High-Frequency Expectations from Asset Prices: A Machine Learning Approach[*]

Aditya Chaudhry[†] Sangmin S. Oh[‡]

September 16, 2020. Comments welcome.

Abstract

We propose a novel reinforcement learning approach to extract high-frequency aggregate growth expectations from asset prices. While much expectations-based research in macroeconomics and finance relies on low-frequency surveys, the multitude of events that pass between survey dates renders identification of causal effects on expectations difficult. Our method allows us to construct a daily time-series of the cross-sectional mean of a panel of GDP growth forecasts. The high-frequency nature of our series enables clean identification in event studies. In particular, we use our estimated daily growth expectations series to test the "Fed information effect" and find little evidence to support its existence. Extensions of our framework can obtain daily expectations series of any macroeconomic variable for which a low-frequency panel of forecasts is available. In this way, our method provides a sharp empirical tool to advance understanding of how expectations are formed.

[*] First draft: September 2019. We thank Ralph Koijen and Stefan Nagel for their invaluable advice and generous help at all stages of this project. For comments, we thank Lubos Pastor, Harald Uhlig, Nicholas Polson, and participants at the Chicago Booth Finance Brownbag. Results are preliminary, and all remaining errors are own. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1746045. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

[†] University of Chicago, Booth School of Business. (aditya.chaudhry@chicagobooth.edu)

[‡] University of Chicago, Booth School of Business. (oh@chicagobooth.edu)

1 Introduction

Investor expectations play a central role in asset pricing, as demonstrated by the equation

$$P_t = \mathbb{E}_t \left[M_{t+1} X_{t+1} \right] \tag{1}$$

where investors price assets based on their beliefs about the joint distribution of the stochastic discount factor M_{t+1} and the asset's cash flows X_{t+1}. One of the key drivers of investor expectations is news of macroeconomic events such as impending trade wars, interest rate changes, or announcements of new tax policy. To examine the impact of such news events, researchers have either examined the behavior of asset prices around announcement dates (Lucca and Moench (2015), Neuhierl and Weber (2018)) or utilized surveys that directly measure expectations (Gennaioli et al. (2016), Fuhrer (2017)). Each approach, however, suffers from certain shortcomings. The diversity of information sources (e.g. news about future interest rates, growth, unemployment, etc.) that affect asset prices limits the first method in targeting a specific type of expectation, while the low frequency of survey data, which is quarterly or monthly at best, inhibits the second.

In this paper, we construct a daily time series of investor expectations of macroeconomic growth. Since surveyed expectations are available at a quarterly frequency, our task is to recover the unobserved daily series of expectations between two quarterly survey releases dates. We then repeat this process quarter by quarter. While previous papers have proposed Kalman filtering (KF) and regression-based approaches to estimate the latent expectations series, we propose an alternate framework based on reinforcement learning (RL), a branch of machine learning with roots in dynamic programming. We utilize daily asset prices that reflect investors' updated beliefs about macroeconomic growth. Thus, as econometricians we tackle the inverse problem of extracting beliefs from daily asset prices.

Our main task of interest resembles that of Ghysels and Wright (2009), in which authors propose a mixed frequency data sampling (MIDAS) approach for using asset price data to predict the forecasts of professional forecasters. There are, however, several important points of departure in our paper. First and foremost, we target a different quantity than Ghysels and Wright (2009) do. At any day t within a quarter, MIDAS yields a prediction of the end-of-quarter survey expectation. On the other hand, we seek to estimate the latent expectation at time t, which need not bear any relation to the time t prediction of the end-of-quarter expectation. To this end, we use reinforcement learning as our method of choice. Second, our approach greatly reduces the number of parameters to be estimated. Third, instead of modeling the evolution of forecasts in a reduced form way as done in Ghysels and Wright (2009), we construct a state-space of growth and returns that distinguishes discount rate shocks from cash flow shocks. This approach enables our algorithm to use multiple assets and extract only the unexpected component of returns in constructing our forecasts.[1]

Two immediate questions arise: why reinforcement learning (RL), and why asset prices? The short answer to the first question, which will be expanded upon in Section 3, is that RL achieves a significant gain in efficiency over traditional filtering techniques such as the Kalman filter. RL avoids an explicit model of the state dynamics and thus requires estimation of far fewer parameters. To see this point more explicitly, consider the following simple model:

$$\begin{aligned} y_{t+1} &= Hy_t + e_{t+1}, \quad e_{t+1} \sim \mathcal{N}\left(0, \Sigma\right) \\ x_{t+1} &= Fx_t + u_{t+1}, \quad u_{t+1} \sim \mathcal{N}\left(0, \Phi\right) \end{aligned} \tag{2}$$

where H, F, Σ, Φ are scalar coefficients. Here $\{y_t\}$ represents the observed series (e.g. asset prices), $\{x_t\}$ represents the latent series (e.g. macroeconomic growth expectations), and e_{t+1} and u_{t+1} are assumed to be independent. In this example, the update rule for the estimate of x in the Kalman Filter is

$$\hat{x}_{t+1|t} = F\left(\hat{x}_{t|t-1} + \left(\frac{H\Omega_{t|t-1}}{\Sigma + H^2\Omega_{t|t-1}}\right)(y_t - \hat{y}_t)\right)$$

where $\Omega_{t|t-1}$ is variance of the state estimate. To compute the update increments, one must estimate the parameters (H, F, Σ, Φ) using maximum likelihood estimation. When the data prove scarce compared to the number of parameters, however, the parameters are estimated inefficiently and the subsequent errors propagated to the state updates. RL avoids this problem by estimating the update function directly:

$$\hat{x}_{t+1|t} = \hat{x}_{t|t-1} + f\left(y_t\right)$$

where f is a parsimoniously parameterized function of the observables, y_t. The efficiency gain from estimating fewer parameters lies at the core of why our reinforcement learning approach outperforms existing methods in the task of interest.

The answer to the question of why asset prices may be useful in this task proves more nuanced. First, since we construct a daily series of expectations *within* each quarter, we cannot use fundamental data such as dividend growth or GDP growth that is released at low frequency. Our data must be available at a daily frequency, a constraint that makes asset prices the prime candidate. However, prices reflect many variables besides growth expectations. Expectations of any variable related to cash flow growth will appear in asset prices, as will discount rates. With a single asset, we cannot extract the component of asset returns driven solely by changes in expectations of macroeconomic growth. But with multiple assets, a suitable linear combination of them can

[1] This is essentially the idea behind economic tracking portfolios in Lamont (2001).

cancel the extraneous sources of return variation and deliver a estimate of the change in growth expectations. In other words, the econometrician's task can be interpreted as finding an optimal combination of asset returns that correlates maximally with the change investors' expectations of future macroeconomic growth.

To illustrate the above intuition and motivate our empirical strategy, we start by providing empirical evidence regarding the relationship between asset returns and expectations of macroeconomic growth. As a proxy for aggregate expectations, we use the mean of a cross-section of GDP growth forecasts obtained from the Survey of Professional Forecasters (SPF).

We then elucidate the differences among our RL algorithm, the Kalman Filter (KF) and MIDAS regression by presenting a stylized economy with Bayesian agents. We consider an economy in which expected returns and dividend growth are linear in common factors, one of which is macroeconomic growth. To complete the analogy to our final empirical task at hand – estimating the average expectation across forecasters – we also generate a panel of growth expectations by explicitly modeling a cross-section of Bayesian forecasters. Each forecaster employs a correctly-specified Kalman Filter with a slightly different calibration of the underlying state-space model. We derive the expressions for the RL, KF, and MIDAS estimators in this setting to demonstrate the greater statistical efficiency of our RL approach.

Next, we take our RL algorithm to the real data. Specifically, we apply the algorithm to forecasts from the Survey of Professional Forecasters (SPF) and construct a daily series of investor expectations and disagreement. In a recursive out-of-sample estimation procedure, we train six models with different lookback windows and average the resulting policy weights. Across the entire out-of-sample period, we find that the constructed daily series of average growth expectations realizes an R^2 of 82.3% against the true quarterly series. These results prove far superior to the results from the KF and MIDAS, which achieve R^2 values of 2.3% and 39.2%, respectively.

Finally, we use our estimated daily series of growth expectations to test the existence of the Fed information effect. Introduced by Romer and Romer (2000), the Fed information effect refers to the notion that perhaps the Fed has private information about the current and future state of the economy that is revealed by its actions. Part of the motivation for the Fed information effect comes from regressions of the change in low-frequency surveyed growth expectations from before and after an FOMC announcement on some measure of the monetary policy shock. These regressions yield coefficients with the "wrong" sign: positive (Nakamura and Steinsson, 2018). Hawkish surprises for interest rates correspond to increases in real GDP growth expectations and dovish surprises correspond to decreases in real GDP growth expectations. The problem with this type of analysis is that it uses low-frequency (e.g. monthly) surveys of expectations. Recent work has suggested that news between the pre-FOMC survey and the FOMC announcement causes omitted variable bias in these regressions. We bypass this omitted variables problem by moving to a higher frequency. Specifically, we regress the FOMC announcement-day change in growth expectations using our

RL-estimated series on the the monetary policy shock of Nakamura and Steinsson (2018). Our high-frequency regressions provide no evidence of the Fed information effect: hawkish surprises correspond to decreases in real GDP growth expectations.

To our knowledge, our paper represents the first serious application of reinforcement learning in asset pricing. Our RL approach can be applied to obtain a daily series of expectations for any macro variable for which a low-frequency panel of forecasts is available. Immediate candidates include interest rate expectations and inflation expectations. Furthermore, the estimated daily series of aggregate expectations has useful practical applications, such as for testing theories of expectations formation and responses to unexpected macroeconomic events.

Econometrically, our paper relates to the literature on measuring latent economic and financial variables in a time-series setting. Popular frameworks include balanced panel regressions (Stock and Watson (1989)), state-space models (Bernanke et al. (1997), Evans (2005), Van Binsbergen and Koijen (2010)), and latent VARs (Brandt and Kang (2004)). While we estimate *expectations* of economic variables rather than the actual variables themselves, our core task shares common features with the existing literature. In our setup, we treat the average forecast across a cross-section of forecasters as a latent variable that our RL approach efficiently recovers.

Our paper contributes to a growing body of work that incorporates machine learning methods in finance. Particularly in the asset pricing literature, researchers have used a wide array of methods including shrinkage and selection (Rapach et al. (2013), Freyberger et al. (2017), Kelly et al. (2017), Giglio and Xiu (2018), and Kozak et al. (2019)), neural networks (Hutchinson et al. (1994), Yao et al. (2000), Sirignano et al. (2016), and Heaton et al. (2017)), and tree-based models (Moritz and Zimmermann (2016)). Distinct from papers using the aforementioned methods, our paper focuses on reinforcement learning. With roots in dynamical systems theory, reinforcement learning tries to maximize a reward signal rather than find hidden structures and features in the data. Our contribution is to show that a parsimonious estimation of the optimal policy function via reinforcement learning can bring a significant efficiency gain relative to traditional filtering techniques.

The remainder of the paper proceeds as follows. Section 2 motivates our empirical strategy by developing a framework for estimating the investor expectations at a daily frequency. Section 3 outlines the three methods we use: RL, KF, and MIDAS. Section 4 contains the results of the empirical estimation of the daily growth expectations series. Section 5 presents our test of the Fed information effect. Section 6 concludes.

2 Empirical Framework

In this section, we consider an economy in which expected returns and dividend growth across assets and over all horizons are linear in common factors, one of which is macroeconomic growth. We establish, both theoretically and empirically, that asset prices are useful in estimating growth expectations.

2.1 Data

We interpret asset prices broadly to include interest rates, spreads, returns, and various measures related to the value of financial assets. Since we seek to construct a daily time-series, we require assets for which liquid daily returns are available. For equities, we consider returns on the market index, Fama-French industry portfolios, and Fama-French factors (size, value, and momentum). For fixed income, we consider returns on treasury bonds, changes in the slope of the yield curve, and changes in credit spreads. For exchange rates, we consider the change in the value of the U.S. dollar versus a basket of other currencies.[2] For derivatives, we consider changes in VIX index. Table 1 summarizes the data sources we use.

For macroeconomic growth, we use realized real GDP growth from the FRED database. Data on real GDP is seasonally adjusted in billions of chained 2012 dollars, obtained via the FRED database (GDPC1). We compute the annualized GDP growth rate as four times the quarterly percentage change in real GDP.

Since we are interested in measuring the expectations of growth, we utilize GDP growth forecasts from the Survey of Professional Forecasters (SPF). The SPF survey occurs every quarter, asking participants for quarterly projections up to five quarters ahead as well as annual projections for the current year and the following year. The forecast variables include GDP growth, various measures of inflation including CPI inflation, and the unemployment rate. This paper focuses on the one-quarter ahead GDP growth forecasts. For instance, we focus on the forecaster's expectation of GDP growth in 2018:Q4 (t) from surveys conducted in mid 2018:Q3 ($t-1$).

Table 2 presents selected summary statistics for the SPF forecasts. We find that the surveys provide decently accurate forecasts of real GDP growth. Figure 1 plots the time-series of cross-sectional mean and standard deviation of SPF forecasts from 1970 to 2018.[3] We verify that the mean of the forecasts is pro-cyclical while the standard deviation is counter-cyclical. The result accords with Kozeniauskas et al. (2018), which documents that cross-sectional disagreement regarding growth is countercyclical.

Growth forecasts also prove more persistent than realized GDP growth. The autocorrelation in SPF one-quarter ahead forecasts from 1990:3Q to 2018:4Q is 0.7337 while the autocorrelation in realized GDP growth during the same period is 0.3944.[4] In addition, as shown in Figures 2 and 3,

[2] This is the variable "DTWEXM" from FRED.

[3] The first survey the Philadelphia Fed conducted in real time occurred in 1990:Q3. Thus, in our estimation restrict our sample of forecasts to 1990:Q3 onward. However, this plot extends back to 1970:Q1 to highlight the procyclicality and countercyclicality of the cross-sectional mean and standard deviation, respectively

[4] This result proves unsurprising. As an extreme example, if real GDP growth follows an i.i.d. process and agents have access to a sufficiently long time series, then their forecasts each quarter will be the unconditional mean of the GDP growth process. Thus, even though growth is serially uncorrelated, mean expectations will be perfectly

we find that the forecast accuracy, as measured by correlation and root mean square error, declines as the forecasting horizon increases.

2.2 Model of the Economy

To motivate our empirical analysis, we consider an economy in which GDP growth is persistent:

$$\theta_{t+1} = \mu + \delta\theta_t + \epsilon_{t+1} \tag{3}$$

where θ_t is the GDP growth at time t and ϵ_{t+1} is a normal shock with variance σ^2_ϵ. There are m assets indexed by i, and we assume that GDP growth affects each asset's dividend growth:

$$d^i_{t+1} - d^i_t = \gamma + \beta^i \theta_{t+1} + \nu_{t+1} \tag{4}$$

where d^i_t is the log dividend of asset i in period t and β^i is the asset's loading on contemporaneous macroeconomic growth θ_{t+1}.

Furthermore, we assume that the conditional expected return of asset i depends linearly on another latent factor ζ_t:

$$\mathbb{E}_t\left[r^i_{t+1}\right] = \alpha + \phi^i \zeta_t \tag{5}$$

and that ζ_t is persistent:

$$\zeta_{t+1} = \tau + \psi\zeta_t + \xi_{t+1} \tag{6}$$

where ξ_{t+1} is a normal shock with variance σ^2_ξ. For generality, we assume that innovations to θ_t and ζ_t are correlated:

$$\mathrm{Corr}\left(\epsilon_t, \zeta_t\right) = \pi \tag{7}$$

Under this setup[5], we prove in Appendix A that applying the approximation in Campbell and Shiller (1988) delivers:

$$\forall i = 1, ...m : r^i_{t+1} = \gamma + \left(\beta^i + \frac{\delta\beta^i}{1-\rho\delta}\right)\theta_{t+1} - \left(\frac{\delta\beta^i}{1-\rho\delta}\right)\theta_t - \frac{\phi^i}{1-\rho\psi}\left(\zeta_{t+1} - \zeta_t\right) + \nu_{t+1} \tag{8}$$

where

$$\rho = \frac{1}{1+\exp\left(\overline{d-p}\right)}$$

and $\overline{d-p}$ is the average log dividend-price ratio. Therefore, the return r_{t+1} is a simple function of $\theta_{t+1}, \theta_t, \zeta_{t+1} - \zeta_t$ and ν_{t+1}. Returns increase with contemporaneous growth θ_{t+1} and the shock to

autocorrelated. In general, Bayesian agents with tight priors will have persistent mean expectations.

[5] This setup is equivalent to the present value system in Kelly and Pruitt (2013) with θ_t and ζ_t as two underlying common factors. Factor models are sufficiently general to include a wide range of models in asset pricing that link asset-specific expected returns and dividend growth to aggregate variables.

the dividend process ν_{t+1}, and decrease with previous period's growth θ_t and the change in ζ_{t+1}.

2.3 Relationship between Asset Prices and Growth Expectations

Equation (8) implies that asset prices should be useful for understanding changes in investor expectations. To test this question empirically, we examine whether asset returns can explain innovations in the average growth forecast. Specifically, we define the forecast innovation for period t as the difference between the mean SPF forecast of period t-growth reported in period t and the mean SPF forecast of period t-growth reported in period $t-1$, i.e. the difference between the nowcast and the lag-one-period forecast for period t. For example, we compute the difference in the forecaster's expectation of GDP growth in 2018:Q4 (t) from surveys conducted in mid 2018:Q4 (t) (nowcast) and mid 2018:Q3 ($t-1$) (lag-one-period forecast). We then run time-series regressions of innovations in mean growth expectations on asset returns. In the interest of parsimony, we only consider bivariate pairs of assets in our test asset set. We discuss the impact of adding more assets in Section 4.2. We conduct this analysis at the frequency of the SPF forecast releases, which is quarterly.

Table 3 displays the results of these regressions. We find that the following pair of assets explains the greatest amount of variance in the quarterly forecast innovations ($R^2 = 38.3\%$): the CRSP U.S. Treasury five-year fixed-term index and the CRSP value-weighted portfolio. Other pairs of assets involving bond returns, credit spreads, and VIX also yield sizable R^2 values of over 25%. Thus, we find empirically that asset returns contain useful information about forecast innovations.

2.4 Incorporating Bayesian Agents

In Section 2.2, we introduced the data-generating process for our simulated economy. To complete the setup, we now incorporate Bayesian agents who observe realized returns and form expectations of the latent growth process.

We start by generating the latent growth (θ_t) and return series (r_t^i) from the state-space model in Section 2.2. We then instantiate 20 agents who observe returns but cannot observe growth. We assume these agents are Bayesians who form estimates of θ_t via the Kalman Filter.[6] Each agent uses the same state equation (10) and observation equation (11).

Given the substantial dispersion across forecasts in the SPF data, we introduce heterogeneity among the simulated agents along two dimensions. First is *prior-mean heterogeneity*: the mean of each agent's prior belief regarding θ_t at the start of the quarter is drawn from a normal distribution with mean θ_0 and standard deviation $0.5\theta_0$ where θ_0 is a calibrated parameter. Second is *learning heterogeneity*: for each parameter in equations (10) and (11), each agent draws his value of the parameter from a normal distribution centered at the baseline parameter value with variance

[6] The output of the KF yields both estimates of θ_t and ζ_t, but empirically we are only interested in θ_t.

parameterized by a fixed signal-to-noise ratio. Consequently, each agent updates his belief via correctly specified KF equations but with misspecified parameters. While prior-mean heterogeneity decays monotonically over time, learning heterogeneity introduces a persistent level of heterogeneity across forecasts in simulation. Having set up this environment, we next examine how to estimate the cross-sectional moments of growth forecasts.

2.5 Learning the Cross-sectional Moments

We now discuss the task of an econometrician who seeks to estimate the daily mean from a cross-section of GDP forecasts given observed returns on assets.

Consider an approach via the KF. One possibility is to estimate the parameters for each individual agent via maximum likelihood using the agent-specific time-series of reported expectations. While this approach enables the construction of an entire panel of expectations and the derived cross-sectional moments, it inevitably requires estimating a large number of parameters. The resulting inefficiency proves concerning given the dearth of growth expectations survey data. Therefore, we consider a more parsimonious approach.

