,
	p =0.76)	p =0.75)	p =0.78)	p =0.76)	p =0.76)	p =0.74)
Adjusted R^2	0.14%	0.21%	-0.13	-0.18%	-0.25%	-0.20%
Signal-to-Noise Ratio	0.00	0.01	0.00	0.00	0.00	0.00
Panel F: Predictability of Unexpected Changes in FFF, FOMC Days						
State Variables	(0.48,	(2.16,	(1.29,	(1.30,	(1.58,	(0.80,
	p =0.39)	p =0.79)	p =0.73)	p =0.83)	p =0.34)	p =0.81)
Adjusted R^2	-5.25%	10.06%	2.75%	2.81%	5.26%	-1.91%
Signal-to-Noise Ratio	0.05	0.23	0.14	0.14	0.17	0.09

Notes: Same empirical exercises as in Table 1.3 using subsample prior to zero-lower-bound period. See notes for Table 1.3 for more details.

Table 1.5: Robustness Check - Using Bandwidths Five Times of Optimal Values

	1 Month Ahead	2 Months Ahead	3 Months Ahead	4 Months Ahead	5 Months Ahead	6 Months Ahead
Panel A: Predictability of HF Monetary Shocks by Expected Daily Changes						
Expected Changes in FFF	0.90**	1.77***	0.96***	1.51***	0.32**	0.94***
	(p =0.01)	(p =0.00)	(p =0.00)	(p =0.00)	(p =0.04)	(p =0.00)
Adjusted R^2	12.33%	36.10%	23.06%	40.67%	3.65%	30.26%
Signal-to-Noise Ratio	0.15	0.58	0.31	0.70	0.05	0.45
Panel C: Predictability of HF Monetary Shocks by Expected Daily Changes, Conditional on State Variables						
Expected Changes in FFF	0.90***	1.83***	0.97***	1.54***	0.32**	0.92***
	(p <0.01)	(p =0.00)	(p =0.00)	(p =0.00)	(p <0.05)	(p =0.00)
Adjusted R^2	13.28%	36.34%	20.72%	39.51%	4.23%	30.35%
Signal-to-Noise Ratio	0.22	0.66	0.33	0.75	0.10	0.52
Panel D: Predictability of HF Monetary Shocks by Index of State Variables						
Expected Changes in FFF	0.00*	0.00*	0.00	0.00	0.00	0.00
	(p =0.07)	(p =0.09)	(p =0.17)	(p =0.21)	(p =0.18)	(p =0.13)
Adjusted R^2	2.02%	1.69%	0.99%	0.70%	1.06%	1.52%
Signal-to-Noise Ratio	0.03	0.03	0.02	0.02	0.02	0.02
Panel E: Predictability of Unexpected Changes in FFF, All Days						
State Variables	(5.29,	(7.41,	(7.73,	(7.65,	(6.03,	(6.51,
	p =0.46)	p =0.43)	p =0.41)	p =0.42)	p =0.42)	p =0.41)
Adjusted R^2	0.56%	0.83%	0.87%	0.86%	0.65%	0.71%
Signal-to-Noise Ratio	0.01	0.01	0.01	0.01	0.01	0.01
Panel F: Predictability of Unexpected Changes in FFF, FOMC Days						
State Variables	(0.55,	(1.96,	(0.83,	(1.42,	(0.82,	(1.16,
	p =0.56)	p =0.47)	p =0.46)	p =0.50)	p =0.47)	p =0.46)
Adjusted R^2	-2.12%	4.24%	-0.79%	1.89%	-0.84%	0.01
Signal-to-Noise Ratio	0.03	0.09	0.04	0.07	0.04	0.06

Notes: Same empirical exercises as in Table 1.3 except using bandwidths five times of their optimal values selected by the method outlined in Appendix B. See notes for Table 1.3 for more details.

Table 1.6: Robustness Check - Using Median Bandwidth across Horizons

	1 Month Ahead	2 Months Ahead	3 Months Ahead	4 Months Ahead	5 Months Ahead	6 Months Ahead
Panel A: Predictability of HF Monetary Shocks by Expected Daily Changes						
Expected Changes in FFF	0.59***	0.17	0.19	0.60***	0.81***	0.63***
	(p =0.00)	(p =0.23)	(p =0.21)	(p <0.01)	(p =0.00)	(p =0.00)
Adjusted R^2	15.01%	0.27%	0.45%	8.18%	12.94%	7.85%
Signal-to-Noise Ratio	0.19	0.01	0.01	0.10	0.16	0.10
Panel C: Predictability of HF Monetary Shocks by Expected Daily Changes, Conditional on State Variables						
Expected Changes in FFF	0.90***	1.83***	0.97***	1.54***	0.32**	0.92
	(p <0.01)	(p =0.00)	(p =0.00)	(p =0.00)	(p <0.05)	(p =0.00)
Adjusted R^2	13.28%	36.34%	20.72%	39.51%	4.23%	30.35%
Signal-to-Noise Ratio	0.22	0.66	0.33	0.75	0.10	0.52
Panel D: Predictability of HF Monetary Shocks by Index of State Variables						
Expected Changes in FFF	0.00*	0.00*	0.00	0.00	0.00	0.00
	(p =0.07)	(p =0.09)	(p =0.17)	(p =0.21)	(p =0.18)	(p =0.13)
Adjusted R^2	2.02%	1.69%	0.99%	0.70%	1.06%	1.52%
Signal-to-Noise Ratio	0.03	0.03	0.02	0.02	0.02	0.02
Panel E: Predictability of Unexpected Changes in FFF, All Days						
State Variables	(5.29,	(7.41,	(7.73,	(7.65,	(6.03,	(6.51,
	p =0.46)	p =0.43)	p =0.41)	p =0.42)	p =0.42)	p =0.41)
Adjusted R^2	0.56%	0.83%	0.87%	0.86%	0.65%	0.71%
Signal-to-Noise Ratio	0.01	0.01	0.01	0.01	0.01	0.01
Panel F: Predictability of Unexpected Changes in FFF, FOMC Days						
State Variables	(0.55,	(1.96,	(0.83,	(1.42,	(0.82,	(1.16,
	p =0.56)	p =0.47)	p =0.46)	p =0.50)	p =0.47)	p =0.46)
Adjusted R^2	-2.12%	4.24%	-0.79%	1.89%	-0.84%	0.71%
Signal-to-Noise Ratio	0.03	0.09	0.04	0.07	0.04	0.06

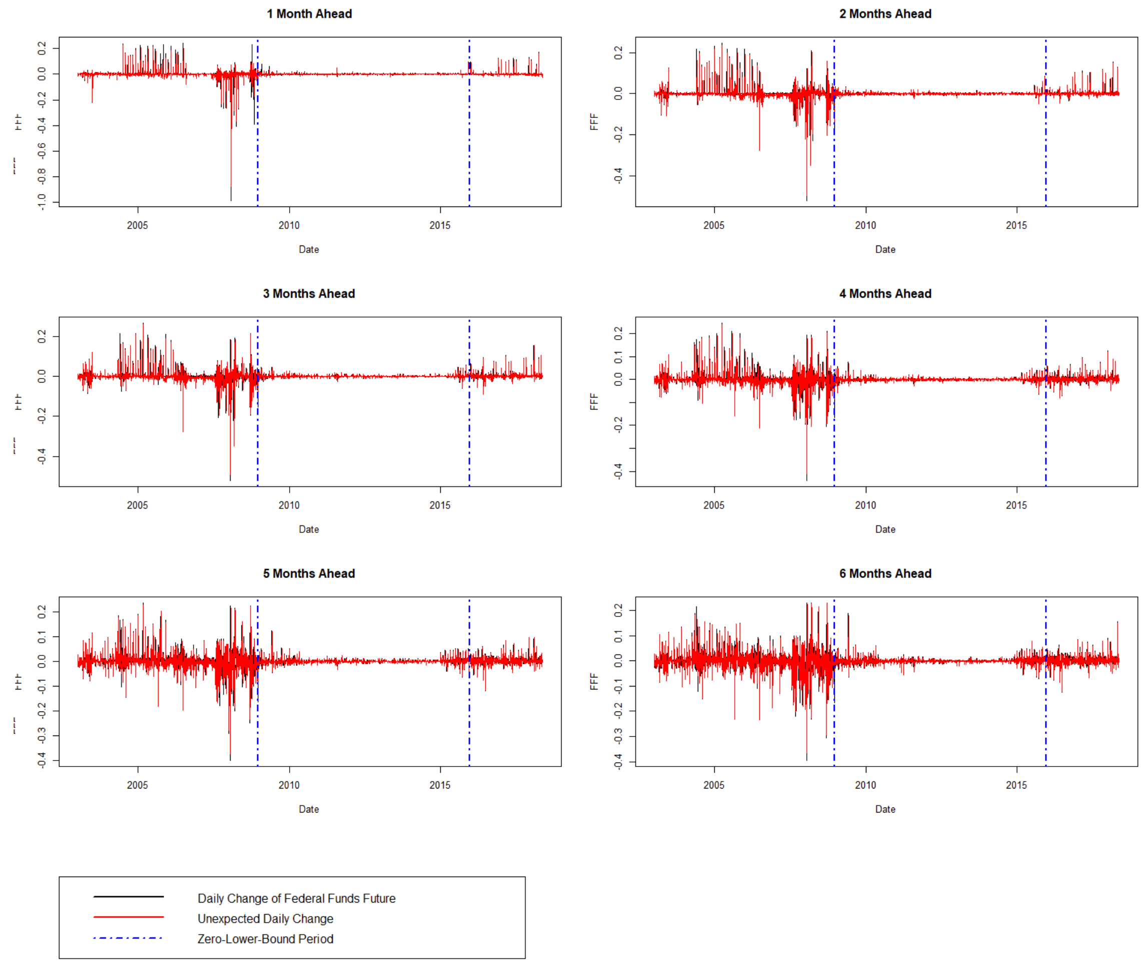
Notes: Same empirical exercises as in Table 1.3 except bandwidth for each horizon is taken as the median of the optimal bandwidths across horizons. See notes for Table 1.3 for more details.

Table 1.7: Robustness Check - Using First Differences of State Variables

	1 Month Ahead	2 Months Ahead	3 Months Ahead	4 Months Ahead	5 Months Ahead	6 Months Ahead
Panel A: Predictability of HF Monetary Shocks by Expected Daily Changes						
Expected Changes in FFF	-0.42	0.71	0.65	0.22	0.82	0.29
	(p =0.11)	(p =0.16)	(p =0.43)	(p =0.71)	(p =0.35)	(p =0.68)
Adjusted R^2	1.81%	0.60%	-0.40%	-0.84%	-0.00%	-0.83%
Signal-to-Noise Ratio	0.03	0.02	0.01	0.00	0.01	0.00
Panel C: Predictability of HF Monetary Shocks by Expected Daily Changes, Conditional on State Variables						
Expected Changes in FFF	-0.48*	0.20	1.36	0.49	1.26	0.68
	(p =0.06)	(p =0.68)	(p =0.14)	(p =0.44)	(p =0.11)	(p =0.40)
Adjusted R^2	7.33%	6.21%	0.09	4.19%	2.84%	0.95%
Signal-to-Noise Ratio	0.14	0.13	0.16	0.11	0.09	0.07
Panel D: Predictability of HF Monetary Shocks by Index of State Variables						
Expected Changes in FFF	0.00	0.01	0.00	0.00	-0.00	-0.00
	(p =0.37)	(p =0.33)	(p =0.61)	(p =0.80)	(p =0.86)	(p =0.93)
Adjusted R^2	-0.30%	-0.15%	-0.74%	-0.88%	-0.91%	-0.93%
Signal-to-Noise Ratio	0.01	0.01	0.00	0.00	0.00	0.00
Panel E: Predictability of Unexpected Changes in FFF, All Days						
State Variables	(1.28,	(2.94,	(1.11,	(2.48,	(0.88,	(0.90,
	p =0.46)	p =0.52)	p =0.52)	p =0.48)	p =0.51)	p =0.57)
Adjusted R^2	0.04%	0.26%	0.01%	0.19%	-0.02%	-0.00
Signal-to-Noise Ratio	0.00	0.00	0.00	0.00	0.00	0.00
Panel F: Predictability of Unexpected Changes in FFF, FOMC Days						
State Variables	(1.01,	(1.75,	(1.16,	(1.37,	(1.73,	(2.06,
	p =0.40)	p =0.60)	p =0.61)	p =0.66)	p =0.46)	p =0.44)
Adjusted R^2	0.06%	3.37%	0.76%	1.70%	3.28%	4.68%
Signal-to-Noise Ratio	0.05	0.09	0.06	0.07	0.08	0.10

Notes: Same empirical exercises as in Table 1.3 using first differences of the state variables used previously as the new set of state variables. See notes for Table 1.3 for more details.

Figure 1.1: First Differences of Federal Funds Futures and Their Unexpected Components



Chapter 2: Unspanned Monetary Shocks in the Yield Curve

I Introduction

Identifying forecastable components in bond risk premia is a central problem in finance and macroeconomics. A predominant view in the literature is that the cross section of the yield curve, which is effectively spanned by its first three principal components, contains all available information necessary to forecast bond yields (Bauer and Hamilton, 2018). A large literature, however, has been looking at macroeconomic variables that predict bond yields beyond information in the yield curve. Among others, Ludvigson and Ng (2009) identify common factors in a large panel of US macroeconomic variables using dynamics factor analysis, and find that these common factors exhibit strong forecasting power on bond risk premia conditional on the Cochrane-Piazzesi (2005) yield curve factor, and account for a large fraction of the business cycle variations in long-term Treasury yields. Joslin, Priebsch, and Singleton (2014) develop an affine term structure model in which real economic activities and inflation drive the market prices of yield curve risk factors. Cieslak and Povala (2015) find that a measure of trend inflation has strong forecasting power on bond excess returns. Variables that exhibit such forecasting power on top of information contained in the yield curve are commonly termed as "unspanned factors".

Whether these macro factors are truly unspanned is still left open to debate. Specifically, Bauer and Hamilton (2018) argue that the overlapping nature of the excess returns coupled with the high persistence of the yield curve factors can lead to severe size distortion. After taking this into account by a bootstrap procedure, they find that these so called unspanned macro factors documented in the literature have little additional forecasting power conditional on the first three principal components of the yield curve.

Given that monetary policy is a major disturbance to the economy, a natural question to ask is whether unexpected news about future monetary policy, which I term as "monetary shocks", is unspanned, and whether its forecasting power is robust to unspanned macro factors. To avoid the econometric complications caused by the overlapping nature of bond excess returns, I instead look into Treasury carry trade strategies that buy a long-maturity bond and borrow from a short-maturity bond by equal amount, and unwind the positions one day later, whose returns are serially uncorrelated by construction. Following the literature on high-frequency identification of monetary shocks pertaining to FOMC announcements (e.g. Kuttner, 2001; Cochrane and Piazzesi, 2002; Hanson and Stein, 2015), I measure monetary shocks as the daily changes in federal funds futures (FFFs).

