Weighted Least Squares Estimates of Return Predictability Regressions

Travis L. Johnson*
McCombs School of Business
The University of Texas at Austin

November 2016

Abstract

Time varying volatility causes substantial heteroskedasticity in return predictability regressions, making OLS estimates less efficient than least squares estimates weighted by ex-ante return variance (WLS-EV). In small sample simulations, I show that using WLS-EV instead of OLS results in large efficiency gains, fewer false negatives, and avoids the bias associated with ex-post weighting schemes. Using WLS-EV changes several important conclusions based on OLS estimates: traditional predictors such as the dividend-to-price ratio perform better in- and out-of-sample, whereas WLS-EV estimates of the predictability afforded by the variance risk premium, politics, the weather, and the stars are not significant.

*Thanks to Svetlana Bryzgalova, John Griffin, Yufeng Han (SFA discussant), Bryan Kelly, Arthur Korteweg, Xiang (Nicole) Liu, Zack Liu, and seminar participants at the University of California San Diego, Massachusetts Institute of Technology, the 2015 Southern Finance Association Meeting, and The University of Texas at Austin for their helpful comments. I conducted part of this research while visiting the MIT Sloan School of Management. Send correspondences to travis.johnson@mccombs.utexas.edu or Speedway Stop B6600, Austin, TX 78712.

1. Introduction

Most return predictability regressions in financial economics take the form:

$$r_{t+1} = X_t \cdot \beta + \epsilon_{t+1}, \tag{1}$$

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} \left(r_{t+1} - X_t \cdot \beta\right)^2, \tag{2}$$

where X_t is a vector of predictor(s) that includes a constant. These regressions are typically estimated using ordinary least squares (OLS, equation (2)) with standard errors adjusted for any autocorrelation and heteroskedasticity in ϵ_{t+1}. Asymptotically, this approach results in point estimates and standard errors for β that are unbiased.

Despite their popularity, OLS estimates of return predictability regressions are inefficient, meaning β is estimated with more error than is necessary. In fact, OLS is only the most efficient linear unbiased estimator when the ϵ_{t+1} have no autocorrelation or heteroskedasticity. If we know the covariance matrix of the ϵ_{t+1}, Σ, then generalized least squares (GLS) is the most efficient linear unbiased estimator and should be used instead of OLS. However, the problem in most fields of economics is that Σ is unobservable and difficult to estimate. This view is summarized well by:

> "Many studies ... do not take advantage of the potential efficiency gains of GLS, for reasons of convenience and because the efficiency gains may be felt to be relatively small." (Cameron and Trivedi, 2005, page 81)

In this paper, I provide and implement a method for applying GLS to return predictability regressions that is convenient to use, results in large efficiency gains, and yields substantially different conclusions about frequently-studied return predictors than OLS. The reason GLS is so effective in return predictability regressions is that, unlike most fields in economics, finance has excellent estimates for both the conditional variance of returns and the autocorrelation of returns that inform us about the covariance matrix Σ. Specifically, there is an entire subfield of finance devoted to finding good estimates of the conditional variance of returns $\sigma_t^2 \equiv \mathrm{Var}_t(r_{t+1})$. There is also strong economic reasoning and empirical evidence that, as long as the return windows do not overlap, there is little to no autocorrelation in returns. In this case, the GLS estimator becomes:

$$\hat{\beta}_{\text{WLS-EV}} = \arg\min_{\beta} \sum_{t=1}^{T} \left(\frac{r_{t+1} - X_t \cdot \beta}{\sigma_t} \right)^2, \tag{3}$$

a weighted least squares estimator where the observations are weighted by ex-ante return variance $\frac{1}{\sigma_t^2}$. I abbreviate this procedure WLS-EV, and implement it using estimates of return variance $\hat{\sigma}_t^2$ suggested by the literature.

It is important to note WLS-EV downweights volatile observations *econometrically* and not *economically*. It estimates the same linear relation:[1]

$$\mathbb{E}_t(r_{t+1}) = a + b \cdot x_t, \tag{4}$$

just more efficiently by downweighting volatile observations. This is not to be confused with economically distinct alternatives, including a regression of $\frac{r_{t+1}}{\sigma_t}$ on x_t and a constant (a linear relation between Sharpe Ratio and x_t) or $\frac{r_{t+1}}{\sigma_t}$ on $\frac{x_t}{\sigma_t}$ and a constant (a linear relation between Sharpe Ratio and scaled predictors). The WLS-EV estimator in Equation (3) is equivalent to an OLS regression of $\frac{r_{t+1}}{\sigma_t}$ on $\frac{x_t}{\sigma_t}$ with the constant scaled to $\frac{1}{\sigma_t}$, making WLS-EV easy to implement. However, unlike economically distinct alternatives, volatile observations have the same linear relation between x_t and $\mathbb{E}(r_{t+1})$ given in Equation (4), they are just downweighted econometrically to produce more efficient estimates of a and b.

To illustrate the efficiency gains from using WLS-EV instead of OLS, I simulate samples designed to mimick the return predictability settings typically studied in the literature. In samples mimicking those used for the dividend-to-price ratio and variance risk premium, WLS-EV estimates are 24% and 28% less volatile, respectively, than OLS estimates.

[1] To illustrate economically distinct alternatives, I break X_t into a constant and a vector of predictors x_t.

The source of these efficiency gains is downweighting observations that occur in extremely volatile times. For example, in October 2008 the VIX index peaked at 80%, indicating next-month returns had a risk-neutral volatility of around 23% – more than four times the median level. In such extreme instances of volatility, realized returns are particularly noisy proxies for expected returns, making the signal-to-noise ratio low and the OLS weighting inefficiently high. Scaling the ϵ_{t+1} by σ_t "standardizes" them in units of ex-ante standard deviation and therefore makes them comparable in terms of information about expected returns.

My simulations also demonstrate that using WLS-EV improves small-sample hypothesis testing in two ways. First, using WLS-EV produces fewer false negatives than OLS in small samples because the point estimates are closer to the true (significant) value and standard errors are lower. Second, using WLS-EV procures fewer false positives than least squares estimates weighted using ex-post volatility information, for example the "robust least squares" approach used in Drechsler and Yaron (2011). The reason is that using ex-post volatility information introduces a bias due to the strong correlation between realized variance and ϵ_{t+1}. Because negative returns are more volatile than positive returns, negative ϵ have larger magnitudes and smaller weights than positive ϵ. As a result, when a predictor is positively (negatively) correlated with return variance, the coefficient estimated with RLS or any ex-post weighting scheme will be biased upwards (downwards).

The idea of weighting return predictability regressions by ex-ante variance is not new to the literature. Singleton (2006) discusses the econometric basis for this approach in Section 3.6.2. French, Schwert, and Stambaugh (1987) uses this procedure in the context of the "risk-return tradeoff" regression $r_{t+1} = a + b \cdot \sigma_t^2 + \epsilon_{t+1}$. The GARCH-in-mean framework estimated in Engle, Lilien, and Robins (1987) and Glosten, Jagannathan, and Runkle (1993), as well as the MIDAS framework in Ghysels, Santa-Clara, and Valkanov (2005), are structural approaches to incorporating conditional variance in estimating the risk-return tradeoff. I add to this literature by documenting the size and benefits of the efficiency gains WLS-EV affords, comparing it to alternatives, and applying it to predictors other than σ_t^2.

My primary contribution is to show WLS-EV produces three substantially different conclusions regarding return predictability than OLS. First, I show that using WLS-EV strengthens the in-sample and out-of-sample predictability afforded by the variables studied in Goyal and Welch (2008), lowering both the asymptotic and small-sample simulated standard errors without substantially changing the point estimates relative to OLS. For example, after adjusting for the Stambaugh (1999) bias, WLS-EV estimates indicate 8 of the 16 predictors I test significantly predict next-month returns at the 5% level, whereas OLS estimates indicate only 2 of the 16 are significant predictors.

Using WLS-EV also consistently improves upon the out-of-sample performance of OLS. Across 16 predictors, the average out-of-sample R^2 (OOS R^2 hereafter) improves for both next-month and next-year returns, as does the average out-of-sample R^2 achieved by the Campbell and Thompson (2008) approach and the Pettenuzzo, Timmermann, and Valkanov (2014) approach. The increase in OOS R^2 afforded by WLS-EV is not driven by a few outlier predictors, with 11 and 12 of the 16 experiencing increased OOS R^2 for next-month and next-year returns, respectively. The increase is also economically substantial, representing between 50% and 90% of in-sample OLS R^2.

Compared to other approaches to improving the out-of-sample performance of return predictors,[2] using WLS-EV has the advantage of being a minimal extension to OLS, making it easy to understand and implement. This approach also highlights one reason out-of-sample estimates based on OLS perform poorly: they are inefficient because they give full weight to extremely volatile observations with low signal-to-noise ratios.

The second contribution I make to the return predictability literature is showing the predictability afforded by proxies for the variance risk premium, documented in Bollerslev, Tauchen, and Zhou (2009) and Drechsler and Yaron (2011), is not robust to the WLS-EV approach. Across many alternate specifications, I find WLS-EV estimates of the relation

[2]For example, Campbell and Thompson (2008) and Pettenuzzo, Timmermann, and Valkanov (2014) impose economic restrictions on return forecasts, and Johannes, Korteweg, and Polson (2014) uses Bayesian estimates that incorporate estimation risk and time-varying volatility.

between variance risk premia and future market returns are statistically and economically insignificant. The insignificance of WLS-EV estimates indicate the OLS relation between the variance risk premium and future returns arises from a few observations with extreme values of the variance risk premium and high return volatility. As further evidence this is the case, even OLS estimates often become insignificant when standard errors are based on heteroskedastic simulations retaining the observed variance risk premia and return variances.

My results indicate that the empirical proxies and small sample we have do not provide compelling evidence the variance risk premium comoves with the equity premium. It remains possible that variance risk premia are indeed related to equity risk premia in the way described by the models in Bollerslev, Tauchen, and Zhou (2009) and Drechsler and Yaron (2011). We are limited empirically by the 25-year history of volatility indices, a very short sample for time-series return predictability analysis. Furthermore, even the WLS-EV estimates are often economically quite significant, especially for US data. For these reasons, I view the WLS-EV results as indicating we need more data before we can reach a conclusion about the predictability afforded by proxies for the variance risk premium.

My third contribution to the return predictability literature is showing the surprising predictability afforded by politics, the weather, and the stars, documented in Novy-Marx (2014), is insignificant when estimated using WLS-EV. One interpretation of the evidence in Novy-Marx (2014) is that the standard OLS methodology, when combined with sufficient data mining, may over-reject the null hypothesis of no predictability. If this is the case, WLS-EV can be useful as a partially-independent test of the same null hypothesis. Consistent with the false-positive interpretation, all three of the significant market return predictors in Novy-Marx (2014) are weakened by using WLS-EV instead of OLS, and WLS-EV indicates the nine proposed predictors are jointly insignificant.

2. Weighted least squares with ex-ante return variance

The weighted least squares with ex-ante return variance (WLS-EV) approach estimates the linear regression:

$$r_{t+1} = X_t \cdot \beta + \epsilon_{t+1}. \tag{5}$$

The returns r_{t+1} can be raw or log returns, can be overlapping or non-overlapping, and can be adjusted for the risk-free rate or unadjusted. There can be multiple return predictors along with an optional constant in the X_t vector.

There are two steps to estimating β in Equation (5) using WLS-EV:

1. Estimate σ_t^2, the conditional variance of next-period unexpected returns ϵ_{t+1}.

2. Estimate $\hat{\beta}_{\text{WLS-EV}}$ using:

$$\hat{\beta}_{\text{WLS-EV}} = \arg\min_{\beta} \sum_{t=1}^{T} \left(\frac{r_{t+1} - X_t \cdot \beta}{\hat{\sigma}_t} \right)^2, \tag{6}$$

   where $\hat{\sigma}_t$ is the empirical estimate of σ_t. This estimator can be implemented using any OLS package by regressing $\frac{r_{t+1}}{\hat{\sigma}_t}$ on $\frac{X_t}{\hat{\sigma}_t}$. Note that, since the constant is in X_t, this OLS regression has no constant term.

There are many different potential approaches in the literature for estimating σ_t^2, the ex-ante variance of next-period returns, any of which can be used to estimate WLS-EV.

Standard errors for WLS-EV are the same as OLS standard errors when regressing the weighted returns $\frac{r_{t+1}}{\hat{\sigma}_t}$ on the weighted constant $\frac{1}{\hat{\sigma}_t}$ and regressors $\frac{x_t}{\hat{\sigma}_t}$. These standard errors can be adjusted for remaining heteroskedasticity and autocorrelation using the standard Newey and West (1987) HAC adjustment or a simulation approach I describe in Section 3.

2.1. Estimating $\hat{\sigma}_t^2$

The estimator in Equation (6) is the perfectly-efficient GLS estimator if and only if:

$$\text{Var}(\epsilon_{t+1}) = \hat{\sigma}_t^2, \text{ and} \tag{7}$$

$$\text{Cov}(\epsilon_s, \epsilon_t) = 0 \qquad \forall s \neq t. \tag{8}$$

The latter condition requires that any autocorrelation in returns arises through the X_t variables, making unexpected returns uncorrelated at any lag. Rational asset pricing models predict that, given the right X_t variables, non-overlapping returns satisfy this property. I assume this is the case in this section, and discuss overlapping returns in Section 2.2.

The condition in Equation (7) requires that the $\hat{\sigma}_t^2$ used empirically are the true variances for future unexpected returns σ_t^2. Since the true σ_t^2 are unobservable, I strive to find ex-ante proxies that are as accurate as possible. Proxies based on ex-post information about realized returns could more-accurately reflect the true variance of ϵ_{t+1}, but by using time $t+1$ information they introduce a substantial bias in $\hat{\beta}$ I discuss in Section 2.3.

I use two simple and effective proxies for conditional variance $\hat{\sigma}_t^2$, both fitted values from regressions of realized variance on past variance and the VIX index when it is available (starting in 1990).[3] Specifically, for monthly samples pre-dating 1990, I use fitted values from a first-stage regression of the form:

$$\text{RV}_{t+1} = a + b \cdot \text{RV}_t + c \cdot \text{RV}_{t-11,t} + \gamma_{t+1}, \tag{9}$$

where RV_t is the annualized sum of squared daily log index returns in month t, and $\text{RV}_{t-11,t}$ is the sum of squared daily log index returns in months $t-11$ through t. For daily samples

[3]A Black-Scholes version of VIX, calculated for the S&P 100, is available starting in 1986, but I restrict my attention to the model-free calculation of the VIX which is available starting in 1990.

using post-1990 data, I use fitted values from first-stage regressions of the form:

$$\text{FutRV}_{t+1} = a + b \cdot \text{FutRV}_t + c \cdot \text{VIX}_t^2 + \gamma_{t+1}, \tag{10}$$

where VIX_t is the VIX index on day t and FutRV_t is the annualized sum of squared five-minute log S&P 500 futures returns on day t. In estimating both (9) and (10), I restrict the intercept and coefficients to be positive so the fitted values I use for $\hat{\sigma}_t$ are always positive.

Fitted values from these regressions have the advantage of being simple to compute while still leveraging four key conclusions from the literature on return variance:

1. Variance is mean-reverting (Merville and Pieptea (1989)). I therefore include a constant in (9) and (10) instead of assuming future variance is proportional to past variance.

2. Within-period realized variance is a better proxy for realized variance than squared full-period returns (Merton (1980)). I therefore use within-month and within-day realized variance as my outcome variables.

3. Past variance is a better predictor of realized variance than structural estimates from models like GARCH (Ghysels, Santa-Clara, and Valkanov (2005)). I therefore use past realized variance as my primary predictor.