We propose estimating the moments directly rather than keeping track of the entire cross-section. To emphasize that the target variable here is aggregate expectations of growth and not growth itself, denote $\mu_{i,t} \equiv \mathbb{E}_t^i\left[\theta_{t+1}\right]$ as agent i's period t expectation of growth at period $t+1$, $\forall i = 1, ..., N$. Under the model in Section 2, the expression for the optimal Kalman gain implies the following relationship:

$$\mu_{i,t} = c_{0,t}^i + c_{1,t}^i \mu_{i,t-1} + \left(\mathbf{c}_{2,t}^i\right)' \mathbf{r}_t$$

where $c_{0,t}^i, c_{1,t}^i$ are scalars; $\mathbf{c}_{2,t}^i$ is a vector of scalars; and $\mathbf{r}_t$ is a vector of m asset returns in period t. Averaging across all agents, we get the following expression for the cross-sectional mean of growth expectations at period t, denoted as μ_t:

$$\mu_t \equiv \frac{1}{N}\sum_{i=1}^{N} \mu_{i,t} = \frac{1}{N}\sum_{i=1}^{N} c_{0,t}^i + \frac{1}{N}\sum_{i=1}^{N} c_{1,t}^i \mu_{i,t-1} + \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{c}_{2,t}^i\right)' \mathbf{r}_t$$

Motivated by the expression, we use the following approximating moment:

$$\mu_t = c_0 + c_1 \mu_{t-1} + \mathbf{c}_2' \mathbf{r}_t \approx c_1 \mu_{t-1} + \mathbf{c}_2' \mathbf{r}_t \tag{9}$$

where the second approximation follows since $c_0 \approx 0$ in our calibration in Appendix C. Note that the quality of this approximation depends on the degree of learning heterogeneity since $c_{0,t}^i, c_{1,t}^i$ and $\mathbf{c}_{2,t}^i$ are functions of the underlying structural parameters. Having established the approximation that underlies the estimation in remainder of this paper, we now proceed to characterize the three approaches for estimating the cross-sectional mean.

3 Three Approaches to Estimation: KF, RL, and MIDAS

In this section, we provide a comparison of three approaches – Kalman filtering (KF), reinforcement learning (RL), and mixed data sampling (MIDAS) – through which an econometrician can estimate the latent factor processes. We also describe our RL method in detail and compare its features to those of existing methods to provide a clear comparison.

3.1 The Kalman Filtering (KF) Approach

Rearranging the state-space and observation equations from our simulated economy yields a final system of state and observation equations:

$$\begin{aligned}\theta_{t+1} &= \mu + \delta\theta_t + \epsilon_{t+1}\\ \zeta_{t+1} &= \tau + \psi\zeta_t + \xi_{t+1}\end{aligned} \tag{10}$$

$$\forall i = 1, ..., m : r^i_{t+1} = \gamma + \left(\beta^i + \frac{\delta\beta^i}{1-\rho\delta}\right)\theta_{t+1} - \frac{\delta\beta^i}{1-\rho\delta}\theta_t - \frac{\phi^i}{1-\rho\psi}\left(\zeta_{t+1} - \zeta_t\right) + \nu_{t+1} \tag{11}$$

The KF approach models the cross-sectional moments as latent variables and uses the Kalmlan Filter for estimation. Substituting (8) into the approximation (9) yields:

$$\begin{aligned}\mu_t &= c_1\mu_{t-1} + \mathbf{c}_2'\mathbf{r}_t\\ &= c_1\mu_{t-1} + \mathbf{c}_2'\left[\mathbf{1}\gamma + \mathbf{a}\theta_t + \mathbf{b}\theta_{t-1} + \mathbf{c}\left(\zeta_t - \zeta_{t-1}\right) + \boldsymbol{\nu}_t\right]\end{aligned}$$

where the elements $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are

$$a^i = \beta^i + \frac{\delta\beta^i}{1-\rho\delta}, \quad b^i = -\frac{\delta\beta^i}{1-\rho\delta}, \quad c^i = -\frac{\phi^i}{1-\rho\psi}$$

Adding μ_t as another latent variable to the state equation (10) and the observation equation (11) yields the corresponding state equations:

$$\begin{aligned}\theta_{t+1} &= \mu + \delta\theta_t + \epsilon_{t+1}\\ \zeta_{t+1} &= \tau + \psi\zeta_t + \xi_{t+1}\\ \mu_{t+1} &= \mathbf{c}_2'\left(\mathbf{1}\gamma + \mathbf{a}\mu + \mathbf{c}\tau\right) + \mathbf{c}_2'\left(\mathbf{a}\delta + \mathbf{b}\right)\theta_t + \mathbf{c}_2'\mathbf{c}\left(\psi - 1\right)\zeta_t + c_1\mu_t\end{aligned} \tag{12}$$

and the observation equation:

$$\mathbf{c}_2'\mathbf{r}_t = \mu_t - c_1\mu_{t-1} \tag{13}$$

There are $3m + 11$ parameters to be estimated where m is the number of assets used. We fit the model in-sample using maximum likelihood and then use the estimated KF to obtain daily estimates out-of-sample. Note that by comparison, as will be detailed later, the cross-sectional RL method requires estimation of only $m + 1$ parameters.

Note that the KF described above is misspecified since approximation (9) does not describe the true law of motion for the cross-sectional mean. The correctly specified filter would require estimating $N\,(2m + 10)$ parameters, where N is the number of forecasters in the cross section. Given the relatively short time-series available for surveyed expectations, this method would prove futile from an efficiency perspective. Instead we admit bias into the KF by way of model misspecification so that it achieves the same order of magnitude of efficiency as the cross-sectional RL method. This step is explicitly undertaken to provide a fair comparison of KF against the RL approach.

3.2 The Mixed Data Sampling (MIDAS) Approach

We also implement the MIDAS regression forecasting method from Ghysels and Wright (2009) as a benchmark to our RL approach. MIDAS regressions forecast low-frequency variables from higher-frequency predictors. To be concrete, assume there is a low-frequency variable of interest denoted y_t for which we have quarterly observations. Let d_t denote the day we observe y_t. Additionally, there are several high-frequency predictors $r^i_\tau, i = 1, \ldots, m$, that we observe daily. We seek to forecast y_t on day τ where $d_{t-1} < \tau < d_t$, so τ represents a day between the quarterly observation dates of y_t. To this end, for each day τ we fit the following model from Ghysels and Wright (2009):

$$y_{t,} = \alpha^\tau + \rho^\tau y_{t-1} + \sum_{i=1}^{m} \beta^\tau_i \gamma^\tau(L) r^i_\tau + \epsilon_t, \tag{14}$$

where $\gamma^\tau(L)$ is a lag-polynomial of order l and the superscripts on the coefficients indicate that they can vary across days τ. Therefore we have,

$$\gamma^\tau(L) r^i_\tau = \sum_{d=\tau-l+1}^{\tau} \gamma^\tau_d r^i_d.$$

To limit the number of parameters to estimate, we follow Ghysels and Wright (2009) and use the beta lag specification from Ghysels et al. (2007), which parameterizes $\gamma^\tau(L)$ by two parameters: κ_1 and κ_2. Moreover, we follow Ghysels and Wright (2009) and use a maximal lag of $l = 90$ days. In our setting, y_t is μ_t, the quarterly observed cross-sectional mean survey expectation, and the high-frequency predictors are daily asset returns. Each MIDAS regression involves estimating $m + 4$

parameters.

3.3 The Reinforcement Learning (RL) Approach

In this subsection, we first describe the intuition behind our proposed RL algorithm, and then formalize the RL setup in the context of our cross-sectional estimation and explain why RL achieves more efficient estimation . We delegate the details of our RL algorithm to the online appendix.

3.3.1 Intuition

In general, RL algorithms enable an agent to learn the optimal *policy* that dictates what *action* to take given the current *state*. In our setting, the agent's state is the current expectation of next quarter growth and the current asset returns. The policy is the function of the current state that yields the agent's new growth expectation, and action is the agent's updated growth expectation.

The KF uses the same framework: the agent updates the growth expectation based on a linear combination of the observed asset returns. Unlike the KF, however, RL is *model-free* in that it does not require an explicit model of the underlying state transition dynamics of the environment. Instead of using maximum likelihood to fit model parameters and then computing the optimal Kalman gain, RL directly learns the policy function. Therefore, RL enables more efficient estimation by omitting the model of the state transitions.

Specifically, let s_t denote the state in period t and

$$\varphi\left(s_t\right)=\left[\begin{array}{c}\hat{\mu}_{t|t-1}\\ \mathbf{r}_t\end{array}\right]\in\mathbb{R}^{m+1}$$

be the state features in the agent's information set in time t, where m is the number of assets. Recall that $\hat{\mu}_{t|t-1}$ is the agent's expectation from period $t-1$ of period t growth, and $\mathbf{r}_t$ is the vector of period t asset returns. We can also write the policy function as

$$a_t=g_{\boldsymbol{\lambda}}\left(s_t\right)$$

where a_t is the action taken by the agent in period t and $\boldsymbol{\lambda}$ parameterizes the policy function. As discussed previously, the action in our setup is the agent's updated growth expectations: $a_t=\hat{\mu}_{t+1|t}$.

The KF and RL diverge in how they learn the optimal policy $g_{\boldsymbol{\lambda}}\left(\cdot\right)$. In the KF, the expression for the optimal Kalman gain gives rises to $g_{\boldsymbol{\lambda}}\left(\cdot\right)$ as a linear function with $\boldsymbol{\lambda}\in\mathbb{R}^{m+1}$ having a closed-form expression that is a function of the structural parameters given in Section 3.1. Thus, to compute $g_{\boldsymbol{\lambda}}$, the KF requires estimation of $3m+11$ parameters. On the other hand, in our proposed

RL algorithm, we do maintain the linearity restriction on $g_{\boldsymbol{\lambda}}$, namely:

$$g_{\boldsymbol{\lambda}}(s) = \varphi(s)^{\top} \boldsymbol{\lambda}$$

but instead the algorithm learns $\boldsymbol{\lambda}$ directly from historical data. This approach requires estimation of only $m+1$ parameters.

Translating this approach into our stylized setting, we consider an agent who starts with $\hat{\mu}_{1|0} = \mu_0$ and follows $g_{\boldsymbol{\lambda}}$ each day for one quarter, ultimately arriving at $\hat{\mu}_{T|T-1}$ where T is the number of days in a quarter. Our goal is to minimize the Euclidean distance between $\hat{\mu}_{T|T-1}$ and μ_T, i.e. solve the following minimization problem:

$$\min_{\lambda} \quad \left|\left| g_{\boldsymbol{\lambda}}^{T-1}(s_1) - \mu_T \right|\right| \tag{15}$$

where $g_{\boldsymbol{\lambda}}^{T-1}(s_1)$ is the value of $\hat{\mu}_{T|T-1}$ achieved by following the policy $g_{\boldsymbol{\lambda}}$ for $T-1$ periods starting with s_1.

In our setting, μ_0 and μ_T are the observed cross-sectional mean expectations of a survey of forecasters in two consecutive quarters, and $g_{\boldsymbol{\lambda}}$ is a policy function that yields daily estimates of the latent cross-sectional mean between these two quarterly releases. Specifically, the RL agent observes μ_0 on the survey release date at the start of quarter j and iteratively uses observed asset returns to construct daily estimates $\hat{\theta}_{t|t-1}$ of the unobservable mean expectation for each day t in quarter j. The agent continues this process until the next survey release date at the start of quarter $j+1$, at which point θ_T is revealed and the loss function in (15) can be computed. The optimal policy minimizes the average end-of-quarter j loss. In Appendix B, we formalize this intuition in order to discuss how our RL algorithm learns the optimal policy.

3.3.2 Cross-sectional Estimation: RL Approach

We apply our RL algorithm to learn the optimal policy for estimating μ_t. The state vector for period t includes the period $t-1$ cross-sectional mean, variance, and period t asset returns:

$$\varphi(s_t) = \begin{pmatrix} \hat{\mu}_{t-1} \\ \hat{\sigma}_{t-1}^2 \\ \mathbf{r}_t' \end{pmatrix} \in \mathbb{R}^{m+2}$$

where $\hat{\mu}_{t-1}$ and $\hat{\sigma}_t^2$ are the estimated cross-sectional mean and variance at period $t-1$. The initial state is

$$\varphi(s_1) = \begin{pmatrix} \mu_0 \\ \sigma_0^2 \\ \mathbf{r}_1' \end{pmatrix}$$

where μ_0 and σ_0^2 are the true cross-sectional mean and variance from the previous quarter's survey release. Following the approximation (9), we use the following policy function:

$$g_{\boldsymbol{\lambda}}\left(s_t\right) \equiv \begin{pmatrix} \mu_t \\ \sigma_t \end{pmatrix} = \begin{pmatrix} c_1\mu_{t-1} + \mathbf{c}_2'\mathbf{r}_t \\ \sqrt{c_3\sigma_{t-1}^2 + \mathbf{c}_4'\mathbf{r}_t\mathbf{r}_t'\mathbf{c}_4 + \mathbf{c}_5'\mathbf{r}_t\mu_{t-1}} \end{pmatrix} \in \mathbb{R}^2$$

where

$$\boldsymbol{\lambda} = \begin{pmatrix} c_1 \\ \mathbf{c}_2 \\ c_3 \\ \mathbf{c}_4 \\ \mathbf{c}_5 \end{pmatrix} \in \mathbb{R}^{3m+2}$$

Notice that we are estimating both μ_t and σ_t jointly. In unreported set of results, we find that estimating both moments jointly yields better estimates of μ_t. We derive the form of the policy for updating σ_t through a similar set of approximations as thouse used to estimate μ_t.[7] The rewards in this setting are defined as

$$r_t\left(s^t\right) = \begin{cases} 0 & \text{if } t < T \\ -\left\|\begin{pmatrix} \hat{\mu}_{T|T-1} \\ \hat{\sigma}_{T|T-1} \end{pmatrix} - \begin{pmatrix} \mu_T \\ \sigma_T \end{pmatrix}\right\| & \text{if } t = T \end{cases}$$

where μ_T and σ_T are the observed moments computed from the cross-section of forecasts released quarterly. With this formalism, we can directly apply the algorithm to learn the optimal parameters $\boldsymbol{\lambda}$ in-sample and subsequently use the learned $g_{\boldsymbol{\lambda}}$ to estimate the daily cross-sectional mean.

[7] Since we have the following expression for the cross-sectional variance:

$$\sigma_t^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\mu_{i,t} - \mu_t\right)^2$$

substituting in the approximations

$$\mu_{i,t} = c_{0,t}^i + c_{1,t}^i\mu_{i,t-1} + \left(\mathbf{c}_{2,t}^i\right)'\mathbf{r}_t \approx c_{1,t}^i\mu_{i,t-1} + \left(\mathbf{c}_{2,t}^i\right)'\mathbf{r}_t$$
$$\mu_t = c_0 + c_1\mu_{t-1} + \mathbf{c}_2'\mathbf{r}_t \approx c_1\mu_{t-1} + \mathbf{c}_2'\mathbf{r}_t$$

from 2.5 yields:

$$\begin{aligned}\sigma_t^2 &= \frac{1}{N}\sum_{i=1}^{N}\left[\left(c_{1,t}^i\mu_{i,t-1} - c_1\mu_{t-1}\right)^2 + \left(\left(\mathbf{c}_{2,t}^i\right)'\mathbf{r}_t - \mathbf{c}_2'\mathbf{r_t}\right)^2 + 2\left(c_{1,t}^i\mu_{i,t-1} - c_1\mu_{t-1}\right)\left(\left(\mathbf{c}_{2,t}^i\right)'\mathbf{r}_t - \mathbf{c}_2'\mathbf{r}_t\right)\right] \\ &\approx c_3\sigma_{t-1}^2 + \mathbf{c}_4'\mathbf{r}_t\mathbf{r}_t'\mathbf{c}_4 + \mathbf{c}_5'\mathbf{r}_t\mu_{t-1}\end{aligned}$$

3.4 A Three-Way Comparison

In this section we provide a brief comparison of the three cross-sectional estimation methods discussed above. The two main differences among the methods are 1) the interpretation of the output of each method and 2) the bias-variance tradeoff each method incurs.

Whereas the RL and KF approaches yield daily estimates of the current latent cross-sectional mean expectation, MIDAS produces a prediction of the end-of-quarter cross-sectional mean expectation. To be concrete, let $\mathcal{F}_t^F$ and $\mathcal{F}_t^E$ represent the time t information sets of the panel of forecasters and the econometrician, respectively. In this notation, on date t within a quarter between two consecutive survey releases, the cross-sectional mean expectation μ_t is in $\mathcal{F}_t^F$ but is not in $\mathcal{F}_t^E$. On day t the RL and KF approaches generate estimates of $\mathbb{E}\left[\mu_t \mid \mathcal{F}_t^E\right]$ while MIDAS yields an estimate of $\mathbb{E}\left[\mu_T \mid \mathcal{F}_t^E\right]$. Note the difference in the subscripts on μ.[8] In this setting, we explicitly seek to estimate a daily series of $\mathbb{E}\left[\mu_t \mid \mathcal{F}_t^E\right]$, not $\mathbb{E}\left[\mu_T \mid \mathcal{F}_t^E\right]$. Thus, the RL and KF approaches prove better suited to our setting than the MIDAS approach. Indeed, as we verify in simulations in the next section, the $\{\mathbb{E}\left[\mu_T \mid \mathcal{F}_t^E\right]\}_{t=0}^T$ series output by MIDAS aligns poorly with the true underlying $\{\mu_t\}_{t=0}^T$ series in a quarter.

Moreover, the RL approach proves more efficient in this setting due to its parsimony. All three methods prove biased in this setting, as they all rely on the approximate law of motion of the cross-sectional mean given in (9) as opposed to keeping track of the entire panel of individual forecaster expectations. Nonetheless, the KF approach most closely exploits the structure of the state-space system, and so should prove least biased.[9] However, this reduction in bias comes at the expense of greater variance. The KF approach necessitates estimation of $3m + 11$ parameters while the RL method requires estimating only $m + 1$ parameters, where m is the number of assets. On the other hand, since we fit a separate MIDAS regression for each day in the quarter, the MIDAS approach involves estimating $60\,(m + 4)$ parameters, assuming 60 trading days within a quarter. Thus, due to its parsimony in the number of estimated parameters, the RL approach proves far more efficient than the other two methods. In Appendix E, we verify in simulations of a stylized filtering task that while the KF achieves lower bias than the RL approach, the RL method attains greater efficiency.

In Appendix C, we compare the performance of these three methods in a simulated version of our cross-sectional expectation filtering task.

[8] Note that (9) implies that μ_t is not a $\mathcal{F}_t^E$-adapted martingale, so $\mathbb{E}\left[\mu_t \mid \mathcal{F}_t^E\right] \neq \mathbb{E}\left[\mu_T \mid \mathcal{F}_t^E\right]$.

[9] More specifically, the Kalman filter should prove least biased asymptotically, but since maximum-likelihood estimation of the filter parameters is biased in finite samples, which of the three methods proves least biased in our setting remains an empirical question.

4 Empirical Performance of RL

In this section, we use the cross-sectional RL approach from Section 3.3 to estimate a daily series of cross-sectional mean of GDP growth forecasts. We also compare its performance to that of three benchmarks: KF, MIDAS, and a "Naive" policy.

4.1 Results

We follow the estimation of our RL policy as described in Section 3.3 with a few minor modifications. Recall that the policy function of interest is of the following form:

$$g_{\boldsymbol{\lambda}}\left(s_t\right) \equiv \begin{pmatrix} \mu_t \\ \sigma_t \end{pmatrix} = \begin{pmatrix} c_1\mu_{t-1} + \mathbf{c}_2'\mathbf{r}_t \\ \sqrt{c_3\sigma_{t-1}^2 + \mathbf{c}_4'\mathbf{r}_t\mathbf{r}_t'\mathbf{c}_4 + \mathbf{c}_5'\mathbf{r}_t\mu_{t-1}} \end{pmatrix} \in \mathbb{R}^2$$

In a slight deviation from the setup in Section 3.3, we fix the coefficient on the one-day lag cross-sectional mean (c_1) to be equal to one. In unreported set of results, we find that this approach performs better than freely estimating c_1. This result proves reminiscent of Meese and Rogoff (1983) in which the authors find that a random walk outperforms more sophisticated methods in forecasting exchange rates out-of-sample.

We conduct a recursive out-of-sample estimation procedure as illustrated in Figure 4. For each out-of-sample quarter t between 2005:Q3 and 2018:Q4, the RL policy is trained via COPDAC-LSTDQ on a rolling lookback window of T quarters, from the SPF release date in quarter $t-T$ to the release date in quarter t. With the initial state set to the cross-sectional mean and standard deviation of the SPF release at the end of quarter t, the trained policy updates the state each day based on observed asset returns until the SPF release at the end of quarter $t+1$.

We compare the performance of our RL approach to that of three benchmarks. The first two are KF and MIDAS where we follow the procedures described in Sections 3.1 and 3.2 on the Survey of Professional Forecasters (SPF) data. The third benchmark is a naive policy that assumes that the cross-sectional mean expectation only updates on survey release days; it is constant within each quarter. For all benchmarks, to reduce estimation variance, we train six models on overlapping lookback windows of $T = 40, 44, 48, 52, 56,$ and 60 quarters and average the resulting output series.[10] As discussed in Section 2, we use returns on the CRSP value-weighted portfolio and the CRSP U.S. Treasury five-year fixed-term index as our test assets.

Overall, we find that the RL approach yields a daily estimated cross-sectional mean series that accurately reflects the true quarterly mean series. We judge the accuracy of our daily series by comparing the estimated cross-sectional mean from the RL policy on SPF release dates to the

[10] For the RL approach, averaging the output series is equivalent to averaging the policy weights. Therefore the averaging is akin to a simplified version of a method known as bootstrap aggregating or "bagging" in the machine learning literature.

actual counterparts of those values on those dates. For example, the true cross-sectional mean from the SPF release on August 15th, 2005 is 3.46%, whereas the daily estimated value on that day is 3.79% from the RL policy. This yields an absolute error of 0.33%.