I find that monetary shocks exhibit both statistically and economically significant forecasting power on Treasury carry trade returns across a variety of maturity pairs on top of information contained in the cross section of bond yields. Morever, such forecasting power can be effectively summarized by a single factor constructed from changes in the first 24 FFFs, which is not specific to each maturity pair. The forecasting power of the single factor is robust to macro factors à la Ludvigson and Ng (2009).

The remainder of this chapter is organized as follows. Section 2 outlines the econometric framework that bypasses the critique of Bauer and Hamilton (2018). Section 3 discusses the data and results. Finally, Section 4 concludes.

II Econometric Framework

The literature on bond risk premia predictability almost entirely focuses on strategies that buy one unit of n-period bond at time t and sell it at time $t+k$ as one unit of $(n-k)$-period bond, meanwhile financing this position by borrowing at the k-period bond yield. This is equivalent to a k-period forward contract on one unit of $(n-k)$-period bond. Denote the time-t price of an n-period zero coupon bond as $P_t^{(n)}$, and the corresponding yield as $y_t^{(n)} \equiv -\frac{1}{n}\log P_t^{(n)}$. The excess return of this strategy is

$$rx_{t+k}^{(n)} \equiv -(n-k)\, y_{t+k}^{(n-1)} + ny_t^{(n)} - ky_t^{(k)} \tag{2.1}$$

which I shall refer to as "bond excess return". The spanning hypothesis asserts that the cross section of yield curve contains all information available for forecasting future bond yields. This is conventionally tested by regressing $rx_{t+k}^{(n)}$ on the suggested variable(s) that may contain additional return-forecasting information while controlling for variables that are supposed to span the cross section of yield curve. An econometric complication, as is argued by Bauer and Hamilton (2018), is that the overlapping nature of the excess return $rx_{t+k}^{(n)}$ coupled with the high persistence of the control variables can result severe finite-sample distortion, leading to spuriously large t-value(s) of the additional variable(s) to be tested as well as spurious increase in R^2. Such distortion may persist even if the variables are cointegrated, and cannot be corrected by HAC estimators that take into account serial correlation in the error term. They thus suggest using a bootstrap procedure for statistical inference.

I instead avoid this problem by looking at the predictability of Treasury "carry trade" returns at daily frequency, which are strategies in the line of, say, buying an n_1-month bond and financing it by borrowing at the n_2-month interest rate with $n_2 < n_1$ on day t, and then unwinding these positions one day later. Let $\Delta \equiv 1/21$ denote the time interval of one trading day, and t be in monthly unit. The realized return is

$$rx_{t+\Delta}^{+n_1/-n_2} \equiv -\left(n_1 - \Delta\right) y_{t+\Delta}^{(n_1-\Delta)} + n_1 y_{t+\Delta}^{(n_1)} + \left(n_2 - \Delta\right) y_{t+\Delta}^{(n_2-\Delta)} - n_2 y_{t+\Delta}^{(n_2)} \tag{2.2}$$

which, unlike the conventional measure $rx_{t+k}^{(n)}$, is nonoverlapping. Note that any excess return in the form of equation (2.1) can be approximated by some nonoverlapping return as in equation (2.2), and thus testing the predictability of daily Treasury carry trades is equivalent to testing that of bond excess returns with much longer holding periods, except now without the econometric complication. For example, the excess return of holding a two-year bond for one year, $rx_{t+12}^{(24)}$, can be replicated by the cumulative return of a carry trade of buying a 24-month bond and selling a 12-month bond on the first day, a carry trade of buying a $(24 - \Delta)$-month bond and selling a $(11 - \Delta)$-month on the second day, etc. such that

$$rx_{t+12}^{(24)} = rx_{t+\Delta}^{+24/-12} + rx_{t+2\Delta}^{+23/-11} + \cdots + rx_{t+12}^{+13/-1} \tag{2.3}$$

$$\simeq \sum_{s=1}^{21} rx_{(t+1)+(s-1)\Delta}^{+18/-6} \tag{2.4}$$

The approximation (2.4) is a result of the fact that yields of adjacent maturities are similar so that on average the cumulative return of the long legs of the terms on the RHS of equation (2.3) is equal to the 12-month cumulative return of repeatedly holding a 18-month bond for one day, which also applies to the short legs. I look at five maturity pairs with (n_1, n_2) = (12-month, 6-month), (18-month, 12-month), (24-month, 18-month), (30-month, 24-month) or (36-month, 30-month). Forecasts of the returns of any other maturity pairs in-between can be conveniently inferred from these these five benchmark cases. For example, forecast of the returns at the (18-month, 6-month) horizon is the sum of the forecasts of the (12-month, 6-month) and (18-month, 12-month) returns.

Let X_t be the vector of state variables spanning the cross section of the yield curve, which is to be specified in the following, and $\Delta f_t^{(n)}$ be the daily change in the n-th FFF from $t - \Delta$ to t. I take $\Delta f_t^{(n)}$ as the measure of monetary shocks and run the regression

$$rx_{t+\Delta}^{+n_1/-n_2} = \alpha^{+n_1/-n_2} + \beta^{+n_1/-n_2}\Delta F_t + \gamma^{+n_1/-n_2} X_t + \epsilon_{t+\Delta}^{+n_1/-n_2} \tag{2.5}$$

for each maturity pair $(n_1,\ n_2)$, where $\Delta F_t \equiv \begin{pmatrix} \Delta f_t^{(1)} & \dots & \Delta f_t^{(24)} \end{pmatrix}'$. The null hypothesis that monetary shocks have no forecasting power on bond yields on top of the information already priced in the current yield curve is thus $H_0 :\ \beta^{+n_1/-n_2} = 0$. Under the null hypothesis, all the serial correlation in $rx_{t+\Delta}^{+n_1/-n_2}$ net of measurement errors is captured by the time series behavior of X_t. Morever, the fact that $rx_{t+\Delta}^{+n_1/-n_2}$ is nonoverlapping means that the measurement errors do not accumulate over time. Thus, the error term $\epsilon_{t+\Delta}^{+n_1/-n_2}$ is serially uncorrelated under the null hypothesis.

The set of control variables X_t is commonly taken as the first three principal components of bond yields, which are termed as the level, slope and curvature factors (e.g. Bauer and Hamilton, 2018). It is widely documented that they almost entirely explain the cross sectional variations of the yield curve (e.g. Litterman and Scheinkman, 1991). This naturally raises the critique that the additional forecasting power found within the new variable(s) may merely come from the principal components that are left out. I address this critique by taking X_t as the collection of five out of the ten bond yields used to construct the returns of the five maturity pairs $(n_1,\ n_2)$, namely the 6-month, 12-month, 18-

month, 24-month, 30-month and 36-month bond yields. The $(6 - \Delta)$-month yield is very similar to the 6-month yield and is thus excluded. The same applies to the $(12 - \Delta)$-month, $(18 - \Delta)$-month, $(24 - \Delta)$-month, $(30 - \Delta)$-month and $(36 - \Delta)$-month yields.

III Data and Results

I use the US Treasury yield dataset assembled by Gürkaynak, Sack, and Wright (2007), which is published and updated daily on the New York Fed website. They fit the daily yield curve using the Svensson (1994) method and report the estimated parameters for each day. I use these parameters to interpolate the yields needed to construct the daily returns of Treasury carry trades. The FFF data for the first 24 months are downloaded from Bloomberg.[1] To form a balanced panel, I take December 8, 2004 as the start of the sample. The sample ends on May 18, 2018.

Figure 2.1 plots the time series of the five carry trade returns. As expected, they are more tranquil in the zero-lower-bound (ZLB) period due to the expansionary monetary policy implemented at an immense scale, making them easier to predict. As a result, wherever possible I repeat the empirical exercises in the following over the ZLB period, non-ZLB period and full sample, respectively, for robustness check.

III.I Unspanned Monetary Shocks in the Yield Curve

Table 2.1 summarizes the results for regressions (2.5), where Wald test is conducted to test against the null hypotheis that $\beta^{+n_1/-n_2} = 0$ for each maturity pair (n_1, n_2). The null hypothese are rejected at the 1% level across all maturity pairs and in all three different sample periods, except the +30-month/-24-month and +36-month/-30-month horizons in

[1] The 25th to 36th FFF contracts were not available for trading until February 28, 2011. They are not included in the dataset given the very short sample period they would otherwise result.

the ZLB period. In the two exceptions, however, the results are still statistical significant at the 5% and 10% levels, respectively. The results are not only statistically significant, but also economically significant. In the full sample, the adjusted R^2s for the +12-month/-6-month and +18-month/-12-month horizons increase for more than 1% from 2.89% and 1.77%, respectively, compared with the baseline model including only the control variables as independent variables. They also increase for about 50% of their baseline values for the +24-month/-18-month and +30-month/-24-month horizons. In the ZLB period, the adjusted R^2s are more than doubled for all but the +30-month/-24-month horizon compared with the baseline model. For the +30-month/-24-month horizon, it is nonetheless almost doubled from 0.71% to 1.36%. In the non-ZLB period, the adjusted R^2s increase for about 1.7% from 2.16%-3.02% for the +12-month/-6-month, +18-month/-12-month and +24-month/-18-month horizons, for about 1.4% from 2.37% for the +30-month/-24-month horizon, and for about 0.9% from 2.39% for the +36-month/-30-month horizon. These are strong evidences that monetary shocks as identified from FFFs are unspanned factors, and have strong influences on the variations of the yield curve.

Figure 2.2 plots the values of the slope coefficients $\beta^{+n_1/-n_2}$ for each maturity pair (n_1, n_2). While there is some heterogeneity at the short end in full sample and non-ZLB period, and at the 6th to 7th horizons as well as the long end in the ZLB period, the slope coefficients $\beta^{+n_1/-n_2}$ are largely proportional across maturity pairs. This suggests that there is a single factor in the cross section of monetary shocks, which counts for a large proportion of the additional forecasting power of monetary shocks for each maturity pair. Following Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009), I construct such a single return-forecasting factor, MP_t, by regressing the average of the returns of different maturity pairs, $\bar{rx}_t \equiv \frac{1}{5}\sum_{(n_1,n_2)} rx_t^{+n_1/-n_2}$, on the same set of independent variables in model (2.5), and taking the proportion of the fitted value due to monetary shocks as the new factor. I.e.

$$\bar{rx}_{t+\Delta} = \bar{\alpha} + \bar{\beta}\Delta F_t + \bar{\gamma} X_t + \bar{\epsilon}_{t+\Delta} \tag{2.6}$$

$$MP_t \equiv \hat{\bar{\beta}}\Delta F_t \tag{2.7}$$

where $\hat{\bar{\beta}}$ denotes the OLS estimate of $\bar{\beta}$. To evaluate the empirical performance of MP_t, I run predictive regression

$$rx_{t+\Delta}^{+n_1/-n_2} = \alpha_{MP}^{+n_1/-n_2} + \beta_{MP}^{+n_1/-n_2} MP_t + \gamma_{MP}^{+n_1/-n_2} X_t + \epsilon_{MP,\,t+\Delta}^{+n_1/-n_2} \tag{2.8}$$

The results are reported in the columns named "Model 2" in Table 2.1. The slope coefficients $\beta_{MP}^{+n_1/-n_2}$ are significant at the 1% level for all maturity pairs and in all the three different sample periods. The adjusted R^2s see small increases compared with the results for model (2.5) taking individual monetary shocks as predictive variables. These confirm the conjecture that the joint forecasting power of ΔF_t is effectively summarized in a single factor.

This, however, does not imply that the entirety of the forecasting power of ΔF_t is captured by MP_t, which amounts to the hypothesis that the matrix of slope coefficients for model (2.5) across maturity pairs, $\beta \equiv \begin{pmatrix} \beta^{+12/\text{-}6} & \dots & \beta^{+36/\text{-}30} \end{pmatrix}$, is of rank one, which corresponds to the collection of linear constraints (e.g. Hansen and Hodrick, 1983; Campbell, 1987; Cochrane and Piazzesi, 2005)

$$\frac{\beta_{1,1}}{\beta_{i,1}} = \frac{\beta_{1,2}}{\beta_{i,2}} = \cdots = \frac{\beta_{1,24}}{\beta_{i,24}}, \quad \text{for } i = 1, \dots, 5 \tag{2.9}$$

where $\beta_{i,j}$ denotes the (i, j)-th element of β. Allowing both cross-sectional correlation and heteroskedasticity between equations, I test the null hypothesis (2.9) using Wald statistic in each sample period, which is asymptotically Chi-squared distributed with 92 degrees of freedom. As is shown in Panel F of Table 2.1, the resulting Wald statistics all have p-values being effectively zero across the three sample periods.

III.II Robustness to Unspanned Macro Factors

A large class of unspanned factors documented in the literature is macro variables (e.g. Ludvigson and Ng, 2009; Joslin, Priebsch, and Singleton, 2014). Given the strong impacts of monetary policy on the economy, it is quite possible that there is significant amount of overlap between the monetary factor MP_t and macro factors. On the other hand, monetary policy affects the economy with some lag, suggesting that there may be sizeable information in MP_t that has not yet been incorporated by macro variables at time t. The objective of this section is to test whether MP_t contains extra unspanned return-forecasting information on top of macro factors.

I choose the set of factors studied by Ludvigson and Ng (2009) as it nests macro variables studied in other research. The set of macro variables in Ludvigson and Ng (2009) are constructed from the FRED-MD dataset, which includes time series of 134 macro variables of the US economy at monthly frequency. Following their procedure, I first estimate a factor model for the entire panel using principal component analysis, and determine the appropriate number of factors to include using the information criterion developed by Bai and Ng (2002). Denote the set of factors chosen by the information criterion as LN, and the vector containing the j-th order of each element in LN as LN^j. For each return process, say $rx_t^{+n_1/-n_2}$, I conduct forward variable selection based on BIC criterion for the regression at monthly frequency

$$rx^{+n_1/-n_2} = \phi^{+n_1/-n_2} + LN\theta_1^{+n_1/-n_2} + LN^2\theta_2^{+n_1/-n_2} + LN^3\theta_3^{+n_1/-n_2} + \epsilon^{+n_1/-n_2} \tag{2.10}$$

where the monthly values of $rx_t^{+n_1/-n_2}$ are taken as its value on the last trading day of each month, and the higher-order terms are included to capture potential nonlinearity in the relationship. I denote the replicated factors chosen from LN, LN^2 and LN^3 by the variable selection procedure collectively as $LN^{+n_1/-n_2}$.