4. When available, option-implied variance is the best variance predictor, capturing most economically significant variation in conditional volatility (Christensen and Prabhala (1998)). I therefore include the VIX, a model-free estimate of S&P 500 option-implied volatility, as my only predictor other than past realized variance.

Table 1 shows estimates of Equations (9) and (10) along with other potential first-stage regressions. For both monthly and daily realized variance, these ex-ante variables explain a significant proportion of realized variance, with R^2 between 25% and 50%, indicating WLS-EV could provide substantial efficiency gains relative to OLS. For next-month variance, the unconstrained intercepts and coefficients are always strictly positive, resulting in positive $\hat{\sigma}_t^2$

without any further adjustment. For next-day variance, Columns (3), (5), and (7) of Panel B show that when VIX_t^2 is included, the intercept and some fitted values become negative. The constrained estimates in Column (4), (6), and (8) have no intercept, positive coefficients, and suffer very little reduction in R^2.

I use Columns (6) in Panel A and (7) in Panel B, which include all potential predictors, to guide my choice of specification for computing $\hat{\sigma}_t^2$. In Panel A, only prior month and prior year realized variance are statistically significant predictors, and as Column (5) illustrates they combine to provide nearly all the predictability afforded by all four lags of realized variance. For this reason, I use the more-parsimonious specification given in Equation (9) to produce fitted values for my main ex-ante variance proxy, which I refer to as RV $\hat{\sigma}_t^2$ hereafter:[4]

$$\text{RV } \hat{\sigma}_t^2 \equiv \hat{a} + \hat{b} \cdot \text{RV}_t + \hat{c} \cdot \text{RV}_{t-11,t}. \tag{11}$$

Similarly, only VIX_t^2 and FutRV_t are significant incremental predictors in Column (7) of Panel B, and so I use the specification given in Equation (10), constrained so that $a \geq 0$, to produce fitted values for VIXF $\hat{\sigma}_t^2$, my ex-ante variance proxy for post-1990 samples:

$$\text{VIXF } \hat{\sigma}_t^2 \equiv \hat{a} + \hat{b} \cdot \text{FutRV}_t + \hat{c} \cdot \text{VIX}_t^2. \tag{12}$$

While these ex-ante variance proxies are effective empirically, other proxies may predict realized variance as well or even better. As discussed above, any of these can be used in WLS-EV as long as they are constructed from ex-ante information. Fortunately, these proxies are strongly correlated with each other, and in untabulated tests I find my results are not sensitive to using other predictors in Table 1, MIDAS estimates following Ghysels, Santa-Clara, and Valkanov (2005), or a variety of other proxies. As a robustness check and to illustrate the effectiveness of a simple alternative, in some of my tests I supplement VIXF $\hat{\sigma}_t^2$

[4] I do not include the incrementally-insignificant $\text{RV}_{t-2,t}$ and $\text{RV}_{t-5,t}$ in the first stage regression for parsimony and to improve out-of-sample performance.

with VIX $\hat{\sigma}_t^2$, the fitted value from a regression with only VIX_t^2 as a predictor:

$$\text{VIX } \hat{\sigma}_t^2 \equiv \hat{a} + \hat{b} \cdot \text{VIX}_t^2. \tag{13}$$

Figure 1 plots RV $\hat{\sigma}_t^2$ for my 1927-2013 sample, and VIXF $\hat{\sigma}_t^2$ for 1990-2013, both displayed as annualized standard deviations. Like other conditional volatility estimates, RV $\hat{\sigma}_t$ is small and steady in normal times but spikes upwards during market downturns, particularly in 1929, 1987, and 2008. These episodes have conditional return volatility higher than 50%, approximately three times the typical values between 15% and 20%. The more-recent sample shows similar patterns but with even more extreme values during the 2008 crisis. Together with the R^2 in Table 1, the extreme movements in conditional volatility shown in Figure 1 indicate the first-stage regressions I use to compute $\hat{\sigma}_t$ capture a substantial fraction of heteroskedasticity in returns, allowing WLS-EV to substantially improve efficiency.

2.2. Overlapping returns

To maximize power from relatively short samples, many return predictability studies use sampling frequencies greater than their forecast horizon, resulting in overlapping returns. The standard approach in this case is to estimate $\hat{\beta}$ using OLS and adjust the standard errors using the procedures suggested by Newey and West (1987) or Hodrick (1992).

To apply GLS in this setting, I rely on the insight in Hodrick (1992) that the overlapping return predictability coefficient is isomorphic to the coefficient in a non-overlapping regression of returns on a rolling sum of X_t. Writing log returns r_t, consider a regression of next h-period returns on X_t:

$$r_{t+1,t+h} = X_t \cdot \beta + \epsilon_{t+1,t+h} \tag{14}$$

$$\Rightarrow \hat{\beta}_{\text{OLS}} = \mathbb{E}_T(X_t' X_t)^{-1} \mathbb{E}_T(X_t' r_{t+1,t+h}), \tag{15}$$

where $\mathbb{E}_T$ represents the sample average. Substituting in $r_{t+1,t+h} = \sum_{s=1}^{h} r_{t+s}$, we have:

$$\hat{\beta}_{\text{OLS}} = \mathbb{E}_T(X_t'X_t)^{-1}\mathbb{E}_T\left(\sum_{s=1}^{h} X_t' r_{t+s}\right) = \mathbb{E}_T(X_t'X_t)^{-1}\mathbb{E}_T(\overline{X}_t'\overline{X}_t)\hat{\beta}_{\text{OLS}}^{\text{roll}}, \tag{16}$$

$$\hat{\beta}_{\text{OLS}}^{\text{roll}} \equiv \mathbb{E}_T(\overline{X}_t'\overline{X}_t)^{-1}\mathbb{E}_T(\overline{X}_t' r_{t+1}), \tag{17}$$

$$\overline{X}_t \equiv \sum_{s=0}^{h-1} X_{t-s}. \tag{18}$$

In words, the overlapping OLS β is identical to the OLS β in a *non-overlapping* regression of r_{t+1} on a rolling sum of X_t, scaled by matrix of constants. When X_t includes a constant and a univariate x_t, Equation (17) reduces to:

$$\hat{b}_{\text{OLS}} = \frac{\text{Var}_T(\overline{x}_t)}{\text{Var}_T\,(x_t)} \cdot \hat{b}_{\text{OLS}}^{\text{roll}} \tag{19}$$

$$\hat{b}_{\text{OLS}}^{\text{roll}} \equiv \frac{\text{Cov}_T(r_{t+1}, \overline{x}_t)}{\text{Var}_T(\overline{x}_t)}. \tag{20}$$

I use this insight to estimate $\hat{\beta}$ in overlapping samples using OLS or WLS-EV as follows:

1. Estimate the non-overlapping regression $r_{t+1} = \left(\sum_{s=0}^{h-1} X_{t-s}\right) \cdot \beta + \epsilon_{t+1}$ using either OLS or WLS-EV. Use Newey and West (1987) standard errors with lags selected following Newey and West (1994) to adjust for remaining heteroskedasticity or autocorrelation.

2. Scale the resulting coefficients and standard errors by $\mathbb{E}_T(X_t'X_t)^{-1}\mathbb{E}_T(\overline{X}_t'\overline{X}_t)$, which simplifies to $\frac{\text{Var}_T(\overline{x}_t)}{\text{Var}_T(x_t)}$ when X_t has a constant and univariate predictor.

Note that while the resulting $\hat{\beta}_{\text{OLS}}$ are identical to the overlapping regression $\hat{\beta}$, the standard errors are different because they adjust for the autocorrelation in $\epsilon_{t+1,t+h}$ by specifying its structure as function of the overlap rather than estimating it using Newey and West (1987). Simulations in Hodrick (1992) show these standard errors have better small-sample properties for overlapping return regressions than Newey and West (1987) standard errors.

By transforming an overlapping return regression into a non-overlapping regression, the modified Hodrick (1992) procedure I use assures the WLS-EV estimates are the most-efficient

GLS estimates under the assumptions described in Section 2.1. Without this transformation, GLS would require specifying the full covariance matrix of the errors and to estimate β. For example, if we estimate the variance of each daily return and assume these returns are independent, we can compute the full covariance matrix for the overlapping returns and use that as a proxy for the covariance matrix of the errors. However, this approach requires using variance information from time $t+1$ to weight observations with time t variables on the right-hand side, creating a potential bias I discuss in Section 2.3.

A possible alternative to transforming the regression using Hodrick (1992) is to use least squares weighted by conditional next-h period variance to account for heteroskedasticity in estimating β, and HAC consistent standard errors from Newey and West (1987) or simulations to account for any remaining heteroskedasticity and autocorrelation driven by the overlap. This approach suffers from at least three problems. The first is the conditional next-h period variance measures do not predict realized variance as well as conditional next-period variance measures, reducing the efficiency gains associated with WLS. The second is the overlapping conditional variances are often inconsistent with eachother in the sense that no path of per-period conditional variances would justify them, making it impossible to simulate returns under the null that the conditional variances are correct. The third is that the same small-sample bias in Newey and West (1987) standard errors for overlapping return regressions documented in Hodrick (1992) applies here.[5]

2.3. WLS-EV compared to other weighting functions

Previous papers studying market-level return predictability use "robust least squares" (RLS) estimates (e.g., Drechsler and Yaron (2011)), which weight observations using some function of estimated values of $|\epsilon_{t+1}|$. Observations with larger $|\epsilon_{t+1}|$ presumably also have more volatile ϵ_{t+1} on average, and therefore receive smaller weights. These weights use information from the time period returns are realized, $t+1$, rather than the ex-ante variance

[5] The Hodrick (1992) technique for computing standard errors without transforming to non-overlapping regressions cannot be directly here because each overlapping observation is weighted by a different $\sigma^2_{t,h}$.

measures available at time t I use in WLS-EV. The advantage of using time $t+1$ information is that return volatility changes over time, meaning $t+1$ information can provide more accurate estimates of the true variance of ϵ_{t+1}.

However, there is a critical disadvantage to using time $t+1$ information that, to my knowledge, is not discussed in any previous papers using ex-post weighting schemes: there is a strong correlation between realized variance and the directional realization of ϵ_{t+1} that biases the coefficient estimates. Because negative returns are more volatile than positive returns, negative ϵ have larger variance and smaller weights than positive ϵ. As a result, when the predictor X_t is positively (negatively) correlated with return variance, the coefficient estimated with RLS or any ex-post weighting scheme will be biased upwards (downwards). It is therefore unsurprising, given the variance risk premium is positively correlated with return variance, that Drechsler and Yaron (2011) finds RLS coefficients are more positive than OLS coefficients.

By comparison, the WLS-EV approach uses weights based exclusively on ex-ante information, avoiding the mechanical connection between weights and the average ϵ_{t+1}. However, there could potentially still be a correlation between WLS-EV weights and ϵ_{t+1} if ex-ante variance predicted future returns. Empirically, weights based on both the RV $\hat{\sigma}_t$ and VIXF $\hat{\sigma}_t$ have near-zero correlation with next-period market returns, detailed in Appendix A. More importantly, any such bias can be corrected for by adding the weights $1/\hat{\sigma}_t^2$ to the right-hand side of the regression, assuring that the regression residuals are independent of the weights. Regression weights using time $t+1$ information, by contrast, cannot be added as independent variables in predictive regressions.

I formalize this discussion in Appendix A by deriving the the estimation error $\hat{\beta} - \beta$ in a general weighted least squares setting and showing conditions under which the average estimation error is zero (i.e., the estimator is consistent). I also provide evidence WLS-EV weights meet these conditions while RLS weights do not. More practically, the small sample simulations in Section 3 show that only RLS estimates are biased in realistic settings.

3. Small sample simulations

I use small sample simulations to illustrate the relative efficiency and bias of three different linear estimation techniques: ordinary least squares (OLS), robust least squares (RLS), and weighted least squares using ex-ante variance (WLS-EV). I find that WLS-EV is unbiased and substantially more efficient than OLS. RLS estimates are also more efficient than OLS estimates, but are less efficient than WLS-EV estimates and suffer severely from the bias described in Section 2.3. Furthermore, RLS standard errors are understated in small samples, resulting in frequent false positives.

The efficiency and bias of each estimation procedure depends critically on variability of return variance, the asymmetry in the return distribution, the time-series distribution of the predictor, and the correlations among these variables. Rather than attempting to model these distributions, I use observed return predictors and conditional variances but re-sample the realized return innovations. Specifically, given observed excess returns r_t^{data} and ex-ante return volatilities $\hat{\sigma}_t^{\text{data}}$, I compute the standardized next-period return for each observation:

$$\psi_{t+1}^{\text{data}} \equiv \frac{r_{t+1}^{\text{data}} - \mu_r}{\hat{\sigma}_t^{\text{data}}}, \tag{21}$$

where μ_r is chosen so that $\mathbb{E}(\psi_{t+1}^{\text{data}}) = 0$. I then create 100,000 simulated samples by re-sampling the ψ_t^{data} (with replacement) and computing the next-month returns as follows:

$$r_{t+1}^{\text{sim}} = \mu_r + b \cdot x_t^{\text{data}} + \hat{\sigma}_t^{\text{data}} \psi_{t+1}^{\text{re-sampled}}, \tag{22}$$

where x_t^{data} are the observed values of a predictor variable, and I specify the population prediction coefficient b. These simulated returns inherit the skewness, any heteroskedasticity not captured by $\hat{\sigma}_t^{\text{data}}$, and other properties of the observed return distribution while still having conditional mean $\mu_r + b \cdot x_t^{\text{data}}$ and conditional volatility $\hat{\sigma}_t^{\text{data}}$. For each simulated return sample, I regress the redrawn excess returns r_{t+1}^{sim} on x_t^{data} and a constant using each of the three regression techniques, and record the resulting coefficients ($\hat{b}$) and standard errors (SE $\hat{b}$). For RLS, I use the bisquare weighting function and HAC consistent standard errors.

I first implement this procedure on a monthly sample from 1927-2013 using the log dividend-to-price ratio dp as the predictor x_t and RV $\hat{\sigma}_t^2$ for the WLS-EV estimates. Panel A of Table 2 shows summary statistics for these simulations under the no-predictability null $b = 0$. Using WLS-EV rather than OLS results in large efficiency gains, reducing the standard deviation of $\hat{b}$ from 0.440 to 0.333, a 24% decrease. RLS $\hat{b}$ have a standard deviation of 0.369, making them nearly as efficient as WLS-EV regressions. However, while the mean $\hat{b}$ is zero for OLS and WLS-EV, the mean $\hat{b}$ from RLS is positive, reflecting the aforementioned bias that arises because RLS weights are positively correlated with both ϵ_{t+1} and x_t.[6]

Given the true b is zero, an effective estimator rejects the $b = 0$ null (a "false positive") as infrequently as possible. There are two potential reasons why an estimator would reject with a 5% critical value in more than 5% of simulations: downward bias in asymptotic standard errors and directional bias in the average $\hat{b}$. In addition to having unbiased $\hat{b}$, both OLS and WLS-EV have average standard errors very close to the standard deviation of $\hat{b}$, indicating that the asymptotic heteroskedasticity-consistent standard errors are unbiased and quite accurate for dp in this sample. As a result, OLS and WLS-EV t-tests reject the null at the 5% level in 5.11% and 5.12% of simulations, respectively. RLS, by contrast, has significant downward bias in standard errors in addition to an upward bias in $\hat{b}$, resulting in false positives in 14.38% of simulations, 7.45% (6.93%) with positive (negative) coefficients.