Table 4 and Figures 5–8 present these out-of-sample estimation results for the RL approach and the three benchmarks. Across the entire out-of-sample period, the RL-estimated series realizes a RMSE of 0.449 percentage points and an R^2 of 82.3% versus the true quarterly series. These results prove far superior to those attained by our benchmarks: naive policy (RMSE of 0.588 percentage points and R^2 of 64.7%), MIDAS (RMSE of 0.916 percentage points and R^2 of 39.2%), and KF (RMSE of 39.103 percentage points and R^2 of 2.37%).

Figure 5 compares the daily cross-sectional mean series from the RL approach to the true SPF quarterly counterpart, and Table 5 displays the summary statistics of the daily series. Comparing the values in Table 5 to those in Table 2, we see that the RL daily series has a similar mean and standard deviations to its true quarterly counterpart. However, the RL daily series prove sfar more persistent than the quarterly version. Moreover, the daily series is highly kurtotic, both in levels (7.006) and in changes (17.282) due to the high kurtosis of the underlying daily asset returns.

In Table 6, we compute the correlations among the estimated daily series and the returns of assets used in our estimation. For the RL approach, we see that the daily cross-sectional mean innovations from the RL approach are positively correlated with U.S. equity returns (0.87) and negatively correlated with 5-year bond returns (-0.16). The KF series does not seem to load on either asset. The MIDAS series loads slightly on U.S. equity returns (.11).

Figure 9 plots the one-year rolling correlation between changes in the RL-estimated daily cross-sectional mean series and returns to our two assets: CRSP value-weighted portfolio and the CRSP five-year fixed-term index. These correlations display significant time-variation, which may explain significant time-variation in the policy weights from the RL approach as depicted in Figure 10.

As a further validation of our measure, we compare the most significant innovations in our cross-sectional mean series and compare them to macroeconomic events. Table 7 exhibits the ten largest absolute daily innovations in the cross-sectional mean series.[11] We find that most of these days correspond to significant macroeconomic events in response to which investors likely did update their growth expectations. As an example, the most significant update in our sample is August 8th, 2011 or the "Black Monday," the first trading day after Standard & Poor's downgraded the United States' sovereign debt rating. On that day, our series indicates that investors lowered their expectations for the next quarter's GDP growth by 0.65%. The next day when the Federal Reserve released a statement pledging to keep rates "exceptionally low" until mid 2013, investor expectations of next quarter's GDP growth rose by 0.51%. Most of other dates occur in late 2008 and early 2009 and correspond to developments in the Great Recession.

[11] We exclude SPF release dates from this table since, by construction, the daily series jump on those days to match the true cross-sectional moments.

4.2 Discussion

Origins of RL's Outperformance The core difficulty of our task is obtaining a *daily* law of motion for expectations given quarterly training data. The KF approach accomplishes this task by imposing parametric assumptions and using maximum likelihood. The RL approach builds upon the KF approach by directly estimating the Kalman gain using a linear learning rule, and we have illustrated the bias-efficiency trade-off that occurs at this step. The MIDAS benchmark also estimates a linear learning rule from quarterly data.

Recall that in Section 4, we fix the RL policy coefficient on the lag-one-day mean at 1. Thus, at any day in the quarter, the current estimate is the sum of the value from the start of the quarter and a series of daily asset returns. On the other hand, if we do not fix the coefficient at 1, then current estimate is an exponentially-weighted moving average of the value from the start of the quarter and the daily returns since its release.

As discussed in Section 3.4, unlike the KF and RL approaches, MIDAS estimates $\{\mathbb{E}\left[\mu_T \mid \mathcal{F}_t^E\right]\}_{t=0}^T$ instead of $\{\mathbb{E}\left[\mu_t \mid \mathcal{F}_t^E\right]\}_{t=0}^T$. However, this feature does not hinder the performance of MIDAS in the context of the results in Table 4 since we do not observe the intra-quarter expectations and so only compare end-of-quarter estimates to end-of quarter observed values (i.e. we compare $\mathbb{E}\left[\mu_T \mid \mathcal{F}_T^E\right]$ versus μ_t for all four methods). Thus, we cannot attribute MIDAS's poorer performance here to the fact that it estimates a quantity different from that of interest. Indeed, focusing just on the last-day-of-quarter estimates, the RL and MIDAS approaches involve estimating similar numbers of parameters ($m + 1$ for RL and $m + 4$ for MIDAS), so we cannot attribute the RL approach's significant outperformance here to a large efficiency advantage.[12] Instead, the primary difference between the two methods in the context of Table 4 appears to be how each approach treats lagged asset returns. The RL approach uses only asset returns since the start of the last survey release, weighting them uniformly, while the MIDAS approach applies a non-monotonic weighting scheme to 90 days of lagged asset returns, which extend back roughly 1.5 quarters.[13] Empirically, the RL approach's treatment of lagged asset returns proves more useful.

On the other hand, we do attribute the RL approach's superior performance versus the KF to the large efficiency advantage the RL approach realizes by estimating fewer parameters than the KF. This point was illustrated in detail in Section 3.4 and the Appendix.

Hyperparameters Our RL algorithm involves two hyper-parameters: the step size and the noise in behavioral policy.

The step size appears in each iteration of the COPDAC-LSTDQ algorithm, which is based on

[12] Of course, when it comes to estimating the daily series of expectations across the entire quarter, the RL approach proves much more efficient than the MIDAS approach, since a separate MIDAS model must be fit for each day in the quarter.

[13] We repeated the analysis using 30 days but the results were not materially different.

gradient ascent. If the step size is too small, one may get stuck in a local maximum. If the step size is too large, the algorithm may have trouble converging.

The noise in the behavioral policy regulates the tradeoff between exploration and exploitation. In RL, exploration refers to the notion that performing ex-ante suboptimal actions can improve long-term model performance due to the information gained from pursuing these actions. Exploitation refers to performing the ex-ante optimal action to maximize short term reward. In our setting, adding noise to the behavioral policy allows the algorithm to explore a wider set of potentially policy weights. Too little exploration will result in a suboptimal policy; too much exploration will prevent the algorithm from making proper gradient updates to the weights because any reduction in loss from changing the parameters will be drowned out in the noise.

While our baseline model currently uses a fixed step size and noise, a proper hyperparameter optimization procedure would involve the following steps:

1. Divide the sample into a training subsample and a pseudo-testing subsample.

2. From a grid of hyperparameters, train a model at each grid point on the training subsample and test on the pseudo-testing subsample.

3. Choose the set of hyperparameters that performs best in the pseudo-testing subsample to use in the RL algorithm for fitting the policy to the entire in-sample series.

Asset Selection Since the optimal policy in the RL approach is a linear function of the returns, a reasonable bound on the number of parameters is critical for efficiency. One way of imposing such a bound is to run a LASSO regression of forecast innovations on asset returns. Then one can consider modeling the optimal policy as a function of returns on these assets that survive the selection process.

An alternate approach is "boosting." We would start with only the five-year fixed term index, which has the highest R^2 in univariate regressions of quarterly forecast innovations on asset returns, apply the RL algorithm and construct quarterly residuals between the true SPF data and the estimated mean series. Then we would run univariate regressions of the residuals on quarterly returns of the remaining assets and select the asset with the highest R^2. Continued application of these steps yields an iterative approach in selecting assets to be used in our RL policy function.

5 Testing the "Fed Information Effect"

In this section, we use our obtained daily time-series of real GDP expectations to sharply identify the effect of FOMC announcements on growth expectations. Using our measure, we find no evidence of the "Fed information effect."

Nakamura and Steinsson (2018) regress the change in real GDP growth expectations from before and after an FOMC announcement on the monetary policy shock from the announcement. Specifically, they use the monthly Blue Chip survey for growth expectations and the first principal component of the thirty-minute changes around FOMC announcements for five interest rate instruments.[14] This regression yields the "wrong" sign: positive. Hawkish surprises for interest rates correspond to increases in real GDP growth expectations and dovish surprises correspond to decreases in real GDP growth expectations. This empirical result proves contrary to standard macro models in which positive shocks to interest rates are viewed as contractionary. Some researchers have viewed the wrong sign in these regressions as evidence that forecasters view hawkish surprises as a signal that the Fed's private forecasts of GDP growth are more positive than their own. Thus, forecasters revise their growth expectations upward.

The shortcoming of this line of analysis is that the low frequency of surveys prevents identification of the effect of monetary policy shocks on growth expectations. Nakamura and Steinsson (2018) use monthly changes in growth expectations. But surely the monetary policy shock is not the only relevant macro development in any month; investors may update their expectations due to many other factors. Bauer and Swanson (2020) provide anecdotal evidence of the absence of Fed information effects using the daily forecasts of a single forecaster. Bauer and Swanson (2020) also argues that the observed positive coefficient in the usual Fed information effect regressions is also consistent with both the forecasters and the Fed responding to common macro news released between survey dates.

Indeed, the usual Fed information effect regression is

$$\mathbb{E}_{t+15}\left[g_Q\right] - \mathbb{E}_{t-15}\left[g_Q\right] = \beta_0 + \beta_1 \text{Shock}_t + \epsilon_t, \tag{16}$$

where $\mathbb{E}_{t+15}\left[g_Q\right] - \mathbb{E}_{t-15}\left[g_Q\right]$ is the one-month change in surveyed expectations around the FOMC announcement on day t and Shock_t is the monetary policy shock on the FOMC announcement. In this regression, there is an omitted variable: economic news released between day $t - 15$ and day $t - 1$. If the omitted economic news is positively correlated with both the montary policy shock and the monthly change in growth expectations, then the estimated β_1 in (16) will be positively biased. Bauer and Swanson (2020) provide evidence of these positive correlations by demonstrating that non-farm payroll numbers released between $t - 15$ and day $t - 1$ correlate positively with both the monthly change in Blue Chip growth expectations and with the Nakamura and Steinsson (2018) monetary policy shock.

Therefore, we have reason to believe that the puzzling positive sign of the estimated β_1 in (16) may simply arise due to omitted variable bias instead of from a Fed information effect. The

[14] The five instruments are: the expected Fed funds rate for immediately following the current and next FOMC meetings (extracted from Fed funds futures) and the expected three-month eurodollar interest rates at horizons of two, three, and four quarters (extracted from eurodollar futures).

announcement-day change (from $t-1$ to t) in our daily series of growth expectations is uncorrelated with omitted economic news since the growth expectation at time $t-1$ already incorporates all news from $t-15$ and day $t-1$. Hence, we can run regression (16) without fear of omitted variable bias.

In Table 8, we report the response of our expected growth series to policy and federal funds rate shocks. Specifically, we estimate a regression of the following form:

$$\Delta CXMean_t = \beta_0 + \beta_1 \text{Shock}_t + \epsilon_t \tag{17}$$

at a daily frequency where the $CXMean_t$ is the daily series of growth expectations obtained from our RL algorithm. The policy news and federal funds rate shocks are obtained from Nakamura and Steinsson (2018).

For the full sample (January 2004 to March 2014), we find that the estimated coefficient β_1 is statistically significant with magnitude of -0.83. Note that Nakamura and Steinsson (2018) scale their policy news shock variable such that its effect on the one-year nominal Treasury yield is equal to one. Therefore, the coefficient of -0.83 is interpreted as saying a policy news shock that is equivalent to a 100 bps increase in one-year Treasury yield leads to on average 83 bps decrease in real GDP growth expectations. The size of this effect is economically meaningful as a policy news shock as large as a 100 bps increase in one-year Treasury yield is quite rare. Note that Nakamura and Steinsson (2018) find a significantly positive coefficient over a similar time period.

In the remaining cells of the table, we report results for a subsample that excludes the height of the Great Recession and an analysis using the fed funds rate shock instead of the policy rate shock. The sign and magnitude of the coefficient, as well as its statistical significance, implies that investors adjust their growth expectations downward in responses to hawkish surprises. Thus, we do not find evidence of the Fed information effect.

Furthermore, in Figures 11 (full-sample) and 12 (excluding 2008:06 – 2009:06), we display the results from running regression (17) within each year from 2004 to 2013. We find significantly negative or insignificant coefficients across all years (with the exception of 2008 and 2009 in Figure 12, in which the regressions exclude the height of the Great Recession 2008:06 – 2009:06). Thus, we do not find evidence of time-varying Fed information effects across the business cycle.

6 Conclusion

We propose a simple reinforcement learning approach using asset prices to estimate high-frequency expectations of macroeconomic growth. Specifically, we provide a framework for constructing daily series of the cross-sectional mean of growth forecasts and find that our method proves efficient and robust to model specifications. Our approach achieves greater efficiency than the traditional

Kalman filter and the MIDAS regression by estimating the optimal gain directly rather than via the structural parameters of an underlying state-space model.

We apply our RL approach to obtain a daily time-series of growth expectations, which we use to test the existence of the Fed information effect. While traditional tests of the Fed information effect use low-frequency changes in surveyed expectations around FOMC announcements, our RL-estimated daily time series allows us to capture the high-frequency change in growth expectations on these days. Doing so obviates the omitted variable bias that plagues regressions with low-frequency changes in expectations. Using our RL-estimated series, we find no evidence of the Fed information effect: hawkish monetary surprises correspond to decreases in growth expectations.

Our paper is the first serious application of reinforcement learning in the growing literature that uses machine learning methods in finance. We have presented reinforcement learning as a more efficient improvement over traditional filtering methods. Given the low frequency of surveyed expectations and abundance of asset price data, our RL approach proves promising for extracting investor expectations of macroeconomic variables.

Our RL approach can obtain a daily series of expectations for any macroeconomic variable with a low-frequency panel of forecasts. The accuracy of the daily series would depend on the availability of training data and the relevance of the macroeconomic variable for asset returns. Thus, interest rates and inflation expectations represent good candidates given their long time-series and impact on government bond returns. Existing literature uses derivatives to construct high-frequency interest rate expectation series, so it would be interesting to see how our approach compares to existing methods.

Testing theories of expectations formation represents another application of our framework. For example, suppose we want to test the hypothesis that agents update their expectations about growth after observing commodity prices. One way to empirically test this hypothesis is to include commodity returns in the state vector and fit the optimal policy for updating growth expectations. The coefficient on commodity returns in the optimal policy reflects if an agent would find it optimal to use commodity returns in updating his expectation of growth.

Finally, our method can be used to construct firm-specific cash flow expectations at a daily frequency. Instead of using value-weighted equity and bond returns, we can use firm's equity returns and corporate bond returns. We can also use quarterly expectations from analyst forecasts to conduct the same exercise and obtain daily expectations.

A Appendix: Campbell-Shiller Approximation

Recall the state-space model:

$$\begin{aligned}
\theta_{t+1} &= \mu + \delta\theta_t + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N\left(0, \sigma_\epsilon^2\right) \\
\zeta_{t+1} &= \tau + \psi\zeta_t + \xi_{t+1}, \quad \xi_{t+1} \sim N\left(0, \sigma_\xi^2\right) \\
d_{t+1}^i - d_t^i &= \gamma + \beta^i\theta_{t+1} + \nu_{t+1}^i, \quad \nu_{t+1}^i \sim N\left(0, \sigma_\nu^2\right) \\
\mathbb{E}_t\left[r_{t+1}^i\right] &= \alpha + \phi^i\zeta_t \\
\operatorname{Corr}\left(\epsilon_t, \xi_t\right) &= \pi
\end{aligned}$$

Campbell and Shiller (1988) develop a useful approximation to the present-value formula:

$$p_t^i \equiv \log P_t^i = \frac{\kappa}{1-\rho} + (1-\rho)\sum_{j=0}^{\infty}\rho^j E_t\left[d_{t+j+1}^i\right] - \sum_{j=0}^{\infty}\rho^j E_t\left[r_{t+j+1}^i\right] \tag{18}$$

where

$$\rho = \frac{1}{1+\exp\left(\overline{d-p}\right)}, \quad \kappa = -\log\rho - (1-\rho)\log\left(\frac{1}{\rho} - 1\right)$$

From the law of motion for θ_t, we obtain:

$$\mathbb{E}_t\left[\theta_{t+j}\right] = \left(\mu + \delta\mu + \cdots + \delta^{j-1}\mu\right) + \delta^j\theta_t = \mu\frac{1-\delta^j}{1-\delta} + \delta^j\theta_t$$

Similarly iterating $\mathbb{E}_t\left[r_{t+1}^i\right]$ yields:

$$\mathbb{E}_t[r_{t+j+1}^i] = \begin{cases} \alpha + \phi^i\tau\frac{1-\psi^j}{1-\psi} + \phi^i\psi^j\zeta_t & j > 1 \\ \alpha + \phi^i\zeta_t & j = 1 \end{cases}$$

Therefore:

$$\begin{aligned}
\sum_{j=0}^{\infty}\rho^j E_t[r_{t+j+1}^i] &= E_t\left[r_{t+1}^i\right] + \sum_{j=1}^{\infty}\rho^j\left[\alpha + \phi^i\tau\frac{1-\psi^j}{1-\psi} + \phi^i\psi^j\zeta_t\right] \\
&= \left(\alpha + \phi^i\zeta_t\right) + \frac{\alpha\rho}{1-\rho} + \frac{\phi^i\tau\rho}{(1-\rho)(1-\psi)} - \frac{\psi\rho\phi^i\tau}{(1-\psi)(1-\psi\rho)} + \frac{\phi^i\rho\psi}{(1-\rho\psi)}\zeta_t \\
&= \frac{\alpha}{1-\rho} + \frac{\phi^i\tau\rho}{(1-\rho)(1-\psi)} - \frac{\psi\rho\phi^i\tau}{(1-\psi)(1-\psi\rho)} + \frac{\phi^i\zeta_t}{(1-\rho\psi)} \qquad (19)
\end{aligned}$$

Proceeding analogously for $\mathbb{E}\left[d_{t+j+1}\right]$, denote $g_{t+1} = d_{t+1} - d_t$ and write:

$$\begin{aligned}
\mathbb{E}_t\left[d_{t+j+1}\right] &= d_t + \sum_{k=0}^{j} \mathbb{E}_t\left[g_{t+k+1}\right] \\
E_t\left[g_{t+k+1}\right] &= \gamma + \beta^i \mu \left(\frac{1-\delta^{k+1}}{1-\delta}\right) + \beta^i \delta^{k+1} \theta_t
\end{aligned}$$

Combining the expressions:

$$\begin{aligned}
\sum_{j=0}^{\infty} \rho^j E_t\left[d_{t+j+1}\right] &= \sum_{j=0}^{\infty} \rho^j \left\{ d_t + \sum_{k=0}^{j} E_t\left[g_{t+k+1}\right] \right\} \\
&= \frac{d_t}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \sum_{k=0}^{j} E_t\left[g_{t+k+1}\right] \\
&= \frac{d_t}{1-\rho} + \sum_{j=0}^{\infty} \rho^j \sum_{k=0}^{j} \left\{ \gamma + \beta^i \mu \left(\frac{1-\delta^{k+1}}{1-\delta}\right) + \beta^i \delta^{k+1} \theta_t \right\}
\end{aligned}$$

Simplifying the algebra yields:

$$\sum_{j=0}^{\infty} \rho^j E_t\left[d_{t+j+1}\right] = \frac{d_t}{1-\rho} + \frac{\gamma}{(1-\rho)^2} + \frac{\beta^i \mu}{(1-\delta)(1-\rho)^2} - \frac{\delta \beta^i \mu}{1-\delta} \frac{1}{(1-\rho)(1-\rho\delta)} + \frac{\delta \beta^i \theta_t}{(1-\rho)(1-\rho\delta)} \tag{20}$$

Using (19) and (20) to simplify (18), we arrive at:

$$\begin{aligned}
r_{t+1}^i &\equiv p_{t+1}^i - p_t^i \\
&= \left(d_{t+1}^i - d_t^i\right) + \frac{\delta \beta^i}{1-\rho\delta} (\theta_{t+1} - \theta_t) - \frac{\phi^i}{1-\rho\psi} (\zeta_{t+1} - \zeta_t)
\end{aligned}$$

Using the expression for $d_{t+1} - d_t$, we arrive at the desired expression:

$$r_{t+1}^i = \gamma + \left(\beta^i + \frac{\delta \beta^i}{1-\rho\delta}\right) \theta_{t+1} - \frac{\delta \beta^i}{1-\rho\delta} \theta_t - \frac{\phi^i}{1-\rho\psi} (\zeta_{t+1} - \zeta_t) + \nu_{t+1} \tag{21}$$

■

B Appendix: Formal Representation of the RL Algorithm

We now formalize the intuition provided above using the familiar notation of a generic dynamic programming problem. In general, RL algorithms address the following maximization problem:

$$V\left(s_0\right) = \max_{\{a_t\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \gamma^t r_t\left(s^t\right) \text{ such that } s_{t+1} = \mathcal{T}\left(a_t, s_t\right)$$

where γ is the discount factor, $s^t = \left(s_1, ..., s_t\right)$ is a history of states, $r_t\left(s^t\right)$ is the reward at period t given history s^t, and $\mathcal{T}$ is the state transition function. Rewriting this problem in the recursive form yields:

$$V\left(s_t\right) = \max_{a_t}\left\{r_t\left(s^t\right) + \gamma V\left(\mathcal{T}\left(a_t, s_t\right)\right)\right\}$$

Now define the *action-value* function:

$$Q\left(a_t, s_t\right) = r_t\left(s^t\right) + \gamma V\left(\mathcal{T}\left(a_t, s_t\right)\right) \tag{22}$$

which gives the value of taking action a_t in state s_t. Using Q, the optimal policy is defined as

$$g\left(s_t\right) \equiv \arg\max_{a_t} Q\left(a_t, s_t\right) \tag{23}$$

In our setting, a_t is the agent's updated growth expectation in period t. We correspondingly define the rewards as:

$$r_t\left(s^t\right) = \begin{cases} 0 & \text{if } t < T \\ -\left|\left|\hat{\mu}_{T|T-1} - \mu_T\right|\right| & \text{if } t = T \end{cases}$$

Thus, the agent only earns a reward at the end of the quarter based on the distance between his terminal growth expectation $\left(\hat{\mu}_{T|T-1}\right)$ and the surveyed growth expectation $\left(\mu_T\right)$ that he observes at the end of the quarter.