Given the relatively short time span of my dataset, simply rolling all daily variables to monthly frequency may leave too few observations to conduct meaningful statistical inferences. I instead adopt the mixed frequency Granger causality test à la Ghysels, Hills, and Motegi (2016), which takes the values of MP over the last k trading days within month m, $\tilde{M}P_m = \begin{pmatrix} MP_m & MP_{m-\Delta} & \dots & MP_{m-(k-1)\Delta} \end{pmatrix}'$, into the monthly regression

$$rx_{m+1}^{+n_1/-n_2} = \tilde{\alpha}_{MP}^{+n_1/-n_2} + \tilde{\beta}_{MP}^{+n_1/-n_2}\tilde{M}P_m + \tilde{\gamma}_{MP}^{+n_1/-n_2}X_m + \tilde{\theta}_{MP}^{+n_1/-n_2}LN_m^{+n_1/-n_2} + \tilde{\epsilon}_{MP,m+1}^{+n_1/-n_2} \tag{2.11}$$

The null hypothesis is $H_0: \ \tilde{\beta}_{MP}^{+n_1/-n_2} = 0$, which can be examined by Wald test. I take $k = 15$ (i.e. three trading weeks). Compared with rolling all variables to lower frequency and then running OLS, this test also utilitzes high-frequency information and is less vulnerable to misspecification.

The results are summarized in Table 2.2. Across all maturity pairs $+n_1/-n_2$, the Wald statistics for the joint significance of the 15 daily lags of the single monetary variable are significant at the 1% level.

IV Conclusions

By focusing on the return-forecastability of Treasury carry trade returns, I resolve the econometric complication caused by the overlapping nature of bond excess returns, a problem that underlies most of the literature on bond return predictability. I find that monetary shocks, which are proxied by daily changes in FFFs, have additional forecasting power on bond yields on top of information contained in the cross section of the yield curve. Such forecasting power is effectively captured by a single factor, and is robust to unspanned macro factors.

The fact that such a single monetary factor is unspanned by the cross section of Treasury yields may or may not be a result of irrational behaviors of investors. It is possible that investors are aware of this monetary policy factor, but do not wish to be exposed to it. If the disutility of exposure to this factor exactly cancels out the utility of exploiting it to better forecast bond returns, then it is not priced in the Treasury market. In other words, while the monetary policy factor does not Granger cause the yield curve factors in the risk-neutral measure, it does Granger cause them in the physical measure. A setting like this can be effectively realized in the reduced-form affine term structure model developed by Joslin, Priebsch, and Singleton (2014).

While the literature in the past 15-20 years has been adopting high-frequency changes in FFFs as the standard measure of monetary shocks, one potential problem is that these changes may have forecastable components, making them endogeneous and thus invalid measure. In fact, in Chapter 1, I find that changes in the first six FFFs in a 30-minute window bracketing each FOMC announcement are strongly forecastable by information in the yield curve, Treasury yield volatility and real business conditions. Subsequent research may reexamine the findings documented in this research using alternative measure of monetary shocks that do not contain such endogeneous information.

Tables and Figures

Table 2.1: Predictability of Treasury Carry Trade Returns

	Full Sample			ZLB Period			Non-ZLB Period		
	Baseline	Model 1	Model 2	Baseline	Model 1	Model 2	Baseline	Model 1	Model 2
Panel A: +12-Month/-6-Month Horizon									
ΔF_t		(6.16)***			(3.28)***			(3.71)***	
ΔMP_t			0.99***			0.56***			0.96***
			(0.13)			(0.08)			(0.17)
Adjusted R^2	2.89%	4.35%	4.48%	1.31%	4.21%	4.01%	3.02%	4.71%	4.86%
Panel B: +18-Month/-12-Month Horizon									
ΔF_t		(4.24)***			(2.66)***			(3.20)***	
ΔMP_t			1.16***			0.79***			1.50***
			(0.17)			(0.12)			(0.20)
Adjusted R^2	1.77%	2.86%	3.19%	1.23%	3.00%	3.48%	2.16%	4.01%	4.22%
Panel C: +24-Month/-18-Month Horizon									
ΔF_t		(3.59)***			(2.07)***			(3.19)***	
ΔMP_t			1.89***			0.94***			1.19***
			(0.19)			(0.15)			(0.21)
Adjusted R^2	1.46%	2.26%	2.64%	0.95%	2.02%	2.99%	2.26%	4.00%	4.26%
Panel D: +30-Month/-24-Month Horizon									
ΔF_t		(3.05)***			(1.67)**			(3.00)***	
ΔMP_t			1.38***			1.03***			1.15***
			(0.20)			(0.17)			(0.21)
Adjusted R^2	1.28%	1.75%	2.24%	0.71%	1.36%	2.64%	2.37%	3.73%	4.25%
Panel E: +36-Month/-30-Month Horizon									
ΔF_t		(2.62)***			(1.46)*			(2.68)***	
ΔMP_t			1.06***			1.09***			1.08***
			(0.21)			(0.19)			(0.20)
Adjusted R^2	1.19%	1.35%	1.96%	0.06%	0.98%	2.42%	2.39%	3.28%	4.10%

	Full Sample	ZLB Period	Non-ZLB Period
Panel F: Test on Single Factor Restriction across Horizons			
Wald Statistic	(428.71)***	(5732.28)***	(1861.06)***
Degree of Freedom	92	92	92

Notes: Panels A-E - predictive regressions of daily Treasury carry trade returns on the collection of daily changes in the first 24 federal funds futures, ΔF_t, or on the single monetary factor ΔMP_t, conditional on the 6-month, 12-month, 18-month, 24-month, 30-month and 36-month yields. For each test on the joint significance of ΔF_t, the Wald statistic is reported in brackets, which has 24 degrees of freedom. For each test on the significance of ΔMP_t, point estimate of the slope coefficient is reported in the first line, whereas the corresponding standard error is reported in brackets in the second line. Panel F - Wald tests on the restriction that the slope coefficients of ΔF_t are proportional across horizons. The Wald statistics are reported in brackets.

"***": Significance at 1% level; "**": Significance at 5% level; "*": Significance at 10% level.

Table 2.2: Robustness of Single Monetary Factor to Macro Factors, Full Sample

	+12-Month/-6-Month	+18-Month/-12-Month	+24-Month/-18-Month	+30-Month/-24-Month	+36-Month/-30-Month
Wald Statistic	(5.34)***	(3.90)***	(3.35)***	(2.91)***	(2.58)***

Note: predictive regressions of end-of-month Treasury carry trade returns on the first 15 daily lags of monetary factor, conditional on the Ludvigson-Ng (2009) macro factors and the 6-month, 12-month, 18-month, 24-month, 30-month and 36-month yields. For each maturity pair, Wald test on the joint significance of the lags of monetary factor is conducted. The degree of freedom is 15. This procedure is in line with the mixed-frequency Granger causality test in Ghysels, Hills, and Motegi (2016).

"***": 1% significant; "**": 5% significant; "*": 10% significant.

Figure 2.1: Time Series of Treasury Carry Trade Returns

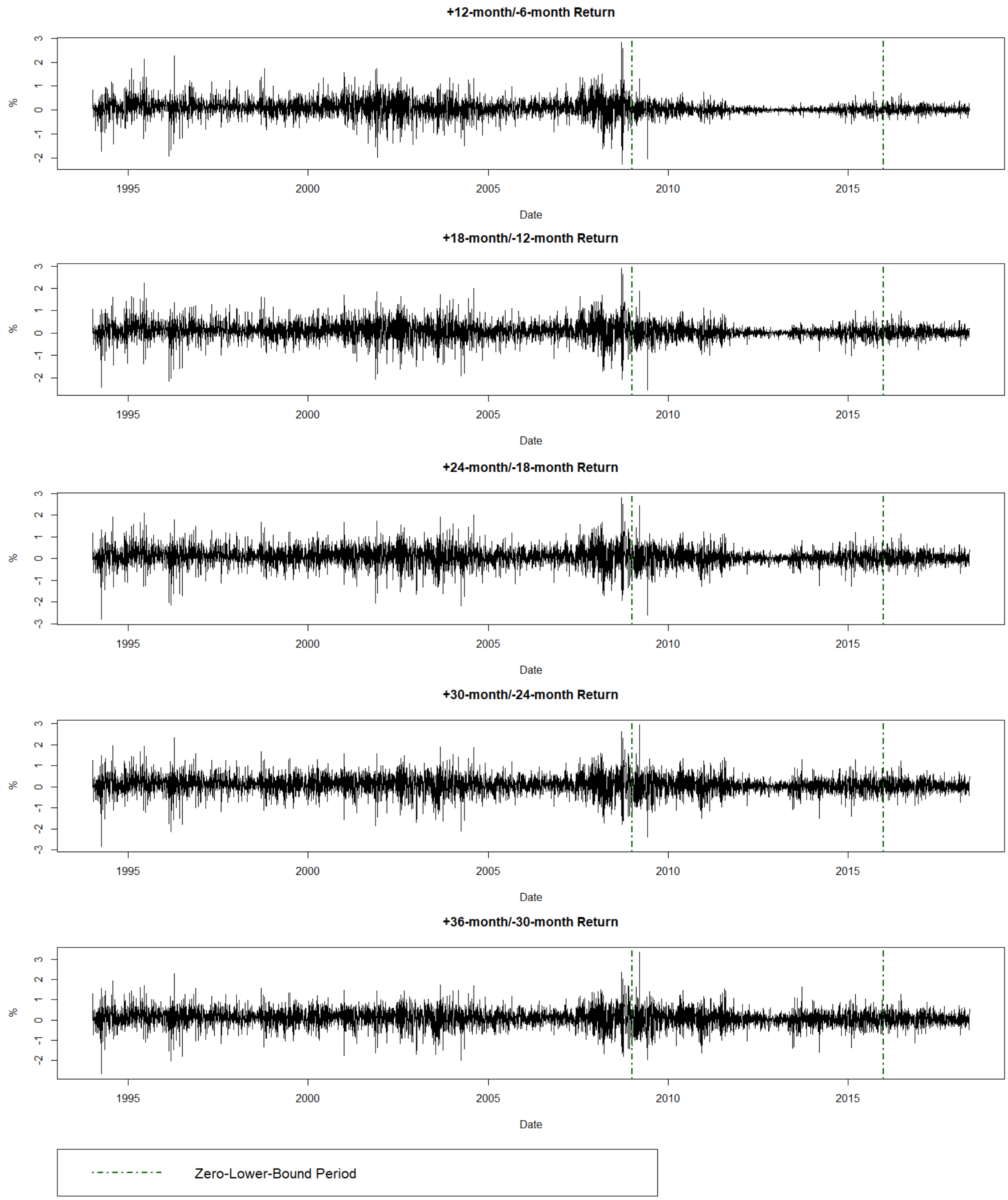
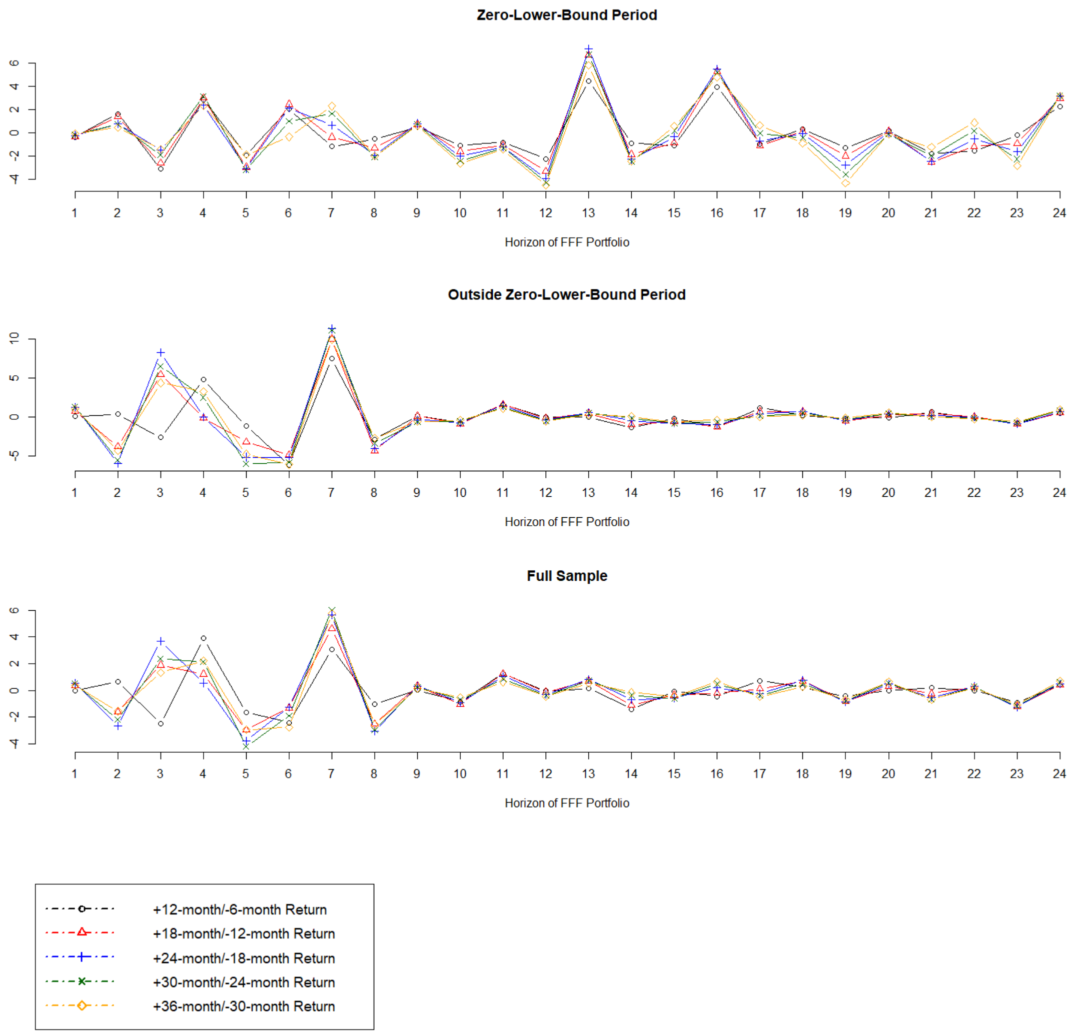


Figure 2.2: Slope Coefficients of Monetary Shocks across Horizons



Chapter 3: Macro Factors in Currency Risk Premia and Cross-Country Differentials in Risk Pricing

I Introduction

Currency trading strategies pursued by practitioners, such as carry trade and momentum, typically involve exploiting deviations from the well-known uncovered interest rate parity (UIP). Such strategies can generate sizeable profit but are also subject to significant downside risk under periods of market stress.

While a large body of the literature has been focusing on explaining the occasional crashes of carry trades, a phenomenon termed as the "peso problem" (e.g. Burnside, Eichenbaum, and Rebelo, 2011), and on identifying risk factors determining the cross section of currency returns (e.g. Ang and Chen, 2010; Lustig, Roussanov, and Verdelhan, 2011; Menkhoff, Sarno, Schmeling, and Schrimpf, 2012), questions that have not been sufficiently addressed are their time series predictability, as well as the underlying mechanism driving such predictability. I investigate these problems through the lense of the risk premia of zero-net investment strategies involving the US dollar and a foreign currency. Predictability of carry trade returns involving, say, the Australian Dollar and Japanese Yen in the long and the short legs, respectively, can be inferred from the difference between their respective risk premia relative to the US dollar. Specifically, I study the predictive power of macro factors on currency risk premia.

