Journal of Financial Economics 144 (2022) 732–760

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Validity, tightness, and forecasting power of risk premium bounds

Kerry Back[a,b], Kevin Crotty[a,*], Seyed Mohammad Kazempour[a]

[a] Jones Graduate School of Business, Rice University, Houston, TX 77005, USA
[b] School of Social Sciences, Rice University, Houston, TX 77005, USA

ARTICLE INFO

Article history:
Received 29 November 2021
Revised 17 February 2022
Accepted 17 February 2022
Available online 28 March 2022

JEL classification:
G11
G12
G17

Keywords:
Risk premia
Bounds
Conditional tests
Predictability
Forecasting

ABSTRACT

Recent work uses option prices to derive lower bounds for the risk premia of the market portfolio and individual stocks. We test the bounds conditionally. We cannot reject that they are valid, but we do reject that they are tight. Using the market bounds as forecasts appears unreasonable in many cases due to their high slackness. Adding past mean slackness is a potential improvement but is hampered by the brevity of the available data series. The correlation of the stock bounds with subsequent returns stems primarily from the time series rather than the cross section.

An innovative series of recent papers, beginning with the seminal paper by Martin (2017), uses option prices and economic theory to place ex ante lower bounds on risk premia of the market portfolio and individual stocks. These bounds are potentially important advances in forecasting returns. Using an extended sample, we analyze the bounds on the market risk premium developed by Martin (2017) and (Chabi-Yo and Loudis, 2020) and the bounds on the risk premia of individual stocks developed by Martin and Wagner (2019) and Kadan and Tang (2020). Using conditional tests, we cannot reject validity, but we do reject tightness. We find some evidence that the bounds are correlated with subsequent returns. We also perform out-of-sample forecasting tests. The bounds sometimes outperform the historical market mean for forecasting, but the outperformance is statistically insignificant in our rather short data series.

We test validity of the bounds by following Boudoukh et al. (1993), who test whether zero is a lower bound on the conditional market risk premium. We interact realized bound slackness (the excess return minus the bound) with positive predictive variables studied by Welch and Goyal (2008). By iterated expectations, the mean interactions are nonnegative if the bound is valid. We test if the mean interactions are nonnegative against an unrestricted alternative. Unlike Boudoukh et al., we generally do not reject bound validity (our sample begins after theirs ends due to the availability of option prices). Because we do not reject bound validity, we proceed to test bound tightness by testing the null hy-

* Corresponding author.
E-mail addresses: kerry.e.back@rice.edu (K. Back), kevin.p.crotty@rice.edu (K. Crotty), smkazempour@rice.edu (S.M. Kazempour).

https://doi.org/10.1016/j.jfineco.2022.02.003
0304-405X/

pothesis that the mean interactions are zero against the alternative that they are nonnegative. We find strong evidence against tightness.

To analyze the in-sample predictive power of the bounds, we run time-series regressions for market bounds and Fama–MacBeth and panel regressions for stock bounds. In univariate market regressions, the Martin bound is significant only at the 6 month horizon, and the Chabi-Yo/Loudis bound is significant only at 6 and 12-month horizons. However, both bounds are significant at all horizons when we control for other standard market return predictors. In the Fama–MacBeth regressions, the stock bounds are insignificant predictors of returns, and the signs are even wrong in the full sample; i.e., larger bound realizations are associated with lower subsequent realized returns. However, the bounds are significant predictors in panel regressions with stock fixed effects, so we conclude that the predictive power of the stock bounds is primarily in the time series. We find additional evidence that the cross-sectional predictability is limited by examining portfolio returns from sorts on the stock bounds.

For both the market and the stock bounds, we calculate out-of-sample R^2s and run Diebold and Mariano (1995) tests of out-of-sample forecasting power relative to the historical market mean as a benchmark. There is some outperformance (especially for the Chabi-Yo/Loudis market bound and the Martin–Wagner stock bound) but it is statistically insignificant. Adding past mean slackness to the bounds appears to be a promising forecasting approach, but we have a much longer historical period for estimating the mean market return than we do for estimating mean slackness of the bounds. Simulations for the market bounds show that we may need another century or more of data before the 'bound + past mean slackness' forecast can be expected to consistently generate positive out-of-sample R^2s relative to the market-mean benchmark.

There are other related bounds/formulas for risk premia that were developed more recently that we do not test. Bakshi et al. (2019) derive a formula for the market risk premium. Chabi-Yo et al. (2022) derive a lower bound on the risk premia of individual stocks, which they call a 'generalized lower bound.' Among other things, they follow our lead in using the Kodde–Palm/Boudoukh–Richardson–Smith methodology with the Goyal–Welch variables to test the bound. They apply the test to stock portfolios. They do not reject validity of their bound, nor do they reject tightness at shorter horizons (1 and 3 months), but they do reject tightness at longer horizons (6 and 12 months). They also find that their bound outperforms the Martin–Wagner and Kadan–Tang bounds in predicting stock returns; thus, their paper is an important complement to ours.