Given the true b is non-zero, an effective estimator fails to reject the $b = 0$ (a "false negative") as little as possible. To assess the frequency of false negatives, I repeat the simulation exercise assuming $b = 1$. Panel B of Table 2 presents the results. Because the only difference from the simulations in Panel A is the added $b \cdot x_t^{\text{data}}$ to Equation (22), the efficiency and bias of the estimators are identical to those in Panel A.

[6] There is no Stambaugh (1999) bias here because the re-drawn standardized returns are uncorrelated with innovations in dp. This allows me to examine the efficiency and bias associated with heteroskedasticity alone. I correct for the Stambaugh (1999) bias in Section 4.

The main result in Panel B, however, is the fraction of simulations for which each estimator fails to recognize the predictive power of x_t. For each simulation, I compute both the asymptotic t-stat as well as the small-sample p-value based on the distribution of $\hat{b}$ for each estimator under the no-predictability null. A "false negative" is a case in which the asymptotic t-stat is less than 1.96 or the simulated p-value is more than 5%. For OLS and WLS-EV, because the asymptotic standard errors are almost identical to the simulated ones, asymptotic and small-sample tests have the same false negative rates. Furthermore, false negatives occur much less often for WLS-EV (15% of simulations) than OLS (37% of simulations). For RLS, the asymptotic t-stats are less than 1.96 quite infrequently (9% of the simulations) because of the upward bias in $\hat{b}$ and downward bias in the asymptotic standard errors. However, when using simulation-based p-values, the less-efficient RLS estimator results in false negative rates around 21%, between OLS and WLS-EV.

To assess efficiency and bias in shorter samples and using a predictor more directly related to ex-ante variance, I also implement this procedure on an overlapping daily sample from 1990-2013 using the variance risk premium proxy defined in Drechsler and Yaron (2011) as x_t to predict next-month returns.[7] The results are in the second column of Table 2. The conclusions are largely the same as for dividend yields, but the effects are bigger because of the stronger correlation between x_t and ex-ante variance. WLS-EV estimates have a standard deviation of 0.162, 28% more efficient than the 0.225 standard deviation of OLS estimates and 15% more efficent than the 0.190 standard deviation of RLS estimates. More importantly, the upward bias in the RLS coefficients is much larger for this x_t, almost three times the asymptotic standard error, while OLS and WLS-EV $\hat{b}$ remain unbiased. To make matters worse, the asymptotic RLS standard errors are dramatically understated, 40% smaller than the cross-simulation standard deviation. The upward bias in $\hat{b}$ together with the downward bias in asymptotic standard errors combine make false positives are extremely likely for RLS, with resulting t-stats are above 1.96 in 71% of simulations under the $b = 0$ null. In light of

[7]See Section 5 for detailed description of this proxy.

this upward bias in t-stats, it is not surprising that Drechsler and Yaron (2011) find RLS coefficients are both larger and more significant than OLS coefficients.

Finally, I assess the false negative rate of the three estimators in the variance risk premium setting with $b = 0.4$. Because of the severe downward bias in asymptotic standard errors, it is important to do hypothesis tests using simulated p-values rather than asymptotic standard errors. Mirroring the results in Panel B for dividend yield, WLS-EV has false negatives in 30.2% of simulations, compared to 56.6% for OLS and 44.3% for RLS.

A potential concern about the simulations described in Equation (22) is that the $\hat{\sigma}_t^{\text{data}}$ I use for WLS-EV are the exact conditional variances of returns, perhaps resulting in an over-estimate of the efficiency gains. I address this concern by redrawing the ψ_{t+1}^{data}, which reflect any residual heteroskedasticity not corrected by $\hat{\sigma}_t^{\text{data}}$ and therefore produce simulated samples with as much uncorrected heteroskedasticity as the observed samples. However, to provide additional reassurance, I redo the simulations with the same return generating process but using WLS-EV with log-normal noise multiplying the variance estimates:

$$\hat{\sigma}_t^{\text{WLS-EV}} = \hat{\sigma}_t^{\text{data}} e^{z_t}, \qquad \sigma(z_t) = \frac{1}{2}. \tag{23}$$

Panel C of Table 2 shows that even when WLS-EV is estimated using a noisy $\hat{\sigma}_t$, the efficiency gains are substantial relative to OLS, though smaller than the gains in Panel A. The cross-simulation standard deviation of WLS-EV estimates is 0.378 for dp simulations and 0.184 for variance risk premium siumulations, representing efficiency gains of 14% and 18%, respectively. More importantly, the asymptotic HAC standard errors I use for WLS-EV reflect this decrease in efficiency, averaging 0.374 and 0.183. This indicates the HAC standard errors detect how much noise is in the WLS-EV weights and correct the standard errors appropriately. It also suggests my simulation approach is effective in carrying through any residual heteroskedasticity caused by noise in the $\hat{\sigma}_t^{\text{data}}$, giving credence to the use of WLS-EV for efficient hypothesis testing even when ex-ante variance proxies are imperfect.

4. Traditional predictors

My first application of the WLS-EV methodology is to re-assess the return predictability afforded by the 16 variables studied in Goyal and Welch (2008). Overall, I find the evidence for return predictability both in-sample and out-of-sample is substantially stronger with WLS-EV than the marginal OLS evidence.

The 16 predictors I study are: the log dividend-to-price ratio (dp), the dividend-to-price ratio (DP), the log earnings-to-price ratio (ep), the log dividend-to-earnings ratio (de), the conditional variance of returns estimated using rolling estimates of Equation (9) (RV $\hat{\sigma}_t^2$), the treasury bill yield (tbl), the long-term treasury bond yield (lty), the return of long-term bonds (ltr), the term spread (tms), the default yield spread (dfy), inflation (infl), the log book-to-market ratio (bm), the cross-sectional beta premium (csp), net equity expansion (ntis), the log net payout yield (lpy), and the consumption wealth ratio (cay). To improve the readability of the coefficients, I divide dp, ep, de, bm, and lpy by 100. I compute RV $\hat{\sigma}_t^2$, and retrieve lpy from Michael Roberts' website, cay from Martin Lettau's website, and the remaining 13 predictors from Amit Goyal's website. Detailed definitions of the predictors are in Boudoukh et al. (2007) for lpy, Lettau and Ludvigson (2001) for cay, and Goyal and Welch (2008) for the remaining 13 predictors.

4.1. In-sample predictability

For each of the 16 predictors, I estimate univariate predictability regressions of the form:

$$r_{t+1,t+h} = a + b \cdot x_t + \epsilon_{t+1,t+h}, \tag{24}$$

where $r_{t+1,t+h}$ is the log excess return of the CRSP value-weighted index in months $t+1$ through $t+h$. I use both the standard OLS and WLS-EV to estimate the coefficients a and b. I assess next-month ($h = 1$) and next-year ($h = 12$) predictability, and adjust for the overlap when $h = 12$ using the procedure in Section 2.2. I also compute simulated standard errors using simulations identical to the ones described in Section 3.

To accurately assess the predictability afforded by these candidate variables, I account for the small-sample bias described in Stambaugh (1999) by simulating both the x_t and subsequent returns r_{t+1} under the no-predictability null, as suggested in Goyal and Welch (2008).[8] Specifically, I generate r_{t+1} and x_t using the following processes:

$$r_{t+1}^{\text{sim}} = \mu_r + \sigma_t^{\text{sim}} \epsilon_{t+1}^{\text{re-sampled}} \tag{25}$$

$$x_{t+1}^{\text{sim}} - \mu_x = \rho_x (x_t^{\text{sim}} - \mu_x) + \delta_{t+1}^{\text{re-sampled}} \tag{26}$$

$$\log \sigma_{t+1}^{\text{sim}} - \mu_\sigma = \rho_\sigma (\log \sigma_t^{\text{sim}} - \mu_\sigma) + \gamma_{t+1}^{\text{re-sampled}}, \tag{27}$$

where μ_r, μ_x, μ_σ, ρ_x, and ρ_σ are estimated from the data for the predictor in question, and x_0 and σ_0 are chosen from a random date in the sample period. To preserve the correlations among innovations in r, x, and σ^2, I jointly re-sample (with replacement) the innovations vector $\begin{bmatrix} \epsilon_{t+1} & \delta_{t+1} & \gamma_{t+1} \end{bmatrix}'$ from the innovations observed in the data.

The only difference from the approach I use to estimate the Stambaugh (1999) bias and the Goyal and Welch (2008) approach is the addition of stochastic volatility as modeled by σ_t. The reason for this addition is to allow me to use WLS-EV in the simulated samples, which I do using $\left(\sigma_t^{\text{sim}}\right)^2$ as the ex-ante variance proxy. For each simulated sample, I estimate the predictive regression in Equation (24) using both OLS and WLS-EV and record the average $\hat{b}$. To the extent the Stambaugh (1999) bias affects each combination of estimator, return predictor, and covariance matrix of r, x, and σ^2, and forecast horizon h, the average simulated $\hat{b}$ will be non-zero despite the no-predictability null. For this reason, each predictor's OLS and WLS-EV Stambaugh (1999) bias-corrected coefficients are defined as:

$$\text{OLS Stambaugh } \hat{b}_{\text{adj}} \equiv \text{OLS } \hat{b} - \mathbb{E}_{\text{Stambaugh sim}} \left(\text{OLS } \hat{b} \right) \tag{28}$$

$$\text{WLS-EV Stambaugh } \hat{b}_{\text{adj}} \equiv \text{WLS-EV } \hat{b} - \mathbb{E}_{\text{Stambaugh sim}} \left(\text{WLS-EV } \hat{b} \right) \tag{29}$$

[8] My simulations in Section 3 only redraw returns and therefore do not reflect the Stambaugh (1999) bias.

The results of these in-sample tests are in Table 3, beginning with a one-month prediction horizon ($h = 1$) in Panel A. As summarized at the bottom of the panel, the WLS-EV estimates have asymptotic and simulated standard errors an average of 26% smaller than their OLS counterparts. Furthermore, the WLS-EV point estimates are generally consistent with the OLS point estimates in most cases, and substantially larger for tbl, lty, ltr and infl. Combining these features strengthens the overall in-sample evidence of return predictability. Using 1%, 5%, and 10% critical values, WLS-EV results in statistical significance for nine, eight, and five of the predictors, compared to only three, two, and one for OLS.

I assess the predictive power of these 16 variables for next-year returns ($h = 12$) in Panel B of Table 3. The results are largely consistent with the next-month return results in Panel A, indicating stronger but not overwhelming in-sample evidence of return predictability. The WLS-EV approach yields 25% smaller simulated standard errors and largely unchanged point estimates, making the WLS-EV evidence for return predictability stronger than the OLS evidence. Using 1%, 5%, and 10% critical values, WLS-EV results in statistical significance for seven, six, and two predictors compared to five, three, and one for OLS.

4.2. Out-of-sample predictability

There are a few potential concerns with the evidence supporting return predictability in Table 3. The first is data mining: the predictive variables are not chosen at random, but instead selected among many potential predictors based on their statistical significance. The second concern is a bias in the standard errors not captured by the asymptotic HAC or simulation standard errors I use to test the no-predictability null hypothesis. The third concern is introduced by my heteroskedastic simulations and WLS-EV approach, namely that the RV $\hat{\sigma}_t^2$ I use are noisy proxies for the true conditional variance of returns. As discussed above, this third concern is diminished by the evidence in Table 2 and the use of re-drawn regression errors that retain any remaining heteroskedasticity. However, without observing the true σ_t^2, in-sample tests cannot completely rule out the possibility that errors in RV $\hat{\sigma}_t^2$ cause my simulations to understate the WLS-EV standard errors.

To address these concerns, I examine the out-of-sample predictive power of these regressors using both OLS and WLS-EV. As discussed in Goyal and Welch (2008), out-of-sample tests provide an additional falsifiable implication of the no-predictablity null that we can test for the variables that predict future returns in-sample. While there is some debate (e.g. in Cochrane (2008) or Campbell and Thompson (2008)) about the power of out-of-sample tests for rejecting the null, making a failure to reject hard to interpret, any significant out-of-sample return predictability would be strong evidence in favor predictability because it cannot be explained by the three aforementioned concerns. Data mining cannot explain out-of-sample predictability because these predictors were not selected for publication based on out-of-sample performance. Biases or noise in standard errors, point estimates, and the RV $\hat{\sigma}_t^2$ also cannot explain out-of-sample predictability because it does not use the standard errors and would be impaired by any bias in the point estimates or RV $\hat{\sigma}_t^2$.

In addition to providing researchers with an alternative test of the no-predictability null that avoids the aforementioned biases, out-of-sample predictability provides a simple measure of the practical value a predictor offers to investors. As discussed in Campbell and Thompson (2008), Johannes, Korteweg, and Polson (2014), and elsewhere, investors may use more sophisticated techniques in forming expectations about future market returns and their portfolios. Nevertheless, out-of-sample R^2 provides a good indicator of which predictors would have benefited investors if used in "real-time" over the past century.

I compute the out-of-sample R^2 for each predictor using a procedure very similar to the one in Goyal and Welch (2008). Specifically, for each date τ in my 1927-2013 sample,[9] starting 20 years after the first month the predictor is available, I compute the conditional expected future return over the next h months, $\mathbb{E}_\tau \left(r_{\tau+1,\tau+h} | x \right)$ as follows:

[9] Unlike Goyal and Welch (2008), my sample starts in 1927 because I require daily return data to compute the RV $\hat{\sigma}_t^2$, and ends in 2013 rather than 2005. The only exceptions are csp (available 1937-2002), lpy (available 1927-2010), and caya (the ex-ante version of cay, available 1952-2013).

1. Estimate coefficients $\hat{a}_\tau$ and $\hat{b}_\tau$ in the regression:

$$r_{t+1,t+h} = a_\tau + b_\tau \cdot x_t + \epsilon_{t+1,t+h}, \tag{30}$$

using OLS or WLS-EV, and only data available as of τ, i.e. $t \leq \tau - h$.[10] To maximize power, I use overlapping monthly regressions instead of the annual regressions used in Goyal and Welch (2008).

2. Use estimated coefficients and current predictor values to compute:

$$\mathbb{E}_\tau \left(r_{\tau+1,\tau+h} | x \right) \equiv \hat{a}_\tau + \hat{b}_\tau x_\tau. \tag{31}$$

As a benchmark, I also compute the unconditional out-of-sample return prediction based on a simple average of past returns:

$$\mathbb{E}_\tau \left(r_{\tau+1,\tau+h} \right) \equiv \frac{1}{\tau - h} \sum_{t=1}^{\tau-h} r_{t+1,t+h}. \tag{32}$$

Given time-series of out-of-sample return predictions $\mathbb{E}_\tau \left(r_{\tau+1,\tau+h} | x \right)$ and $\mathbb{E}_\tau (r_{\tau+1,\tau+h})$, I compute the out-of-sample R^2 and adjusted R^2 as in Goyal and Welch (2008):

$$R^2 \equiv 1 - \frac{\text{MSE}_A}{\text{MSE}_N}, \qquad \text{Adj. } R^2 \equiv R^2 - (1 - R^2) \frac{K}{T - K - 1}, \tag{33}$$

$$\text{MSE}_A \equiv \frac{1}{T} \sum_{\tau=1}^{T} e_A(\tau, x)^2 \qquad \text{MSE}_N \equiv \frac{1}{T} \sum_{\tau=1}^{T} e_N(\tau)^2 \tag{34}$$

$$e_A(\tau, x) \equiv r_{\tau+1,\tau+h} - \mathbb{E}_\tau \left(r_{\tau+1,\tau+h} | x \right) \qquad e_N(\tau) \equiv r_{\tau+1,\tau+h} - \mathbb{E}_\tau \left(r_{\tau+1,\tau+h} \right) \tag{35}$$

where T is the number of observations in the post-training sample period and K is the number of regressors (including the constant). Following Goyal and Welch (2008), I focus my analysis on in- and out-of-sample adjusted R^2.