It is shown under some minimalistic assumptions that currency risk premium is predictable by a risk factor only if the domestic and the foreign economies assign different market prices to that same underlying source of risk. This interesting connection is the motivation behind my focus on the predictability of currency risk premia. To incorporate as much information about the global economy as possible, and meanwhile to minimize the impact of measurement errors in individual variables and to avoid running into the curse of dimensionality, I adopt dynamic factor analysis to effectively summarize the information contained in a large number of time series variables by a small number of factors. The method is applied to a dataset consisting of 129 macroeconomic variables of the US economy at monthly frequency. Foreign variables that mimic the economic interpretations of these US factors are then chosen to capture the foreign macroeconomic conditions. For each currency and horizon, a variable selection scheme is conducted on some transformations of these variables, and then the resulting set of variables being chosen is used as regressors in the predictive regression. To guard the results against finite sample distortions as illustrated by Bauer and Hamilton (2018), a semiparametric bootstrap procedure is conducted for statistical inferences.

Several interesting phenomena are observed. First, currency risk premia exhibit strong long-horizon predictability. For each currency, as the maturity increases from 1 month to 12 months, more predictors are selected and are significant at the 5% level or above, and the predictors that are selected and significant at shorter maturities tend to stay selected and significant at longer maturities. This suggests that the pricing differentials of macro risks become more prevalent at longer horizons. Moreover, the adjusted R^2s increase dramatically from 0-4% at 1-month maturity to 13-29% at 12-month maturity, suggesting that a large proportion of the variation in currency risk premia at longer maturities are driven by macro risks. Second, conditional on the US factors being selected, foreign variables seldom contribute to the predictability of currency risk premia, suggesting that much of the return-forecasting information in the foreign variables is already incorporated in the US factors.

The single most important predictor for currency risk premia is a US factor that loads heavily on nominal interest rates, yield spreads, and measures of housing market activities. It is significant at the 1% level at the 6-month and 12-month maturities across all currencies, and at the 5% significance level or above at the 2-month and 3-month maturities for 4 currencies. Its predictive power is also nonlinear in that its higher-order transformations are also selected and significant across various currencies and maturities. This is striking in that no previous research has documented such a single source of risk that exhibits uniform forecasting power across all currencies.

The strong in-sample predictability is preserved out-of-sample. To assess the out-of-sample predictive power of the factors selected in-sample, the factor loadings and the set of factors being selected for each currency and maturity are held the same as those in the in-sample analyses, whereas the parameters in the predictive regressions are recursively estimated. It it found that in 36 out of the 45 cases across different currencies and maturities, the regression models generate lower mean squared prediction errors than the baseline models that predict returns using their historic means. A formal statistical test of the out-of-sample performance using the ENC-T statistic as in Clark and McCracken (2001) is conducted and is significant at the 5% level or above in all but only 4 cases at the 3-month, 6-month and 12-month maturities across different foreign currencies.

Their out-of-sample predictive power becomes even stronger when estimations of the factor loadings and the varaible selection procedures are also conducted recursively, which mimics the real-time implementation of these strategies. it is found that in 42 out of the 45 cases the regression models generate lower mean squared prediction errors, and the ENC-T statistics are significant at the 5% level or above in all but only 3 cases at the 3-month, 6-month, and 12-month maturities. The reduction in root mean squared prediction errors can be as high as 40% compared with the benchmark specifications.

While the literature devoted to currency returns is huge, this research differs from others in either objective or methodology. While Ang and Chen (2010), Lustig, Roussanov and Verdelhan (2011), and Menkhoff, Sarno, Schmeling, and Schrimpf (2012) put emphasis on the cross section of carry trade returns, this research focuses on the time series dimension. Brunnermeier, Nagel, and Pedersen (2009) find that a higher level of VIX predicts higher return of the long leg and lower return of the short leg of a carry trade. I also include VIX as a potential predictor in the dataset but do not put special emphasis on it. Instead, I let the model decide which variables are more important. While Jordà and Taylor (2012) tackle the same problem by focusing on exchange rate predictability in a vector error correction model (VECM), this research focuses on the predictability of currency risk premia and the role of macro factors. While Bakshi and Panayotov (2013) study the predictability of carry trade using three variables that capture the global financial conditions, this research focuses on predictability of individual currencies and the roles of macroeconomic variables. The method of dynamic factor analysis has previously been used by Ludvigson and Ng (2007), and Ludvigson and Ng (2009) to study the predictability of equity and bond risk premia, respectively.

The remainder of this chapter is organized as follows. Section 2 fixates terminologies and notations, and develops a theoretic framework demonstrating how currency risk premia emerge. Section 3 outlines the econometric framework for in-sample and out-of-sample analyses. Section 4 presents and discusses the empirical results. Finally, Section 5 concludes.

II Economic Origin of Currency Risk Premia

I take US Dollar (USD) as the domestic currency and study the risk premia of nine commonly traded currencies, namely Australian Dollar (AUD), Canadian Dollar (CAD), Swiss Franc (CHF), Deutsche Mark (DEM), British Pound (GBP), Japanese Yen (JPY), New Zealand Dollar (NZD), Norwegian Krone (NOK) and Swedish Krona (SEK). Carry trade on a pair of currencies purchases 1-USD worth of a high interest rate currency, say AUD, to earn the *h*-period risk-free rate in Australia, and finances this position by selling 1-USD worth of a low interest rate currency, say JPY, at the cost of the *h*-period risk free rate in Japan. The return of this strategy is

$$R_{t,t+h}^{+AUD/-JPY} = R_{t,t+h}^{+AUD/-USD} - R_{t,t+h}^{+JPY/-USD} \tag{3.1}$$

where the plus (minus) sign in the superscript indicates which currency to buy (sell). I shall formulate separate forecast of each term on the RHS, from which the return of the overall strategy can be inferred. The return of such a strategy whose short leg involves the domestic currency, USD, is

$$R_{t,t+h}^{+FCU/-USD} = \frac{S_{t+h}^{USD/FCU}}{S_t^{USD/FCU}} R_{t,t+h}^{f,FCU} - R_{t,t+h}^{f} \tag{3.2}$$

where the gross return of the domestic zero-coupon bond with maturity h is denoted as $R_{t,t+h}^{f}$, that of the foreign zero-coupon bond of maturity h as $R_{t,t+h}^{f,FCU}$, and the spot exchange rate, i.e. the dollar value of one foreign currency unit (FCU), as $S_t^{USD/FCU}$.

Alternatively, this strategy can be implemented in the currency forward market. Let $F_{t,t+h}^{USD/i}$ denote the forward exchange rate of currency i, i.e. the exchange rate at which the two parties in the contract agree to trade the underlying currency i at time $t+h$. Using the covered interest rate parity (CIP),

$$\frac{F_{t,t+h}^{USD/i}}{S_t^{USD/i}} R_{t,t+h}^{f,i} = R_{t,t+h}^{f} \tag{3.3}$$

it is straightforward to verify that the return (2) is porportional to that of a strategy that enters one unit of the forward contract $F_{t,t+h}^{USD/i}$:

$$\tilde{R}_{t,t+h}^{+i/-USD} = \frac{S_{t+h}^{USD/i}}{F_{t,t+h}^{USD/i}} = \frac{S_{t+h}^{USD/i}}{S_t^{USD/i}} \frac{R_{t,t+h}^{f,i}}{R_{t,t+h}^{f}} \tag{3.4}$$

I define the corresponding yield as

$$rx_{t,t+h}^{i} \equiv \frac{1}{h} \log\left(\tilde{R}_{t,t+h}^{+i/-USD}\right) = \frac{1}{h} \log\left(\frac{S_{t+h}^{USD/i}}{S_t^{USD/i}} \frac{R_{t,t+h}^{f,i}}{R_{t,t+h}^{f}}\right) \tag{3.5}$$

What determine the predictability of currency risk premia? Consider a simple model featuring two countries, home and foreign. To simplify notations, for now I drop the superscript of the spot exchange rate S_t, which is supposed to denote the dollar value of one unit of the foreign currency. Suppose that the world economy is driven by a vector of M state variables X_t, which follows some generic diffusion process under the physical measure $\mathbb{P}$. I.e.

$$dX_t = \gamma(t, X_t)dt + \eta(t, X_t)dB \tag{3.6}$$

where B is a vector of $M \times 1$ independent Brownian motions that capture all sources of uncertainty in the global economy, and the matrix $\eta(t, X_t)$ has full rank. Denote the $\mathbb{P}$-

dynamics or the domestic riskless asset as

$$\frac{dR_t^f}{R_t^f} = r(t, X_t)dt \tag{3.7}$$

where $r(t, X_t)$ is the short rate, then no-arbitrage condition implies that the stochastic discount factor (SDF) of the domestic economy takes the form

$$\frac{dM_t}{M_t} = -r(t, X_t)dt - \theta(t, X_t)'dB \tag{3.8}$$

where $\theta(t, X_t)$ is the $M \times 1$ vector of market prices of risk endowed in the state variables X_t. These assumptions are minimal in that no specific restrictions on what exact processes the state variables and the risk-free rate have to follow are placed.

The equivalent martingale measure of home, $\mathbb{Q}$, is defined by the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}} \equiv M_t R_t^f$. Girsanov theorem implies that the dynamics of the state variables under the $\mathbb{Q}$ measure is

$$dX_t = \left[\gamma(t, X_t) - \eta(t, X_t)\theta(t, X_t)\right] dt + \eta(t, X_t)dB \tag{3.9}$$

The following expression of the market prices of risk then follows immediately:

$$\theta(t, X_t) = \eta(t, X_t)^{-1} \left[E_t^{\mathbb{P}}\left(dX_t\right) - E_t^{\mathbb{Q}}\left(dX_t\right)\right] \tag{3.10}$$

where $E_t^{\mathbb{P}}\left(\cdot\right)$ denotes the expectation under the $\mathbb{P}$ measure conditional on all information up to time t, and $E_t^{\mathbb{Q}}\left(\cdot\right)$ denotes that under the domestic $\mathbb{Q}$ measure.

Similarly, suppose that the foreign SDF follows the diffusion process

$$\frac{dM_t^*}{M_t^*} = -r^*(t, X_t)dt - \theta^*(t, X_t)' dB \tag{3.11}$$

where the foreign market prices of risk $\theta^*(t, X_t)$ are expressed as

$$\theta^*(t, X_t) = \eta(t, X_t)^{-1} \left[E_t^{\mathbb{P}} (dX_t) - E_t^{\mathbb{Q}*} (dX_t) \right] \tag{3.12}$$

and $E_t^{\mathbb{Q}*}(\cdot)$ denotes the conditional expectation under the foreign $\mathbb{Q}$ measure.

Backus, Foresi, and Telmer (2001) first establish the result that in discrete time, the assumption of market completeness implies that currency appreciation risk is equal to the ratio of the underlying foreign and domestic SDFs. The continuous-time analog under the setup presented here is

$$S_t = \frac{M_t^*}{M_t} \tag{3.13}$$

To see this, consider an asset that is traded both within both countries and across the border. Denote its price denominated in the foreign currency as P_t^*, then its price denominated in the domestic currency is $P_t = S_t P_t^*$. No-arbitrage condition requires that $M_t P_t$ and $M_t^* R_t^*$ are $\mathbb{P}$-martingales, which implies that

$$E^{\mathbb{P}} \left[\frac{M_{t+dt}}{M_t} \frac{P_{t+dt}}{P_t} \right] = 1 = E^{\mathbb{P}} \left[\frac{M_{t+dt}^*}{M_t^*} \frac{S_t}{S_{t+dt}} \frac{P_{t+dt}}{P_t} \right] \tag{3.14}$$

Note that given any feasible domestic SDF M_t, the foreign SDF M_t^* as defined by equation (13) obviously satisfies (14). If the markets are dynamically complete, then the domestic and the foreign SDFs are unique. Since equation (14) holds for any assets traded across the border, equation (13) is the unique solution.

The potential controversy is of course whether the markets really are dynamically complete. While this assumption is appealing in that it simplifies many potential complications and is widely adopted in the literature, in some occasions it is not fully in line with the empirical findings. In fact, Backus, Foresi, and Telmer (2001) find that under the assumption of market completeness, their model has the difficulty in accounting for the empirical characteristics that the exchange rates are more volatile than the interest rates. Anderson, Hammond, and Ramezani (2010) show that in addition to the canonical assumptions of affine term structure models of interest rates (e.g. Dai and Singleton, 2000) that the state variables evolve according to an affine diffusion under the equivalent martingale measure $\mathbb{Q}$ and that the short rate is an affine function of the state variables, if in addition the log of the exchange rate is an affine function of the state variables, then equation (13) still holds. In the setup presented here, suppose that the exchange rate follows some diffusion process under the $\mathbb{P}$ measure:

$$\frac{dS_t}{S_t} = \mu(t, X_t)dt + \sigma(t, X_t)dB \tag{3.15}$$

this set of assumptions translates to that the drift and diffusion terms of the state variables and the exchange rate, $\gamma(t, X_t)$, $\eta(t, X_t)$, $\mu(t, X_t)$ and $\sigma(t, X_t)$, and the domestic and foreign short rates $r(t, X_t)$ and $r^*(t, X_t)$ are all affine functions of the state variables X_t. Equation (13) under either set of assumptions has been widely adpoted in the fixed income and international finance literature.

Given the focus of this research, I assume that equation (13) holds without entering the debate of which set of assumptions is more realistic. The objective here is simply to demonstrate that the risk-pricing-differential interpretation of currency risk premium predictability holds in a very general context, for which it should be noted that both sets of assumptions are quite generic and are applied to a large body of research.

Combining equations (8) and (10)-(13), it is possible to write the conditional expectation of currency yield as in equation (5) as

$$
\begin{aligned}
E_t^{\mathbb{P}}\left[rx_{t,t+h}^*\right] &= \frac{1}{2}\int_t^{t+h} \eta(s, X_s)^{-2} \left(\left[E_t^{\mathbb{P}}(X_s) - E_t^{\mathbb{Q}}(X_s)\right]' \left[E_t^{\mathbb{P}}(X_s) - E_t^{\mathbb{Q}}(X_s)\right] \right. \\
&\quad \left. - \left[E_t^{\mathbb{P}}(X_s) - E_t^{\mathbb{Q}*}(X_s)\right]' \left[E_t^{\mathbb{P}}(X_s) - E_t^{\mathbb{Q}*}(X_s)\right] \right) ds
\end{aligned}
\tag{3.16}
$$

The discrete-time analog is

$$
\begin{aligned}
E_t^{\mathbb{P}}\left[rx_{t,t+h}^*\right] &= \frac{1}{2}\eta(t, X_t)^{-2} \left(\left[E_t^{\mathbb{P}}(X_{t+h}) - E_t^{\mathbb{Q}}(X_{t+h})\right]' \left[E_t^{\mathbb{P}}(X_{t+h}) - E_t^{\mathbb{Q}}(X_{t+h})\right] \right. \\
&\quad \left. - \left[E_t^{\mathbb{P}}(X_{t+h}) - E_t^{\mathbb{Q}*}(X_{t+h})\right]' \left[E_t^{\mathbb{P}}(X_{t+h}) - E_t^{\mathbb{Q}*}(X_{t+h})\right] \right) \\
&= \frac{1}{2}\sum_{m=1}^{M} \left[\theta_m(t, t+h, X_t)^2 - \theta_m^*(t, t+h, X_t)^2\right]
\end{aligned}
\tag{3.17}
$$

where $\theta_m(t, t+h, X_t)$ is the domestic market price of risk of the state variable X_m at time $t+h$. Similarly, $\theta_m^*(t, t+h, X_t)$ denotes that in the foreign economy. Hence, any state variable X_m exhibits predictive power on currency risk premium at time t only if it is priced differently between the two countries, in which case the conditional mean of its future realization under the domestic $\mathbb{Q}$ measure is different from that under the foreign $\mathbb{Q}^*$ measure. If no such variable exists, then currency risk premium is only driven by noise and thus unpredictable. This implies that testing whether a macro variable has the same degree of impact on home and foreign in terms of its shadow prices can be conducted by testing whther it predicts the currency risk premium between the two countries.