The most basic RL algorithms discretize the state and action spaces into a finite grid and then estimate $Q\left(a_t, s_t\right)$ nonparametrically by randomly exploring different actions in different states and observing the resulting rewards. Note that it is Q that is being estimated, not V. After having estimated Q, the optimal policy directly comes from (23). These methods, however, prove inefficient in high-dimensional and continuous state and action spaces, such as our setting since $a_t = \hat{\mu}_{t+1|t}$ is continuous.

We instead use linear approximations to the optimal policy and action-value functions to improve the estimation efficiency. To efficiently learn the optimal policy weights $\boldsymbol{\lambda}$, we propose the COPDAC-LSTDQ algorithm, inspired by the COPDAC-Q algorithm of Silver et al. (2014). We delegate the details of COPDAC-LSTDQ to Appendix D. Here we discuss the intuition behind the original COPDAC-Q algorithm and our subsequent improvement.

The COPDAC-Q algorithm performs gradient ascent on a linear approximation of the action-value function Q with respect to the policy parameter vector $\boldsymbol{\lambda}$. Every time the agent performs an action and earns a reward, the algorithm uses that reward to 1) update the approximate Q function and 2) move $\boldsymbol{\lambda}$ in the direction of $\nabla_{\boldsymbol{\lambda}} Q$. Unfortunately, as noted by Silver et al. (2014) this iterative approach known as "Q-learning" has poor convergence properties due to noise in the individual updates (i.e. the learned $g_{\boldsymbol{\lambda}}$ may not converge to the optimal policy). Thus, to reduce the update variance, we propose a modified method that we call COPDAC-LSTDQ. To be concrete, instead of updating the action-value and policy functions after each action, we let the agent take a sequence of actions, observe the resulting rewards, and essentially update the action-value and policy function parameters based on the average reward realized. While the reward to a single action may prove noisy, averaging the rewards over a batch of actions should yield lower variance parameter updates.[15] We refer the reader to the online appendix D for a more detailed discussion.

For the remainder of the paper, we refer to the application of the COPDAC-LSTDQ to learn the optimal policy function $g_{\boldsymbol{\lambda}}$ as the "RL approach."

C Appendix: Comparison of KF, RL, and MIDAS in Simulations

In this section, we compare the performance of the three methods in a simulated version of our cross-sectional expectation filtering task.

The main task of interest is to estimate the mean of a cross-section of GDP growth forecasts at a daily frequency, given corresponding moments from a quarterly survey. In this section, we illustrate in a simulated economy that RL outperforms KF and MIDAS in this task. The measures of performance are the RMSE and correlation between the estimated and true moment series.

To start the simulation, we first calibrate the system by estimating equations (3) – (6) and (8) using annual data. The calibration is summarized in Table 9. Parameters for GDP growth $(\mu, \delta, \sigma_{\epsilon}^2,)$ are estimated by fitting an AR(1) to annual real GDP growth from 1931 to 2018. Parameters for dividend growth $(\gamma, \mathbb{E}\left[\beta^i\right], \sigma_{\nu}^2)$ are estimated by regressing annual dividend growth of the S&P 500 on contemporaneous annual GDP growth from 1994 to 2018. Parameters for dynamics of the latent factor $\left(\tau, \psi, \sigma_{\xi}^2\right)$ are estimated by fitting an AR(1) to the year-end market-to-book ratio of the S&P 500 from 1977 to 2018. Parameters for the dynamics of conditional expected returns $(\alpha, \mathbb{E}\left[\phi^i\right])$ are obtained from the following regression: $R_{t+1} = \alpha + \beta\left(B/M\right)_t + u_{t+1}$ using S&P 500 returns and book-to-market ratio. The term from the Campbell-Shiller expression (ρ) is estimated from the dividends and price of S&P500 from 1994 to 2018, and the correlation between innovations (π) is

[15] The name "COPDAC-LSTDQ" stems from our use of batch "LSTDQ updates" to the action-value approximating function instead of iterative "Q updates."

computed from covariance of θ_t and ζ_t.

Next, we simulate panels of growth expectations and returns as described in Section 3. We first generate a forty-quarter series of daily growth expectations and asset returns for $m = 10$ assets. We generate a panel of daily growth expectations by having each of $N = 20$ agents form his expectations about growth from observed asset returns via a Kalman filter, as described in Section 2.4. For asset returns, we calibrate the first asset to have $\beta^i = \mathbb{E}[\beta^i]$ and $\phi^i = \mathbb{E}[\phi^i]$ from the calibration; for the remaining nine assets, we draw random pairs of (β^i, ϕ^i) where β^i is independently uniformly distributed between 0 and 1 and ϕ^i is independently uniformly distributed between -0.1 and 0.[16] As done in the real SPF survey data, we compute the daily cross-sectional means from of the expectations panel and sample those series quarterly. Thus, we have a 40-quarter in-sample training series of quarterly cross-sectional means and daily asset returns. We then simulate 1,000 out-of-sample testing quarters of daily cross-sectional means and asset returns in this same manner.

With the constructed samples, we estimate four different update policies in-sample, and then test each on the 1,000 out-of-sample quarters. The first three methods for estimating the update policy are the RL, KF, and MIDAS approaches discussed in Section 3. As a baseline, we also consider a "naive" approach that estimates today's cross-sectional mean as its lag value. So each daily cross-sectional mean series updates only on the SPF release day and retains that value for the entire quarter until the next release. For each out-of-sample quarter and for each method, we compute the RMSE and correlation between the true and estimated daily cross-sectional mean series.

To examine the impact of heterogeneity, we conduct this simulation exercise for different degrees of learning heterogeneity.[17] We parameterize learning heterogeneity with a single signal-to-noise ratio that determines the variance of the distribution from which each agent draws his KF parameters. Specifically, for a signal-to-noise ratio of s and a true parameter value of θ_0, the distribution from which agents draw their values of this parameter is $\mathcal{N}\left(\theta_0, (\theta_0/s)^2\right)$. A lower signal-to-noise ratio therefore implies higher learning heterogeneity.

Figures 13 and 14 display the cross-sectional mean estimation results of these simulations. The x-axis is the signal-to-noise ratio, ranked from smallest learning heterogeneity (value of 10) to highest learning heterogeneity (value of 1). Figure 13 exhibits the average relative RMSE, defined as the RMSE divided by the average absolute cross-sectional mean, across all 1,000 out-of-sample quarters

[16] Given the expression for returns:

$$r^i_{t+1} = \gamma + \left(\beta^i + \frac{\delta\beta^i}{1-\rho\delta}\right)\theta_{t+1} - \frac{\delta\beta^i}{1-\rho\delta}\theta_t - \frac{\phi^i}{1-\rho\psi}\left(\zeta_{t+1} - \zeta_t\right) + \nu_{t+1}$$

β^i and ϕ^i along with the calibrated parameters are sufficient to simulate returns.

[17] Earlier, we defined *learning heterogeneity* to refer to the fact that each agent draws his value of the parameter from a normal distribution centered at the true parameter value.

at varying levels of learning heterogeneity. We see that the RL approach achieves relative RMSEs of at most 1.02 across all signal-to-noise ratios. The MIDAS and KF approaches realize much higher relative RMSEs (consistently about 1.40 for MIDAS and between 2.74 and 6.87 for the KF). Indeed, the KF performs more poorly than the naive approach of setting each day's expectation estimate equal to the observed value at the start of the quarter.

Figure 14 exhibits the average correlation across all 1,000 out-of-sample quarters at different levels of learning heterogeneity. For signal-to-noise ratios of two and below, the RL approach achieves correlations of greater than 0.79 while MIDAS fails to deliver a positive correlation, the KF fails to deliver a correlation greater than 0.15, and the naive approach achieves a correlation of zero (by construction).

In terms of both RMSE and correlation, the RL approach's performance degrades with higher learning heterogeneity, while the MIDAS and KF performances remain unaffected but poor. The degradation in RL's performance can be attributed to the poorer approximations in (9) as noise increases. MIDAS performs so poorly, especially in terms of correlation, precisely for the reason discussed in Section 3.4: MIDAS estimates $\{\mathbb{E}\left[\mu_T \mid \mathcal{F}_t^E\right]\}_{t=0}^T$, which in this setting need not bear any relation to the series of interest $\{\mu_t\}_{t=0}^T$. Moreover, the MIDAS approach estimates a very large number of parameters, thereby rendering its output high-variance. The KF's poor performance versus the RL approach also derives from its lack of efficiency.

Thus, we see that the RL approach outperforms the other approaches in cross-sectional mean estimation across essentially all levels of learning heterogeneity. The RL approach's outperformance derives, as we continue to emphasize, from its greater estimation efficiency. For m assets, the KF and MIDAS approaches must estimate $3m+11$ and $60(m+4)$ parameters, respectively, while the RL approach need only estimate $m+1$ parameters. Overall, these results indicate that the RL approach proves most useful for estimating cross-sectional forecast moments due to its greater efficiency and ability to perform well at relatively high levels of learning heterogeneity. Motivated by these results, we take these methods to actual forecasts from the Survey of Professional Forecasters (SPF) and construct daily series of expectations from asset returns.

References

Bauer, M. D. and E. T. Swanson (2020). The fed's response to economic newss explains the "fed information effect". *Working Paper*.

Bernanke, B. S., M. Gertler, and M. Watson (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings papers on economic activity 1997*(1), 91–157.

Brandt, M. W. and Q. Kang (2004). On the relationship between the conditional mean and volatility of stock returns: A latent var approach. *Journal of Financial Economics 72*(2), 217–257.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies 1*(3), 195–228.

Evans, M. D. (2005). Where are we now? real-time estimates of the macro economy. Technical report, National Bureau of Economic Research.

Freyberger, J., A. Neuhierl, and M. Weber (2017). Dissecting characteristics nonparametrically. Technical report, National Bureau of Economic Research.

Fuhrer, J. (2017). Expectations as a source of macroeconomic persistence: Evidence from survey expectations in a dynamic macro model. *Journal of Monetary Economics 86*, 22–35.

Gennaioli, N., Y. Ma, and A. Shleifer (2016). Expectations and investment. *NBER Macroeconomics Annual 30*(1), 379–431.

Ghysels, E., A. Sinko, and R. Valkanov (2007). Midas regressions: Further results and new directions. *Econometric Reviews 26*(1), 53–90.

Ghysels, E. and J. H. Wright (2009). Forecasting professional forecasters. *Journal of Business & Economic Statistics 27*(4), 504–516.

Giglio, S. and D. Xiu (2018). Asset pricing with omitted factors. *Chicago Booth Research Paper* (16-21).

Heaton, J., N. Polson, and J. H. Witte (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry 33*(1), 3–12.

Hutchinson, J. M., A. W. Lo, and T. Poggio (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance 49*(3), 851–889.

Kelly, B. and S. Pruitt (2013). Market expectations in the cross-section of present values. *The Journal of Finance 68*(5), 1721–1756.

Kelly, B. T., S. Pruitt, and Y. Su (2017). Instrumented principal component analysis. *Available at SSRN 2983919*.

Kozak, S., S. Nagel, and S. Santosh (2019). Shrinking the cross-section. *Journal of Financial Economics*.

Kozeniauskas, N., A. Orlik, and L. Veldkamp (2018). What are uncertainty shocks? *Journal of Monetary Economics 100*, 1–15.

Lagoudakis, M. G. and R. Parr (2003). Least-squares policy iteration. *Journal of machine learning research 4*(Dec), 1107–1149.

Lamont, O. A. (2001). Economic tracking portfolios. *Journal of Econometrics 105*(1), 161–184.

Lucca, D. O. and E. Moench (2015). The pre-fomc announcement drift. *The Journal of Finance 70*(1), 329–371.

Meese, R. A. and K. Rogoff (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics 14*(1-2), 3–24.

Moritz, B. and T. Zimmermann (2016). Tree-based conditional portfolio sorts: The relation between past and future stock returns. *Available at SSRN 2740751*.

Nakamura, E. and J. Steinsson (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics 133*(3), 1283–1330.

Neuhierl, A. and M. Weber (2018). Monetary momentum. Technical report, National Bureau of Economic Research.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International stock return predictability: what is the role of the united states? *The Journal of Finance 68*(4), 1633–1662.

Romer, C. D. and D. H. Romer (2000, June). Federal reserve information and the behavior of interest rates. *American Economic Review 90*(3), 429–457.

Silver, D., G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller (2014). Deterministic policy gradient algorithms.

Sirignano, J., A. Sadhwani, and K. Giesecke (2016). Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470*.

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

Van Binsbergen, J. H. and R. S. Koijen (2010). Predictive regressions: A present-value approach. *The Journal of Finance 65*(4), 1439–1471.

Watkins, C. J. C. H. (1989). Learning from delayed rewards.

Yao, J., Y. Li, and C. L. Tan (2000). Option price forecasting using neural networks. *Omega 28*(4), 455–466.

Table 1: **Summary of Data Sources (US)**

Asset Type	Data	Source
Equity		
Market	CRSP Value-weighted Return	CRSP
Industry	Industry Returns	Fama-French Library
Factors	Factor Returns	Fama-French Library
Fixed Income		
Government Bonds	Return on Fixed-term Indices	CRSP Treasuries
Government Bonds	Slope of Yield Curve	CRSP Treasuries
Credit	Change in AAA-10Y Spread	FRED
Credit	Change in BAA-10Y Spread	FRED
Exchange Rates		
USD	Change in Weighted Average of Forex Value	FRED
Derivatives		
Options	Change in VIX Index	CRSP Treasuries
Surveys		
Forecasts	Survey of Professional Forecasters (SPF)	Philadelphia Fed

The table summarizes the data sources. For equities, we consider daily returns on the CRSP value-weighted portfolio, Fama-French industry portfolios, and Fama-French factors. For government bonds, we consider daily returns on U.S. Treasury fixed-term indexes from CRSP, which comprise fully taxable, non-callable, non-flower bonds that best represent each term. The slope of the yield curve is defined as the return on the ten-year index minus five times the return on the two-year index. For credit, we consider the changes in spread betweeen the yields on corporate bonds with AAA and BAA Ratings. For exchange rates, we consider change in the weighted average of the foreign exchange value of the U.S. dollar (variable 'DTWEXM' from FRED), and for derivatives we use the change in the VIX index. For growth forecasts, we use quarterly forecast data from the Survey of Professional Forecasters (SPF).

Source: CRSP, FRED, Philadelphia Fed

Table 2: **Summary Statistics of GDP Growth and Forecasts**

Panel A: No Winsorization

	# Forecasters	# Months	CX Mean	CX Median	CX Std	Real GDP Growth
Mean	38.816	114	2.542	2.510	0.825	2.520
Std Dev	6.846	114	0.927	0.946	0.331	2.307
Autocorr(1)	·	114	0.734	0.735	0.449	0.359

Panel B: Winsorization at 5% Level

	# Forecasters	# Months	CX Mean	CX Median	CX Std	Real GDP Growth
Mean	35.956	114	2.531	2.510	0.683	2.520
Std Dev	6.014	114	0.929	0.946	0.228	2.307
Autocorr(1)	·	114	0.742	0.735	0.730	0.359

Panel C: Winsorization at 10% Level

	# Forecasters	# Months	CX Mean	CX Median	CX Std	Real GDP Growth
Mean	31.860	114	2.527	2.510	0.586	2.520
Std Dev	5.445	114	0.932	0.946	0.203	2.307
Autocorr(1)	·	114	0.745	0.735	0.722	0.359

The table reports the summary statistics for selected cross-sectional moments of one-quarter ahead forecasts from the Survey of Professional Forecasters (SPF) and realized GDP growth from FRED. In constructing the time-series, we perform the winsorization on each quarterly cross section.

Source: Survey of Professional Forecasters

Table 3: **Regressions of Forecast Innovations on Asset Returns**

Asset 1	Asset 2	Coeff. on Asset 1	Coeff. on Asset 2	R-squared
CRSP Value-weighted Return	5YR Fixed-term Index Return	0.036	-0.199	0.383
5YR Fixed-term Index Return	Change in BAA-10Y Spread	-0.178	-0.013	0.344
5YR Fixed-term Index Return	Change in Weighted Average of Forex Value	-0.242	-0.036	0.334
5YR Fixed-term Index Return	Change in VIX index	-0.225	-0.003	0.328
5YR Fixed-term Index Return	Slope of Yield Curve	-0.221	0.013	0.317
5YR Fixed-term Index Return	Change in AAA-10Y Spread	-0.213	-0.003	0.316
CRSP Value-weighted Return	Change in BAA-10Y Spread	0.030	-0.022	0.261
CRSP Value-weighted Return	Change in AAA-10Y Spread	0.038	-0.016	0.253
Slope of Yield Curve	Change in BAA-10Y Spread	0.029	-0.027	0.234
Change in BAA-10Y Spread	Change in VIX index	-0.029	-0.001	0.224
Change in AAA-10Y Spread	Change in BAA-10Y Spread	-0.002	-0.027	0.222
Change in BAA-10Y Spread	Change in Weighted Average of Forex Value	-0.030	0.005	0.222
CRSP Value-weighted Return	Slope of Yield Curve	0.050	0.054	0.217
Slope of Yield Curve	Change in AAA-10Y Spread	0.029	-0.019	0.188
Change in AAA-10Y Spread	Change in VIX index	-0.021	-0.002	0.188
Change in AAA-10Y Spread	Change in Weighted Average of Forex Value	-0.022	-0.023	0.184
CRSP Value-weighted Return	Change in Weighted Average of Forex Value	0.055	0.019	0.172
CRSP Value-weighted Return	Change in VIX index	0.059	0.002	0.172
Slope of Yield Curve	Change in VIX index	0.060	-0.003	0.092
Slope of Yield Curve	Change in Weighted Average of Forex Value	0.064	-0.005	0.071
Change in VIX index	Change in Weighted Average of Forex Value	-0.004	0.001	0.030

The table summarizes the results from time-series regressions of forecast innovations on asset returns from 1990:Q4 to 2018:Q4. Forecast innovation at quarter t defined as the difference between average SPF forecast of t growth made at quarter t $\left(\bar{f}_{t|t}\right)$ and the average forecast of t growth made at quarter $t-1$ $\left(\bar{f}_{t|t-1}\right)$. The average is computed after winsorizing the data at 5% level. We use returns between the consecutive release dates for SPF forecasts corresponding to the innovation. We compute the regressions for pairs of available assets.

Source: Compustat, CRSP, FRED

Table 4: **Recursive Out-of-Sample Estimation Results**

	RL Approach	Naive	MIDAS	KF
RMSE	0.449	0.588	0.916	39.103
R^2	0.823	0.647	0.392	0.0237

The table reports the RMSEs and correlations between the true SPF cross-sectional moment series and the estimated daily series on the SPF release dates. For all four methods we compute the corresponding daily series with the initial state set to the cross-sectional moments from the quarter t release and compare the daily values on the release date for quarter $t+1$ to the true values from that release. The RL approach learns the weights using our COPDAC-LSTDQ algorithm. The naive method places a weight of one on the lag cross-sectional moments and zero on all of the asset return terms. The RL approach uses For assets, we use returns on the CRSP value-weighted portfolio and the CRSP five-year fixed-term index. The MIDAS approach fits a separate model for each day where we use the "Beta Lag" specification from Ghysels et al. (2007). In the KF approach, we fit the model's structural parameters to the generated time-series via maximum likelihood, and the estimated parameters are used to compute the Kalman gain. For all methods, the resulting output series are averaged across outputs from estimation based on 40, 44, 48, 52, 56, and 60 quarters of training periods.