[10] For WLS-EV, I estimate the first-stage variance prediction regression using only data available as of τ.

Table 4 presents the adjusted out-of-sample R^2 (OOS R^2 hereafter) afforded by the 16 traditional predictors I study. Panel A examines out-of-sample next-month return predictions, Panel B next-year return predictions. While the longer overlapping sample I use results in somewhat better out-of-sample performance than documented in Goyal and Welch (2008), the takeaway remains that these predictors combined with OLS do not produce significant OOS R^2. Only four predictors have positive OOS R^2 for next-month returns (DP, tms, infl, and caya), and only two have positive OOS R^2 for next-year returns (ltr and caya).

Table 4 also shows the out-of-sample performance of WLS-EV is substantially better than OLS. WLS-EV OOS R^2 are higher than their OLS counterparts for 11 of the next-month and 12 of the next-year predictors. Mean OOS R^2 is -0.22% and -3.83% for next-month and next-year returns for WLS-EV, compared to -0.39% and -6.72% for OLS. This increases are economically large relative to the average in-sample OLS R^2 of 0.33% and 3.08%, respectively. Finally, and most importantly, four predictors offer positive OOS R^2 using WLS-EV for both next-month returns and next-year returns (DP, ltr, tms, and infl). Some of these OOS R^2 are also quite substantial, varying from 0.05% to 0.55% for next-month returns and 0.81% to 2.05% for next-year returns, largely comparable to the average in-sample OLS R^2.

To illustrate the source of the out-of-sample performance gains, I examine the dividend-price ratio predictor (DP) in more detail. In addition to being the most widely-studied predictor, DP is illustrative because it has substantially negative OOS R^2 using OLS but positive OOS R^2 using WLS-EV for next-year returns. Figure 2 shows the evolution of $\hat{b}_\tau$ over the post-training period for both OLS and WLS-EV estimates. For next-year returns, although the full-sample estimates are very similar for OLS and WLS-EV (both around 2.3, as presented in Panel B of Table 3), the rolling WLS-EV estimates are much closer to the full-sample estimate early on in the sample, and much more stable over time. The reason for this improvement is the WLS-EV estimators react more efficiently to extreme return observations that occur in periods of high ex-ante variance. For example, the OLS estimates "read too much" into the high returns in the mid 1930s that follow high DP but also high

$\hat{\sigma}_t^2$, making WLS-EV downweight them. The fact that WLS-EV rolling $\hat{b}_\tau$ are closer to full-sample $\hat{b}_T$ than their OLS counterparts results in WLS-EV OOS R^2 closer to the IS R^2.

Despite the improved OOS performance of WLS-EV relative to OLS, many predictors that are significant in-sample still have negative OOS R^2 when using WLS-EV. Campbell and Thompson (2008) provides a method for improving the OOS performance of these predictors. Specifically, Campbell and Thompson (2008) suggests two economically-motivated restrictions on the $\hat{b}_\tau$ and $\mathbb{E}_\tau\left(r_{\tau+1,\tau+h}|x\right)$:

1. For each predictor, economic intuition suggests the correct sign of b. For example, dp should have a positive b because higher discount rates lead to larger dp. If $\hat{b}_\tau$ has the economically incorrect sign, set $\hat{b}_\tau = 0$ and $\mathbb{E}_\tau\left(r_{\tau+1,\tau+h}|x\right) = \mathbb{E}_\tau\left(r_{\tau+1,\tau+h}\right)$.

2. The expected equity risk premia $\mathbb{E}_\tau\left(r_{\tau+1,\tau+h}|x\right)$ should always be positive. If it is not, use $\mathbb{E}_\tau\left(r_{\tau+1,\tau+h}|x\right) = 0$.

I apply these restrictions for each of the return predictors using both OLS and WLS-EV, and compute resulting adjusted OOS R^2 as defined in Equation (34).

Table 4 shows the OOS R^2 using the Campbell and Thompson (2008) approach (CT OOS R^2) for each predictor. For both OLS and WLS-EV, CT OOS R^2 are substantially higher than OOS R^2, with 8 (8) predictors offering positive CT OOS R^2 for next-month (next-year) returns using OLS. However, even with the Campbell and Thompson (2008) restrictions, WLS-EV outperforms OLS out-of-sample. WLS-EV CT OOS R^2 are higher than their OLS counterparts for 9 (12) of the next-month (next-year) predictors. Similarly, mean CT OOS R^2 is 0.11% for next-month and 0.34% for next-year returns using WLS-EV, compared to 0.01% and -1.30% for OLS.

Another method for improving out-of-sample performance of these predictors is using the economic restrictions in Pettenuzzo, Timmermann, and Valkanov (2014). Specifically, if we assume conditional Sharpe Ratios for the market are bounded between 0 and 1, we can estimate coefficients $\hat{a}_\tau$ and $\hat{b}_\tau$ using the regression in Equation (30) constrained so that:

$$0 \leq \frac{\hat{a}_\tau + \hat{b}_\tau x_t}{\hat{\sigma}_t} \leq 1 \quad \forall t \leq \tau - h. \tag{36}$$

Note that this is different from the Campbell and Thompson (2008) approach because it requires that the conditional risk premia is positive for all $t \leq \tau - h$, rather than just τ, and also adds an upper bound on the conditional Sharpe Ratio.

Table 4 shows that, as demonstrated in Pettenuzzo, Timmermann, and Valkanov (2014), their economic restrictions result in out-of-sample R^2 (PTV OOS R^2) substantially larger than even the CT OOS R^2. Because the Pettenuzzo, Timmermann, and Valkanov (2014) approach already dampens the influence of extreme observations on OLS point estimates, the additional improvement afforded by WLS-EV is smaller for PTV OOS R^2 than for OOS R^2 or CT OOS R^2. Nevertheless, using WLS-EV instead of OLS results in slightly higher average PTV OOS R^2 and more predictors with positive PTV OOS R^2.

As an alternative measure of out-of-sample predictive performance, I also compute certainty equivalents (CEs) for a hypothetical investor optimizing their portfolio using estimated conditional means and variances, an approach used in Campbell and Thompson (2008) and Johannes, Korteweg, and Polson (2014). Compared to OOS R^2, certainty equivalents have the advantage of a natural economic interpretation but the disadvantage of begin dependent of the function form of the investor's utility functions. Online Appendix A, available at bit.ly/wlsapp, shows that CEs follow largely the same pattern as OOS R^2, with WLS-EV offering a 20bp-50bp increase in per-year CE over OLS.

5. Variance risk premium as a predictor

5.1. Methodology

As a second application of WLS-EV, I revisit the empirical relation between future returns and the variance risk premium proxies in Bollerslev, Tauchen, and Zhou (2009) and Drechsler and Yaron (2011), BTZ and DY hereafter, and show it is not robust to the WLS-EV approach. BTZ and DY both show that the difference between VIX^2 and an estimate of statistical-measure variance positively predicts equity returns. Both papers motivate this result by modeling variance and equity risk premia in a setting with stochastic volatility and volatility-of-volatility, resulting in a positive correlation between equity and variance risk premia.

BTZ and DY use slightly different proxies for the variance risk premium, both of which I replicate. The BTZ proxy is:

$$\text{BTZ } \hat{\text{VRP}}_t \equiv \text{VIX}_t^2 - \text{IndRV}_{t-20,t}, \tag{37}$$

where VIX_t is the CBOE VIX index on day t and $\text{IndRV}_{t-20,t}$ is the realized variance of S&P 500 index returns over the 21 trading days ending on day t. I follow BTZ and compute IndRV from realized five-minute log S&P 500 index returns, and scale both VIX_t^2 and $\text{IndRV}_{t-20,t}$ to monthly percents squared. The DY proxy for the variance risk premium is:

$$\text{DY } \hat{\text{VRP}}_t \equiv \text{VIX}_t^2 - \hat{\mathbb{E}}_t(\text{FutRV}_{t+1,t+21}), \tag{38}$$

where $\text{FutRV}_{t+1,t+21}$ is the sum of squared five-minute log S&P 500 futures returns on the 21 trading days following t. I follow DY and use the fitted value from a full-sample time-series regression of $\text{FutRV}_{t+1,t+21}$ on $\text{IndRV}_{t-20,t}$ and VIX_t^2 as $\hat{\mathbb{E}}_t\left(\text{FutRV}_{t+1}\right)$.

I use the two $\hat{\text{VRP}}_t$ to predict $r_{t+1,t+h}$, the log excess return of the CRSP value-weighted index over the h days following the measurement of $\hat{\text{VRP}}_t$. Because the results in BTZ and DY indicate that these proxies' predictive power is strongest at one-month and one-quarter horizons, I consider $h = 21$ and $h = 63$. Also following BTZ and DY, I scale log returns to annualized percentages. I adjust the point estimates for the Stambaugh (1999) bias using the simulation procedure described in Section 4, and account for the overlap using the modified Hodrick (1992) approach I describe in Section 2.2. I also compute simulated standard errors and p-values using the heteroskedastic simulations (Sim) described in Section 3.

For both the observed and simulated samples, I compute WLS-EV using the VIXF $\hat{\sigma}_t^2$ and VIX $\hat{\sigma}_t^2$ defined in Section 2. Since the $\hat{\text{VRP}}_t$ themselves use VIX, they are only available from 1990-2013, making them a perfect candidate for the VIX-based $\hat{\sigma}_t^2$ that Section 2 shows are strong predictors of realized variance.

5.2. Results

Table 5 presents the results for all twelve combinations of variance risk premia proxy, forecast horizon, and estimator. In all cases, the OLS coefficients are much larger than the corresponding asymptotic standard errors, resulting in asymptotic p-values of 6.3%, 4.6%, 2.8%, and 0.6%. This indicates that the OLS estimates documented in BTZ and DY are slightly less significant in the post-publication period (2008-2013) and when using an overlapping daily sample. As suggested by the simulations in Table 2, using the Hodrick (1992) approach to estimating overlapping regressions results in daily OLS asymtotic standard errors that match the simulated standard errors, meaning the OLS evidence for return predictability is robust to heteroskedastic simulations of the standard errors.

The WLS-EV results, which account for heteroskedasticity in point estimates as well as standard errors, are much more more pessimistic than the OLS. In all cases, the WLS-VIXF and WLS-VIX point estimates are smaller than the OLS ones, and statistically insignificant using simulated or asymptotic p-values.

To help understand why WLS-EV point estimates are so much smaller than OLS estimates of the predictability afforded by proxies for the variance risk premium, Figure 3 plots the observed $\hat{\text{VRP}}_t$ and next-month returns r_{t+1}, along with the OLS and WLS-VIX regression lines, for DY in the top panel and BTZ in the bottom panel. The darkness of each point represents its weight in the WLS-VIX regressions, where the observation with the highest weight is black and other points are on the grayscale based on what fraction of the maximum weight the corresponding observation receives.

For both the DY and BTZ proxies, Figure 3 indicates that $\hat{\text{VRP}}_t$ is near zero for most observations but extremely positive or negative for a small subset. These extreme observations also have high VIX $\hat{\sigma}_t^2$, resulting in a low weight in the WLS-VIX regressions, represented by their light gray color. The observations with extremely negative VRP tend to have negative future return realizations, and those with extremely positive VRP tend to have positive future return realizations, and so when these points receive full weight in OLS the coefficient is strongly positive. However, WLS-EV downweights these points because the return realizations are so volatile, and instead fits mostly on the darker points towards the middle of the distribution, which do not significantly support return predictability.[11]

5.3. Robustness

To ensure the failure of $\hat{\text{VRP}}$ to significantly predict returns in Table 5 is not driven by the extended sample or overlapping daily returns, I repeat my analysis on a monthly sample from 1990-2007 designed to mimick the original sample used in BTZ and DY. For the BTZ analysis, I use the proxy as provided on Hao Zhou's website to assure that my results are not driven by an error in my calculation of $\hat{\text{VRP}}$.[12] To match BTZ and DY, I use S&P 500 returns rather than the CRSP index returns I use elsewhere, and consider both quarterly and monthly forecast horizons.

The results of my replication analysis are in Table 6, along with the point estimates and standard errors from the original DY and BTZ papers for comparison. In both cases, my replication is quite close the original papers in terms of t-stats and p-values. For BTZ, with the benefit of the authors' data, the point estimates are also nearly identical. I extend this replication by estimating heteroskedastic simulated (Sim) standard errors, as well as WLS-EV using VIX $\hat{\sigma}_t^2$, on the original BTZ and DY samples. Both approaches substantially weaken the evidence of return predictability. Sim standard errors and p-values are much higher for OLS than their asymptotic counterparts, making the one-month predictability evidence insignificant and the one-quarter evidence marginal.

[11] Online Appendix B shows that other approaches to mitigating the influence of these observations, for example by using deciles of $\hat{\text{VRP}}_t$ or winsorizing $\hat{\text{VRP}}_t$ below at zero, result in even weaker evidence of return predictability in both OLS and WLS-EV regressions.

[12] The proxy I use in Table 5 is identical to the downloadable version but measured daily using a rolling 21 day window rather than calendar months.

More importantly, even in the original papers' sample, the WLS-EV estimates are only about half as large as the OLS estimates, making them statistically insignificant despite lower simulated standard errors. This failure holds across both VRP proxies and both prediction horizons, in all cases with simulated p-values above 22.6%. Table 6 also shows results for simple and log S&P 500 returns, 1990-2007 daily samples, extended monthly samples, and combinations thereof. In only one of the ten alternative procedures (BTZ VRP, 1990-2013, monthly sampling interval, log S&P 500 returns) are WLS-EV estimates statistically significant.

As a final robustness check, in Online Appendix C I revisit the analysis of VRP's predictability around the world in Bollerslev et al. (2014) using WLS-EV. While I replicate the OLS evidence of predictability in many countries for certain forecast horizons, this evidence disappears when using WLS-EV or small-sample standard errors based on heteroskedastic simulations. With either methodology, none of the 56 country-forecast horizon pairs in Bollerslev et al. (2014) yield statistically significant evidence of return predictability.

Combined, the results in Tables 5 and 6, as well as Online Appendices B and C, indicate that in the relatively short post-1990 sample, there is no evidence for a relation between conditional variance and equity risk premia when using more-efficient WLS-EV regressions.

6. Politics, the weather, and the stars as predictors

Novy-Marx (2014) discusses nine variables that predict future factor returns in OLS regressions: the political party of the President of the United States, the monthly highest temperature in New York City, the global temperature, the rolling average global temperature, the quadiperiodic Pacific temperature anomaly (El Niño), the rolling average Pacific Ocean temperature, the angle between Mars and Saturn, the angle between Jupiter and Saturn, and the observed number of sunspots. Novy-Marx (2014) shows these nine variables often predict returns for 22 factors or anomalies, forcing readers to either accept implausible predictive relations or question the standard OLS hypothesis testing methodology.

As a final application of the WLS-EV methodology, I revisit the surprising predictability evidence in Novy-Marx (2014) and show it is insignificant when using WLS-EV instead of OLS. For each of the nine predictors Novy-Marx (2014) considers, I estimate monthly predictive regressions using data from 1961 through 2012. I find that small-sample p-values based on the simulation approach in Section 3 indicate OLS estimates are significant for the same three variables Novy-Marx (2014) finds predict market returns: the political party of the president, NYC Weather, and the Mars/Saturn Angle. However, WLS-EV estimates of the predictability afforded by these three variables are all closer to zero and all have higher p-values than their OLS counterparts, with only Mars/Saturn remaining statistically significant. Moreover, WLS-EV estimates of the other six variables remain insignificant despite smaller simulated standard errors.