The problem here is of course that without further parametric assumptions, it is impossible to tell whether a macro factor is priced differently merely from equation (17). As a result, I undertake a reduced-form approach to tackle this problem, which is outlined in the next section.

III Econometric Framework

III.I In-Sample Predictability

To start with, consider a regression of some currency risk premium, $rx^i_{t,t+h}$, on a set of N macro variables of the domestic and foreign economies denoted as Z_t and Z^i_t, respectively:

$$rx^i_{t,t+h} = \alpha^i + \beta^{i\,\prime} Z_t + \gamma^{i\,\prime} Z^i_t + \epsilon^i_{t+h} \tag{3.18}$$

where $i = 1, \ldots, I$ is the index of foreign currencies. In principal, one would want to include as many predictors as possible in order to take into account more relevant information. But doing it in the naive way in which one simply puts these many predictors into the regression will quickly run into the degrees-of-freedom problem as N increases, and identification will eventually become infeasible as N becomes larger than the sample size T. This is indeed the case in this research, where Z_t consists of 129 US macro variables, and Z^i_t consists of 8 macro variables of foreign country i. Another concern is that macro variables are in general imperfectly measured with a fair amount of noise. To deal with these problems, I undertake the method of dynamic factor analysis by assuming that Z_t has a factor structure of the form:

$$Z_t = \lambda_i' f_t + e_t \tag{3.19}$$

where f_t is an $r \times 1$ vector of latent common factors with $r \ll N$, λ_i is the corresponding $r \times 1$ factor loadings, and e_t is the vector of error terms. The crucial assumption here is $r \ll N$ so that substaintial dimension reduction can be achieved by replacing Z_t with a subset of factors $PC_t \subset f_t$. I follow common practice in the literature by estimating f_t using principal components analysis (PCA) on the panel of Z_t, and choose the optimal number of principal components (PCs) to keep using the panel information criterion developed by Bai and Ng (2002).

It remains to narrow down a subset of factors $\hat{PC}_t$ from the estimated factors $\hat{f}_t$, as well as a subset of foreign macroeconomic variables FPC_t^i from Z_t^i to include in the predictive regression. This step is important in that it further reduces the problem of multicolinearity and overfitting, and in that factors that are pervasive in the panel of Z_t and Z_t^i need not be important predictors of the currency risk premium. According to Stock and Watson (2002), conducting such a variable selection scheme by minimizing the BIC leads to the preferred set of predictors. To come up with such a suitable variable selection scheme, several issues have to be taken care of. First, while by construction the variables in $\hat{f}_t$ are orthogonal to each other, they may still be correlated with the foreign variables Z_t^i. Thus, variable selection should be conducted among the factors in $\hat{f}_t$. Second, while Z_t is a balanced panel, the foreign variables are not and in general span shorter periods of time. Thus, variable selection should be conducted among $\hat{f}_t$ first in order to fully exploit their predictive power. Finally, to capture any possible nonlinear effect, variable selection is also conducted on the second and third order terms of the domestic and foreign factors. The resulting variable selection scheme is as follows:

1. Choose a subset of $\hat{f}_t$ using stepwise variable selection routine such that the BIC value of the regression of $rx_{t,t+h}^i$ on this chosen subset, $\hat{PC}_t$, is mininal;

2. Keeping the set of variables chosen in step 1, $\hat{PC}_t$, choose a subset of Z_t^i in the same way as in step 1;

3. Keeping the set of variables chosen in steps 1 and 2, $\hat{PC}_t$ and FPC_t^i, choose among the second and third order terms of $\hat{PC}_t$ and FPC_t^i in the same way as in step 1.

Denote the chosen sets of second and third order terms of $\hat{PC}_t$ as $\hat{PC}2_t$ and $\hat{PC}3_t$, and those of FPC_t^i as $FPC2_t^i$ and $FPC3_t^i$, respectively. The resulting predictive regression is

$$rx_{t,t+h}^i = \alpha^i + \beta_1^i{}'\hat{PC}_t + \beta_2^i{}'\hat{PC}2_t + \beta_3^i{}'\hat{PC}3_t + \gamma_1^i{}'FPC_t^i + \gamma_2^i{}'FPC2_t^i + \gamma_3^i{}'FPC3_t^i + \epsilon_{t+h}^i \quad (3.20)$$

It is shown in Bai and Ng (2006) that as $N,\ T \to \infty$ with $\sqrt{T}/N \to 0$, the estimated PCs, $\hat{PC}_t$, $\hat{PC}2_t$ and $\hat{PC}3_t$, can be treated as if they are observed in the second-stage regression. In other words, the fact that the factors are estimated does not affect the asymptotic properties of the least square estimates of equation (20). As a result, in the follow I shall slightly abuse the notations by rewriting the estimated factors as PC, $PC2$, and $PC3$ to emphasize that asymptotically they can be regarded as observed.

Note that by the overlapping nature of $rx_{t,t+h}^i$, the error terms ϵ_{t+h} are serially correlated with lag h. I thus use the Newey-West (1994) covariance matrix estimator, where the number of lags is chosen automatically.

III.II Bootstrap Procedure

Bauer and Hamilton (2018) demonstrate in the context of forecasting bond risk premia that when examining the predictive power of a predictive variable on certain asset return, the overlapping nature of the return process coupled with highly persistent control variables can result significant size distortions in small sample, leading to spuriously large increase in R^2 and large t-statistic. As is seen in Appendix, the second macro variable for each foreign country, FPC_2^i, which is constructed as the difference between the 10-year bond yield and the policy rate, is indeed highly persistent. To guard against this problem, I conduct statistical inferences based on the following bootstrap procedure:

1. Estimate the coefficients in equation (20), the corresponding t-statistics constructed from the Newey-West (1994) covariance matrix estimator, and the adjusted R^2 . Store the parameters and the residuals.

2. Estimate a VAR(1) model on PC_t and another VAR(1) model on FPC_t^i. Store the parameters and the residuals.[1]

3. Form a matrix of the residuals from steps 1 and 2, and resample, with replacement, the rows of the matrix. This preserves potential cross-sectional dependencies between the regression residuals and the predictors, as well as possible heteroskedasticity in the regression residuals (e.g. MacKinnon, 2006).

4. Given the k-th bootstrapped sample of the error terms, $k = 1, \ldots, K$, build the bootstrapped series of currency premium under the null hypothesis of no predictability, i.e. $rx_{t,t+h}^{i\,(k)} \equiv \hat{\alpha}^i + \epsilon_{t+h}^{i\,(k)}$.

5. Given the k-th bootstrapped sample of the error terms, $k = 1, \ldots, K$, build the bootstrapped series of PC_t and FPC_t^i using the parameters estimated in step 2.

6. Run the regression as in equation (20) using the k-th bootstrapped series of each variable. Store the resulting t-statistics and the adjusted R^2.

[1]While one might argue that fitting a single VAR(1) model on both PC_t and FPC_t^i for all i could better capture the true data generating process, this would lead to too many variables to estimate given the small sample size. The step implemented here is in fact a compromise.

7. For each parameter in equation (20), calculate the porportion of the absolute values of the bootstrapped t-statistics that exceed the absolute value of the t-statistic estimated in step 1. This is its bootstrapped p-value of the parameter. For the adjusted R^2, its p-value is taken as the the porportion of the bootstrapped adjusted R^2s that exceed the adjusted R^2 estimated in step 1.

I take the number of bootstrap samples $K = 1,000$.

III.III Out-of-Sample Predictability

Another interesting question to address is the out-of-sample predictability of currency risk premia. In the first exercise, I study whether the factors as estimated and selected in-sample exhibit out-of-sample forecasting power. This involves fixing the factor loadings and the set of variables selected in-sample, and recursively fit the predictive regression (20) to forecast the h-period ahead yield. In the second exercise, I also run PCA and variable selection recursively. The first exercise is useful in that under assumption (19), the underlying US factors are most precisely estimated in the full sample, and thus it provides a more convincing answer to their out-of-sample predictive power. The second exercise is useful in that it mimics real-time trading.

Specifically, in the out-of-sample analyses the data of length T is splitted into two parts, where the first part spanning $t = 1, \ldots, R$ is used as the training sample. In the first round of iteration, for the second exercise the PCA is conducted from $t = 1$ to $t = R$, and then the variable selection scheme as well as the estimation of equation (20) is conducted using data up to time $t = R - h$; whereas for the first exercise only the estimation of equation (20) is conducted.[2] The resulting model (20) is then used to forecast the return

[2]Note that the latest return an investor at time R is able to observe is that of the trade initiated at time $R - h$. Hence the subsample for variable selection and estimation of equation (20) has to stop at time $R - h$.

of the trade that is initiated at time R and is to mature at time $R + h$. The training period is then extended for one time unit, and the same procedure is repeated to forecast the the yield of the trade initiated at time $R + 1$ and to mature at time $R + 1 + h$. This routine is repeated until the end of the sample is reached.

The same iterative routine is also conducted on the regression of currency risk premium on a constant:

$$rx^i_{t,t+h} = \kappa^i + u^i_{t+h} \tag{3.21}$$

such that in the j-th out-of-sample iteration, the point estimate of κ^i is simply the historic mean of $rx^i_{t,t+h}$, i.e. $\frac{1}{R+j-h}\left(\sum_{t=1}^{R+j-h} rx^i_{t,t+h}\right)$.

Denote the forecasting errors of the two out-of-sample exercises as $\left\{\tilde{\epsilon}^i_{R+h}, \ldots, \tilde{\epsilon}^i_T\right\}$ and $\left\{\tilde{u}^i_{R+h}, \ldots, \tilde{u}^i_T\right\}$, respectively. Two measures are used to evaluate the out-of-sample predictive performance. The first one simply looks at the ratio between the mean squared prediction error (MSPE) of the unrestricted model (20) and that of the restrictive model (21). I.e.

$$\frac{MSPE_u}{MSPE_r} = \frac{\sum_{t=R+h}^{T}\left(\tilde{\epsilon}^i_t\right)^2}{\sum_{t=R+h}^{T}\left(\tilde{u}^i_t\right)^2} \tag{3.22}$$

where a value less than 1 indicates that the unrestricted model performs better than the restrictive benchmark.

The second one employs the ENC-T test as in Clark and McCracken (2001) for the null hypothesis that the benchmark model encompasses the unrestrictive model with additional predictors, whereas the alternative hypothesis is that the unrestrictive model contains additional information that predicts the risk premium. Let $P \equiv T - R$. It is shown that if $P/R \to 0$ as $P,\, R \to \infty$, then the test statistic ENC-T$\to_d N(0, 1)$.

Some final comments have to be made on what out-of-sample predictability really implies. While a common perception in applied works is that in-sample tests are more prone to spurious predictability than out-of-sample tests, it is argued by Inoue and Kilian (2004), Campbell and Thompson (2007), and Lettau and Ludvigson (2010), among others, that this is generally not true in that the discrepency between in-sample and out-of-sample predictability can simply be the result of bias-variance tradeoff in small sample. Suppose the predictors in the unrestricted model (20) do have predictive power, then the historic mean κ^i is a biased predictor of the yield, which blows up $MSPE_r$. On the other hand, estimation errors of the parameters in the unrestritive model (20) tend to increase the variance of the forecast, which leads to larger $MSPE_u$. Whether $MSPE_r$ is larger or smaller than $MSPE_u$ depends on how much the reduction in bias is relative to the increase in variance. This intuition is formalized in Inoue and Kilian (2004), who demonstrate that in-sample and out-of-sample tests of predictability are asymptotically equally reliable under the null hypothesis of no predictability, where a test is defined to be unreliable if its effective size is larger than its nominal size. As a result, I simply regard out-of-sample predictability as a test of whether it is feasible to implement the predictability discovered in-sample into a trading strategy instead of another test of whether the predictability exists.

IV Data and Results

As is discussed, I focus on the predictability of risk premia of the G10 currencies, where USD is taken as the domestic currency. Euro (EUR) is taken as a proxy of DEM starting from 1999:01 after being converted at the rate 1.95583 DEM/EUR. I implement the strategies using currency forward contracts, where the spot and forward exchange rates are taken as the means of their respective bid and ask rates. The data is retrieved from Datastream, which spans the period 1983:10-2016:03 for CHF, DEM, GBP and JPY, and 1984:12-2016:03 for AUD, CAD, NOK, NZD and SEK. Five different maturities are available for forward contracts on each foreign currency, namely 1 month, 2 months, 3 months, 6 months and 12 months.

The macroeconomic variables of the US economy are taken from the FRED-MD dataset, in which 5 out of the 134 available time series are deleted to form a balanced panel spanning the period 1963:10-2015:11.[3] These 129 variables are broadly classified into 8 groups, namely output and income (G1), labor market (G2), housing (G3), consumption, orders and inventories (G4), money and credit (G5), interest and exchange rates (G6), prices (G7), and stock market (G8). These variables are stationarized by the corresponding transformation methods described in the appendix of McCracken and Ng (2016). The panel information criterion developed in Bai and Ng (2002) suggests that the factor structure is optimally described by the first 8 PCs.

Table 3.1 presents summary statistics for the 8 chosen PCs. The first PC explains the largest fraction of the total variation in the panel, the second PC explains the second largest fraction, and so on. The first column reports estimates of the first-order autocorrelation (AR1) coefficients, whereas the second column reports the p-values of Phillips-Perron (1988) test of the null hypothesis that the variable under consideration is a unit root process. The first PC has the highest AR1 coefficient in absolute term, which is still far less than 0.9. The p-values of Phillips-Perron test are all less than 0.01. These are strong evidences that the 8 PCs are stationary.

[3] Details on the variables included in the dataset are discussed in McCracken and Ng (2016). Note that these variables start earlier than the carry trade yields. To fully capture the variations in the panel of US variables, PCA is conducted from the beginning of the US macro data instead of the beginning of the carry trade data.