Source: CRSP, Survey of Professional Forecasters (SPF)

Table 5: **Summary Statistics of Estimated Daily Series**

	RL	Naive	MIDAS	KF
Panel A: Daily Series				
Mean	2.476	2.401	2.445	32.888
Std Dev	1.044	0.946	0.911	19.139
Autocorr(1)	0.997	0.997	0.673	0.960
Skewness	-2.431	-2.475	-2.912	2.652
Excess Kurtosis	7.006	7.208	27.095	12.608
Panel B: Change in Daily Series				
Mean	0.002	0.000	-.001	0.570
Mean of Absolute Values	0.036	0.000	0.385	0.597
Std. Dev.	0.060	0.000	0.736	1.923
Skewness	-0.670	0.000	0.920	6.998
Kurtosis	17.282	0.000	75.606	61.905

The table reports summary statistics for the daily cross-sectional mean estimated using the four approaches. Panel A provides the values computed for the levels of the daily series and Panel B provides the values computed for the daily changes in each series. The RL approach learns the weights using our COPDAC-LSTDQ algorithm. The naive method places a weight of one on the lag cross-sectional moments and zero on all of the asset return terms. The RL approach uses For assets, we use returns on the CRSP value-weighted portfolio and the CRSP five-year fixed-term index. The MIDAS approach fits a separate model for each day where we use the "Beta Lag" specification from Ghysels et al. (2007). In the KF approach, we fit the model's structural parameters to the generated time-series via maximum likelihood, and the estimated parameters are used to compute the Kalman gain. For all methods, the resulting output series are averaged across outputs from estimation based on 40, 44, 48, 52, 56, and 60 quarters of training periods.

Source: CRSP, Survey of Professional Forecasters (SPF)

Table 6: **Correlation Structure of Time-Series Changes in Daily Series**

	USA	ret_5yr	RL	MIDAS	KF	Naive
USA	1.00	-0.41	0.87	0.11	0.01	-
ret_5yr	-0.41	1.00	-0.16	-0.03	0.02	-
RL	0.87	-0.16	1.00	0.14	0.01	-
MIDAS	0.11	-0.03	0.14	1.00	-0.01	-
KF	0.01	0.02	0.01	-0.01	1.00	-
Naive	-	-	-	-	-	-

This table reports the daily correlations computed for the following series: CRSP value-weighted portfolio returns, CRSP five-year fixed-term-index returns, daily changes in estimated cross-sectional means from RL, naive, MIDAS, and KF approaches. We exclude the SPF release dates in our computation. The RL approach learns the weights using our COPDAC-LSTDQ algorithm and uses policy weights averaged across 40, 44, 48, 52, 56, and 60 quarters of training periods. The naive method places a weight of one on the lag cross-sectional moments and zero on all of the asset return terms. The RL approach uses For assets, we use returns on the CRSP value-weighted portfolio and the CRSP five-year fixed-term index. The MIDAS approach fits a separate model for each day where we use the "Beta Lag" specification from Ghysels et al. (2007). In the KF approach, we fit the model's structural parameters to the generated time-series via maximum likelihood, and the estimated parameters are used to compute the Kalman gain.

Source: CRSP, Survey of Professional Forecasters (SPF)

Table 7: **Top Ten Changes in RL Daily CX Mean Series with Corresponding Events**

	Mean	Event
2011-08-08	-0.65	U.S. credit rating downgrade
2011-08-09	0.51	Fed promises to keep interest rates near zero for two years
2008-10-15	-0.51	Weak Fed economic forecasts, Bernanke comments
2008-10-28	0.50	Unclear
2011-08-04	-0.45	Weak jobs data, Japan weakens Yen, ECB re-enters bond market
2008-10-09	-0.44	Unclear
2009-03-23	0.44	Treasury announces TARP
2008-09-29	-0.43	House rejects bank bailout plan
2011-08-11	0.40	Jobless claims fall, strong earnings
2009-03-10	0.38	Citi earnings positive (were expected to be negative)

The table reports the dates of the ten largest absolute changes in the daily cross-sectional mean series and any significant macroeconomic events that occurred on those days. We also report the estimated daily standard deviation change on these dates. We exclude the SPF release dates.

Source: News releases, Simulations

Table 8: **Response of Expected Output Growth to Policy and Federal Funds Rate Shock**

	Full Sample	Full Sample	NS (2018)
	$(2004:01\text{–}2014:12)$	Ex. 2008:06-2009:06	$(2000:01\text{–}2014:12)$
Panel A. Response to Policy News Shock			
Policy news shock	-0.83	-0.82	1.04
	$(t:-3.632)$	$(t:-2.938)$	$(t:2.971)$
Observations	71	63	90
Panel B. Response to Fed Funds Rate (FFR) Shock			
FFR Shock	-0.39	-0.38	
	$(t:-1.914)$	$(t:-2.028)$	N/A
Observations	71	63	

This table reports the coefficients from a regression of daily changes in growth expectations on policy news and federal funds rate shocks from 2004 to 2014. The dependent variable is the change in daily growth expectations obtained from our method. The policy news shock and fed funds rate (FFR) shocks are identical to the ones used in Nakamura and Steinsson (2018). The policy news shock is the first principal component of the unanticipated change over the 30-minute windows in a set of interest rates.

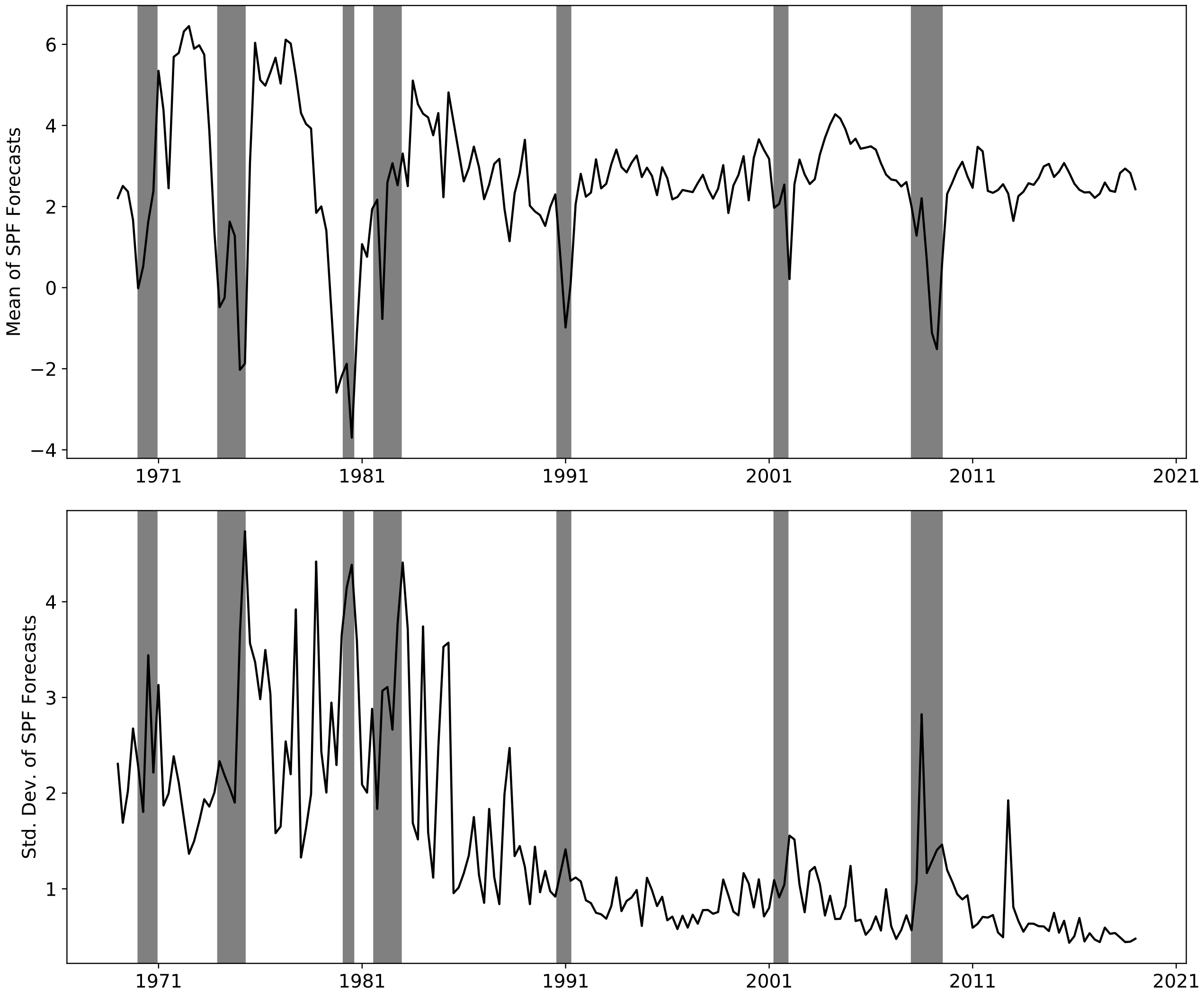
Table 9: **Calibrated Parameters of the Simulated Economy**

Growth Dynamics: $\theta_{t+1} = \mu + \delta\theta_t + \epsilon_{t+1}$	
μ	0.0174
δ	0.5291
σ^2_ϵ	0.001501
Dividend Growth Dynamics: $d^i_{t+1} - d^i_t = \gamma + \beta^i\theta_{t+1} + \nu_{t+1}$	
γ	0.0408
$\mathbb{E}[\beta^i]$	0.6835
σ^2_ν	0.00745
Market-to-Book Dynamics: $\zeta_{t+1} = \tau + \psi\zeta_t + \xi_{t+1}$	
τ	0.2952
ψ	0.8970
σ^2_ξ	0.1234
Conditional Expected Returns Dynamics: $\mathbb{E}_t[r^i_{t+1}] = \alpha + \phi^i\zeta_t$	
α	0.1900
$\mathbb{E}[\phi^i]$	-0.0387
Term from Campbell-Shiller Expression: $\rho = 1/(1 + \exp(\overline{d-p}))$	
ρ	0.9845
Correlation in Error Terms: $\pi = \mathrm{Corr}(\epsilon_{t+1}, \xi_{t+1})$	
π	0.0762

The table reports the calibrated parameters of the model. Here we present the calibration to annual data, but we convert these values to their daily counterparts in the simulations. Parameters for GDP growth $(\mu, \delta, \sigma^2_\epsilon,)$ are estimated by fitting an AR(1) to annual real GDP growth from 1931 to 2018. Parameters for dividend growth $(\gamma, \mathbb{E}[\beta^i], \sigma^2_\nu)$ are estimated by regressing annual dividend growth of the S&P500 on contemporaneous annual GDP growth from 1994 to 2018. Parameters for dynamics of the market-to-book ratio $\left(\tau, \psi, \sigma^2_\xi\right)$ are estimated by fitting an AR(1) to the year-end market-to-book ratio of the S&P500 from 1977 to 2018. Parameters for the dynamics of conditional expected returns $(\alpha, \mathbb{E}[\phi^i])$ are obtained from the following regression: $R_{t+1} = \alpha + \beta\,(B/M)_t + u_{t+1}$ using S&P500 returns and book-to-market ratio. The term from the Campbell-Shiller expression (ρ) is estimated from the dividends and price of S&P500 from 1994 to 2018, and the correlation between innovations (π) is computed from covariance of θ_t and ζ_t.

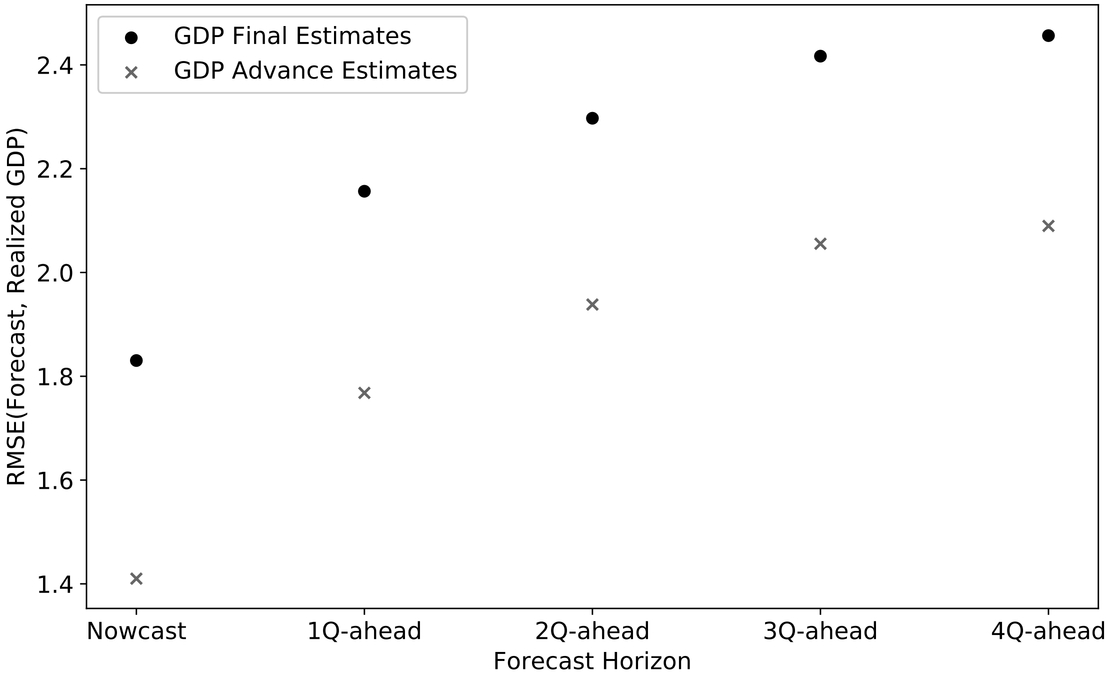
Source: Compustat, CRSP, FRED

Figure 1: **Cyclicality / Counter-cyclicality of SPF Forecasts**



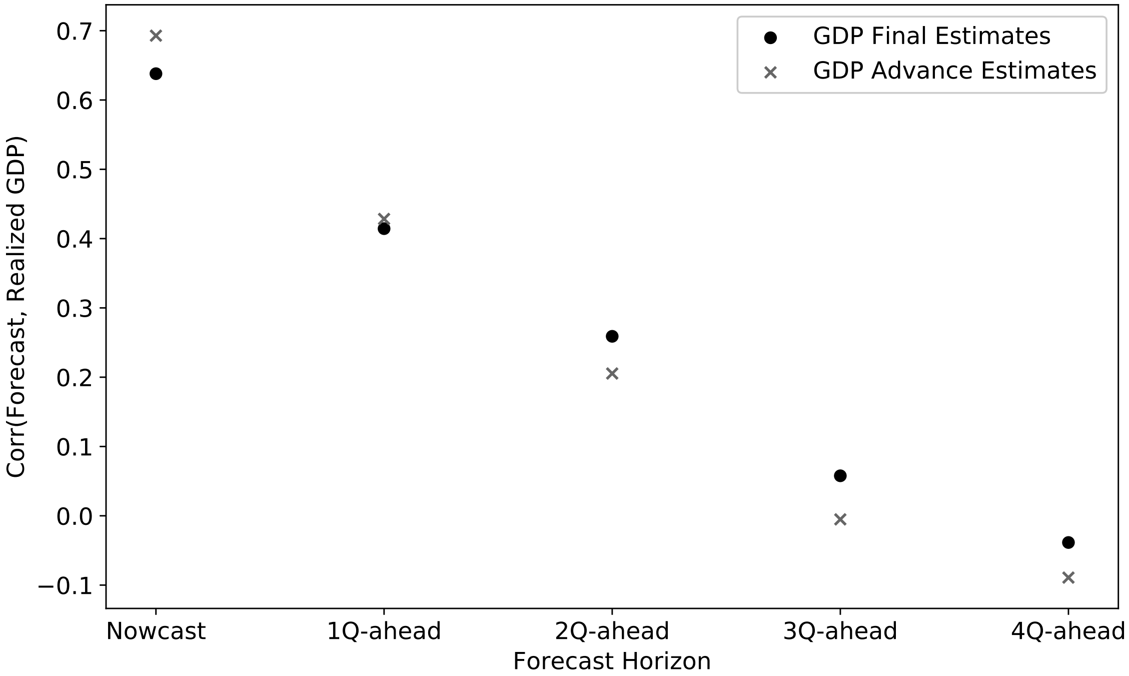
The graph plots the cross-sectional mean and standard deviation of current-quarter forecasts from the entire SPF sample. Each cross-section is winsorized at the 5% level.

Figure 2: **SPF Accuracy by Forecast Horizon (RMSE)**



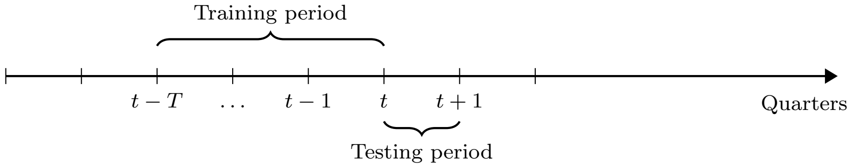
The graph plots the Root Mean Square Error (RMSE) between the mean forecast and the realized GDP growth in the corresponding quarter for different forecasting horizons. For realized GDP growth, we use both the advance and the final estimates. Each cross-sectional mean is computed after winsorization at the 5% level.

Figure 3: **SPF Accuracy by Forecast Horizon (Correlation)**



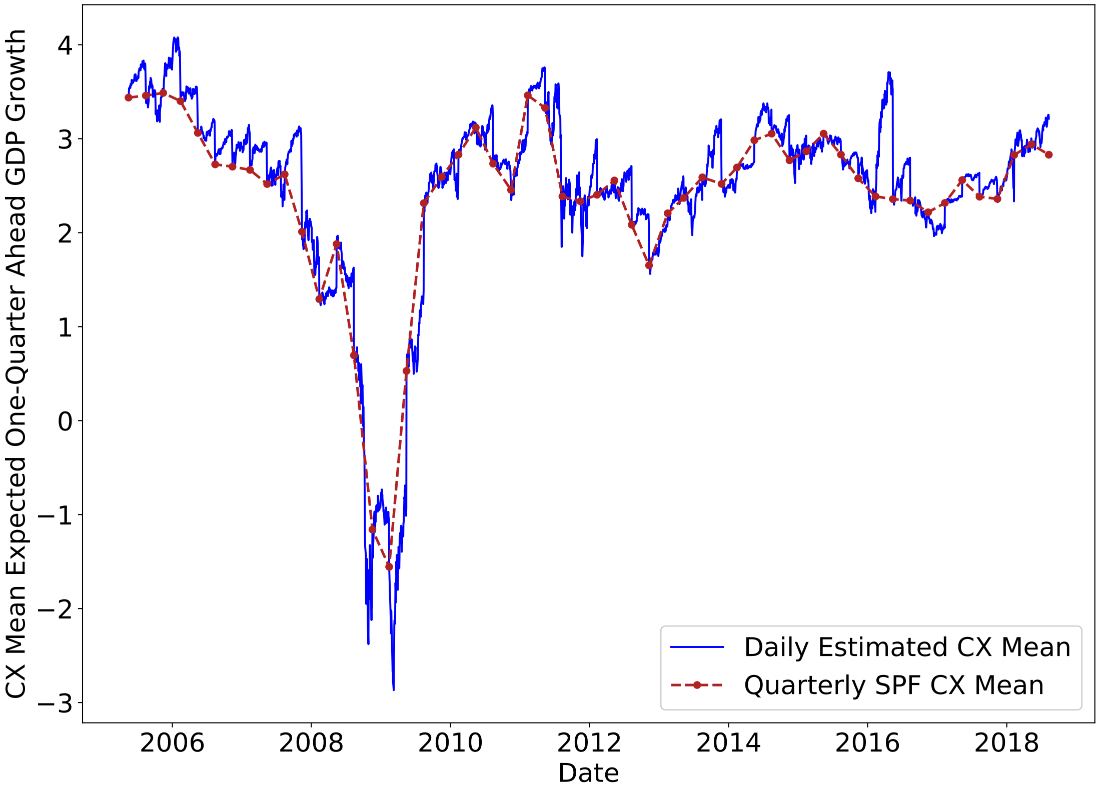
The graph plots the correlation between the mean forecast and the realized GDP growth in the corresponding quarter for different forecasting horizons. For realized GDP growth, we use both the advance and the final estimates. Each cross-sectional mean is computed after winsorization at the 5% level.