I also estimate the joint significance of these predictors using the same simulations. For each simulated sample of returns, I estimate univariate predictability regressions for each predictors. I then compute the following χ^2 statistic for each simulated sample:

$$\chi^2 \text{ statistic} \equiv \hat{\mathbf{b}}'\hat{\Sigma}^{-1}\hat{\mathbf{b}}, \tag{39}$$

where $\hat{\mathbf{b}}$ is a vector of the nine predictability coefficients for the given sample, and $\hat{\Sigma}$ is the covariance matrix of $\hat{\mathbf{b}}$ estimated across simulated samples. Table 7 presents the χ^2 statistics for OLS and WLS-EV based on the observed sample, as well as their simulated p-values based on the distribution of χ^2 statistics in the simulated samples. For example, the OLS p-value is 8.52%, meaning that the χ^2 statistic was higher than the observed sample's 15.22 in 8.52% of simulated samples.

Table 7 shows that while these nine predictors have weak joint significance in OLS estimates, they are not jointly significant in WLS-EV estimates. The insignificant WLS-EV evidence of predictability is consistent with the hypothesis that the Novy-Marx (2014) results are driven by a tendency of the standard OLS methodology, when combined with sufficient data mining, to over-reject the null hypothesis of no predictability. Because WLS-EV was not used to select the predictors or factors in Novy-Marx (2014) and is not as sensitive to highly volatile observations, it naturally produces weaker predictability evidence than OLS.

7. Conclusion

I study a WLS approach to estimating return predictability regressions where the weights are ex-ante estimates of return variance. Relying on insights from the volatility literature, I implement this approach using ex-ante estimates for the variance of returns based on first-stage predictive regressions. The WLS-EV approach is convenient, substantially more efficient than OLS, and does not suffer from the bias introduced by ex-post weighting schemes.

The more-efficient WLS-EV estimates in return predictability regressions have many benefits to researchers and investors. Because their standard errors are smaller, they have fewer false negatives. Because their estimates are partially independent of OLS estimates, they provide an additional test of the no-predictability null that may fail to reject for truly insignificant predictors. They also result in better out-of-sample return predictions, giving researchers more power and investors a better indication of average future returns.

Empirically, using WLS-EV strengthens the in- and out-of-sample evidence of return predictability for traditional predictors such as the dividend-to-price ratio. On the other hand, using WLS-EV results in no significant evidence the variance risk premium, politics, the weather, or the stars predict future market returns.

Appendix A. Consistency of WLS estimator

I show in this appendix that a weighted least squares estimator using arbitrary weights w_t for each observation will converge asymptotically to $\hat{\beta}_w^\infty$, which may be different than the true β generating the data. I also provide conditions under which $\hat{\beta}_w^\infty = \beta$, and show suggestive evidence that WLS-EV weights meet these conditions while RLS weights do not.

Assume the data are generated by:

$$r_{t+1} = X_t \cdot \beta + \epsilon_{t+1}, \tag{40}$$

where $\mathbb{E}(X_t' \epsilon_{t+1}) = 0$. This implies the population (asymptotic) OLS estimate satisfies:

$$\begin{aligned}
\hat{\beta}_{\text{OLS}}^\infty &\equiv \mathbb{E}(X_t' X_t)^{-1} \mathbb{E}(X_t' r_{t+1}) && (41)\\
&= \mathbb{E}(X_t' X_t)^{-1} \mathbb{E}(X_t'(X_t \cdot \beta + \epsilon_{t+1})) && (42)\\
&= \beta + \mathbb{E}(X_t' X_t)^{-1} \mathbb{E}(X_t' \epsilon_{t+1}) = \beta. && (43)
\end{aligned}$$

Since $\hat{\beta}_{\text{OLS}}^\infty = \beta$, the OLS estimator is consistent.

Consider WLS estimates of β using arbitrary weights w_t that may be a function of both past and future data. The population WLS estimator is:

$$\begin{aligned}
\hat{\beta}_{\text{WLS}}^\infty &\equiv \mathbb{E}(w_t^2 X_t' X_t)^{-1} \mathbb{E}(w_t^2 X_t' r_{t+1}) && (44)\\
&= \mathbb{E}(w_t^2 X_t' X_t)^{-1} \mathbb{E}(w_t^2 X_t'(X_t \cdot \beta + \epsilon_{t+1})) && (45)\\
&= \beta + \mathbb{E}(w_t^2 X_t' X_t)^{-1} \mathbb{E}(w_t^2 X_t' \epsilon_{t+1}), && (46)
\end{aligned}$$

making WLS a consistent estimator if and only if:

$$\mathbb{E}(w_t^2 X_t' \epsilon_{t+1}) = 0. \tag{47}$$

One condition under which the WLS consistency condition (47) holds is if:

$$\begin{aligned}
\mathbb{E}\left(r_{t+1} | X_t, w_t\right) &= X_t \cdot \beta && (48)\\
\Rightarrow \mathbb{E}\left(\epsilon_{t+1} \middle| X_t, w_t^2\right) &= 0 && (49)\\
\Rightarrow \mathbb{E}(w_t^2 X_t' \epsilon_{t+1}) &= 0. && (50)
\end{aligned}$$

Economically, equation (48) says that expected returns conditional on X_t and w_t are in fact a linear function of X_t. This is plausible for weights w_t based on ex-ante return variance since, empirically, ex-ante return variance is unrelated to future returns. However, it is clearly false for w_t using time $t+1$ information about realized volatility, which are correlated with future returns and therefore should be included on the right-hand side of (48).

Empirically, I test (48) using a regression of next-month returns on a constant, a test x_t, and w_t^2 for either RLS or WLS-EV. For x_t, I use the log dividend-to-price ratio dp and the Drechsler and Yaron (2011) $\hat{\text{VRP}}$. The results, tabulated below, indicate that because the RLS w_t^2 use information from $t+1$, and returns are negatively correlated with volatility (and thus positively related to RLS weights), they are statistically significant predictors of r_{t+1}. By contrast, the WLS-EV w_t^2 are unrelated to future returns, having near-zero predictive coefficients and zero or negative adjusted R^2. Together, these results indicate that RLS estimates are likely to be biased while WLS-EV estimates are likely to be unbiased.

Dep. Variable:	$\hat{\epsilon}_{t+1}$ using $x_t = \text{dp}_t$		$\hat{\epsilon}_{t+1}$ using $x_t = \text{VRP}_t$	
	(2)	(3)	(2)	(3)
Constant	-5.789***	-0.190*	-40.623***	-3.880
	(1.760)	(0.568)	(10.207)	(3.640)
RLS w_t^2	0.217***	-	2.453***	-
	(0.061)	-	(0.542)	-
WLS-EV w_t^2	-	0.014	-	0.363
	-	(0.033)	-	(0.242)
Adj. R^2	8.30%	-0.10%	1.40%	0.00%

References

Bollerslev, Tim, James Marrone, Lai Xu, and Hao Zhou, 2014, Stock return predictability and variance risk premia: statistical inference and international evidence, *Journal of Financial and Quantitative Analysis* 49, 633–661.

Bollerslev, Tim, George Tauchen, and Hau Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *The Journal of Finance* 62, 877–915.

Cameron, A Colin and Pravin K Trivedi, *Microeconometrics: Methods and Applications* (Cambridge University Press 2005).

Campbell, John Y and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Christensen, Bent J and Nagpurnanand R Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125–150.

Cochrane, John H, 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

Drechsler, Itamar and Amir Yaron, 2011, What's vol got to do with it, *Review of Financial Studies* 24, 1.

Engle, Robert F, David M Lilien, and Russell P Robins, 1987, Estimating time varying risk premia in the term structure: The arch-m model, *Econometrica* 391–407.

French, Kenneth R, G William Schwert, and Robert F Stambaugh, 1987, Expected stock returns and volatility, *Journal of financial Economics* 19, 3–29.

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2005, There is a risk-return trade-off after all, *Journal of Financial Economics* 76, 509–548.

Glosten, Lawrence R, Ravi Jagannathan, and David E Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance* 48, 1779–1801.

Goyal, Amit and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

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

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential learning, predictability, and optimal portfolio returns, *The Journal of Finance* 3, 161–174.

Lettau, Martin and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *The Journal of Finance* 56, 815–849.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

Merton, Robert C, 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Merville, Larry J and Dan R Pieptea, 1989, Stock-price volatility, mean-reverting diffusion, and noise, *Journal of Financial Economics* 24, 193–214.

Newey, Whitney K and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–08.

Newey, Whitney K and Kenneth D West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Novy-Marx, Robert, 2014, Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars, *Journal of Financial Economics* 112, 137–146.

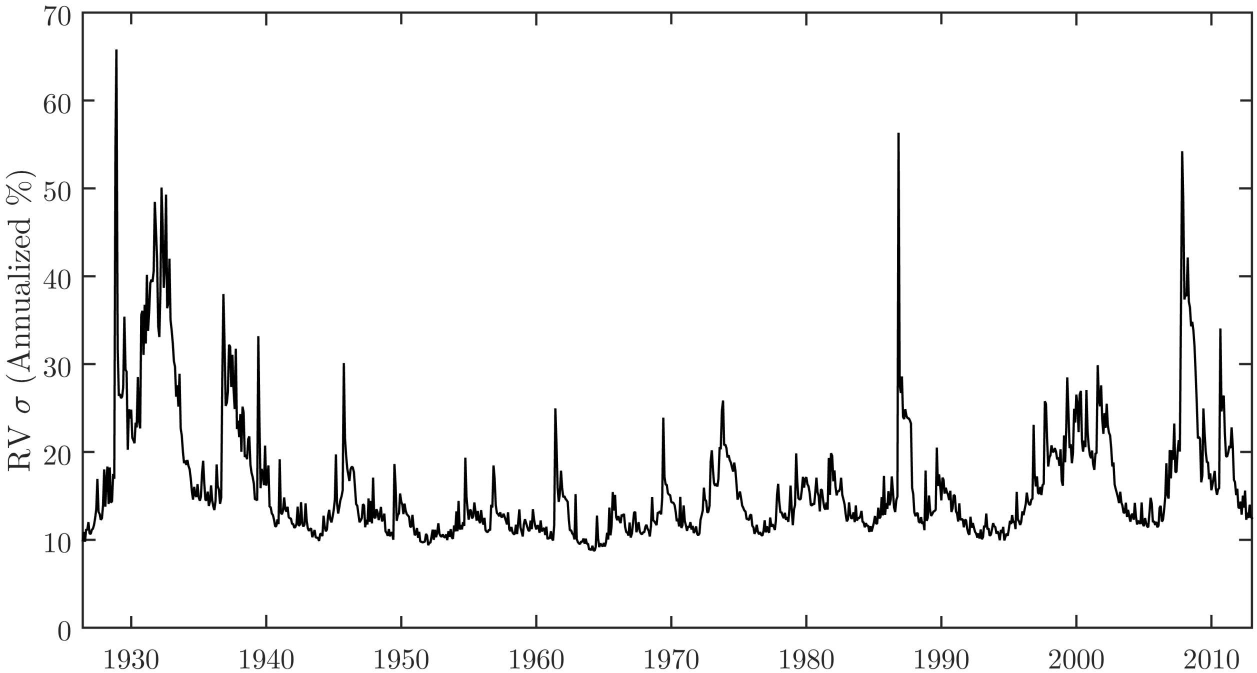
Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov, 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* 114, 517–553.

Singleton, Kenneth J, *Empirical dynamic asset pricing: model specification and econometric assessment* (Princeton University Press 2006).

Stambaugh, Robert F, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Figure 1: Conditional Volatility Measures

The first panel of this figure presents RV $\hat{\sigma}_t$, the volatility of next-month equity market returns conditional on past realized variance, estimated using regressions described in Section 2. The second panel presents VIXF $\hat{\sigma}_t$, the volatility of next-day equity market returns conditional on the VIX and past intraday futures variance. Both volatilities are displayed as an annualized percentage. The monthly sample consists of 1040 observations from 1927 through 2013, the daily sample 6049 observations from 1990 through 2013.



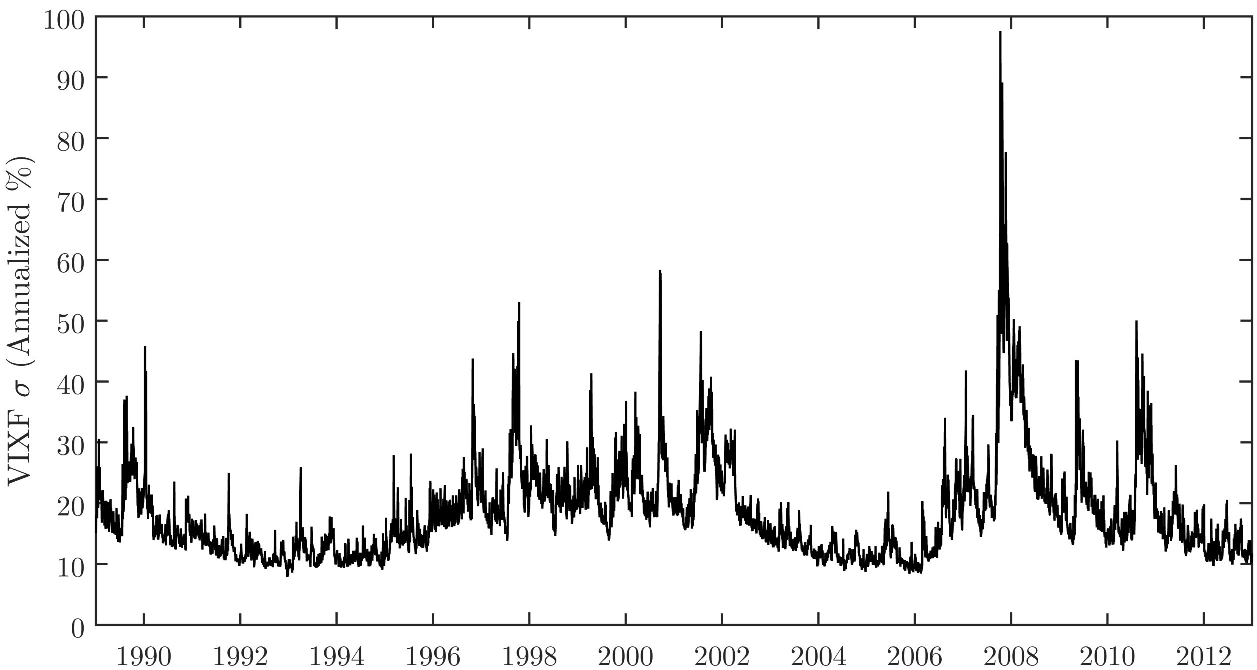


Figure 2: Rolling Estimates of Predictability Coefficient for DP

This figure presents coefficients from rolling regressions of next-year returns on the dividend-to-price ratio, DP. Specifically, for each date τ in my sample following a 20-year training period, using only data available at time τ, i.e. $t \leq \tau - 12$, I estimate the regression:

$$r_{t+1,t+12} = a + b \cdot DP_t + \epsilon_{t+1,t+h}, \tag{51}$$

where $r_{t+1,t+12}$ is the log cumulative dividend-inclusive excess return of the CRSP value-weighted index over the 12 months starting in $t + 1$ and DP_t is the dividend-to-price ratio at time t. For each τ, I compute point estimates $\hat{a}_\tau$ and $\hat{b}_\tau$ using OLS and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2. I plot the resulting OLS and WLS-EV estimates $\hat{b}_\tau$ for each month in my post-training sample period, 1947-2013.