IV.I Factor Interpretation and Choice of Foreign Macro Variables

To provide some economic interpretation of the PCs and some guideline for choosing foreign macroeconomic variables. I regress each of the 129 time series on each of the 8 chosen PCs, and record the resulting R^2 values. Figure 3.1 plots these R^2 values for each PC. The first PC loads heavily on variables in the categories of real economic activities, namely output and income (G1), labor market (G2), housing (G3), and consumption, orders and inventories (G4). It exhibits little correlations with prices and financial variables. I thus refer to it as a real factor. The second PC loads heavily on yield spreads, which I call the yield curve factor. The third PC loads heavily on measures of inflation and price pressure, which I call the inflation factor. The fourth and fifth PCs both load heavily on measures of housing activities, as well as on interest rates. They display little correlation with other measures of real economic activities and measures of inflation. I thus refer to them as the housing and yield curve factors. The sixth PC is most correlated with variables in the categories of output and income (G1), and consumption, orders and inventories (G4), as well as with several yield spreads. I thus refer to it as another real factor. The seventh PC loads almost exclusively on measures of the aggregate stock market, which I call the equity factor. Finally, I call the eighth PC the money and credit factor since it loads almost exclusively on variables within category G5. It should be noted, however, that these interpretations are by no means perfect since each PC loads more or less on all variables in the panel.

Ideally, the same dimension reduction procedure should be conducted on each foreign country so that its macroeconomic conditions can be captured as much and as accurate as possible, but such a dataset is not available. Instead, for each foreign country I pick one representative variable corresponding to each of the 8 broad categories in the FRED-MD dataset. To proceed, for each variable in the FRED-MD dataset, I calculate the average of its R^2 values for the regressions of the PCs that load heavily on the category it belongs to. Within each of the 8 categories, I then look for the foreign counterpart of the variable with the highest average R^2. If such a variable is not available at monthly frequency or is too short, I then look for the one with the second highest average R^2, and so on. The hope is that the foreign countries share similar latent factor structure with the US so that the most relevant variables for the US are also the most relevant for the foreign countries. Although there is no theoretic justification for this procedure, this is perhaps the most reasonable and straightforward way to proceed without imposing further assumptions. These foreign variables are collected from Datastream, OECD database, IMF database, and FRED, and are transformed in the same way as the corresponding US variables whenever possible. Detailed descriptions of these foreign variables, the transformation methods, the data sources, as well as tests of stationarity are presented in Appendix. As is seen, the second variable for each country (FPC_2^i), which is the difference between the 10-year bond yield and the policy rate, is highly persistent and has AR1 coefficient ranging from 0.92 to 0.99 although Phillips-Perron test indicates it is not a unit root process. This is common for yield curve data, and indicates that the bootstrap procedure described above is indeed necessary.

IV.II In-Sample Analysis

Table 3.2 summarizes the results of regressions (20). To save space, only the predictors that are chosen for at least one maturity by the variable selection scheme are shown in the table. A cell is left empty if the predictor is not chosen for that specific maturity.

Note that in general the bootstrapped significance levels of the parameters closely follow those derived from the asymptotic distributions. Moreover, the bootstrap procedure shows that the adjusted R^2s are all significant at the 5% level across different currencies except AUD at 2-month maturity and SEK at 3-month maturity, while most are in fact significant at the 1% level. Thus, the empirical results are not seriously distorted as suggested by Bauer and Hamilton (2018).

Some interesting patterns are observed. First, currency risk premia are better predicted at longer horizons than at shorter horizons. At the 1-month horizon, 5 out of the 10 currency premia, namely AUD, CAD, CHF, GBP and NOK, have no predictors being selected by the variable selection scheme. At the 2-month horizon, 2 out of the 10 strategies, namely CHF and JPY, either have no predictor selected or have no predictor that is significant at the 5% level. As the maturity increases, more predictors are selected and they tend to be significant, the adjusted R^2s increase from about 0-4% at the 1-month horizon to about 13-29% at the 12-month horizon, and the predictors that are selected and/or significant at a shorter horizon tends to stay selected and/or significant at a longer horizon. This pattern holds across different currencies. This implies that the differentials in risk pricing between countries are more prevalent in longer horizon, and is driven by macro factors.

Another interesting pattern is that most of the predictors selected are the US factors, whereas foreign macro variables are selected mostly for the 6-month and 12-month horizons. These include CAD at the 12-month horizon, CHF at the 6-month an 12-month horizons, DEM at the 12-month horizon, GBP at the 6-month horizon, NZD at all horizons, and SEK at the 1-month horizon. Since foreign variables are selected only if they exhibit additional predictive power conditional on the chosen US factors, this implies that in most cases the predictive power of the foreign variables, if any, is already incorporated in the

US factors. This is consistent with a large body of literature showing that country-level macroeconomic and financial conditions can be largely explained by the global factors. For example, Kose, Otrok, and Whiteman (2003) find that a global factor is an important source of variations in output, consumption and investment across countries. Similar pattern in country-level inflation rates is also found by Ciccarelli and Mojon (2010). Moreover, Ehrmann and Fratzcher (2009), and Wongswan (2009) find that equity markets around the world respond strongly to US monetary shocks. Given that the US economy represents a disporportionately part of the world, it is thus reasonable to regard the US macroeconomic factors as global factors.

The single most important predictor is PC_5, a housing and yield curve factor, whose first- and/or higher-order terms exhibit predictive power at the 1% bootstrapped significance level at the 6-month and 12-month maturities across all currencies. They also exhibit predictive power at the 5% bootstrapped significance level at the 2-month and 3-month maturities for currencies DEM, GBP, NOK and NZD, and at the 3-month maturity for CAD. Such predictive power is nonlinear in that the higher order terms are selected across all the 9 currencies and are in general significant at least at the 5% bootstrapped significance level. This is striking in that no previous research has found such a single factor that possesses uniform predictive power across the G10 currencies, or equivalently, that are priced differently across all G10 countries.

IV.III Out-of-Sample Analysis

Is the strong predictability of currency risk premia found in-sample also preserved out-of-sample? To answer this question, I use the sample period 1983:10-2010:10 as the training sample, and 2010:11-2015:11 as the test sample. The training sample is chosen much longer than the test sample so that the asymptotic distribution of the ENC-T statistic is preserved.

As is shown in Table 3.3, the in-sample predictability is well-preserved out-of-sample. Out of the 45 cases across 9 foreign currencies and 5 maturities, only 9 cases have the MSPE ratios being greater than 1 in the first exercise, and 3 cases in the second exercise. Moreover, for the first exercise, out of the 36 cases whose MSPE ratios are less than 1, 26 have significant ENC-T statistics at the 5% level or above. For the second exercise, out of the 43 cases whose MSPE ratios are less than 1, 28 have significant ENC-T statistics at the 5% level or above.

As in the in-sample cases, strategies for longer horizons exhibit stronger out-of-sample predictability. This is seen by that in general, as the maturity increases, the MSPE ratio for each currency decreases, and the number of currencies whose ENC-T statistics are significant at the 5% level or above increases. For the second exercise, at the 6-month maturity, 6 out of the 9 currencies have significant ENC-T statistics at the 1% level, and the remaining 3 have significant ENC-T statistics at the 5% level. At the 12-month maturity, 8 out of the 9 currencies have significant ENC-T statistics at the 1% level, whose MSPE errors can be as low as 0.5215 and no higher than 0.8417.

V Conclusions

This research investigates macro factors in currency risk premia. Motivated by the fact that a macro factor predicts currency risk premium only if there is difference between its market prices in the home and foreign countries, respectively, I tackle this problem by focusing on the predictability of currency risk premia. To effectively summarize information about global macroeconomic conditions, I fit a dynamic factor model to a large panel of US macro variables. Based on the economic interpretations of the estimated factors, I then collect data on their foreign counterparts. A variable selection procedure is then conduted to get a siccinct set of predictors for each currency and each maturity.

Some interesting phenomena are observed. First, currency risk premia exhibit stronger predictability at longer horizons, suggesting that differences in risk pricing between the US and other G10 countries are more prevalent in longer term. Second, foreign macro variables rarely exhibit additional forecasting power on top of the US macro factors, suggesting that the price differentials of foreign factors, if any, are mostly captured by those of the US factors. Third, a factor that loads heavily on activities in the US housing market and bond yields exhibits strong and nonlinear predictive power across all currencies.

The strong in-sample predictive power of macro factors is preserved out-of-sample. It is found that in most cases the predictive models developed in the in-sample analysis perform both economically and statistically better than those corresponding to the null hypotheses of no predictability. Moreover, as in the in-sample analysis, currency risk premia at longer horizons exhibit stronger out-of-sample predictability.

Appendix: List of Non-US Data Collected

Variable Class	Foreign Variable Selected	Period	Transformation Method	Source	AR1	Philips-Perron P-Value
Panel 1: AUD Variables						
G1	Industrial Production	1974:08-2015:11	5	Datastream	0.7147	< 0.01
G2	10-Yr Yield Minus Short Rate	1969:07-2016:02	1	Datastream	0.9226	< 0.01
G3	New Permits for Dwelling	1963:01-2016:01	4	Datastream	-0.3762	< 0.01
G4	Retail Sales	1982:03-2016:01	5	Datastream	-0.1566	< 0.01
G5	M3	1963:01-2016:01	6	IMF IFS	-0.6428	< 0.01
G6	Total Employment	1980:01-2016:01	5	OECD	0.1718	< 0.01
G7	CPI	1963:01-2015:11	6	Datastream	-0.0033	< 0.01
G8	S&P/ASX 200	1971:02-2016:02	5	Datastream	0.0570	< 0.01
Panel 2: CAD Variables						
G1	Industrial Production	1963:01-2015:12	5	OECD	0.0040	< 0.01
G2	10-Yr Yield Minus Short Rate	1963:01-2016:02	1	Datastream	0.9657	< 0.01
G3	New Permits for Dwelling	1963:01-2016:01	4	Datastream	-0.2131	< 0.01
G4	Retail Sales	1991:01-2016:01	5	Datastream	-0.1122	< 0.01
G5	M2	1968:01-2016:01	6	Datastream	-0.5165	< 0.01
G6	Total Employment	1963:01-2016:01	5	OECD	0.2848	< 0.01
G7	CPI	1963:01-2016:02	6	OECD	-0.4686	< 0.01
G8	S&P/TSX	1963:01-2016:02	5	Datastream	0.1181	< 0.01
Panel 3: CHF Variables						
G1	Industrial Production (Growth Rate)	1966:11-2016:02	1	Datastream	0.5814	< 0.01
G2	10-Yr Yield Minus Short Rate	1974:01-2016:02	1	Datastream	0.9596	< 0.01
G3			N/A			
G4	Finished Goods Stocks	1967:01-2016:03	5	Datastream	0.9710	< 0.01
G5	M2	1984:12-2016:02	6	Datastream	-0.5124	< 0.01
G6			N/A			
G7	CPI	1963:01-2016:02	6	OECD	-0.4523	< 0.01
G8	SMI	1988:06-2016:02	5	Datastream	0.1544	< 0.01

Panel 4: DEM Variables						
G1	Industrial Production	1963:01-2015:12	5	OECD	-0.2745	< 0.01
G2	10-Yr Yield Minus Short Rate	1963:01-2016:02	1	Datastream	0.9800	< 0.01
G3	New Permits for Dwelling	1979:01-2015:11	4	Datastream	-0.3648	< 0.01
G4	Retail Sales	1994:01-2016:01	5	Datastream	-0.4605	< 0.01
G5	M2	1963:01-2016:01	6	Datastream	-0.5059	< 0.01
G6	Unemployment Rate	1963:01-2016:02	2	Datastream	0.4577	< 0.01
G7	CPI	1963:01-2016:02	6	OECD	-0.5063	< 0.01
G8	DAX	1964:12-2016:02	5	Datastrean	0.0695	< 0.01
Panel 5: GBP Variables						
G1	Industrial Production	1963:01-2015:12	5	OECD	-0.1645	< 0.01
G2	10-Yr Yield Minus Short Rate	1978:01-2016:02	1	Datastream	0.9717	0.0117
G3			N/A			
G4	Retail Sales	1963:01-2016:02	5	Datastream	-0.1430	< 0.01
G5	M2	1982:07-2016:01	6	Datastream	-0.5913	< 0.01
G6	Total Employment	1971:02-2015:12	5	Datastream	0.7255	< 0.01
G7	CPI	1963:01-2016:02	6	OECD	-0.4280	< 0.01
G8	FTSE 100	1978:01-2016:02	5	Datastream	0.0077	< 0.01
Panel 6: JPY Variables						
G1	Industrial Production	1963:01-2015:12	5	OECD	0.0721	< 0.01
G2	10-Yr Yield Minus Short Rate	1989:01-2016:02	1	Datastream & FRED	0.9879	0.0932
G3	Number of New Constructions	1965:01-2016:01	4	Datastream	-0.2009	< 0.01
G4	Retail Sales	1970:01-2016:01	5	Datastream	-0.3946	< 0.01
G5	M2	1963:01-2016:02	6	Datastream	-0.4823	< 0.01
G6	Total Employment	1963:01-2016:01	5	OECD	-0.0200	< 0.01
G7	CPI	1963:01-2016:01	6	OECD	-0.3275	< 0.01
G8	Nikkei 225	1963:01-2016:02	5	Datastream	0.0730	< 0.01
Panel 7: NOK Variables						
G1	Industrial Production	1963:01-2015:12	5	OECD	-0.4285	< 0.01
G2	10-Yr Yield Minus Short Rate	1985:01-2016:02	1	Datastream	0.9416	< 0.01
G3	New Permits for Dwelling	1990:01-2016:01	4	Datastream	-0.4631	< 0.01
G4	Private Comsumption on Goods (Index)	1979:01-2016:01	5	Datastream	-0.3978	< 0.01
G5	M2	1963:01-2014:02	6	Datastream	-0.6367	< 0.01
G6	Total Employment	1972:02-2015:11	5	Datastream	0.7044	< 0.01
G7	CPI	1963:01-2016:02	6	OECD	-0.5187	< 0.01
G8	OBX	1987:01-2016:02	5	Datastream	0.1812	< 0.01

Panel 8: NZD Variables						
G1			N/A			
G2	10-Yr Yield Minus Short Rate	1973:12-2016:02	1	Datastream	0.9048	< 0.01
G3	New Permits for Dwelling	1973:02-2016:01	4	Datastream	-0.4000	< 0.01
G4	Retail Sales	1992:07-2016:01	5	Datastream	0.9617	< 0.01
G5	M2	1988:02-2016:01	6	Datastream	-0.4802	< 0.01
G6			N/A			
G7	Inflation Rate	1963:01-2015:12	2	Datastream		< 0.01
G8	DJ New Zealand	1992:01-2016:02	5	Datastream	-0.0880	< 0.01
Panel 9: SEK Variables						
G1	Industrial Production	1963:01-2015:12	5	OECD	-0.3339	< 0.01
G2	10-Yr Yield Minus Short Rate	1986:12-2016:02	1	Datastream	0.9444	< 0.01
G3	Volume Index of Bldg Production	1994:01-2015:12	4	Datastream	-0.3461	< 0.01
G4	Retail Sales	1990:01-2016:01	5	Datastream	-0.4382	< 0.01
G5	M3	1963:01-2016:01	6	IMF IFS	-0.6157	< 0.01
G6	Total Employment	1987:01-2016:02	5	Datastream	-0.2514	< 0.01
G7	CPI	1963:01-2016:02	6	OECD	-0.4304	< 0.01
G8	OMXS30	1986:01-2016:02	5	Datastream	0.1408	< 0.01

The tranformation codes are the same as for the FRED-MD dataset. For each series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$.