Figure 4: **Timeline of Recursive Out-of-sample Estimation Procedure**



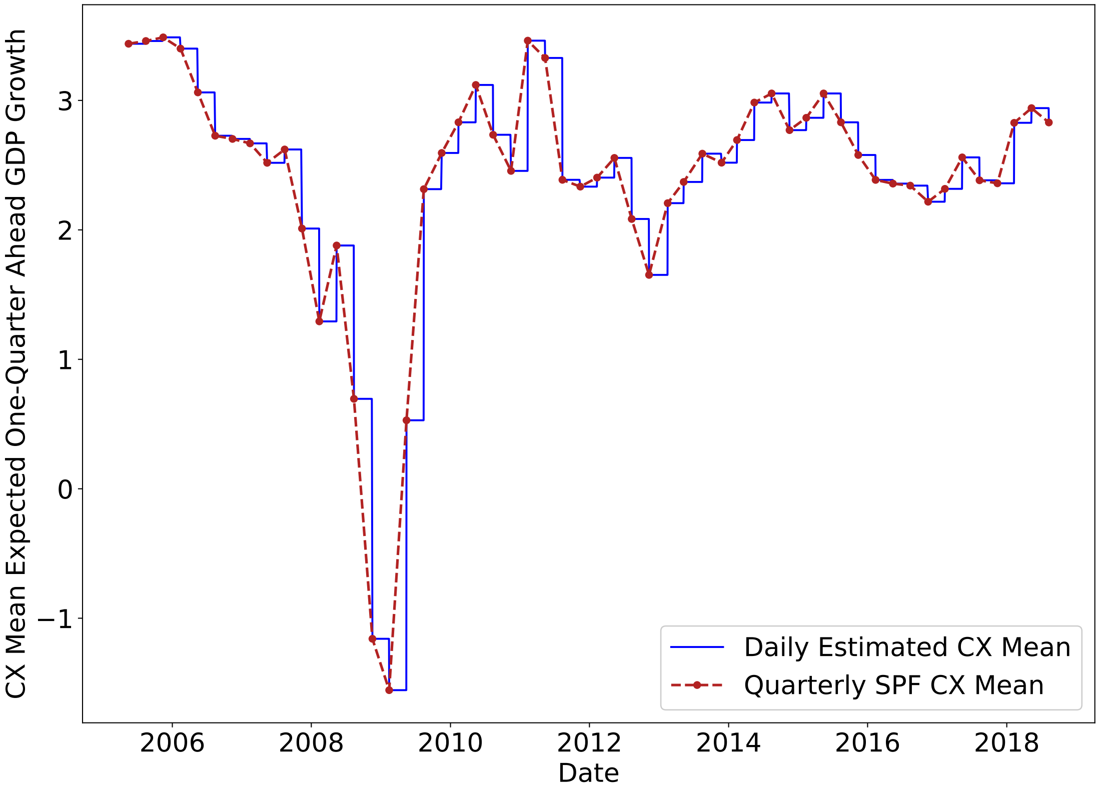
This figure illustrates the timeline of our recursive out-of-sample estimation procedure. The RL policy weights are learned on the in-sample lookback window of T quarters. We then apply these weights to one out-of-sample quarter to construct the daily cross-sectional moment series, with the initial state in the out-of-sample testing period set to the cross-sectional moments from the quarter-t SPF release.

Figure 5: **Estimated Daily Series and True Quarterly Series (RL)**



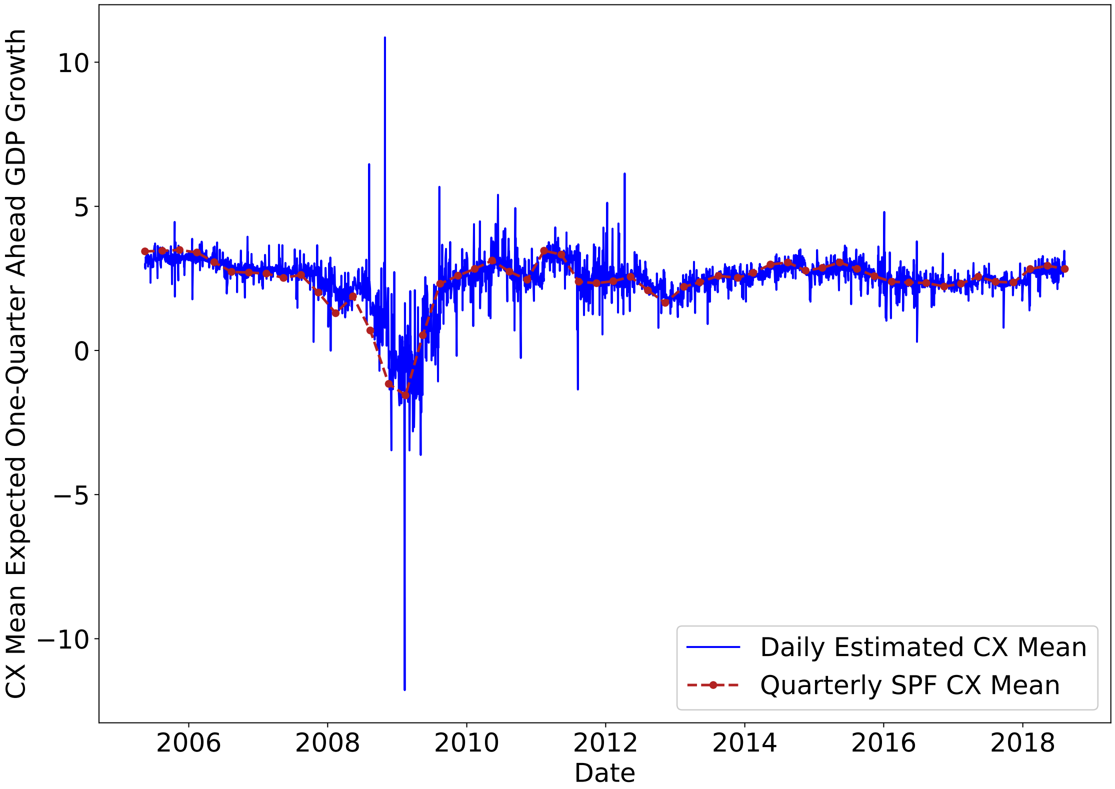
The figure plots the daily cross-sectional mean series estimated using our RL approach and the true quarterly SPF cross-sectional mean series. The daily series is constructed from out-of-sample estimates.

Figure 6: **Estimated Daily Series and True Quarterly Series (Naive)**



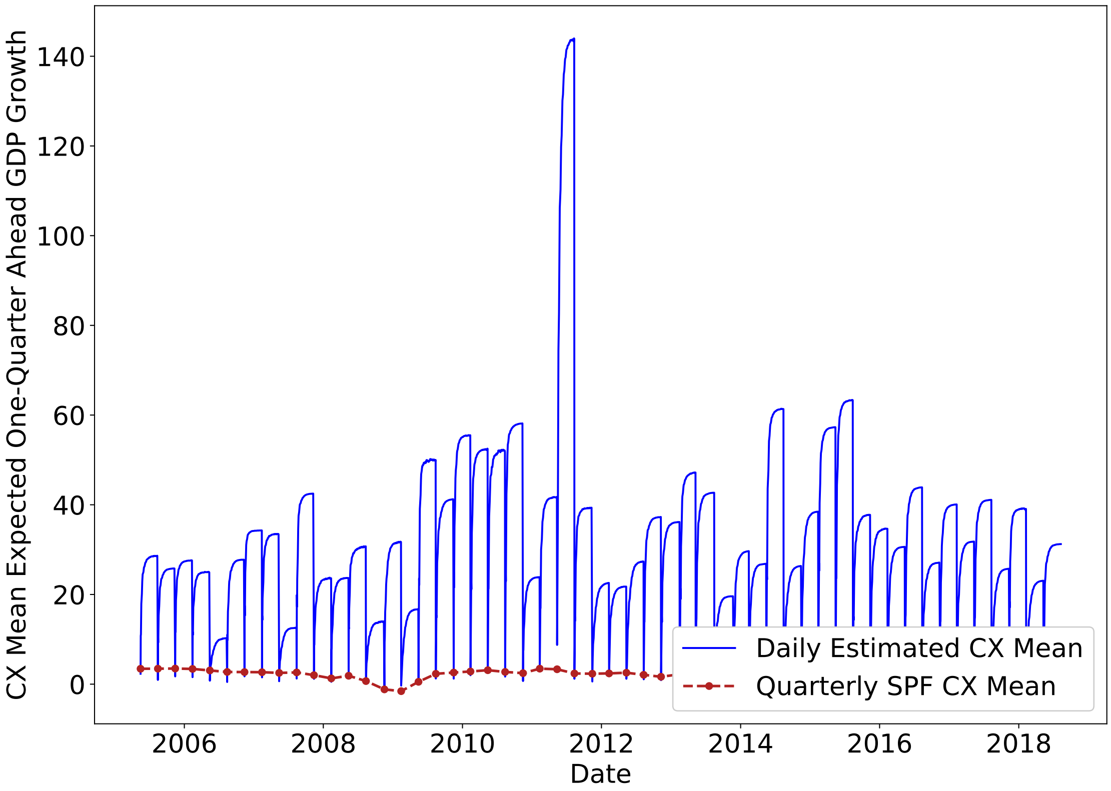
The figure plots the daily cross-sectional mean series estimated using the naive approach and the true quarterly SPF cross-sectional mean series. The daily series is constructed from out-of-sample estimates.

Figure 7: **Estimated Daily Series and True Quarterly Series (MIDAS)**



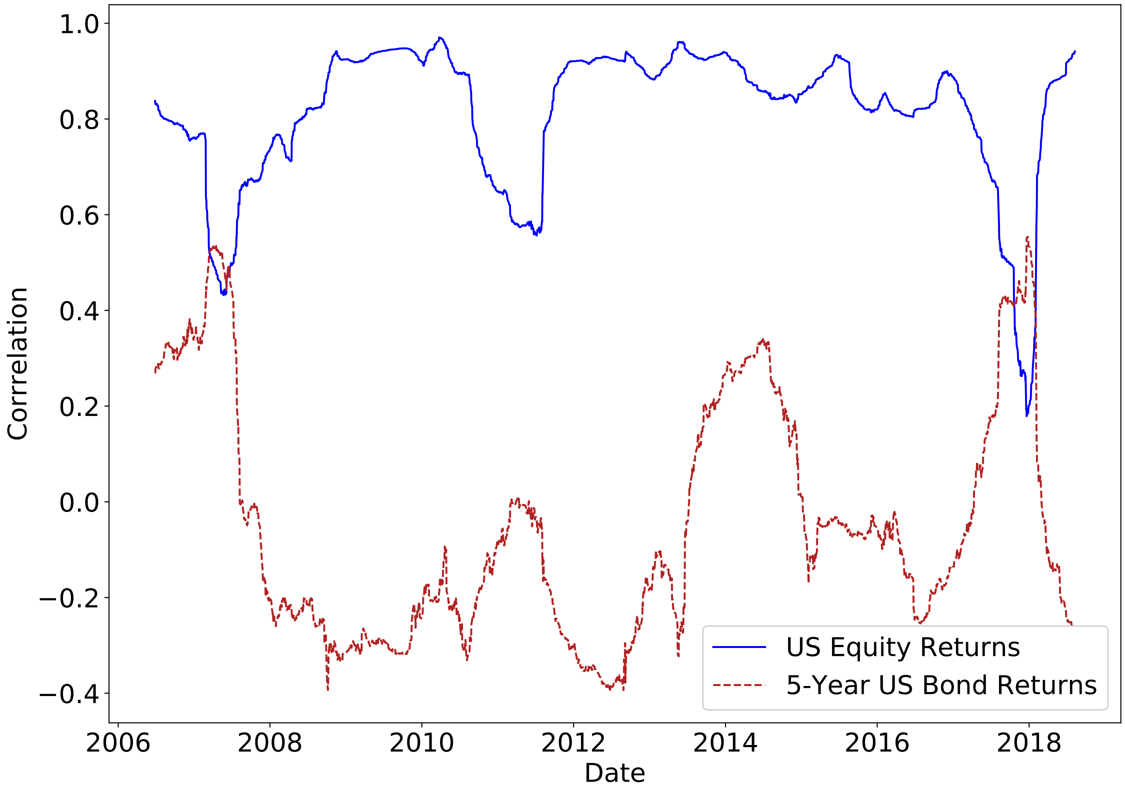
The figure plots the daily cross-sectional mean series estimated using the MIDAS approach and the true quarterly SPF cross-sectional mean series. The daily series is constructed from out-of-sample estimates.

Figure 8: **Estimated Daily Series and True Quarterly Series (KF)**



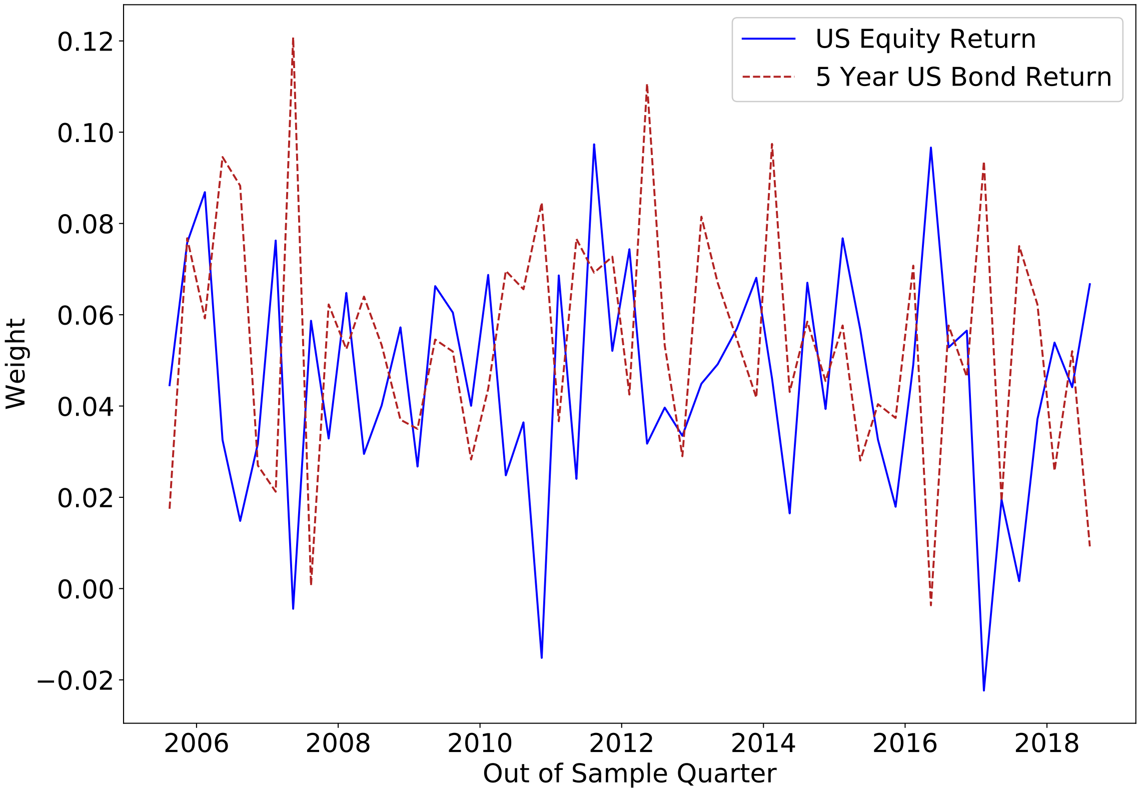
The figure plots the daily cross-sectional mean series estimated using the KF approach and the true quarterly SPF cross-sectional mean series. The daily series is constructed from out-of-sample estimates.

Figure 9: **Correlations between Returns and Changes in Estimated Daily Series from the RL Approach**



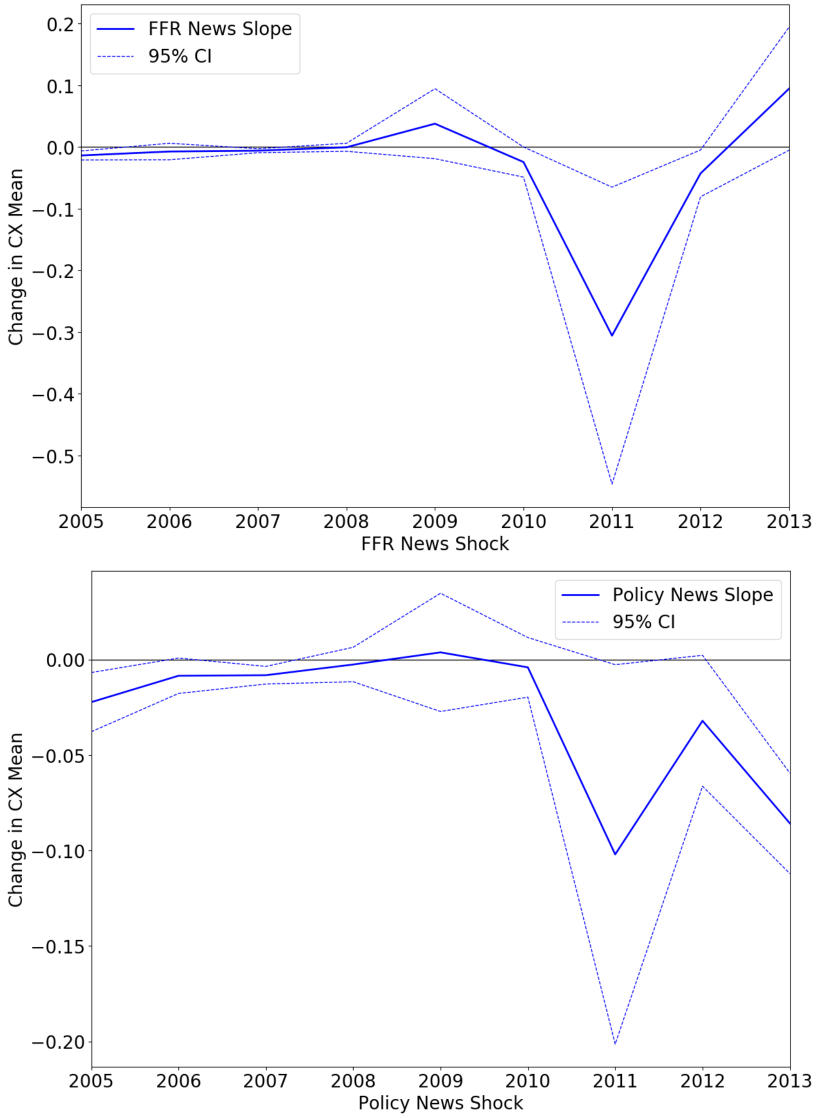
The figure plots the one-year rolling correlation between changes in the daily cross-sectional mean series obtained using the RL approach and returns on the CRSP value-weighted portfolio and the CRSP five-year fixed-term-index. We exclude the SPF release dates in our calculation.

Figure 10: **Policy Weights of RL Approach**



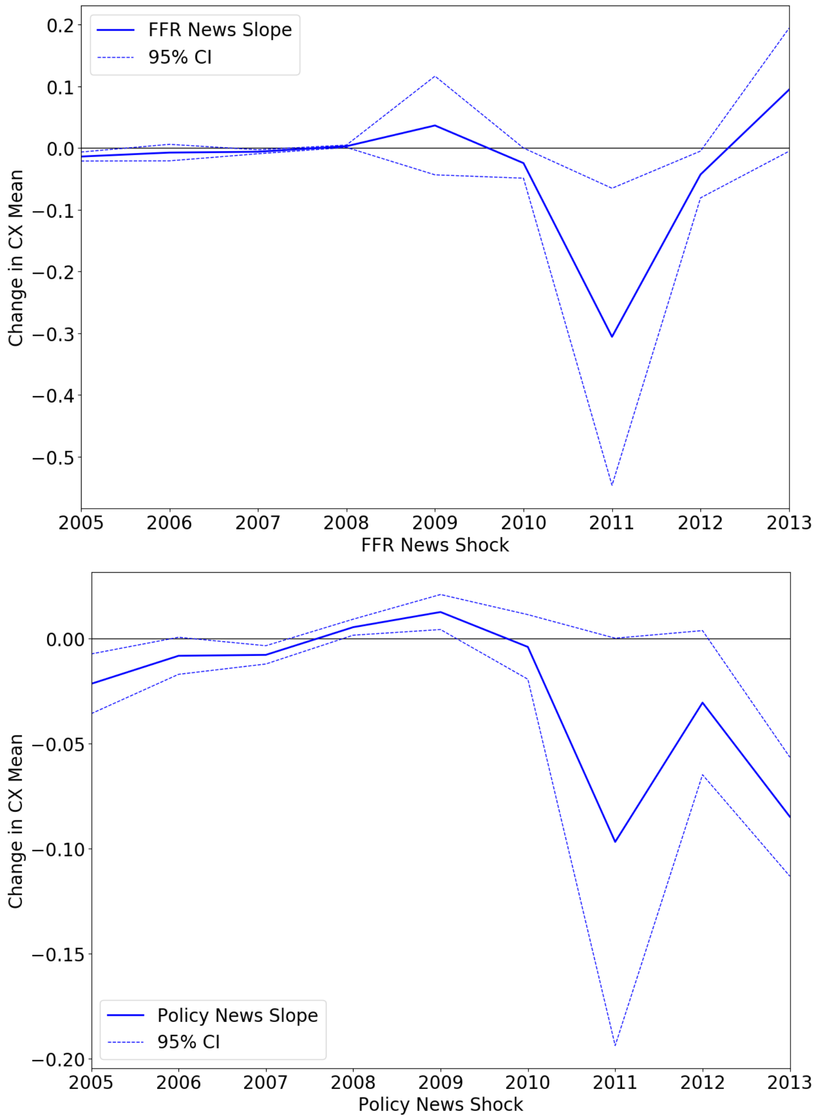
The figure plots the policy weights from the RL approach for estimating the daily cross-sectional mean series. For each out-of-sample quarter, we fit a model from each of the the trailing window of $\in \{40, 44, 48, 52, 56, 60\}$ quarters and average the policy weights across all models.