The bounds are defined in Section 1, and our tests for validity and tightness are described in Section 2. The data is described in Section 3, the tests of validity and tightness are presented in Section 4, full-sample regressions of excess returns on bounds are presented in Section 5, and out-of-sample forecasting results are presented in Section 6.

1. Bounds

For an asset with gross return $R_{t,T}$ over a time period $[t, T]$, define

$$SVIX^2_{t,T} = \text{var}^*\left(\frac{R_{t,T}}{R_{ft,T}}\right), \tag{1}$$

where $R_{f,t,T}$ denotes the gross risk-free return over the same time period, and var* denotes risk-neutral variance. This notation follows Martin (2017), except that, as in Martin and Wagner (2019), we do not annualize $SVIX^2_{t,T}$. Assuming that all dividends paid between t and T are paid at T and are known at t, Martin shows that (1) can be calculated in terms of put and call prices as

$$SVIX^2_{t,T} = \frac{2}{R_{f,t,T}S_t^2}\left[\int_0^{F_{t,T}} put_{t,T}(K)\,\mathrm{d}K + \int_{F_{t,T}}^{\infty} call_{t,T}(K)\,\mathrm{d}K\right], \tag{2}$$

where S denotes the price of the asset and $F_{t,T}$ denotes its forward price at t for a contract maturing at T. Furthermore, under the Negative Correlation Condition (NCC), Martin derives the following lower bound for the risk premium

$$E_t[R_{t,T}] - R_{f,t,T} \geq R_{f,t,T}SVIX^2_{t,T}\,. \tag{3}$$

The NCC is that the return $R_{t,T}$ has a negative correlation with $M_{t,T}R_{t,T}$, where $M_{t,T}$ denotes the stochastic discount factor (SDF) at t for pricing payoffs at T. The NCC is a quite plausible assumption for the market return, as detailed by Martin. For example, if there is a representative investor with constant relative risk aversion ρ and $R_{t,T}$ is the market return, then

$$M_{t,T}R_{t,T} = \frac{R_{t,T}^{1-\rho}}{R_{f,t,T}E\left[R_{t,T}^{-\rho}\right]},$$

which is decreasing in $R_{t,T}$ and hence has a negative correlation with $R_{t,T}$, provided only that $\rho > 1$. When $\rho = 1$ (log utility), $M_{t,T}R_{t,T}$ is nonrandom and hence has a zero correlation with $R_{t,T}$. Thus, the bound (3) is tight when there is a representative investor with log utility.[1]

Chabi-Yo and Loudis (2020) observe that risk premia are proportional to risk-neutral covariances with the reciprocal of the SDF. They assume there is a representative investor and perform a Taylor series expansion of the reciprocal of the representative investor's marginal utility. With an assumption on risk-neutral moments of the market excess return (odd moments are weakly negative) and

[1] For the convenience of the reader, we repeat Martin's reasoning that leads to the bound (3) and which shows that the bound is tight when the correlation is zero. The definition of variance, the definition of risk-neutral expectation as $\mathsf{E}^*[x] = R_{ft,T}\mathsf{E}[M_{t,T}x]$, and the fact that $\mathsf{E}[M_{t,T}R_{t,T}] = 1$ imply

$$\text{var}^*(R_{t,T}) = R_{ft,T}\mathsf{E}[M_{t,T}R_{t,T}^2] - R_{ft,T}^2\,.$$

The definition of covariance and the fact that $\mathsf{E}[M_{t,T}R_{t,T}] = 1$ imply

$$\mathsf{E}[M_{t,T}R_{t,T}^2] = \text{cov}(M_{t,T}R_{t,T}, R_{t,T}) + \mathsf{E}[R_{t,T}]\,.$$

We can substitute this into the previous formula and rearrange to obtain

$$\mathsf{E}[R_{t,T}] - R_{ft,T} = R_{ft,T}SVIX^2_{t,T} - \text{cov}(M_{t,T}R_{t,T}, R_{t,T})\,.$$

some assumptions on the representative investor's tolerances for risk, skewness, and kurtosis, they derive a lower bound on the market risk premium based on the first four risk-neutral moments of the market excess return. The bound involves the tolerance parameters, which Chabi-Yo and Loudis estimate from the data. They call this their unrestricted lower bound. By imposing stronger restrictions on the representative investor's tolerance for risk, skewness, and kurtosis, they obtain a bound that is free of preference parameters, which they call their restricted lower bound. Chabi-Yo and Loudis show empirically that both of their bounds are generally higher than the Martin bound over their sample period. We confirm this for our extended sample. We focus on the restricted Chabi-Yo/Loudis bound in our empirical work, because it is free of preference parameters.