Tables and Figures

Table 3.1: Summary Statistics of US Factors

Principal Component	AR(1) Coefficient	Philips-Perron P-Value	Cumulative R^2
1	0.794	< 0.01	0.165
2	0.728	< 0.01	0.237
3	-0.087	< 0.01	0.304
4	0.546	< 0.01	0.360
5	0.650	< 0.01	0.404
6	0.346	< 0.01	0.441
7	0.153	< 0.01	0.471
8	-0.237	< 0.01	0.494

Notes: Summary statistics of the first eight principal components of the FRED-MD dataset of US macroeconomic variables. The AR(1) coefficient measures the persistence of each time series, and a formal unit-root test is conducted à la Phillips and Perron (1988) under the null hypothesis that the time series is integrated with order 1. Cumulative R^2 indicates the proportion of the cross sectional variation of the macro dataset explained by the first few principal components.

Table 3.2: In-Sample Predictability of Currency Risk Premia

	1-Month	2-Month	3-Month	6-Month	12-Month
Panel 1: AUD					
PC_1			-2.00×10^{-4}	-2.13×10^{-4}	1.58×10^{-4}
			(6.66×10^{-4})	(5.00×10^{-4})	(3.75×10^{-4})
PC_1^2					0.47×10^{-4}
					(0.23×10^{-4}**)
PC_1^3			0.07×10^{-4}***	0.05×10^{-4}***	
			(0.02×10^{-4}***)	(0.01×10^{-4}***)	
PC_5				10.02×10^{-4}***	19.86×10^{-4}***
				(4.18×10^{-4}**)	(8.21×10^{-4}**)
PC_5^2				5.91×10^{-4}***	3.00×10^{-4}***
				(1.38×10^{-4}***)	(0.92×10^{-4}***)
PC_5^3					-0.64×10^{-4}
					(0.31×10^{-4}**)
PC_6		18.34×10^{-4}***	17.23×10^{-4}***		
		(7.85×10^{-4}**)	(7.60×10^{-4}**)		
PC_7				8.55×10^{-4}***	5.72×10^{-4}**
				(3.24×10^{-4}***)	(2.97×10^{-4}*)
R^2_{adj}	N/A	1.51%*	6.73%***	18.30%***	20.38%***
Panel 2: CAD					
PC_5			11.27×10^{-4}***	10.35×10^{-4}***	20.57×10^{-4}***
			(4.53×10^{-4}**)	(3.64×10^{-4}***)	(5.41×10^{-4}***)
PC_5^2			4.00×10^{-4}***	3.38×10^{-4}***	1.73×10^{-4}**
			(1.49×10^{-4}***)	(1.22×10^{-4}***)	(0.68×10^{-4}**)
PC_5^3					-0.62×10^{-4}
					(0.22×10^{-4}***)
PC_6		13.96×10^{-4}***			
		(7.14×10^{-4})			
PC_7				5.44×10^{-4}***	3.50×10^{-4}***
				(2.63×10^{-4}**)	(1.98×10^{-4}*)
FPC_6					6.61×10^{-4}
					(8.83×10^{-4})
R^2_{adj}	N/A	2.21%**	7.15%***	12.88%***	29.49%***
Panel 3: CHF					
PC_2				6.57×10^{-4}***	4.21×10^{-4}**
				(3.83×10^{-4}*)	(1.79×10^{-4}**)
PC_5				12.88×10^{-4}***	7.66×10^{-4}**
				(5.30×10^{-4}**)	(3.89×10^{-4}**)
PC_5^2				4.82×10^{-4}***	3.80×10^{-4}***
				(1.72×10^{-4}***)	(1.24×10^{-4}***)
FPC_4				-1.33×10^{-4}	-2.25
				(1.15)	(0.80***)
FPC_6				24.60×10^{-4}	-21.32×10^{-4}
				(10.57×10^{-4}**)	(7.90×10^{-4}***)
R^2_{adj}	N/A	N/A	N/A	11.07%***	21.02%***

	1-Month	2-Month	3-Month	6-Month	12-Month
Panel 4: DEM					
PC_1					-0.12×10^{-4}
					(2.32×10^{-4})
PC_3		8.64×10^{-4}**	6.29×10^{-4}**		
		(3.67×10^{-4}**)	(2.47×10^{-4}**)		
PC_4					-2.04×10^{-4}
					(4.26×10^{-4})
PC_5	15.48×10^{-4}**	14.36×10^{-4}***	15.05×10^{-4}**	14.30×10^{-4}***	13.40×10^{-4}***
	(7.66×10^{-4}**)	(5.32×10^{-4}***)	(5.36×10^{-4}***)	(5.29×10^{-4}***)	(4.47×10^{-4}***)
PC_5^2	7.72×10^{-4}***	4.55×10^{-4}***	4.45×10^{-4}***	4.13×10^{-4}***	2.50×10^{-4}**
	(2.73×10^{-4}***)	(1.73×10^{-4}***)	(1.75×10^{-4}**)	(1.65×10^{-4}**)	(1.27×10^{-4}*)
FPC_6					16.32×10^{-4}
					(19.31×10^{-4})
FPC_6^2					-5.50×10^{-4}
					(2.83×10^{-4}**)
R^2_{adj}	3.99%**	5.46%**	7.39%***	10.94%***	22.49%***
Panel 5: GBP					
PC_4					6.34×10^{-4}**
					(2.91×10^{-4}**)
PC_4^2					2.33×10^{-4}***
					(0.67×10^{-4}***)
PC_5		12.55×10^{-4}**	11.19×10^{-4}***	13.14×10^{-4}***	11.19×10^{-4}***
		(6.40×10^{-4}*)	(5.66×10^{-4}**)	(3.63×10^{-4}***)	(3.77×10^{-4}***)
PC_5^2				2.80×10^{-4}***	
				(1.12×10^{-4}**)	
PC_6		15.71×10^{-4}*	14.76×10^{-4}*	10.39×10^{-4}**	
		(7.98×10^{-4}**)	(6.78×10^{-4}**)	(4.96×10^{-4}**)	
FPC_1				0.16**	
				(0.07**)	
R^2_{adj}	N/A	3.20%***	4.03%***	9.02***	15.16%***

Panel 6: JPY					
PC_1					9.16×10^{-4}***
					(3.28×10^{-4}***)
PC_1^2					-0.44×10^{-4}
					(0.18×10^{-4}**)
PC_2	22.66×10^{-4}**			18.59×10^{-4}***	9.72×10^{-4}**
	(6.21×10^{-4}***)			(2.79×10^{-4}***)	(4.57×10^{-4}**)
PC_3				6.08×10^{-4}***	
				(1.90×10^{-4}***)	
PC_4					0.57×10^{-4}
					(4.35×10^{-4})
PC_5				33.15×10^{-4}***	20.47×10^{-4}***
				(5.60×10^{-4}***)	(7.75×10^{-4}***)
PC_5^3				-1.10×10^{-4}	-0.71×10^{-4}
				(0.27×10^{-4}***)	(0.29×10^{-4}**)
FPC_6		-1.65			
		(0.36***)			
R^2_{adj}	2.95%***	2.83***	N/A	13.47%***	13.33%***
Panel 7: NOK					
PC_1					2.69×10^{-4}**
					(1.96×10^{-4})
PC_3		8.22×10^{-4}**	7.98×10^{-4}***		
		(3.56×10^{-4}**)	(2.90×10^{-4}***)		
PC_5		10.41×10^{-4}**	12.77×10^{-4}***	12.96×10^{-4}***	13.17×10^{-4}***
		(6.79×10^{-4})	(6.18×10^{-4}**)	(5.19×10^{-4}**)	(4.41×10^{-4}***)
PC_5^2		4.94×10^{-4}***	4.80×10^{-4}***	5.09×10^{-4}***	3.63×10^{-4}***
		(2.13×10^{-4}**)	(2.09×10^{-4}**)	(1.80×10^{-4}***)	(1.32×10^{-4}***)
PC_6		13.52×10^{-4}		6.78×10^{-4}	
		(9.36×10^{-4})		(5.13×10^{-4})	
R^2_{adj}	N/A	4.52%**	5.82%***	10.55%***	18.72%***

Panel 8: NZD					
PC_1					7.02×10^{-4}**
					(2.84×10^{-4}**)
PC_2			15.59×10^{-4}**	17.19×10^{-4}***	10.86×10^{-4}***
			(8.10×10^{-4}*)	(7.08×10^{-4}**)	(6.01×10^{-4}*)
PC_3		3.21×10^{-4}	2.77×10^{-4}	8.61×10^{-4}***	
		(7.31×10^{-4})	(3.66×10^{-4})	(4.08×10^{-4}**)	
PC_3^3		0.11×10^{-4}***	0.12×10^{-4}**		
		(0.03×10^{-4})	(0.03×10^{-4}***)		
PC_5		17.18×10^{-4}***	22.54×10^{-4}***	22.01×10^{-4}***	30.26×10^{-4}***
		(6.82×10^{-4}**)	(7.82×10^{-4}***)	(6.31×10^{-4}***)	(8.34×10^{-4}***)
PC_5^2		7.89×10^{-4}***	5.10×10^{-4}***	4.06×10^{-4}***	
		(2.05×10^{-4}***)	(2.03×10^{-4}**)	(1.50×10^{-4}***)	
PC_5^3					-0.77×10^{-4}
					(0.33×10^{-4}**)
PC_7				14.48×10^{-4}***	6.78×10^{-4}***
				(4.22×10^{-4}***)	(2.80×10^{-4}**)
FPC_4	3.68*	2.88**	2.41*	1.97	
	(1.77**)	(1.31**)	(1.26*)	(1.38)	
FPC_6					9.45×10^{-4}
					(14.48×10^{-4})
R^2_{adj}	2.55%***	15.78%***	19.86%***	23.34***	28.21%***
Panel 9: SEK					
PC_3		9.59×10^{-4}*			
		(6.28×10^{-4})			
PC_5				11.21×10^{-4}**	12.65×10^{-4}***
				(6.03×10^{-4}*)	(5.30×10^{-4}**)
PC_5^2				5.93×10^{-4}***	4.36×10^{-4}***
				(1.90×10^{-4}***)	(1.29×10^{-4}***)
PC_6		13.90×10^{-4}**	9.67×10^{-4}*		
		(10.17×10^{-4})	(8.50×10^{-4})		
FPC_1	0.33***				
	(0.09***)				
R^2_{adj}	3.79%***	3.00%**	0.47%	10.61%***	15.42%***

Notes: Predictive regressions of currency risk premia on macro factors. The domestic currency is taken as USD. For each foreign currency being traded, only the macro factors chosen for at least one horizon by the stepwise variable selection procedure are listed, where PC_i^j (FPC_i^j) denotes the j-th order of the i-th US (foreign) macro factor. The first row in each cell presents the corresponding parameter estimate, followed by marks of its significance level calculated from the bootstrap procedure. The number in paranthese is its Newey-West (1994) standard deviation, followed by marks of its significance level calculated from the asymptotic distribution.

"***": Significance at 1% level; "**": Significance at 5% level; "*": Significance at 10% level.

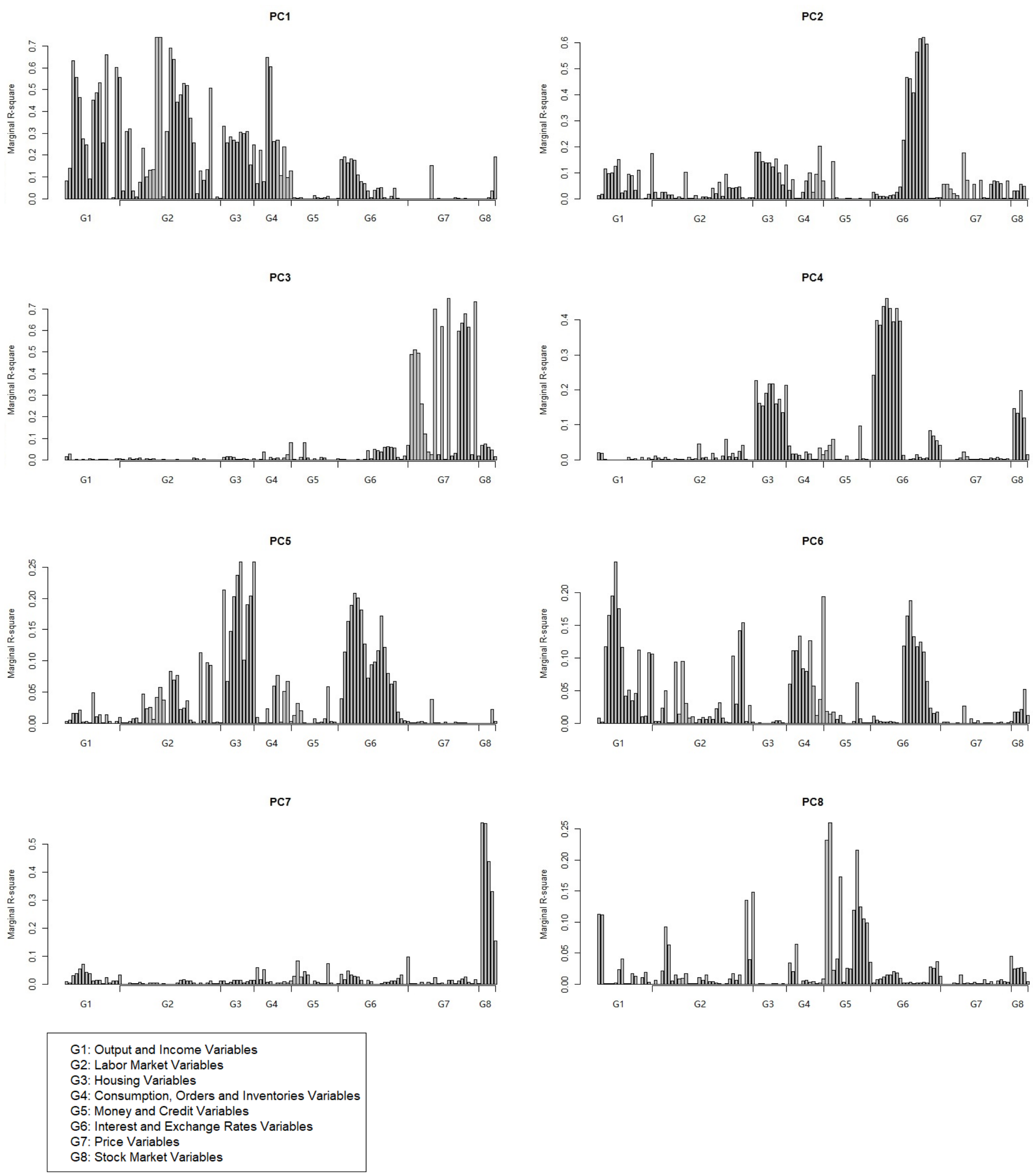
Table 3.3: Out-of-Sample Performance of Macro Variables