Figure 11: **Response of Expected Output Growth to Policy and Federal Funds Rate Shock (Yearly Regressions) - Full Sample**



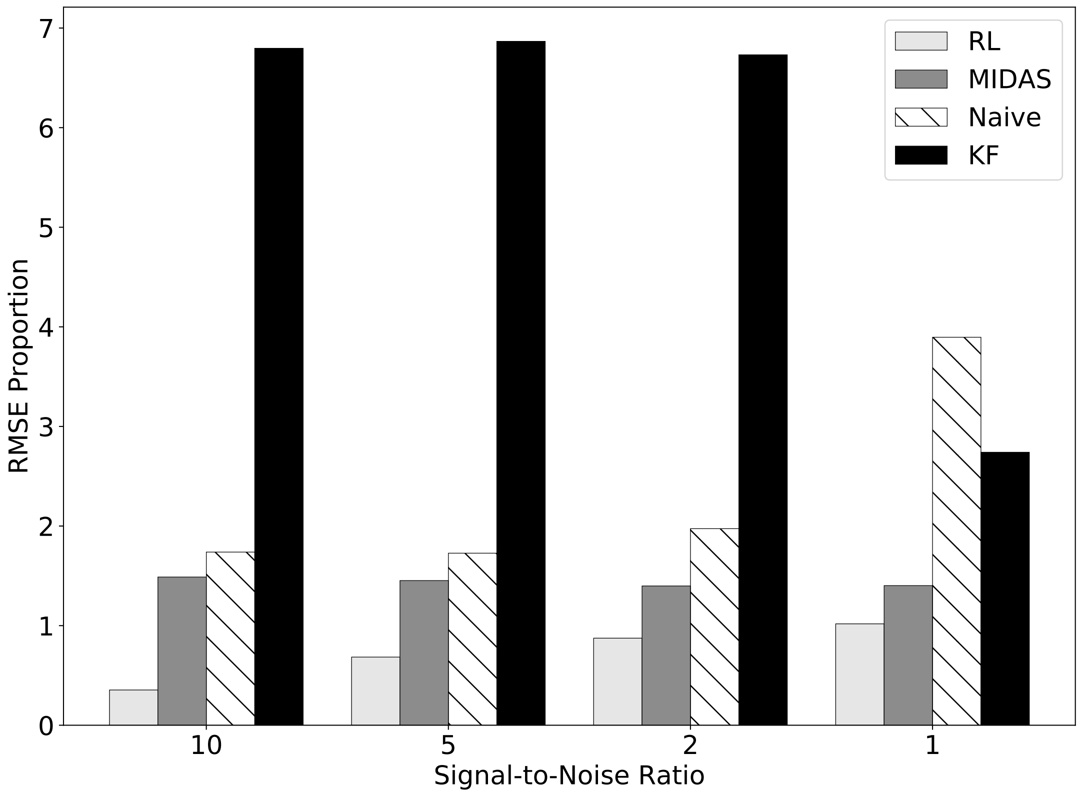
This table reports the coefficients and 95% confidence intervals from a regression each year of daily changes in growth expectations on policy news and federal funds rate shocks from 2004 to 2013. The dependent variable is the change in daily growth expectations obtained from our method. The policy news shock and fed funds rate (FFR) shocks are identical to the ones used in Nakamura and Steinsson (2018). The policy news shock is the first principal component of the unanticipated change over the 30-minute windows in a set of interest rates. Note that in 2014 there are no non-zero policy

Figure 12: **Response of Expected Output Growth to Policy and Federal Funds Rate Shock (Yearly Regressions) - Excluding Height of Great Recession**



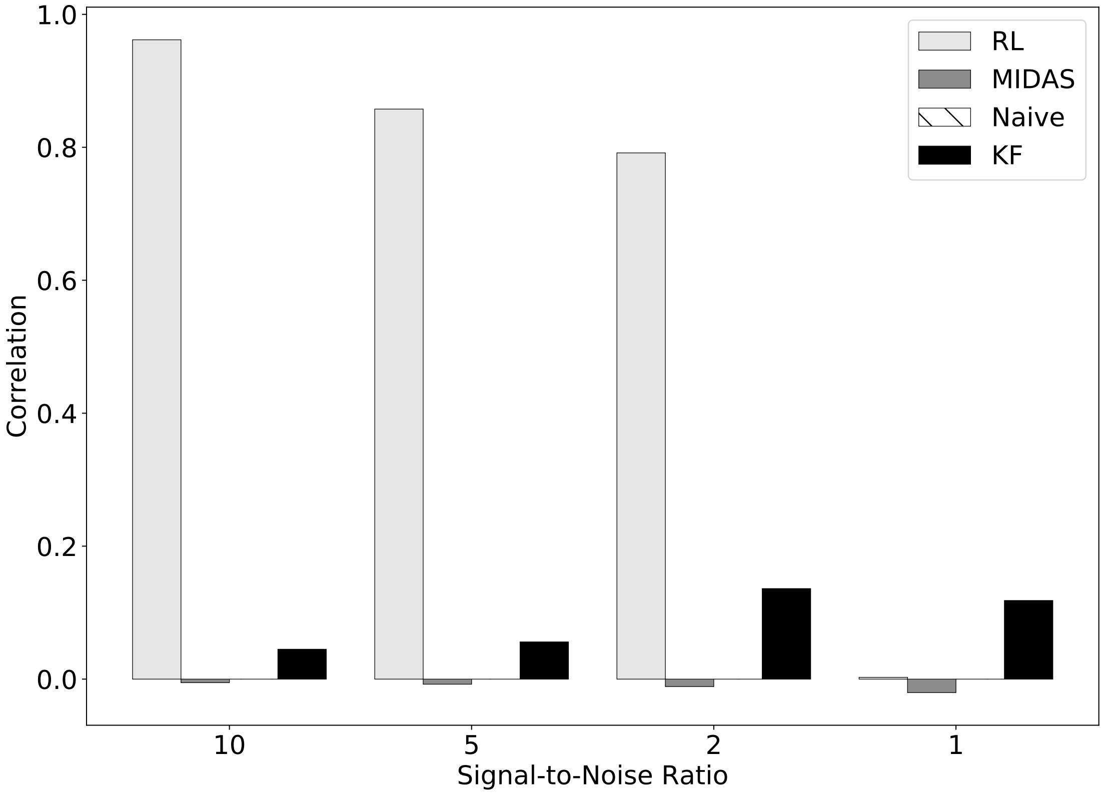
This table reports the coefficients and 95% confidence intervals from a regression each year of daily changes in growth expectations on policy news and federal funds rate shocks from 2004 to 2013, excluding the height of the Great Recession 2008:06 – 2009:06. The dependent variable is the change in daily growth expectations obtained from our method. The policy news shock and fed funds rate (FFR) shocks are identical to the ones used in Nakamura and Steinsson (2018). The policy news shock is the first principal component of the unanticipated change over the 30-minute windows in a set of interest rates. Note that in 2014 there are no non-zero policy

Figure 13: **RMSEs from Simulated Cross-sectional Mean Estimation**



The graph plots the relative RMSE (the average value of the ratio of the RMSE to the average absolute cross-sectional mean) across different levels of learning heterogeneity. Specifically, we train both the Kalman filter and reinforcement learning policy on forty quarters of daily returns and quarterly cross-sectional mean observations generated via the model described in Section 2 using ten assets. We then test both models on 1,000 out-of-sample quarters. For each out-of-sample quarter, we estimate the daily growth series using each model and compute the RMSE versus the true daily latent cross-sectional mean series. We conduct this exercise for different levels of learning heterogeneity, which we parameterize using a single signal-to-noise ratio. Specifically, for a signal-to-noise ratio of s and a true parameter value of θ_0, the distribution from which agents draw their values of this parameter is $\mathcal{N}\left(\theta_0, (\theta_0/s)^2\right)$. A lower signal-to-noise ratio therefore implies higher learning heterogeneity

Figure 14: **Correlations from Simulated Cross-sectional Mean Estimation**



The graph plots the average correlation for different levels of learning heterogeneity parametrized by the signal-to-noise ratio. Specifically, we train both the Kalman filter and reinforcement learning policy on forty quarters of daily returns and quarterly cross-sectional mean observations generated via the model described in Section 2 using ten assets. We then test both models on 1,000 out-of-sample quarters. For each out-of-sample quarter, we estimate the daily growth series using each model and compute the correlation versus the true daily latent cross-sectional mean series. We conduct this exercise for different levels of learning heterogeneity, which we parameterize using a single signal-to-noise ratio. Specifically, for a signal-to-noise ratio of s and a true parameter value of θ_0, the distribution from which agents draw their values of this parameter is $\mathcal{N}\left(\theta_0, (\theta_0/s)^2\right)$. A lower signal-to-noise ratio therefore implies higher learning heterogeneity.

Online Appendix

Aditya Chaudhry Sangmin S. Oh

September 16, 2020

D Online Appendix: Detailed Description of RL Algorithm

To efficiently learn the optimal policy we employ the deterministic policy gradient (DPG) framework of Silver et al. (2014). Specifically, we adapt the Compatible Off-Policy Deterministic Actor Critic-Q Learning (COPDAC-Q) algorithm.

We start with the assumption of a linear policy function, $g_{\boldsymbol{\lambda}}(s) = \varphi(s)^\top \boldsymbol{\lambda}$. Following 23, we seek $\boldsymbol{\lambda}$ that solves $\max_{\boldsymbol{\lambda}} Q(g_{\boldsymbol{\lambda}}(s), s)$. The natural means to solving this optimization problem is to iteratively improve $\boldsymbol{\lambda}$ via gradient ascent in the direction of the gradient of $Q(\cdot,\cdot)$ with respect to $\boldsymbol{\lambda}$. The chain rule decomposes $\nabla_{\boldsymbol{\lambda}} Q(g_{\boldsymbol{\lambda}}(s), s)$ into two factors: $\nabla_a Q(g_{\boldsymbol{\lambda}}(s), s)$ and $\nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}(s)$. However, since we do not know the true action-value function $Q(\cdot,\cdot)$, we replace it with a *compatible function approximator* $Q_{\mathbf{w}}(\cdot,\cdot)$ such that

$$\nabla_a Q(g_{\boldsymbol{\lambda}}(s), s) = \nabla_a Q_{\mathbf{w}}(g_{\boldsymbol{\lambda}}(s), s)$$

As per Silver et al. (2014), one such compatible function approximator is

$$\begin{aligned} Q_{\mathbf{w}}(s,a) &= (a - g_{\boldsymbol{\lambda}}(s))^\top \nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}(s)\mathbf{w} + V_{\mathbf{v}}(s), \\ &= \varphi(s,a)\mathbf{w} + V_{\mathbf{v}}(s), \end{aligned}$$

where $\varphi(s,a) = (a - g_{\boldsymbol{\lambda}}(s))^\top \nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}(s)$ and $V_{\mathbf{v}}(\cdot)$ is any differentiable *baseline function*[18]. We use a linear baseline, $V_{\mathbf{v}}(s) = \varphi(s)^\top \mathbf{v}$.

In brief, the COPDAC-Q algorithm of Silver et al. (2014) operates as follows:

1. Instantiate $\boldsymbol{\lambda}_0, \mathbf{w}_0$, and $\mathbf{v}_0$.

2. For each state s_t:

   (a) Draw action a_t from a *behavioral policy* $\pi_{\boldsymbol{\lambda}}(\cdot|s_t)$. The behavioral policy is the density function of a distribution centered at $g_{\boldsymbol{\lambda}}(\cdot)$. Adding noise in action selection helps ensure adequate exploration of the parameter space in order to prevent convergence to a local maximum. This practice is known as *off-policy learning.*

[18] Baseline functions reduce the variance in gradient updates

(b) Update $\boldsymbol{\lambda}_t, \mathbf{w}_t$, and $\mathbf{v}_t$ via the following *Q-learning updates* (Watkins, 1989):

$$\begin{aligned}
\delta_t &= r_t + \gamma Q_{\mathbf{w}}\left(s_{t+1}, g_{\boldsymbol{\lambda}}\left(s_{t+1}\right)\right) - Q_{\mathbf{w}}\left(s_t, a_t\right) \\
\boldsymbol{\lambda}_{t+1} &= \boldsymbol{\lambda}_t + \alpha_\theta \nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}\left(s_t\right)\left(\nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}\left(s_t\right)^\top w_t\right) \\
w_{t+1} &= w_t + \alpha_w \delta_t \varphi\left(s_t, a_t\right) \\
v_{t+1} &= v_t + \alpha_v \delta_t \varphi\left(s_t\right)
\end{aligned}$$

3. Repeat Step 2 until some stopping criterion (e.g. $\boldsymbol{\lambda}_t$ converges or a maximum number of iterations is reached).

As noted in Silver et al. (2014), off-policy Q-learning with linear function approximation may diverge. To achieve better convergence properties, we thus propose the COPDAC-LSTDQ algorithm in which we replace the Q-learning updates in step 2(b) with batch LSTDQ (least-squares temporal-difference Q-learning) updates (Lagoudakis and Parr, 2003). LSTDQ updates find the $\mathbf{w}$ and $\mathbf{v}$ vectors such that the total Q-update for these vectors is zero:

$$\sum_{t=1}^{T} \alpha_w \delta_t \varphi\left(s_t, a_t\right) = 0 \tag{24}$$

$$\sum_{t=1}^{T} \alpha_w \alpha_v \delta_t \varphi(s_t) = 0. \tag{25}$$

Since we use a linear approximation for $Q_{\mathbf{w}}$, (24) and (25) have the simple analytic solutions. The details of COPDAC-LSTDQ are shown below:

1. Set hyper-parameters $\gamma, \alpha, \boldsymbol{\Sigma}_\pi$

2. Initialize $\mathbf{s}_0, \boldsymbol{\lambda}^0, \mathbf{w}^0, \mathbf{v}^0$

3. For each iteration in $i \in [0, 1, 2, \cdots]$:

   (a) Obtain a history of the states $(s_0, ..., s_T)$ and $(a_0, ..., a_T)$ by iteratively computing:

$$\begin{aligned}
\mathbf{a}_t &\leftarrow \pi_{\boldsymbol{\lambda}^i}\left(\mathbf{s}_t\right) \\
s_{t+1} &\leftarrow T\left(s_t, a_t\right)
\end{aligned}$$

(b) Make the following LSTDQ batch update to $\mathbf{w}^i$ to $\mathbf{w}^{i+1}$:

$$\mathbf{w}^{i+1} \leftarrow \left(\sum_{t=1}^{2500} \varphi\left(\mathbf{s}_t, \mathbf{a}_t\right) \left\{ \varphi\left(\mathbf{s}_t, \mathbf{a}_t\right) - \gamma \varphi\left(\mathbf{s}_{t+1}, \mu_{\boldsymbol{\theta}^i}\left(s_{t+1}\right)\right) \right\}^\top \right)^{-1} \times \left(\sum_{t=1}^{T} \varphi\left(\mathbf{s}_t, \mathbf{a}_t\right) \left(r_t + V^{\mathbf{v}}\left(\mathbf{s}_{t+1}\right) - V^{\mathbf{v}}\left(\mathbf{s}_t\right) \right) \right)$$

(c) Make the following LSTDQ batch update to $\mathbf{v}^i$ to $\mathbf{v}^{i+1}$:

$$\mathbf{v}^{i+1} \leftarrow \left(\sum_{t=1}^{2500} \varphi\left(\mathbf{s}_t, \mathbf{a}_t\right) \left\{ \varphi\left(\mathbf{s}_t\right) - \gamma \varphi\left(\mathbf{s}_{t+1}\right) \right\} \right)^{-1} \times \left(\sum_{t=1}^{T} \varphi\left(\mathbf{s}_t, \mathbf{a}_t\right) \left(r_t + \gamma \varphi\left(\mathbf{s}_{t+1}, \mu_{\boldsymbol{\theta}^i}\left(\mathbf{s}_{t+1}\right)\right)^\top \mathbf{w}^{i+1} - \varphi\left(\mathbf{s}_t, \mathbf{a}_t\right)^\top \mathbf{w}^{i+1} \right) \right)$$

(d) For each period $t \in [1, 2, ..., T]$, iteratively update:

$$\boldsymbol{\lambda}_{t+1} \leftarrow \boldsymbol{\lambda}_t + \alpha \nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}\left(\mathbf{s}_t\right)^\top \left(\nabla_{\boldsymbol{\lambda}} g_{\boldsymbol{\lambda}}\left(\mathbf{s}_t\right)^\top \mathbf{w}_i \right), \quad \boldsymbol{\lambda}_0 = \boldsymbol{\lambda}^i$$

and update $\boldsymbol{\lambda}^i$ to $\boldsymbol{\lambda}^{i+1}$ by:

$$\boldsymbol{\lambda}^{i+1} \leftarrow \boldsymbol{\lambda}_{t+1}$$

4. Repeat Step 3 until some stopping criterion.

This technique of iterating over an entire history of states multiple times is known as *action replay*. Once the algorithm terminates, we can use the learned policy $g_{\boldsymbol{\lambda}}(\cdot)$ out-of-sample.

E Online Appendix: Empirical Comparison of KF and RL

To illustrate the performance of the RL approach and examine its theoretical properties versus the KF approach, we apply both RL and KF to simulated time series generated by the model described in Section 2.2. In this stylized example, we depart from our main task of interest. Instead, each approach seeks to estimate the daily series of growth when it can only observe quarterly growth numbers. As a preview of the results, we find that the RL approach outperforms the KF in terms of root mean-square error (RMSE) and correlation between the estimated and true, latent growth series.

Our procedure to compare the performance of two approaches is as follows.

1. We generate a forty-quarter time series of θ_t and r_t^i for each asset $i = 1, ..., m$. Our default calibration uses $m = 5$ assets.

2. We learn the optimal policy $g_{\boldsymbol{\lambda}}$ via the two approaches.

    - In the KF approach, we fit the model's structural parameters to the generated time-series via maximum likelihood. The estimated parameters can be used to compute the Kalman gain and $g_{\boldsymbol{\lambda}}$.
    - In the RL approach, we learn the policy parameters $\boldsymbol{\lambda}$ directly via COPDAC-LSTDQ.

3. We apply the learned policy to 1,000 out-of-sample quarters.

    - For each out-of-sample quarter and for each approach, we calculate the RMSE and the correlation between the estimated series and the true underlying gorwth series.

To assess the sensitivity of each approach's performance to the precise calibration of the structural parameters, we conduct the simulation exercise for 356 different calibrations, as detailed in Table A.1. Specifically, we consider seven different combinations, each of which involves varying a subset of the structural parameters while holding the others fixed at their baseline values.

Figures A.1 and A.2 display the results of the simulations. They demonstrate that the RL approach outperforms the KF across a wide range of calibrations. Each dot represents the average RMSE / correlation across all 1,000 out-of-sample quarters for a given calibration. Figure A.1 illustrates that the RL approach achieves a median RMSE of 0.00252 across all calibrations, while the KF realizes a median RMSE of 0.007972. Figure A.2 shows that the RL approach achieves a median correlation of 0.5566 across all calibrations compared to a median correlation of 0.3181 for the KF approach. In fact, the KF outperforms RL in only 14 calibrations in terms of RMSE and in 23 calibrations in terms of correlation. While the calibrations for which the KF achieves lower RMSEs follow no distinct pattern, we do find that several of the calibrations in which KF performs better involve high persistence in the latent growth process.[19]

To explicitly demonstrate the greater efficiency of the RL approach, we conduct block-bootstrap simulations to decompose the RMSE into bias and variance terms. By avoiding an explicit model of state dynamics, we expect the RL approach to deliver lower-variance estimates. Moreover, while the correctly specified KF is unbiased, finite-sample maximum-likelihood is not. Thus, for a finite sample, the RL approach may in fact yield a lower bias than the KF.

The block-bookstrap proceeds as follows. We generate a training sample of 60 quarters of quarterly growth observations and daily returns for $m = 5$ assets under the baseline parameterization of the model in Section 2.2. We then draw 100 block-bootstrap subsamples of 30 quarters from the sample, and for each subsample we estimate the optimal policy using both RL and KF. We apply both policies to one out-of-sample quarter, calculate the RMSEs between estimated and true

[19] These calibrations set the quarterly first-order autoregressive coefficient for the growth process to $\delta = 0.9$ instead of the baseline value of $\delta = 0.53$.

growth series, and compute the bias and variance terms via the following decompositions:

$$\underbrace{\mathbb{E}\left[\left(\theta_t - \hat{\theta}_{t|t}\right)^2\right]}_{\text{MSE}} = \underbrace{\left(\theta_t - \mathbb{E}\left[\hat{\theta}_{t|t}\right]\right)^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}\left[\left(\mathbb{E}\left[\hat{\theta}_{t|t}\right] - \hat{\theta}_{t|t}\right)^2\right]}_{\text{Variance}}$$

where all expectations are taken across the 100 bootstrap models.