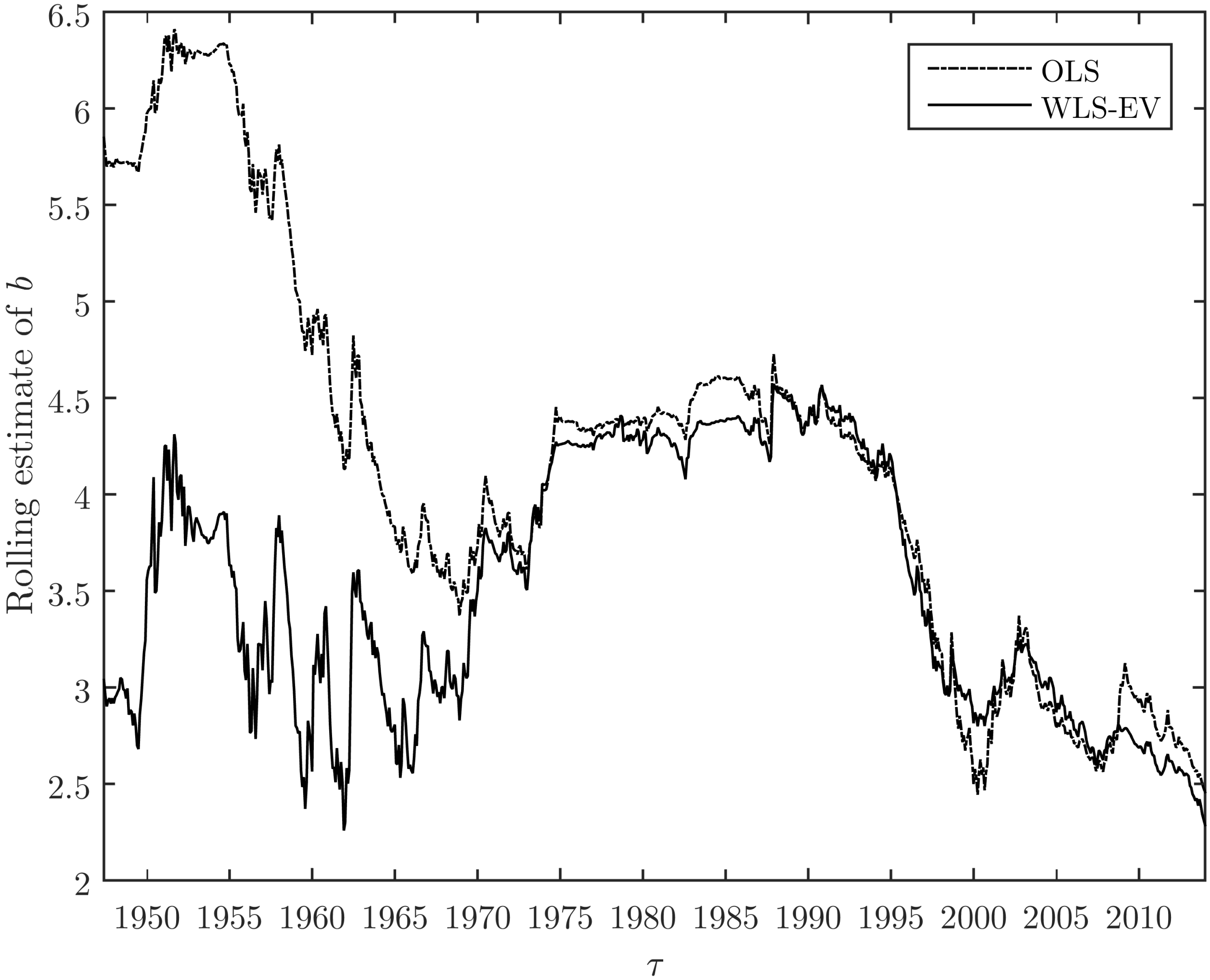


Figure 3: Predicting Returns using the Variance Risk Premia

This figure presents estimates of return predictability regressions of the form:

$$r_{t+1,t+21} = a + b \cdot \hat{\text{VRP}}_t + \epsilon_{t+h},$$

where $r_{t+1,t+21}$ is the log dividend-inclusive excess return of the CRSP value-weighted index over the 21 days starting with $t+1$, annualized and in percent. $\hat{\text{VRP}}_t$ is one of two proxies for the variance risk premium, both expressed as monthly percents squared. The first, DY $\hat{\text{VRP}}_t$, is from Drechsler and Yaron (2011). The second, BTZ $\hat{\text{VRP}}_t$, is from Bollerslev, Tauchen, and Zhou (2009). For each $\hat{\text{VRP}}_t$, I compute point estimates of a and b using OLS and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2, using VIXF $\hat{\sigma}_t^2$. The lines represent the predicted values from the two regressions. The points represent the 6028 daily observations from 1990-2013, with only every fifth observation plotted to improve readability. The darkness of each point represents its weight in the WLS-EV regressions, where the observation with the highest weight is black and other points are on the grayscale based on what fraction of the maximum weight the corresponding observation receives.

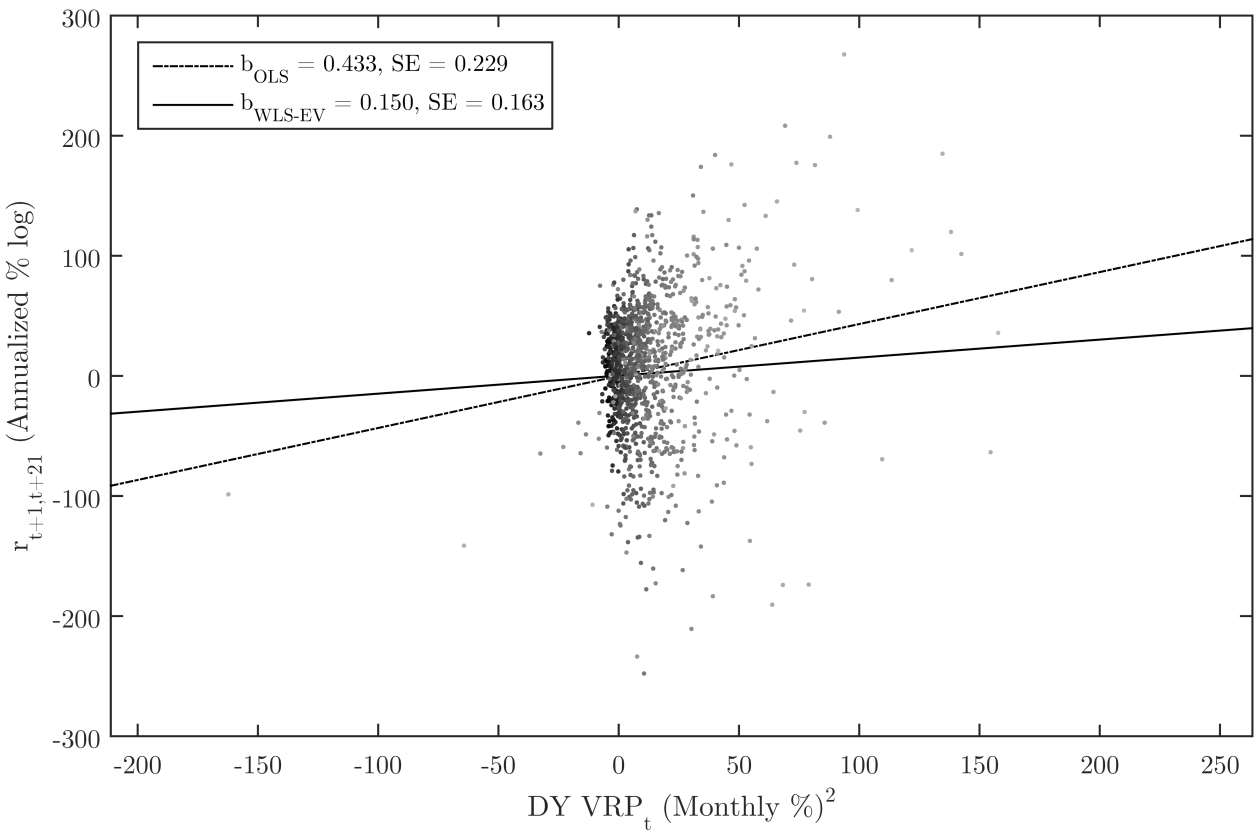


Figure 3: Predicting Returns using the Variance Risk Premia (cont'd)

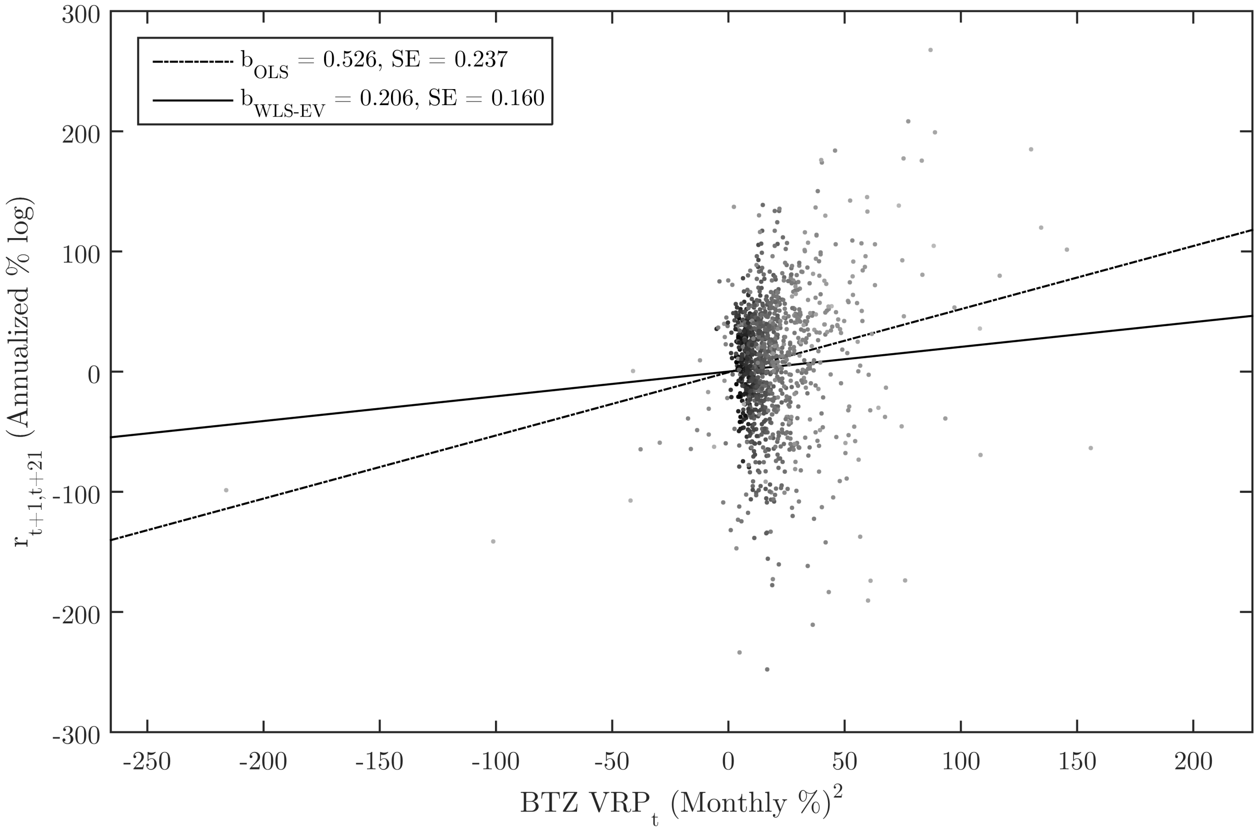


Table 1: Effectiveness of Ex-ante Variance Proxies

This table presents regressions of realized return variance on potential ex-ante variance predictors. For each month t in Panel A, the left-hand side is RV_{t+1}, the annualized realized variance in month $t+1$, where $\mathrm{RV}_m = 12 \cdot \sum_{d \in m} r_d^2$ and r_d is the log dividend-inclusive excess return of the CRSP value-weighted index on day d. Predictors in Panel A are $\mathrm{RV}_{t-a,t} = \frac{1}{a} \sum_{s=0}^{a} \mathrm{RV}_{t-s}$. For each day t in Panel B, the left-hand side is FutRV_{t+1}, the annualized realized variance on day $t+1$, where $\mathrm{FutRV}_d = 252 \cdot \sum_{i \in d} r_{i,\mathrm{fut.}}^2$ and $r_{i,\mathrm{fut.}}$ is the log return of the front-maturity S&P 500 futures contract in five-minute interval i. Predictors in Panel B are $\mathrm{FutRV}_{t-a,t} = \frac{1}{a} \sum_{s=0}^{a} \mathrm{FutRV}_{t-s}$. VIX_t^2 is the square of the VIX index on day t. The sample is 1039 monthly observations from 1927-2013 in Panel A and 6048 daily observations from 1990-2013 in Panel B. Standard errors are in parenthesis and are computed using Newey-West with Newey and West (1994) lag selection. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

Panel A. Predicting next-month variance $\mathbf{RV}_{t+1}$

	(1)	(2)	(3)	(4)	(5)	(6)
Const	0.015***	0.011***	0.009***	0.007**	0.006**	0.006**
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
RV_t	0.527***	-	-	-	0.373***	0.328***
	(0.129)	-	-	-	(0.114)	(0.118)
$\mathrm{RV}_{t-2,t}$	-	0.704***	-	-	-	0.101
	-	(0.102)	-	-	-	(0.090)
$\mathrm{RV}_{t-5,t}$	-	-	0.766***	-	-	0.035
	-	-	(0.115)	-	-	(0.131)
$\mathrm{RV}_{t-11,t}$	-	-	-	0.845***	0.497***	0.407**
	-	-	-	(0.140)	(0.125)	(0.186)
Adj. R^2	29.1%	28.8%	26.3%	24.9%	35.2%	35.3%
					↓ RV $\hat{\sigma}_t^2$	

Panel B. Predicting next-day variance $\mathbf{FutRV}_{t+1}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Const	0.013***	0.005	-0.016***	-	-0.012***	-	-0.012***	-
	(0.002)	(0.003)	(0.004)	-	(0.004)	-	(0.004)	-
FutRV_t	0.600***	-	-	-	0.210***	0.251***	0.211***	0.244***
	(0.052)	-	-	-	(0.050)	(0.046)	(0.050)	(0.045)
$\mathrm{FutRV}_{t-20,t}$	-	0.883***	-	-	-	-	-0.021	0.100
	-	(0.126)	-	-	-	-	(0.096)	(0.102)
VIX_t^2	-	-	1.049***	0.883***	0.819***	0.657***	0.839***	0.580***
	-	-	(0.108)	(0.073)	(0.121)	(0.076)	(0.147)	(0.088)
Adj. R^2	36.0%	38.8%	47.9%	45.5%	50.1%	48.7%	50.0%	48.8%
				↓ VIX $\hat{\sigma}_t^2$		↓ VIXF $\hat{\sigma}_t^2$		

Table 2: Efficiency and Bias of OLS, RLS, and WLS-EV in Simulations

This table presents results of simulations designed to quantify the efficiency and potential bias in OLS, WLS-EV, and RLS regressions. I generate returns according to the null:

$$r_{t+1}^{\text{sim}} = \mu_r + b \cdot x_t^{\text{data}} + \hat{\sigma}_t^{\text{data}} \psi_{t+1}^{\text{re-sampled}},$$

where r_t^{sim} is the return of the CRSP value-weighted index; μ_r is the in-sample mean of r_t; x_t^{data} is a return predictor from the data; $\hat{\sigma}_t^{\text{data}}$ is the conditional volatility of returns from the data; and $\psi_{t+1}^{\text{re-sampled}}$ is randomly redrawn in each simulation. In the first panel, I simulate under the no-predictability null that $b = 0$. In the second, I set $b = 1$ or $b = 0.3$. For dividend yield simulations, I express returns in percentages and use 1039 monthly observations from 1927-2013 of $x_t = d_t - p_t$, the log dividend yield of the market portfolio, along with RV $\hat{\sigma}_t^2$. For the variance risk premia simulations, I express returns in annualized percentages and use 6028 daily observations from 1990-2013 of the Drechsler and Yaron (2011) variance risk premium x_t, detailed in Section 5, along with VIXF $\hat{\sigma}_t^2$. Using each simulated return series, I compute point estimates and Hodrick (1992) standard errors of b using OLS, robust least squares (RLS), and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2. I report summary statistics for the distribution across simulations of $\hat{b}$, and the accompanying asymptotic standard errors and t-stats. In Panel B, I also compute the fraction of simulations for which the estimated coefficient is above a 5% critical value from small-sample simulations under the $b = 0$ null.