Martin and Wagner (2019) derive the following formula for the risk premium of an individual stock, after dropping a stock fixed effect:

$$\mathsf{E}_t\left[R_{i,t,T} - R_{f,t,T}\right] = \mathsf{E}_t[R_{m,t,T} - R_{f,t,T}] + \frac{1}{2}R_{f,t,T}\left(SVIX^2_{i,t,T} - \overline{SVIX}^2_{t,T}\right). \quad (4)$$

Here, $SVIX^2_{i,t,T}$ is defined as in (1) for stock i, and $\overline{SVIX}^2_{t,T}$ is the value-weighted average of (1) across all stocks. Using Martin's formula for the lower bound on the market risk premium, Martin and Wagner deduce that

$$\mathsf{E}_t\left[R_{i,t,T} - R_{f,t,T}\right] \geq R_{f,t,T}SVIX^2_{t,T} + \frac{1}{2}R_{f,t,T}\left(SVIX^2_{i,t,T} - \overline{SVIX}^2_{t,T}\right). \quad (5)$$

with equality if Martin's bound is tight. We refer to the right-hand side of (5) as the Martin–Wagner bound.

Kadan and Tang (2020) show that the NCC holds approximately for an individual stock under conditions similar to those that imply it holds for the market if the parameter

$$\delta_{it} = \frac{\mathrm{var}(R_{i,t,T})}{\mathrm{cov}(R_{i,t,T}, R_{m,t,T})} \quad (6)$$

is small. They conclude that the Martin bound (3) should hold for individual stocks for which δ_{it} is small—specifically, (3) is a lower bound for a stock's expected excess return if δ_{it} is less than risk aversion and if the stock's market beta is positive. They show empirically that δ_{it} tends to be small for stocks with low market betas and for stocks for which the market-model regression has a high R^2. Kadan and Tang also note that their formula may be an upper bound for the risk premia of stocks with high values of δ_{it}, because, if the opposite of the NCC holds—that is, if $R_{t,T}$ and $M_{t,T}R_{t,T}$ are positively correlated—then the inequality in (3) is reversed. We refer to the Martin bound (3) applied to individual stocks as the Kadan–Tang bound.

Simple algebra applied to (3) and (5) shows that

$$KT_{i,t,T} = 2MW_{i,t,T} + \overline{KT}_{t,T} - 2M_{t,T}, \quad (7)$$

where KT denotes the Kadan–Tang bound, MW denotes the Martin–Wagner bound, $\overline{KT}$ denotes the value-weighted average of the Kadan–Tang bounds, and M denotes the Martin bound on the market risk premium. Thus, the Kadan–Tang bound is twice the Martin–Wagner bound plus the term $\overline{KT}_{t,T} - 2M_{t,T}$, which is constant across stocks. We will see that this term is usually positive in our sample (Fig. 3.5 in Section 3), so the Kadan–Tang bound is usually substantially larger than the Martin–Wagner bound. Because the term is constant across stocks, the Kadan–Tang and Martin–Wagner bounds are perfectly correlated in each cross section. Thus, for example, sorting stocks into portfolios based on a bound produces the same portfolios for the two bounds.

2. Multiple inequality tests

We test the bounds by testing inequality restrictions on a vector of moments. Let $R^e_{t,T}$ denote an excess return from t to T, and let $b_{t,T}$ denote a lower bound on the conditional risk premium. So, we have

$$\mathsf{E}_t\left[R^e_{t,T}\right] \geq b_{t,T}. \quad (8)$$

For any vector z_t of nonnegative conditioning variables in the time t information set, inequality (8) implies a vector of inequalities:

$$\mathsf{E}_t\left[\left(R^e_{t,T} - b_{t,T}\right)z_t\right] \geq 0. \quad (9)$$

We will always include the constant 1 as an element of z_t, so inequality (8) is included in the vector of inequalities (9). By the law of iterated expectations,

$$\mathsf{E}\left[\left(R^e_{t,T} - b_{t,T}\right)z_t\right] \geq 0. \quad (10)$$

Let λ_0 denote the population mean on the left-hand side of inequality (10).

We are interested in two questions. First, is the bound valid? To answer this, we test the null hypothesis $\lambda_0 \geq 0$ against the alternative that λ_0 is unrestricted. Second, if it appears that the bound holds, is the bound tight? To answer this, we test the null hypothesis that $\lambda_0 = 0$ against the alternative that $\lambda_0 \geq 0$. The theory of testing inequality restrictions on parameter vectors and functions of parameter vectors has a substantial history, beginning with Perlman (1969). We apply the asymptotic distribution theory developed by Kodde and Palm (1986) and Wolak (1989) for the minimum distance estimators we describe now. Our procedure is the same as that of Boudoukh et al. (1993), except that we test both validity and tightness, whereas they only test validity (of zero as a lower bound for the market risk premium).