Figure A.3 displays the results of these bootstrap simulations. We see that although the RL approach realizes higher bias (0.00687 versus 0.00587), it achieves significantly lower variance (0.0000130 versus 0.0000287) and therefore delivers a lower out-of-sample RMSE. In unreported simulations, we also find that increasing the number of assets and decreasing the training sample length both improve RL performance relative to the KF. In these situations, the RL approach sometimes achieves lower bias than the KF in addition to greater efficiency.

F Online Appendix: Sensitivity to Training Periods

The weights in the learned RL policy prove sensitive to the training sample, so bootstrap aggregating or "bagging" can help reduce estimation variance. In our setting, we average across models learned from different training periods. Typically, bootstrap aggregating involves the following steps:

1. For a given number of training sample, take subsamples and fit a model to each. For example, we would take subsamples of forty quarters from a given sixty-quarter window.

2. In the out-of-sample testing phase, the output of each bootstrapped model is averaged, or alternately the weights from each model would be averaged to yield the policy function.

For the RL, KF, and MIDAS approaches, we train six models – from 40, 44, 48, 52, 56, and 60 quarters of training data – and average the estimated series for each approach to get the final output series.

In Figures A.4 and A.5, we illustrate the benefits of bagging. Having ranked the six RL models from the lowest to highest RMSE, we start with the best model and iteratively "bag" worse models. We see that incrementally adding the worse models increases the R^2 and decreases the RMSE. Bagging improves performance because the individual models perform reasonably well and, importantly, have imperfectly correlated outputs. Intuitively, we can consider each output series or set of policy weights as the true signal plus some noise. Averaging across models averages out the noise, which is akin to raising the Sharpe ratio of a portfolio by adding in non-perfectly correlated assets.

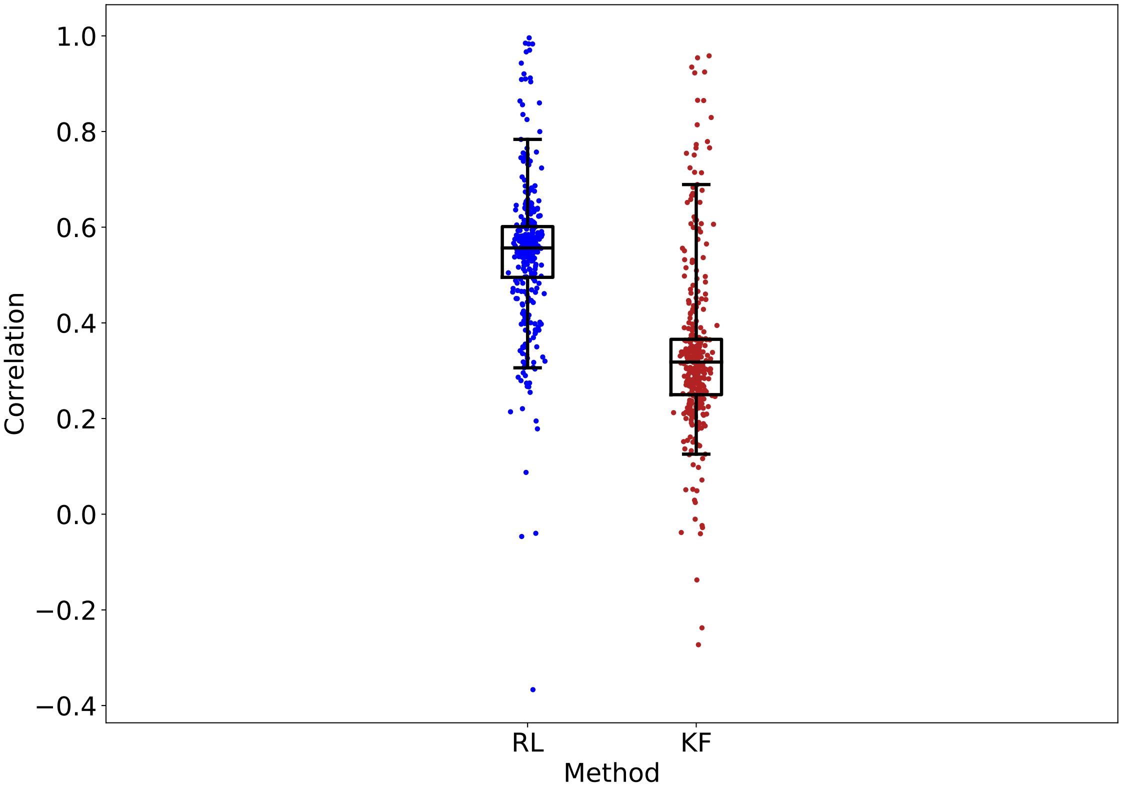
Table A.1: **Variations in Calibrated Parameters of the Simulated Economy**

Experiment	Parameter	Variations
1	μ	4.0e-7, 2.0e-5, 3.9e-5, 6.8e-5, 7.9e-5, 1.2e-4, 1.6e-4
	γ	3.9e-5, 7.9e-5, 1.2e-4, 1.6e-4, 1.9e-4, 2.3e-4, 2.7e-4
2	$\mathbb{E}[\beta^i]$	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
	$\mathbb{E}[\phi^i]$	0.0, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9
3	δ	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
4	$\sigma_\epsilon/\sigma_\nu$	0.01, 0.1, 0.25, 0.5, 0.75, 1, 1.1, 1.25, 1.5, 1.75, 2, 5, 10
5	δ	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
	$\mathbb{E}[\phi^i]$	0.0, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9
6	T	4, 8, 20, 40, 60, 80, 100, 120, 200, 400
7	$\mathbb{E}[\phi^i]$	0.0, -0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8, -0.9
	d	1, 2, 5, 10, 15, 20

This table reports the different calibration variations we use in the simulations from Appendix E. For each experiment, we use all all possible combinations of the corresponding parameters and maintain the baseline values for the remaining ones. The baseline calibration can be found in Table 9. For example, the seven different values of both μ and γ give rise to 49 different calibration variations in Experiment 1. Across all seven experiments we have a total of 356 different calibrations.

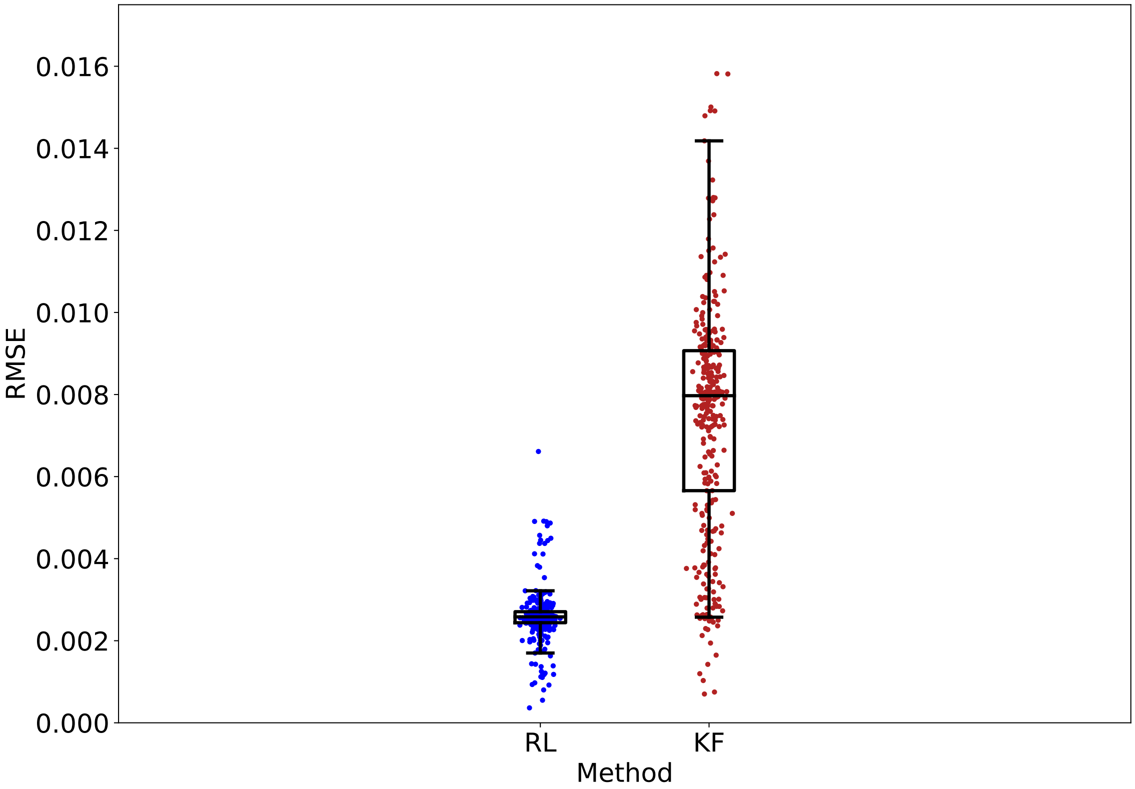
Source: Simulations

Figure A.1: **RMSE from KF and RL Approaches in Simulated Growth Estimation**



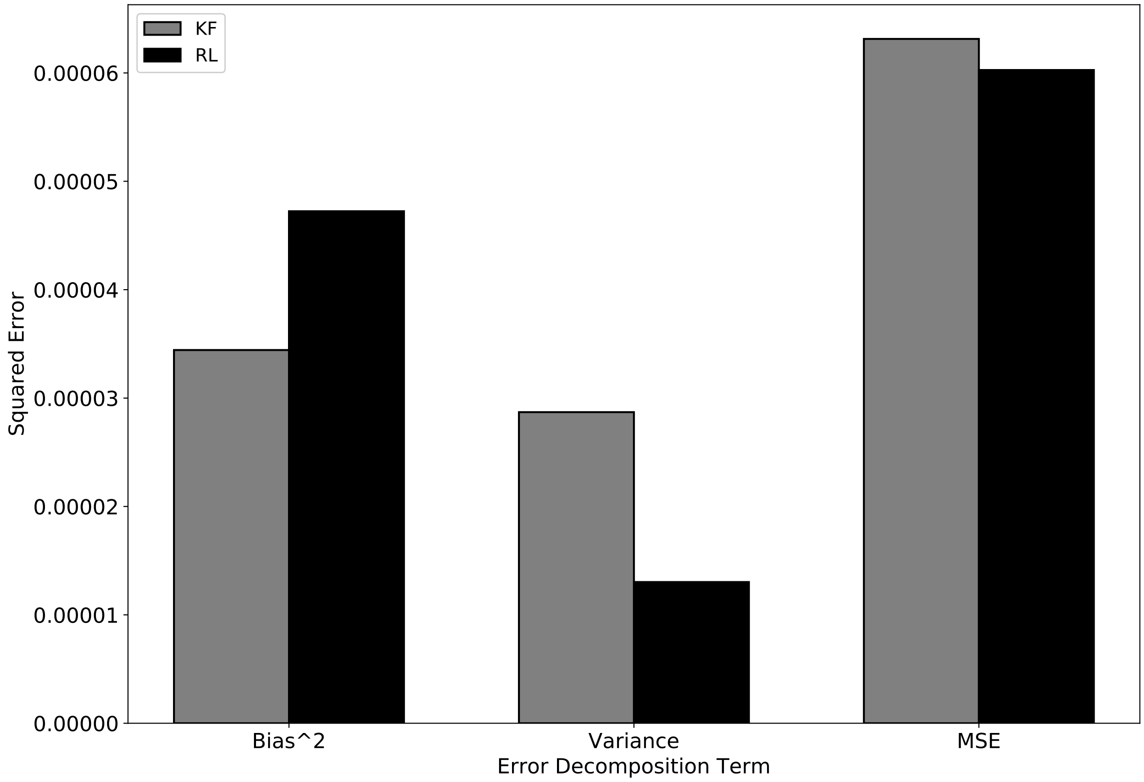
The graph plots the average root mean square errors (RMSEs) between the estimated and true growth rate series across a variety of calibrations of the underlying state space model from Section 2. Specifically, for each calibration, we train both the Kalman filter and reinforcement learning policy on 40 quarters of daily returns and quarterly growth observations. We then test both models on 1,000 out-of-sample quarters. For each out-of-sample quarter we estimate the daily growth series using each model and compute the RMSE versus the true daily latent growth series. Each point represents the average RMSE from all 1,000 out-of-sample quarters for a specific calibration. We test 356 different calibrations in total (detailed in Table A.1). The box horizontal lines correspond to the 25th, 50th, and 75th percentile RMSEs among all 356 values for each method, while the whiskers extend to the 5th and 95th percentiles.

Figure A.2: **Correlations from KF and RL Approaches in Simulated Growth Estimation**



The graph plots the average correlations between the estimated and true growth rate series across a variety of calibrations of the underlying state space model from Section 2. Specifically, for each calibration, we train both the Kalman filter and reinforcement learning policy on 40 quarters of daily returns and quarterly growth observations. We then test both models on 1,000 out-of-sample quarters. For each out-of-sample quarter we estimate the daily growth series using each model and compute the correlation versus the true daily latent growth series. Each point represents the average correlation from all 1,000 out-of-sample quarters for a specific calibration. We test 356 different calibrations in total (detailed in Table A.1). The box horizontal lines correspond to the 25th, 50th, and 75th percentile correlations among all 356 values for each method, while the whiskers extend to the 5th and 95th percentiles.

Figure A.3: **Bias-variance Decomposition of MSE for KF and RL Approaches**

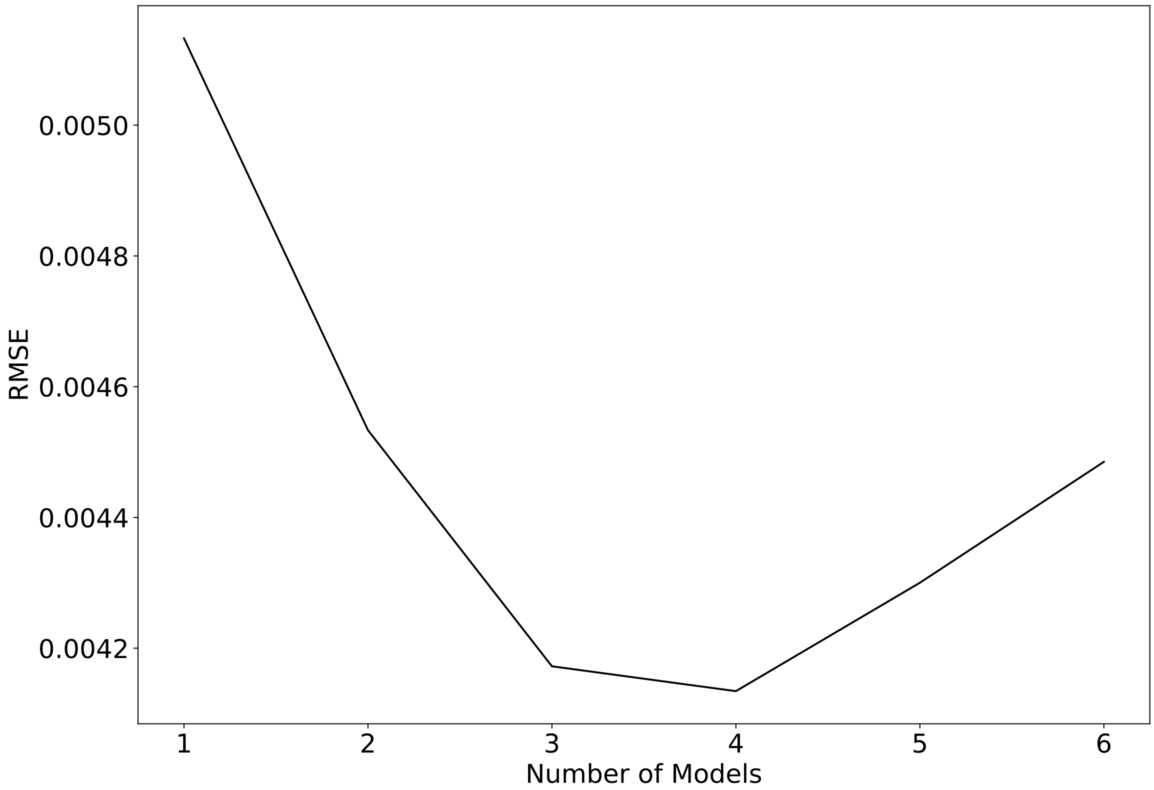


The graph plots the bias-variance decomposition of the mean square error (MSE) for the KF and RL approaches in estimating the growth rate series. Specifically, we generate sixty quarters of daily returns and quarterly growth observations under our baseline calibration of the state-space model from Section 2. From this generated series, we draw 100 block-bootstrap samples of 30 quarters. We train a Kalman filter and reinforcement learning policy on each block-bootstrap sample. We then apply all 100 models for each method to one out-of-sample quarter and compute the MSEs between the estimated daily growth series and the true latent growth series. We then decompose the average MSE across all 100 models into bias and variance terms via the following decomposition:

$$\mathbb{E}\left[\left(\theta_t - \hat{\theta}_{t|t}\right)^2\right] = \left(\theta_t - \mathbb{E}\left[\hat{\theta}_{t|t}\right]\right)^2 + \mathbb{E}\left[\left(\mathbb{E}\left[\hat{\theta}_{t|t}\right] - \hat{\theta}_{t|t}\right)^2\right]$$

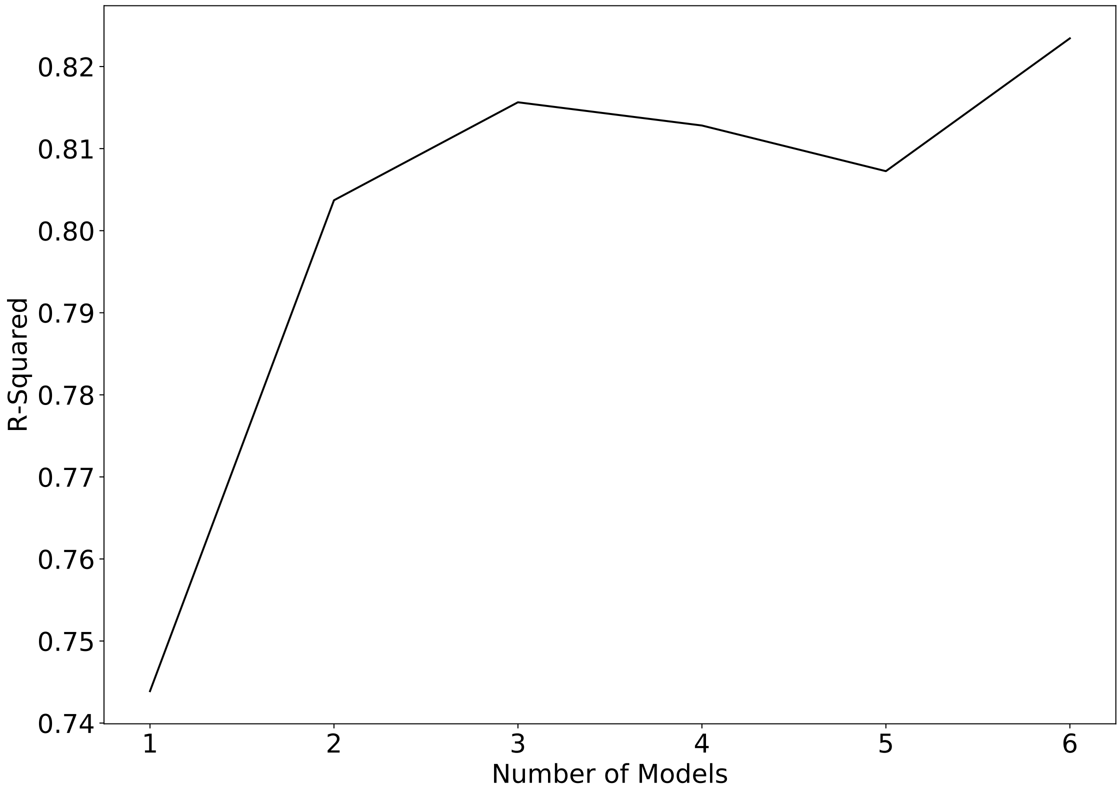
where all expectations are taken across the 100 bootstrapped models. The first term is squared bias and the second is variance.

Figure A.4: **Number of Averaged Models and RMSE from the RL Approach**



The figure plots the RMSE attained using the RL approach as we vary the number of models averaged. Specifically, we first fit a model from the trailing window of $\in \{40, 44, 48, 52, 56, 60\}$ quarters and rank the models in terms of RMSE. Then starting with the best model, we iteratively add in the next best model. For any given set of models, we average the policy weights to obtain the optimal policy.

Figure A.5: **Number of Averaged Models and R^2 from the RL Approach**



The figure plots the R^2 attained using the RL approach as we vary the number of models averaged. Specifically, we first fit a model from the trailing window of $\in \{40, 44, 48, 52, 56, 60\}$ quarters and rank the models in terms of R^2. Then starting with the best model, we iteratively add in the next best model. For any given set of models, we average the policy weights to obtain the optimal policy.