Panel A: No predictability null ($b = 0$)

	Dividend yield sim.			Variance risk premium sim.		
	OLS	RLS	WLS-EV	OLS	RLS	WLS-EV
Mean $\hat{b}$	0.000	0.009	0.000	0.000	0.329	0.000
Standard dev $\hat{b}$	0.440	0.362	0.333	0.225	0.190	0.162
1st percentile $\hat{b}$	-1.027	-0.829	-0.771	-0.528	-0.116	-0.378
10th percentile $\hat{b}$	-0.564	-0.455	-0.426	-0.289	0.085	-0.207
25th percentile $\hat{b}$	-0.295	-0.235	-0.224	-0.151	0.201	-0.109
Median $\hat{b}$	0.000	0.008	0.000	0.000	0.330	0.000
75th percentile $\hat{b}$	0.297	0.252	0.224	0.152	0.457	0.108
90th percentile $\hat{b}$	0.565	0.473	0.426	0.287	0.572	0.207
99th percentile $\hat{b}$	1.022	0.850	0.775	0.520	0.772	0.377
Mean Asy SE $\hat{b}$	0.436	0.270	0.332	0.223	0.113	0.162
Prob(Asy t-stat > 1.96)	2.69%	7.45%	2.54%	2.79%	71.37%	2.57%
Prob(Asy t-stat < -1.96)	2.42%	6.93%	2.58%	2.42%	0.20%	2.46%

Table 2: Efficiency and Bias of OLS, WLS-EV, and RLS in Simulations (cont'd)

Panel B: Predictability null ($b > 0$)

	Dividend yield sim.			Variance risk premium sim.		
	OLS	RLS	WLS-EV	OLS	RLS	WLS-EV
Mean $\hat{b}$	1.000	1.007	1.000	0.400	0.729	0.400
Standard dev $\hat{b}$	0.440	0.362	0.333	0.225	0.190	0.162
1st percentile $\hat{b}$	-0.030	0.167	0.225	-0.129	0.284	0.023
10th percentile $\hat{b}$	0.437	0.543	0.574	0.111	0.486	0.192
25th percentile $\hat{b}$	0.704	0.763	0.775	0.249	0.602	0.291
Median $\hat{b}$	1.001	1.007	0.999	0.401	0.730	0.401
75th percentile $\hat{b}$	1.296	1.252	1.224	0.552	0.858	0.510
90th percentile $\hat{b}$	1.564	1.471	1.428	0.687	0.973	0.608
99th percentile $\hat{b}$	2.020	1.851	1.780	0.918	1.171	0.775
Mean Asy SE $\hat{b}$	0.436	0.270	0.332	0.223	0.113	0.162
Prob(Asy t-stat $<$ 1.96)	36.67%	9.39%	14.71%	55.97%	0.42%	30.34%
Prob(Sim p-value $>$ 5%)	37.24%	21.34%	14.93%	56.55%	44.30%	30.20%

Panel C: No predictability null ($b = 0$), Noisy $\hat{\sigma}_t^2$

	Dividend yield sim.			Variance risk premium sim.		
	OLS	RLS	WLS-EV	OLS	RLS	WLS-EV
Mean $\hat{b}$	0.000	0.009	0.000	0.000	0.329	0.000
Standard dev $\hat{b}$	0.440	0.362	0.378	0.225	0.190	0.184
1st percentile $\hat{b}$	-1.027	-0.829	-0.877	-0.528	-0.116	-0.431
10th percentile $\hat{b}$	-0.564	-0.455	-0.483	-0.289	0.085	-0.238
25th percentile $\hat{b}$	-0.295	-0.235	-0.257	-0.151	0.201	-0.125
Median $\hat{b}$	0.000	0.008	-0.004	0.000	0.330	0.000
75th percentile $\hat{b}$	0.297	0.252	0.253	0.152	0.457	0.124
90th percentile $\hat{b}$	0.565	0.473	0.486	0.287	0.572	0.235
99th percentile $\hat{b}$	1.022	0.850	0.885	0.520	0.772	0.427
Mean Asy SE $\hat{b}$	0.436	0.270	0.374	0.223	0.113	0.183
Prob(Asy t-stat $>$ 1.96)	2.69%	7.45%	2.58%	2.79%	71.37%	2.63%
Prob(Asy t-stat < -1.96)	2.42%	6.93%	2.69%	2.42%	0.20%	2.55%

Table 3: In-Sample Return Predictability

This table presents estimates of in-sample return predictability regressions of the form:

$$r_{t+1,t+h} = a + b \cdot x_t + \epsilon_{t+h},$$

where $r_{t+1,t+h}$ is the log cumulative dividend-inclusive excess return of the CRSP value-weighted index from months $t+1$ through $t+h$; and x_t is a candidate return predictor. The forecast horizons I use are $h = 1$ month in Panel A and $h = 12$ months in Panel B. The x_t I use are the log dividend-to-price ratio (dp), dividend-to-price ratio (DP), the log earnings-to-price ratio (ep), the log dividend-to-earnings ratio (de), the conditional variance of returns RV $\hat{\sigma}_t^2$, the treasury bill yield (tbl), the long-term treasury bond yield (lty), the return of long-term bonds (ltr), the term spread (tms), the default yield spread (dfy), inflation (infl), the log book-to-market ratio (bm), the cross-sectional beta premium (csp), net equity expansion (ntis), the log net payout yield (lpy), and the consumption wealth ratio (cay). To improve the readability of the coefficients, I divide dp, ep, de, bm, and lpy by 100. For each predictor, I compute point estimates and standard errors of b using OLS and weighted least squares with expected variance (WLS-EV), detailed in Section 2. I also compute $\hat{b}$ adjusted for the Stambaugh bias using a simulation procedure. I compute errors and p-values for the bias-adjusted coefficients using Hodrick (1992) (Asy) and the heteroskedastic simulation procedure (Sim) described in Section 3. The sample is 1039 monthly observations from 1927-2013. ***, **, and * indicate simulated p-values below 1%, 5% and 10% level, respectively.

Table 3: In-Sample Return Predictability (cont'd)

Panel A: Predicting Next-Month Returns

Predictor:	dp		DP		ep		de	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	0.145	0.159	0.115	0.105	0.547	0.504	-0.268	-0.148
Unadjusted $\hat{b}$	0.547	0.580	0.169	0.165	0.793	0.763	-0.227	-0.102
SE (Asy)	0.536	0.323	0.201	0.094	0.455	0.370	0.866	0.544
p-value (Asy)	78.62%	62.34%	56.63%	26.34%	22.92%	17.34%	75.70%	78.56%
SE (Sim)	0.445	0.336	0.135	0.090	0.524	0.371	0.818	0.558
p-value (Sim)	74.57%	63.87%	39.04%	23.59%	29.69%	17.54%	74.22%	78.31%
Predictor:	$\hat{\sigma}_t^2$		tbl		lty		ltr	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	-0.361	-0.732	-0.097*	-0.127***	-0.078	-0.105**	0.106	0.171***
Unadjusted $\hat{b}$	-0.339	-0.702	-0.092	-0.122	-0.067	-0.094	0.106	0.171
SE (Asy)	1.288	1.089	0.057	0.051	0.060	0.053	0.067	0.061
p-value (Asy)	77.92%	50.17%	9.17%	1.20%	18.95%	4.92%	11.34%	0.49%
SE (Sim)	1.411	1.051	0.056	0.047	0.057	0.049	0.085	0.061
p-value (Sim)	79.57%	48.32%	8.72%	0.78%	17.05%	3.12%	20.99%	0.52%
Predictor:	tms		dfy		infl		bm	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	0.206	0.199*	0.129	-0.023	-0.353	-0.996***	1.056	0.203
Unadjusted $\hat{b}$	0.204	0.197	0.169	0.019	-0.358	-1.003	1.441	0.610
SE (Asy)	0.133	0.110	0.647	0.361	0.462	0.315	1.210	0.624
p-value (Asy)	12.22%	6.98%	84.15%	95.03%	44.51%	0.16%	38.30%	74.48%
SE (Sim)	0.136	0.106	0.418	0.286	0.453	0.308	0.872	0.581
p-value (Sim)	13.05%	6.17%	75.21%	93.91%	42.81%	0.14%	22.25%	73.06%
Predictor:	csp		ntis		lpy		cay	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	2.106***	1.883***	-0.164	-0.135**	1.643	1.601**	0.190**	0.226***
Unadjusted $\hat{b}$	2.135	1.915	-0.163	-0.134	1.777	1.748	0.197	0.232
SE (Asy)	0.758	0.668	0.107	0.081	1.088	0.795	0.089	0.081
p-value (Asy)	0.54%	0.48%	12.81%	9.65%	13.11%	4.39%	3.26%	0.55%
SE (Sim)	0.787	0.644	0.106	0.066	1.008	0.759	0.094	0.082
p-value (Sim)	0.68%	0.35%	12.30%	4.50%	10.01%	3.39%	4.32%	0.55%

Summary Statistics

	# Statistically Significant (Sim.)				
	10% level	5% level	1% level	Mean$\left(\frac{\text{WLS Asy SE}}{\text{OLS Asy SE}}\right)$	Mean$\left(\frac{\text{WLS Sim SE}}{\text{OLS Sim SE}}\right)$
OLS	3	2	1	-	-
WLS	9	8	5	0.74	0.74

Table 3: In-Sample Return Predictability (cont'd)

Panel B: Predicting Next-Year Returns

Predictor:	dp		DP		ep		de	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	3.497	3.297	1.863	1.655	6.519	6.649	-0.338	-1.534
Unadjusted $\hat{b}$	7.967	7.972	2.454	2.284	9.362	9.593	0.168	-1.031
SE (Asy)	6.036	3.757	2.069	1.026	4.403	4.030	9.499	5.855
p-value (Asy)	56.24%	38.01%	36.78%	10.66%	13.87%	9.90%	97.16%	79.33%
SE (Sim)	5.321	3.956	1.554	1.011	5.351	4.138	8.571	6.112
p-value (Sim)	50.73%	40.03%	23.45%	10.22%	21.86%	10.97%	96.88%	80.51%
Predictor:	$\hat{\sigma}_t^2$		tbl		lty		ltr	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	0.200	-7.808	-0.912	-1.147**	-0.474	-0.764	0.691***	0.644***
Unadjusted $\hat{b}$	0.512	-7.556	-0.844	-1.075	-0.365	-0.659	0.690	0.643
SE (Asy)	11.048	7.866	0.678	0.603	0.694	0.622	0.210	0.187
p-value (Asy)	98.55%	32.09%	17.88%	5.71%	49.46%	21.89%	0.10%	0.06%
SE (Sim)	10.324	8.152	0.646	0.548	0.670	0.575	0.228	0.191
p-value (Sim)	98.51%	33.56%	15.61%	3.65%	48.32%	18.50%	0.26%	0.09%
Predictor:	tms		dfy		infl		bm	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	3.002*	2.503**	2.033	-0.102	-1.823	-5.037*	16.558	7.284
Unadjusted $\hat{b}$	3.015	2.522	2.414	0.296	-1.820	-5.027	21.078	11.954
SE (Asy)	1.499	1.218	6.350	3.564	5.370	3.224	11.843	7.059
p-value (Asy)	4.53%	4.00%	74.89%	97.72%	73.42%	11.82%	16.20%	30.21%
SE (Sim)	1.564	1.192	4.748	3.118	3.870	2.679	10.174	6.905
p-value (Sim)	5.53%	3.39%	66.84%	97.36%	63.97%	5.73%	10.27%	28.92%
Predictor:	csp		ntis		lpy		cay	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Stambaugh $\hat{b}_{\text{adj}}$	5.873	4.130	-2.565**	-1.901**	28.260**	23.954***	1.888*	2.111**
Unadjusted $\hat{b}$	5.975	4.235	-2.535	-1.874	29.625	25.326	1.958	2.180
SE (Asy)	8.238	7.655	1.171	0.975	11.289	8.595	1.079	0.959
p-value (Asy)	47.59%	58.95%	2.85%	5.12%	1.23%	0.53%	8.01%	2.77%
SE (Sim)	8.680	7.351	1.135	0.802	11.301	8.737	1.104	0.960
p-value (Sim)	49.61%	56.34%	2.14%	1.56%	1.33%	0.63%	8.90%	2.77%

Summary Statistics

	# Statistically Significant (Sim.)				
	10% level	5% level	1% level	$\text{Mean}\left(\frac{\text{WLS Asy SE}}{\text{OLS Asy SE}}\right)$	$\text{Mean}\left(\frac{\text{WLS Sim SE}}{\text{OLS Sim SE}}\right)$
OLS	5	3	1	-	-
WLS	7	6	2	0.75	0.76

Table 4: Out-of-Sample Return Predictability

This table presents statistics on the out-of-sample predictability afforded by 16 candidate predictors x_t. In Panel A, I predict next-month log dividend-inclusive excess returns of the CRSP value-weighted index. In Panel B, the forecast horizon is one year. The predictors are identical to those in Table 3 with two exceptions: since cay and RV $\hat{\sigma}_t^2$ require the full-sample of data to construct, I replace them with a rolling estimate of the cay variable, caya, and past realized variance $\mathrm{RV}_{t-251,t}$. For each predictor, I compute out-of-sample return forecasts starting 20 years after the sample begins, using both OLS and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2. Given these out-of-sample forecasts, I compute the out-of-sample R^2 (OOS R^2) using the procedure described in Section 4. I also compute the out-of-sample R^2 using the Campbell and Thompson (2008) approach (CT OOS R^2) and the Pettenuzzo, Timmermann, and Valkanov (2014) (PTV OOS R^2), described in Section 4. In both panels, I use 1039 monthly observations from 1927-2013.