Let $\bar{\lambda}$ denote the sample mean of the vector $(R^e_{t,T} - b_{t,T})z_t$.[2] We assume ergodicity, so $\bar{\lambda}$ is a consistent estimator of λ_0. Let Σ denote a consistent estimator of the asymptotic covariance matrix of $\bar{\lambda}$. Define

$$D_1 = \min_{\lambda \geq 0}\left(\lambda - \bar{\lambda}\right)'\Sigma^{-1}\left(\lambda - \bar{\lambda}\right). \quad (11)$$

The statistic D_1 is the squared distance of $\bar{\lambda}$ from the nonnegative orthant in the norm defined by Σ^{-1}. If $\lambda_0 \geq 0$,

[2] $\bar{\lambda}$ is the sample analogue to the unconditional expectation on the left-hand side of inequality (10). Thus, we test whether the interactions of the bounds with the conditioning variables are nonnegative or zero on average rather than for each time period (inequality (9)).

then it is unlikely that $\bar{\lambda}$ will be far from the nonnegative orthant, so the distance D_1 can be used to test the null that $\lambda_0 \geq 0$ against the alternative that λ_0 is unrestricted (i.e., to test whether the bound is valid). Set

$$\hat{\lambda} \;=\; \text{argmin}_{\lambda \geq 0} \left(\lambda - \bar{\lambda}\right)' \Sigma^{-1} \left(\lambda - \bar{\lambda}\right). \tag{12}$$

Under the null that $\lambda_0 \geq 0$, D_1 is asymptotically distributed as a mixture of chi-square distributions with various degrees of freedom depending on the number of elements of $\hat{\lambda}$ that are strictly positive. See Appendix A for more details.

Set

$$D_0 \;=\; \bar{\lambda}' \Sigma^{-1} \bar{\lambda}. \tag{13}$$

This is the squared distance of $\bar{\lambda}$ from the origin in the norm defined by Σ^{-1}, and it is the standard Wald chi-square statistic for testing $\lambda_0 = 0$ against an unrestricted alternative. Finally, set

$$D_2 \;=\; D_0 - D_1. \tag{14}$$

The statistic D_2 is the squared distance of $\bar{\lambda}$ from the origin minus its squared distance from the nonnegative orthant, all under the norm defined by Σ^{-1}. We use it to test the null that $\lambda_0 = 0$ against the alternative that $\lambda_0 \geq 0$, that is, to test the null that the bound is tight against the alternative that it is valid. Under this null, D_2 is also asymptotically distributed as a mixture of chi-square distributions with various degrees of freedom depending on the number of elements of $\hat{\lambda}$ that are strictly positive. Again, more details are provided in Appendix A. By the usual properties of orthogonal projections, $D_2 = \hat{\lambda}' \Sigma^{-1} \hat{\lambda}$.

In addition to asymptotic p-values based on the limiting chi-square distributions, we also report simulated finite-sample p-values for the market risk premium bounds. We model the conditional mean of daily returns, μ_t, as an AR(1) process. The return-generating process is:

$$\mu_{t+1} = (1-a)\bar{\mu} + a\mu_t + u_{t+1}, \tag{15}$$

$$r_{t+1} = \mu_t + v_{t+1}, \tag{16}$$

where time t is measured in days and where the innovation vectors (u_t, v_t) are mean-zero iid random vectors (with u_t and v_t possibly correlated). We model the forward return as the sum of the daily returns over the horizon, and we model the bound at date t as the mean of the forward return conditional on μ_t. The conditioning variables in our empirical analyses are observed monthly. To be consistent in our simulations, we model the logs of the conditioning variables $x = \log z$ as a VAR(1) with monthly time steps:

$$x_{m+1} = (I - A)\bar{x} + Ax_m + w_{m+1}. \tag{17}$$

We assume the innovation vectors w_m are mean-zero iid random vectors. We calibrate the model allowing for arbitrary empirical slackness (see Appendix B). We calibrate separately for the Martin and Chabi-Yo/Loudis bounds, and we calibrate separately for each horizon. For each of the two bounds and each of the horizons, we run 1000 simulations, each consisting of the same number of days that we have in our sample. In each simulation, we take the bound to be tight, so we estimate p-values under the null of a tight bound.