	1-month	2-month	3-month	6-month	12-month
Panel 1: Out-of-Sample Test with In-Sample Factors					
AUD	1.000	1.002	1.005	0.9741**	0.9367***
	(0.0000)	(-0.6321)	(-0.4865)	(2.244)	(3.575)
CAD	0.9987	0.9995	0.9520***	0.9382***	0.8067***
	(0.4304)	(0.2492)	(3.223)	(3.295)	(6.434)
CHF	1.001	0.9980	0.9939**	0.9980	1.043
	(-0.3589)	(0.8936)	(1.835)	(1.043)	(-0.6119)
DEM	0.9865	0.9725**	0.9519***	0.9194***	1.159
	(1.336*)	(2.020)	(2.675)	(3.391)	(-0.9105)
GBP	1.000	1.008	1.007	0.9705***	0.8875***
	(0.0000)	(-0.8259)	(-0.4207)	(2.878)	(4.265)
JPY	0.9881	0.9831**	0.9767**	0.9123***	0.8358***
	(1.275)	(1.738)	(1.990)	(3.293)	(4.854)
NOK	0.9929	0.9655***	0.9352***	0.8921***	0.8198***
	(1.209)	(2.714)	(3.994)	(4.572)	(5.001)
NZD	0.9956	0.9686***	0.9572***	0.9493***	0.9515***
	(0.9562)	(2.451)	(2.684)	(3.015)	(2.535)
SEK	0.9932	0.9943	0.9787***	0.9086***	0.8654***
	(0.9693)	(0.8669)	(1.790)	(3.751)	(4.100)
Panel 2: Out-of-Sample Test with Recursively Updated Macro Factors					
AUD	0.9975*	0.9990	0.9869**	0.8225***	0.6592***
	(1.4240)	(0.854)	(2.176)	(4.512)	(6.771)
CAD	0.9875**	0.9615*	0.8721***	0.7121***	0.5215***
	(1.822)	(1.610)	(3.597)	(6.475)	(9.225)
CHF	0.9915	0.9966*	0.9907*	0.9409**	0.7700***
	(1.063)	(1.348)	(1.581)	(1.979)	(4.112)
DEM	0.9866	0.9532*	0.9211**	0.8078***	0.8114***
	(0.684)	(1.545)	(2.313)	(3.783)	(4.724)
GBP	0.9981*	1.0149	1.0158	0.9453**	0.8394***
	(1.283)	(-1.425)	(-1.260)	(2.033)	(3.502)
JPY	0.9387**	0.9246***	0.8683***	0.9605**	1.6990
	(2.111)	(2.653)	(3.633)	(1.938)	(1.145)
NOK	0.9905	0.9612**	0.8779***	0.7468***	0.6393***
	(1.243)	(2.073)	(3.601)	(5.690)	(8.077)
NZD	0.9926	0.9430**	0.9175**	0.8411***	0.8417***
	(0.852)	(2.098)	(2.161)	(3.686)	(4.021)
SEK	0.9889	0.9825*	0.9595***	0.7838***	0.6876***
	(0.957)	(1.370)	(2.482)	(3.697)	(5.563)

Notes: Out-of-sample forecasts of currency risk premia by macro factors. The domestic currency is taken as USD. The training period is taken as 1983:10-2010:10, and the prediction period is 2010:11-2015:11. Panel 1 reports results for the specification taking macro factors estimated in-sample as predictive variables, whereas Panel 2 report those for the specification with macro factors estimated recursively. The first row in each cell reports the ratio of the root mean squared prediction errors (RMSE) of the unrestricted and restricted models. The number in paranthese is the test statistic of the ENC-T test on forecasting performance, which is asymptotically $N(0, 1)$.

"***": Significance at 1% level; "**": Significance at 5% level; "*": Significance at 10% level.

Figure 3.1: Loadings of US Macro Factors on Macro Variables



References

Anderson, Bing, Hammond, Peter J., and Ramezani, Cyrus A. 2010. "Affine Models of the Joint Dynamics of Exchange Rates and Interest Rates." *Journal of Financial and Quantitative Analysis*, 45(5): 1341-1365.

Ang, Andrew, and Chen, Joseph. 2010. "Yield Curve Predictors of Foreign Exchange Returns." *Working Paper*.

Angrist, Joshua D., Jordà, Òscar, and Kuersteiner, Guido M. 2018. "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited." *Journal of Business and Economic Statistics*, 36(3): 371-387.

Aruoba, S. Borağan, Diebold, Francis X., and Scotti, Chiara. 2009. "Real-Time Measurement of Business Conditions." *Journal of Business and Economic Statistics*, 27(4): 417-427.

Backus, David K., Foresi, Silverio , and Telmer, Chris I. 2001. "Affine Term Structure Models and the Forward Premium Anomaly." *Journal of Finance*, 56(1): 279-304.

Bai, Jushan, and Ng, Serena. 2002. "Determining the Number of Factors in Approximate Factor Models." *Econometrica*, 70(1): 191-221.

Bai, Jushan, and Ng, Serena. 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions." *Econometrica*, 74(4): 1133-1150.

Bakshi, Gurdip, and Panayotov, George. 2013. "Predictability of Currency Carry Trades and Asset Pricing Implications." *Journal of Financial Economics*, 110(1), 139-163.

Bandi, Federico, Corradi, Valentina, and Moloche, Guillermo. 2009. "Bandwidth Selection for Continuous-Time Markov Processes." *MPRA Paper*, No. 43682.

Bandi, Federico M., and Moloche, Guillermo. 2018. "On the Functional Estimation of Multivariate Diffusion Processes." *Econometric Theory*, 34(4): 896-946.

Bandi, Federico M., and Nguyen, Thong H. 2003. "On the Functional Estimation of Jump–Diffusion Models." *Journal of Econometrics*, 116(1): 293-328.

Bandi, Federico M., and Phillips, Peter C. B. 2003. "Fully Nonparametric Estimation of Scalar Diffusion Models." *Econometrica*, 71(1): 241-283.

Bauer, Michael D., and Hamilton, James D. 2018. "Robust Bond Risk Premia." *Review of Financial Studies*, 31(2): 399-448.

Brémaud, Pierre. 2013. "Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues." Springer Science & Business Media.

Brunnermeier, Markus K., Nagel, Stefan, and Pedersen, Lasse H. 2008. "Carry Trades and Currency Crashes." *NBER Macroeconomics Annual*, 23(1): 313-348.

Burnside, Craig, Eichenbaum, Martin, and Rebelo, Sergio. 2011. "Carry Trade and Momentum in Currency Markets." *Annual Review of Financial Economics*, 3(1): 511-535.

Campbell, John Y. 1987. "Stock Returns and the Term Structure." Journal of Financial Economics 18(2): 373-399.

Campbell, John Y., and Thompson, Samuel B. 2007. "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21(4): 1509-1531.

Christiano, Lawrence J., Eichenbaum, Martin, and Evans, Charles L. 1999. "Monetary Policy Shocks: What Have We Learned and to What End?" In *Handbook of Macroeconomics*, Vol. 1: 65-148. Elsevier.

Ciccarelli, Matteo, and Mojon, Benoit. 2010. "Global Inflation." *Review of Economics and Statistics*, 92(3): 524-535.

Cieslak, Anna, and Povala, Pavol. 2015. "Expected Returns in Treasury Bonds." *Review of Financial Studies*, 28(10): 2859-2901.

Cieslak, Anna, Morse, Adair, and Vissing-Jorgensen, Annette. 2019. "Stock Returns over the FOMC Cycle." *Journal of Finance*, forthcoming.

Clark, Todd E., and McCracken, Michael W. 2001. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics*, 105(1): 85-110.

Cochrane, John H., and Piazzesi, Monika. 2002. "The Fed and Interest Rates – a High-Frequency Identification." *American Economic Review*, 92(2): 90-95.

Cochrane, John H., and Piazzesi, Monika. 2005. "Bond Risk Premia." *American Economic Review*, 95(1): 138-160.

Collin-Dufresne, Pierre, and Goldstein, Robert S. 2002. "Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility." *Journal of Finance*, 57(4): 1685-1730.

Cook, Timothy, and Hahn, Thomas. 1989. "The Effect of Changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s." *Journal of Monetary Economics*, 24(3): 331-351.

Dai, Qiang, and Singleton, Kenneth J. 2000. "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55(5): 1943-1978.

Eagleson, G. K. 1975. "Martingale Convergence to Mixtures of Infinitely Divisible Laws."

Annals of Probability, 3(3): 557-562.

Ehrmann, Michael, & Fratzscher, Marcel. 2009. "Global Financial Transmission of Monetary Policy Shocks." *Oxford Bulletin of Economics and Statistics*, 71(6), 739-759.

Freedman, David A. 1975. "On Tail Probabilities for Martingales." *Annals of Probability*, 3(1): 100-118.

Ghysels, Eric, Hill, Jonathan B., and Motegi, Kaiji. 2016. "Testing for Granger Causality with Mixed Frequency Data." *Journal of Econometrics*, 192(1): 207-230.

Gikhman, I. I., and Skorokhod, A. V. 1972. "Stochastic Differential Equations." Springer Verlag: Berlin.

Gürkaynak, Refet S., Sack, Brian P., and Swanson, Eric T. 2005. "Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking*, 1(1): 55-93.

Gürkaynak, Refet S., Sack, Brian P., and Swanson, Eric T. 2007. "Market-Based Measures of Monetary Policy Expectations." *Journal of Business and Economic Statistics*, 25(2): 201-212.

Hall, P., and Heyde, C. C. 1980. "Martingale Limit Theory and Its Applications." Academic Press: New York.

Hansen, Lars P., and Hodrick, Robert J. 1983. "Risk Averse Speculation in the Forward Foreign Exchange Market: An Econometric Analysis of Linear Models." In *Exchange Rates and International Macroeconomics*. University of Chicago Press.

Hanson, Samuel G., and Stein, Jeremy C. 2015. "Monetary Policy and Long-Term Real Rates." *Journal of Financial Economics*, 115(3): 429-448.

Ichimura, Hidehiko. 1993. "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models." *Journal of Econometrics*, 58(1-2): 71-120.

Inoue, Atsushi, and Kilian, Lutz. 2004. "Bagging Time Series Models." *Working Paper*.

Johannes, Michael. 2004. "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models." *Journal of Finance*, 59(1): 227-260.

Jordà, Òscar, and Taylor, Alan M. 2012. "The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself." *Journal of International Economics*, 88(1): 74-90.

Joslin, Scott, Priebsch, Marcel, and Singleton, Kenneth J. 2014. "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks." *Journal of Finance*, 69(3): 1197-1233.

Kose, M. Ayhan, Otrok, Christopher, and Whiteman, Charles H. 2003. "International Business Cycles: World, Region, and Country-Specific Factors." *American Economic Re-*

view, (93)4: 1216-1239

Kuttner, Kenneth N. 2001. “Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market.” *Journal of Monetary Economics*, 47(3): 523-544.

Lettau, Martin, & Ludvigson, Sydney. C. 2010. “Measuring and Modelling Variation in the Risk-Return Trade-Off.” In *Handbook of Financial Econometrics*, Vol.1: 617-690. Elsevier.

Litterman, Robert, and Scheinkman, Jose. 1991. “Common Factors Affecting Bond Returns.” *Journal of Fixed Income*, 1(1): 54-61.

Lucca, David O., and Moench, Emanuel. 2015. “The Pre-FOMC Announcement Drift.” *Journal of Finance*, 70(1): 329-371.

Ludvigson, Sydney C., and Ng, Serena. 2007. “The Empirical Risk–Return Relation: A Factor Analysis Approach.” *Journal of Financial Economics*, 83(1): 171-222.

Ludvigson, Sydney C., and Ng, Serena. 2009. "Macro Factors in Bond Risk Premia." *Review of Financial Studies*, 22(12): 5027-5067.

Lustig, Hanno, Roussanov, Nikolai, and Verdelhan, Adrien. 2011 “Common Risk Factors in Currency Markets.” *Review of Financial Studies*, 24(11): 3731-3777.

MacKinnon, James G. "Bootstrap Methods in Econometrics." 2006. *Economic Record*, 82(1): S2-S18.

McCracken, Michael W., and Ng, Serena. 2016. "FRED-MD: A Monthly Database for Macroeconomic Research." *Journal of Business and Economic Statistics*, 34(4): 574-589.

Menkhoff, Lukas, Sarno, Lucio, Schmeling, Maik, and Schrimpf, Andreas. 2012. "Carry Trades and Global Foreign Exchange Volatility." *Journal of Finance*, 67(2): 681-718.

Newey, Whitney K., and West, Kenneth D. 1994. "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies*, 61(4): 631-653.

Mueller, Phillippe, Tahbaz-Salehi, Alireza, and Vedolin, Andrea. 2017. "Exchange Rates and Monetary Policy Uncertainty." *Journal of Finance*, 72(3): 1213-1252.

Nakamura, Emi, and Steinsson, Jón. 2018 "High Frequency Identification of Monetary Non-Neutrality: The Information Effect." *Quarterly Journal of Economics*, 133(3): 1283-1330.

Newey, Whitney K., and McFadden, Daniel. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*, Vol.4: 2111-2245. Elsevier.

Phillips, Peter C. B., and Perron, Pierre. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika*, 75(2): 335-346.

Piazzesi, Monika, and Swanson, Eric T. 2008. "Futures Prices as Risk-Adjusted Forecasts of Monetary Policy." *Journal of Monetary Economics*, 55(4): 677-691.

Popoviciu, Tiberiu. 1935. "Sur les Équations Algébriques Ayant toutes Leurs Racines Réelles." *Mathematica*, 9: 129-145.

Protter, Philip E. 1995. "Stochastic Differential Equations." Springer: Berlin, Heidelberg.

Revuz, Daniel, and Yor, Marc. 2013. "Continuous Martingales and Brownian Motion." Springer Science and Business Media.

Stock, James H., and Watson, Mark W. 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association*, 97(460): 1167-1179.

Stout, William F. 1974. "Almost Sure Convergence." Academic Press.

Svensson, Lars E. 1994. "Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994." *NBER Working Paper*, w4871.

Wongswan, Jon. 2009. "The Response of Global Equity Indexes to US Monetary Policy Announcements." *Journal of International Money and Finance*, 28(2): 344-365.