Panel A: Predicting Next-Month Returns

Predictor:	dp		DP		ep		de	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-0.01%	-0.13%	0.14%	0.05%	-0.89%	-0.09%	-1.02%	-0.65%
CT OOS R^2	0.16%	0.14%	0.17%	0.12%	-0.10%	0.34%	0.00%	-0.19%
PTV OOS R^2	0.26%	0.11%	0.30%	0.16%	0.51%	0.41%	-0.21%	-0.11%
Predictor:	$\mathrm{RV}_{t-252,t}$		tbl		lty		ltr	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-0.10%	-0.60%	-0.04%	-0.76%	-0.85%	-1.90%	-0.57%	0.18%
CT OOS R^2	-0.02%	0.00%	0.23%	0.15%	0.22%	0.06%	0.20%	0.11%
PTV OOS R^2	-0.07%	-0.05%	0.60%	0.53%	0.48%	0.35%	0.12%	0.59%
Predictor:	tms		dfy		infl		bm	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	0.23%	0.55%	-0.19%	0.05%	0.15%	0.49%	-1.37%	-0.20%
CT OOS R^2	0.22%	0.46%	-0.17%	-0.03%	0.17%	0.66%	-0.80%	-0.20%
PTV OOS R^2	0.42%	0.56%	-0.17%	0.06%	0.10%	0.10%	0.03%	0.08%
Predictor:	csp		ntis		lpy		caya	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-0.49%	-0.01%	-0.68%	-0.34%	-0.66%	-0.43%	0.15%	0.18%
CT OOS R^2	0.54%	0.51%	-0.67%	-0.34%	-0.01%	0.00%	-0.06%	0.03%
PTV OOS R^2	0.57%	0.57%	0.03%	0.03%	0.24%	0.24%	0.50%	0.50%

Summary Statistics

	IS R^2	OOS R^2			CT OOS R^2			PTV OOS R^2		
	Mean	Mean	#>0	#>OLS	Mean	#>0	#>OLS	Mean	#>0	#>OLS
OLS	0.33%	-0.39%	4	-	0.01%	8	-	0.23%	13	-
WLS	0.22%	-0.22%	6	11	0.11%	11	9	0.26%	14	8

Table 4: Out-of-Sample Return Predictability (cont'd)

Panel B: Predicting Next-Year Returns

Predictor:	dp		DP		ep		de	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-4.08%	-1.92%	-1.22%	0.86%	-8.51%	-0.33%	-3.80%	-2.47%
CT OOS R^2	3.33%	3.22%	2.31%	2.64%	0.66%	4.88%	-0.29%	-0.71%
PTV OOS R^2	3.37%	2.32%	2.64%	2.08%	4.59%	4.55%	-2.77%	-1.25%
Predictor:	$\text{RV}_{t-252,t}$		tbl		lty		ltr	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-0.94%	-5.59%	-10.18%	-13.94%	-12.81%	-20.12%	0.98%	0.81%
CT OOS R^2	-0.09%	0.00%	-1.19%	0.26%	0.35%	-0.36%	1.30%	1.19%
PTV OOS R^2	-0.93%	-0.55%	2.91%	1.86%	0.32%	-0.62%	1.04%	1.22%
Predictor:	tms		dfy		infl		bm	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-0.73%	2.05%	-3.05%	0.45%	-0.55%	1.51%	-21.75%	-6.86%
CT OOS R^2	-0.08%	1.69%	-2.42%	-0.13%	0.02%	1.78%	-8.68%	-3.79%
PTV OOS R^2	3.77%	3.78%	-2.20%	0.52%	-0.43%	1.22%	0.05%	0.64%
Predictor:	csp		ntis		lpy		caya	
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
OOS R^2	-4.49%	-4.39%	-16.14%	-7.47%	-20.97%	-5.65%	0.77%	1.81%
CT OOS R^2	-2.37%	-1.99%	-16.17%	-7.50%	1.62%	2.67%	0.83%	1.62%
PTV OOS R^2	-4.48%	-4.40%	-0.05%	-0.02%	4.53%	4.65%	6.33%	6.28%

Summary Statistics

	IS R^2	OOS R^2			CT OOS R^2			PTV OOS R^2		
	Mean	Mean	#>0	#>OLS	Mean	#>0	#>OLS	Mean	#>0	#>OLS
OLS	3.08%	-6.72%	2	-	-1.30%	8	-	1.17%	9	-
WLS	1.98%	-3.83%	6	12	0.34%	9	12	1.39%	11	10

Table 5: Predicting Returns Using the Variance Risk Premium

This table presents estimates of return predictability regressions of the form:

$$r_{t+1,t+21} = a + b \cdot \hat{\text{VRP}}_t + \epsilon_{t+h},$$

where $r_{t+1,t+21}$ is the log dividend-inclusive excess return of the CRSP value-weighted index over the 21 days starting with $t+1$, annualized and in percent. $\hat{\text{VRP}}_t$ is one of two proxies for the variance risk premium, both expressed as monthly percents squared. The first, from Drechsler and Yaron (2011), is the difference between the squared VIX and an annualized estimate of next-month realized variance $\text{FutRV}^2_{t+1,t+21}$, computed from intraday S&P 500 futures returns. The second, from Bollerslev, Tauchen, and Zhou (2009), is the difference between the squared VIX and annualized past realized variance $\text{FutRV}^2_{t-20,t}$. For each predictor, I compute point estimates of b using OLS and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2, using both VIXF and VIX $\hat{\sigma}^2_t$. I also compute $\hat{b}$ adjusted for the Stambaugh bias using a simulation procedure. I compute standard errors and p-values for the bias-adjusted coefficients using Hodrick (1992) (Asy) and the heteroskedastic simulation procedure (Sim) described in Section 3. The sample is 6028 daily observations from 1990-2013. ***, **, and * indicate Sim p-values below 1%, 5% and 10% level, respectively.

Panel A: Drechsler and Yaron (2011) Approach
$\hat{\text{VRP}}_t = \text{VIX}^2_t - \hat{\mathbb{E}}_t(\text{FutRV}^2_{t+1,t+21})$

Forecast horizon:	*One month*			*Three months*		
	OLS	WLS-VIXF	WLS-VIX	OLS	WLS-VIXF	WLS-VIX
Stambaugh $\hat{b}_{\text{adj}}$	0.426*	0.154	0.133	0.344**	0.143	0.131
Unadjusted $\hat{b}$	0.433	0.166	0.150	0.349	0.153	0.143
SE (Asy)	(0.229)	(0.163)	(0.163)	(0.173)	(0.128)	(0.128)
p-value (Asy)	6.3%	34.3%	41.4%	4.6%	26.1%	30.3%
SE (Sim)	(0.219)	(0.162)	(0.163)	(0.172)	(0.131)	(0.133)
p-value (Sim)	5.5%	34.2%	41.3%	4.7%	27.5%	31.8%

Panel B: Bollerslev, Tauchen, and Zhou (2009) Approach
$\hat{\text{VRP}}_t = \text{VIX}^2_t - \text{IndRV}^2_{t-20,t}$

Forecast horizon:	*One month*			*Three months*		
	OLS	WLS-VIXF	WLS-VIX	OLS	WLS-VIXF	WLS-VIX
Stambaugh $\hat{b}_{\text{adj}}$	0.520**	0.205	0.191	0.416***	0.175	0.170
Unadjusted $\hat{b}$	0.526	0.217	0.206	0.419	0.185	0.181
SE (Asy)	(0.237)	(0.160)	(0.160)	(0.151)	(0.114)	(0.114)
p-value (Asy)	2.8%	19.8%	23.3%	0.6%	12.5%	13.7%
SE (Sim)	(0.221)	(0.159)	(0.163)	(0.153)	(0.120)	(0.120)
p-value (Sim)	1.8%	19.8%	24.1%	0.7%	14.1%	15.7%

Table 6: Predicting Returns Using the Variance Risk Premium: Alternative Procedures

This table presents variations of the regressions in Table 5 but with different data, sample periods, sample frequencies, and return indices. The columns with "Original" data are point estimates from the original papers or using data uploaded by the authors. The remaining columns use variance risk premia proxies I compute with the goal of replicating the original measures. The two potential indices are the CRSP value-weighted index and the S&P 500 index, both inclusive of dividends and net of the risk-free rate. For each variation, I compute point estimates of b using OLS and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2, using VIX $\hat{\sigma}_t^2$. I also compute $\hat{b}$ adjusted for the Stambaugh (1999) bias using a simulation procedure. I compute standard errors and p-values for the bias-adjusted coefficients using Hodrick (1992) (Asy) and a heteroskedastic simulation procedure (Sim). Monthly sampling results in 215 monthly observations from 1990-2007 and 287 from 1990-2013, while daily sampling results in 6028 observations from 1990-2013. ***, **, and * indicate Sim p-values below 1%, 5% and 10% level, respectively.

Panel A: Drechsler and Yaron (2011) Approach

$\hat{\text{VRP}}_t = \text{VIX}_t^2 - \hat{\mathbb{E}}_t(\text{Fut. RV}_{t+1,t+21}^2)$

Data:	Original		Replicated		Replicated		Replicated		Replicated		Replicated	
Sample period:	1990-2007		1990-2007		1990-2007		1990-2007		1990-2013		1990-2013	
Sample interval:	Monthly		Monthly		Monthly		Daily		Monthly		Daily	
Index:	S&P		S&P		Log S&P		Log S&P		Log S&P		Log S&P	
Forecast horizon: one month												
	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS
Stambaugh $\hat{b}_{\text{adj}}$	-	-	0.378	0.196	0.349	0.157	0.446	0.156	0.498*	0.263	0.427*	0.130
Unadjusted $\hat{b}$	0.760***	-	0.403	0.226	0.371	0.184	0.463	0.195	0.507	0.274	0.432	0.144
SE (Asy)	(0.350)	-	(0.179)	(0.222)	(0.195)	(0.230)	(0.298)	(0.238)	(0.191)	(0.188)	(0.228)	(0.162)
p-value (Asy)	2.9%	-	3.4%	37.7%	7.4%	49.6%	13.5%	51.4%	0.9%	16.1%	6.2%	42.4%
SE (Sim)	-	-	(0.316)	(0.252)	(0.320)	(0.255)	(0.285)	(0.235)	(0.279)	(0.207)	(0.222)	(0.164)
p-value (Sim)	-	-	24.5%	44.9%	28.0%	54.3%	11.9%	50.9%	7.4%	20.5%	5.7%	42.8%
Forecast horizon: three months												
	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS
Stambaugh $\hat{b}_{\text{adj}}$	-	-	0.436*	0.214	0.403*	0.170	0.405*	0.142	0.473***	0.239	0.331*	0.114
Unadjusted $\hat{b}$	0.860***	-	0.456	0.239	0.422	0.191	0.418	0.169	0.479	0.245	0.335	0.127
SE (Asy)	(0.270)	-	(0.122)	(0.154)	(0.153)	(0.166)	(0.220)	(0.190)	(0.121)	(0.118)	(0.172)	(0.127)
p-value (Asy)	0.1%	-	0.0%	16.4%	0.8%	30.4%	6.6%	45.5%	0.0%	4.3%	5.4%	37.1%
SE (Sim)	-	-	(0.225)	(0.226)	(0.230)	(0.190)	(0.227)	(0.197)	(0.185)	(0.143)	(0.173)	(0.132)
p-value (Sim)	-	-	5.5%	34.2%	7.8%	36.9%	7.9%	47.7%	0.9%	10.1%	5.8%	38.6%

Table 6: Predicting Returns Using the Variance Risk Premia: Alternative Procedures (cont'd)

Panel B: Bollerslev, Tauchen, and Zhou (2009) Approach

$\hat{\text{VRP}}_t = \text{VIX}_t^2 - \text{Fut. RV}_{t-20,t}^2$

Data:	Original		Original		Replicated		Replicated		Replicated		Replicated	
Sample period:	1990-2007		1990-2007		1990-2007		1990-2007		1990-2013		1990-2013	
Sample interval:	Monthly		Monthly		Monthly		Daily		Monthly		Daily	
Index:	Log S&P		Log S&P		Log S&P		Log S&P		Log S&P		Log S&P	
	Forecast horizon: one month											
	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS
Stambaugh $\hat{b}_{\text{adj}}$	-	-	0.358	0.179	0.386	0.224	0.468	0.161	0.595*	0.405*	0.518**	0.185
Unadjusted $\hat{b}$	0.390*	-	0.372	0.196	0.400	0.235	0.478	0.191	0.598	0.408	0.524	0.200
SE (Asy)	(0.222)	-	(0.212)	(0.247)	(0.211)	(0.247)	(0.300)	(0.241)	(0.142)	(0.188)	(0.236)	(0.160)
p-value (Asy)	7.8%	-	9.1%	46.9%	6.7%	36.4%	11.9%	50.6%	0.0%	3.1%	2.8%	24.6%
SE (Sim)	-	-	(0.337)	(0.271)	(0.343)	(0.276)	(0.293)	(0.241)	(0.308)	(0.222)	(0.221)	(0.162)
p-value (Sim)	-	-	29.0%	51.1%	26.0%	42.0%	11.5%	50.0%	5.3%	6.7%	1.9%	25.1%
	Forecast horizon: three months											
	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS	OLS	WLS
Stambaugh $\hat{b}_{\text{adj}}$	-	-	0.463**	0.231	0.477**	0.248	0.441*	0.171	0.527***	0.321**	0.405*	0.157
Unadjusted $\hat{b}$	0.470***	-	0.479	0.248	0.491	0.263	0.451	0.192	0.529	0.323	0.409	0.169
SE (Asy)	(0.164)	-	(0.149)	(0.163)	(0.148)	(0.163)	(0.218)	(0.189)	(0.086)	(0.103)	(0.150)	(0.114)
p-value (Asy)	0.4%	-	0.2%	15.5%	0.1%	12.7%	4.3%	36.5%	0.0%	0.2%	0.7%	17.0%
SE (Sim)	-	-	(0.231)	(0.191)	(0.230)	(0.191)	(0.227)	(0.196)	(0.187)	(0.135)	(0.151)	(0.121)
p-value (Sim)	-	-	4.5%	22.6%	4.0%	19.5%	5.2%	38.2%	0.5%	1.7%	0.7%	19.5%

Table 7: Predicting Returns Using Politics, the Weather, and the Stars

This figure presents estimates of return predictability regressions of the form:

$$r_{t+1} = a + b \cdot X_t + \epsilon_{t+1},$$

where r_{t+1} is the log dividend-inclusive excess return of the CRSP value-weighted index in month $t+1$, in percent. X_t is one of nine predictors from Novy-Marx (2014): is an indicator for whether the President of the United States in month t is a Democrat (Dem), the monthly highest temperature in New York City (NYC Weather), the global temperature anomaly (Global Temp.), the rolling average global temperature (Roll. Global Temp.), the quadiperiodic Pacific temperature anomaly (El Niño), the rolling average Pacific Ocean temperature (Roll. El Niño), the observed number of sunspots (Sunspots), the angle between Mars and Saturn (Mars/Saturn Angle), and the angle between Jupiter and Saturn (Jupiter/Saturn Angle). For each predictor, I compute point estimates of b using OLS and weighted least squares with ex-ante return variance (WLS-EV), detailed in Section 2. I compute p-values for the coefficients using the heteroskedastic simulation procedure (Sim) described in Section 3. The sample is 624 monthly observations from 1961-2012. ***, **, and * indicate Sim p-values below 1%, 5% and 10% level, respectively.

Predicting Next-Month Returns						
	OLS	WLS-EV	OLS	WLS-EV	OLS	WLS-EV
Predictor:	Dem		NYC Weather		Global Temp.	
$\hat{b}$	0.767**	0.521	-0.026**	-0.018	0.100	0.215
p-value (Sim)	4.07%	10.53%	4.22%	10.66%	85.25%	64.14%
Predictor:	Roll. Global Temp.		El Niño		Roll. El Niño	
$\hat{b}$	0.099	0.267	0.000	0.017	-0.303	-0.058
p-value (Sim)	86.32%	59.66%	99.81%	91.92%	56.44%	89.89%
Predictor:	Sunspots		Mars/Saturn Angle		Jupiter/Saturn Angle	
$\hat{b}$	-0.003	-0.004	0.513***	0.414**	-0.009	-0.079
p-value (Sim)	44.56%	28.83%	0.69%	1.39%	96.55%	66.64%
Joint significance:	OLS	WLS-EV				
χ^2 statistic	15.22*	13.41				
p-value (Sim)	8.52%	14.58%				