Our methodology is quite different from that of Martin (2017) and Chabi-Yo and Loudis (2020), who test tightness in a regression framework and test validity only as a corollary of tightness. They regress the market excess return on their bounds $b_{t,t+h}$:

$$R^e_{t,t+h} = \alpha_h + \beta_h b_{t,t+h} + \varepsilon_{th}. \tag{18}$$

for various horizons h. Both Martin and Chabi-Yo/Loudis observe that they cannot reject the null hypothesis that $\alpha_h = 0$ and $\beta_h = 1$ for any h, i.e., that the bound is tight. However, the standard errors on β_h are quite large in both papers, and the data also fail to reject the null that $\beta_h = 2$. In fact, the point estimates in both papers are above 2 at the six-month horizon. Obviously, if $\beta_h = 2$ in (18) (and α_h is not too negative), then the bound is slack. Thus, with this type of test, we cannot reject that the bounds are tight, but we also cannot reject that they are slack. The test simply does not have much power.

To conclude this section, it seems useful to discuss how our validity and tightness tests relate to regressions of realized slackness on conditioning variables. The validity and tightness tests are unaffected if we rescale each conditioning variable to have a unit mean, so suppose we have done so. Then, for each variable z_j and an excess return $R^e_{t,T}$ (which could be the market or an individual stock), we have

$$\mathsf{E}[(R^e_{t,T} - b_{t,T}) z_{jt}] = \mathsf{E}[R^e_{t,T} - b_{t,T}] + \text{cov}(R^e_{t,T} - b_{t,T}, z_{jt}).$$

Recall that the constant 1 is included in the conditioning variables. So, validity means that

$$\mathsf{E}[R^e_{t,T} - b_{t,T}] \geq 0 \tag{19a}$$

and that

$$\text{cov}(R^e_{t,T} - b_{t,T}, z_{jt}) \geq -\mathsf{E}[R^e_{t,T} - b_{t,T}] \tag{19b}$$

for each j. In other words, a valid bound must be valid unconditionally and moreover covariances of realized slackness with conditioning variables cannot be 'too negative.' Similarly, a tight bound must be tight unconditionally and moreover covariances of realized slackness with conditioning variables must be zero. So, conditional tests of validity and tightness are tantamount to unconditional tests plus tests of regression coefficients from regressing realized slackness on conditioning variables.

3. Bound and return data

3.1. Market bounds and returns

Table 1 presents summary statistics of the market excess return and the bounds. The time-series of the bounds run from January 1990 through December 2020. Appendix C explains the details of how the bounds are calculated.[3] We compute market excess returns from S&P 500

[3] We have verified that our calculations are consistent with the bounds provided by Ian Martin on his website for the period January 1996 through January 2012 and the bounds provided on Fousseni Chabi-Yo's website for the period January 1996 through August 2015. OptionMetrics data starts in 1996; we use option prices from CBOE to extend our sample earlier to 1990.

Table 1
Market summary statistics.
The sample contains daily observations on the bounds from January 1990 through December 2020. Data on subsequent realized market excess returns runs from January 1990 through March 2021. Returns and bounds are annualized and expressed in percent. The market excess return for a given date is the S&P 500 return in excess of the risk-free rate, compounded over the indicated horizon (and annualized). Martin Bound denotes the Martin (2017) bound. CYL Bound denotes the restricted bound from Chabi-Yo and Loudis (2020).

	Mean	SD	P10	P25	P50	P75	P90	N
Panel A. 1-month Horizon								
Market Excess Return	8.76	54.06	−54.48	−17.29	14.22	39.29	64.33	7789
Martin Bound	3.81	3.99	1.20	1.62	2.65	4.54	7.16	7789
CYL Bound	4.39	5.03	1.31	1.80	2.97	5.17	8.25	7789
Panel B. 3-month Horizon								
Market Excess Return	8.73	29.93	−28.17	−4.99	11.39	26.11	40.00	7787
Martin Bound	3.86	3.15	1.47	1.94	2.91	4.79	6.93	7787
CYL Bound	4.77	4.40	1.70	2.30	3.48	5.75	8.73	7787
Panel C. 6-month Horizon								
Market Excess Return	8.67	21.44	−17.66	−1.77	10.33	20.95	32.79	7724
Martin Bound	3.83	2.59	1.68	2.15	3.04	4.79	6.68	7724
CYL Bound	5.06	3.94	2.04	2.73	3.91	6.10	9.03	7724
Panel D. 12-month Horizon								
Market Excess Return	8.86	15.64	−14.24	2.69	10.36	18.40	25.17	7598
Martin Bound	3.68	2.09	1.77	2.25	3.04	4.58	6.12	7598
CYL Bound	5.18	3.47	2.23	3.02	4.24	6.30	8.97	7598

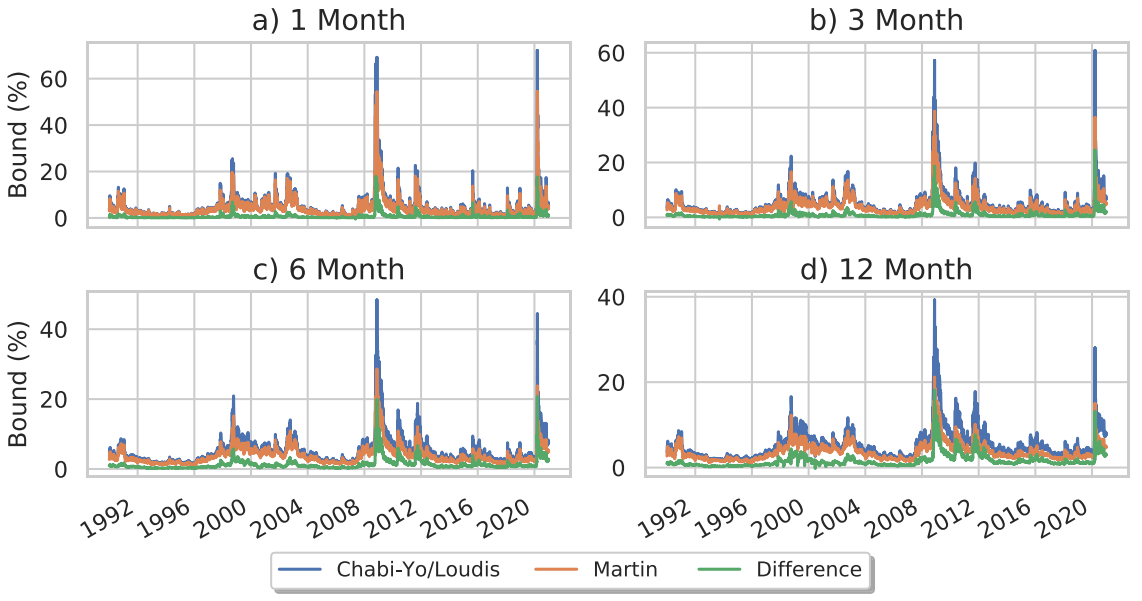

Fig. 3.1. Time series of Martin and Chabi-Yo/Loudis bounds.
The daily time-series of the Martin and Chabi-Yo/Loudis bounds are shown (in % per year) as well as their difference.

returns from January 1990 through March 2021, using the risk-free rate from Ken French's website. The mean market excess return is substantially larger than the mean bounds, by roughly 3.5% to 5% per year, depending on the bound and the horizon. Consistent with the results of Chabi-Yo and Loudis, the Chabi-Yo/Loudis bound is generally higher than Martin's bound, so realized slackness is lower.

Fig. 3.1 shows the time series of the Martin and Chabi-Yo/Loudis bounds. As emphasized by Martin (2017), the bounds are very volatile and are occasionally quite high. The peaks are in periods when measures of market uncertainty like the VIX are also very high. Fig. 3.1 confirms that the Chabi-Yo/Loudis bound is almost always higher then the Martin bound, and their difference is correlated with their levels. Fig. 3.2 presents the same daily bound data in a different format. It confirms that, as is evident also from Fig. 3.1, the two bounds are highly correlated. The correlation ranges from 99.8% at the 1-month horizon to

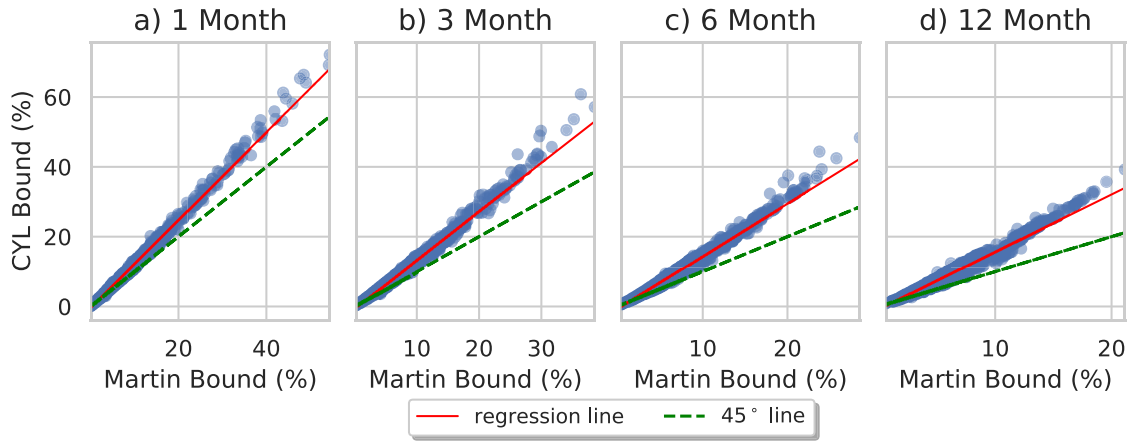

Fig. 3.2. Scatter plot of Martin and Chabi-Yo/Loudis bounds.

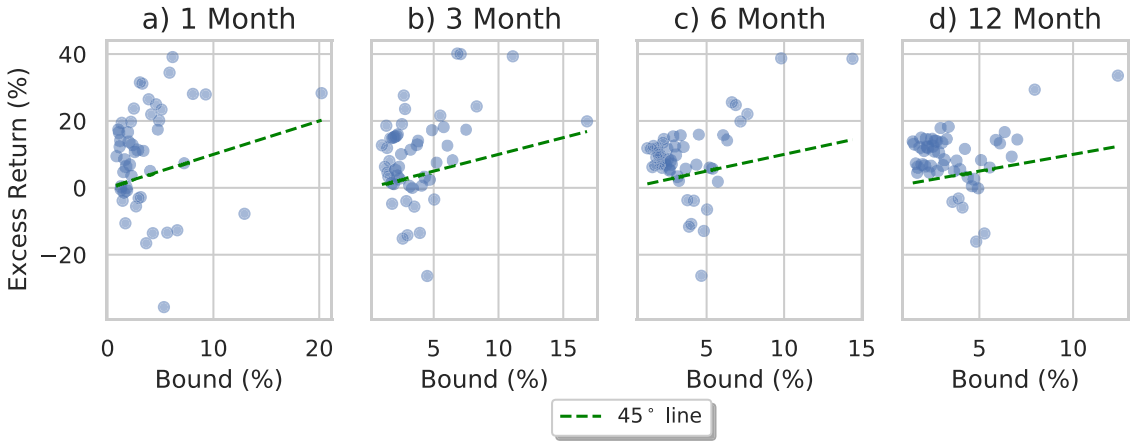

Fig. 3.3. Martin bound and subsequent excess returns.
The Martin bound is calculated at the end of each month for each horizon. The monthly bounds are sorted into 50 groups, and the mean bound and mean subsequent excess return are computed within each group. The green dashed line is the 45° line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

98.9% at the 12-month horizon. As Fig. 3.2 demonstrates, in our sample, the Chabi-Yo/Loudis bound is a multiple greater than one of the Martin bound plus a small amount of noise.

Fig. 3.3 shows binned averages of the Martin bound and subsequent excess returns. A majority of the points in Fig. 3.3 lie above the 45° line, consistent with the bound being a lower bound. Surprisingly, for the longer horizon returns, realized slackness takes some of its largest values when the bound is also high. Due to the high correlation of the Martin and Chabi-Yo/Loudis bounds, the figure is very similar for the Chabi-Yo/Loudis bound, though due to the Chabi-Yo/Loudis bound being generally larger than the Martin bound, fewer points lie above the 45° line for the Chabi-Yo/Loudis bound (see Figure IA.1 in the internet appendix).

3.2. Stock bounds and returns

Our stock panel runs from January 1996 to December 2020 and consists of S&P 500 constituent stocks satisfying the filters discussed in Appendix C. Table 2 reports summary statistics for our stock panel. The Martin–Wagner bound reported in the table is their Eq. (17), which is an exact formula for stock risk premia when the Martin bound is tight (and when the stock fixed effect in Martin and Wagner's Eq. (15) is zero) and a lower bound when the Martin bound is slack; we refer to this as the 'Martin–Wagner bound.' We call the Martin (2017) formula applied to individual stocks the 'Kadan–Tang bound.' The table reports statistics for the full sample as well as for subsamples based on Kadan and Tang's δ parameter. The δ groups are Conservative ($\delta \leq 3$, which is the union of Kadan and

Table 2
Stock panel summary statistics.
The panel contains monthly observations for S&P 500 constituent stocks with traded options passing the filters described in Appendix C from January 1996 through December 2020. The panel contains 1111 distinct stocks. The maximum (minimum) number of distinct stocks in a month is 504 (476). Returns and bounds are annualized and expressed in percent. Martin–Wagner Bound denotes the stock bound from Martin and Wagner (2019) under the assumption that the Martin (2017) market bound is tight. Kadan–Tang Bound denotes the Kadan and Tang (2020) stock bound. Statistics are presented for stock-months binned into δ ranges (δ is the ratio of beta to R^2 from the stock's market-model regression). The δ groups are Conservative ($\delta \leq 3$), Liberal ($3 < \delta \leq 7$) and Other ($\delta > 7$).

	Mean	SD	P10	P25	P50	P75	P90	N
Panel A. 1-month Horizon								
Excess Return								
All	9.76	122.63	−119.71	−49.63	10.84	68.94	134.63	146,922
Conservative	15.11	97.11	−92.93	−35.81	15.51	65.57	121.10	57,638
Liberal	6.59	125.75	−128.28	−56.03	8.19	69.48	136.80	65,814
Other	5.54	162.87	−164.13	−73.07	3.57	79.00	169.09	23,470
Martin–Wagner Bound								
All	5.05	7.99	0.49	1.38	2.90	5.75	11.12	146,922
Conservative	4.17	5.64	0.46	1.22	2.55	4.95	9.19	57,638
Liberal	4.98	8.25	0.55	1.42	2.88	5.60	10.65	65,814
Other	7.38	11.13	0.44	1.92	4.23	8.71	16.87	23,470
Kadan–Tang Bound								
All	12.40	15.93	3.34	4.94	7.98	13.79	24.36	146,922
Conservative	9.47	10.36	2.87	4.17	6.53	10.70	18.19	57,638
Liberal	12.53	16.29	3.65	5.22	8.24	13.89	23.56	65,814
Other	19.25	22.64	4.82	7.55	12.52	22.17	39.26	23,470
Panel B. 3-month Horizon								
Excess Return								
All	9.58	70.11	−67.22	−25.90	10.21	44.60	82.24	145,176
Conservative	14.09	54.75	−48.26	−15.60	14.01	43.45	74.28	56,665
Liberal	6.93	72.62	−74.04	−30.58	7.86	44.57	83.72	65,360
Other	5.97	92.27	−94.19	−41.62	4.07	49.40	103.14	23,151
Martin–Wagner Bound								
All	4.74	6.91	0.74	1.52	2.88	5.42	10.05	145,176
Conservative	3.86	4.41	0.75	1.39	2.54	4.69	8.25	56,665
Liberal	4.68	6.94	0.75	1.56	2.88	5.30	9.73	65,360
Other	7.09	10.44	0.63	1.96	4.10	8.18	16.14	23,151
Kadan–Tang Bound								
All	11.19	13.95	3.22	4.58	7.26	12.44	21.63	145,176
Conservative	8.34	8.22	2.80	3.89	5.88	9.58	15.83	56,665
Liberal	11.29	13.81	3.51	4.86	7.52	12.57	21.19	65,360
Other	17.88	21.28	4.53	6.96	11.50	20.22	36.92	23,151
Panel C. 6-month Horizon								
Excess Return								
All	9.42	50.63	−46.14	−16.87	9.17	34.10	61.75	142,550
Conservative	13.05	40.63	−31.99	−9.00	12.51	33.90	57.30	55,203
Liberal	7.29	52.68	−51.16	−20.86	6.89	33.74	62.18	64,611
Other	6.68	64.26	−63.33	−28.19	4.26	36.03	74.49	22,736
Martin–Wagner Bound								
All	4.61	6.37	0.91	1.67	2.93	5.25	9.42	142,550
Conservative	3.72	3.83	0.94	1.57	2.59	4.52	7.58	55,203
Liberal	4.55	6.26	0.92	1.70	2.95	5.19	9.11	64,611
Other	6.97	10.01	0.76	2.03	4.10	8.03	15.91	22,736
Kadan–Tang Bound								
All	10.68	13.07	3.20	4.50	7.00	11.84	20.40	142,550
Conservative	7.85	7.36	2.81	3.84	5.68	9.03	14.55	55,203
Liberal	10.75	12.65	3.47	4.78	7.28	12.01	20.02	64,611
Other	17.32	20.53	4.41	6.78	11.16	19.51	35.83	22,736
Panel D. 12-month Horizon								
Excess Return								
All	9.23	37.26	−32.05	−11.50	8.02	26.93	47.76	137,230
Conservative	12.31	28.94	−21.12	−3.82	11.78	27.35	44.39	52,684
Liberal	7.44	39.02	−35.70	−15.06	5.53	26.22	48.47	62,593
Other	6.90	47.87	−44.06	−20.63	2.25	27.71	57.24	21,953
Martin–Wagner Bound								
All	4.50	6.04	1.00	1.75	2.92	5.08	8.99	137,230
Conservative	3.50	3.39	1.04	1.65	2.55	4.18	6.83	52,684
Liberal	4.45	5.82	1.00	1.77	2.99	5.12	8.83	62,593
Other	7.07	9.74	0.86	2.12	4.16	8.13	16.27	21,953
Kadan–Tang Bound								
All	10.51	12.57	3.27	4.55	6.94	11.61	19.87	137,230
Conservative	7.45	6.54	2.90	3.89	5.58	8.57	13.35	52,684
Liberal	10.60	11.90	3.54	4.86	7.30	11.97	19.70	62,593
Other	17.57	20.20	4.49	6.85	11.22	19.78	36.95	21,953

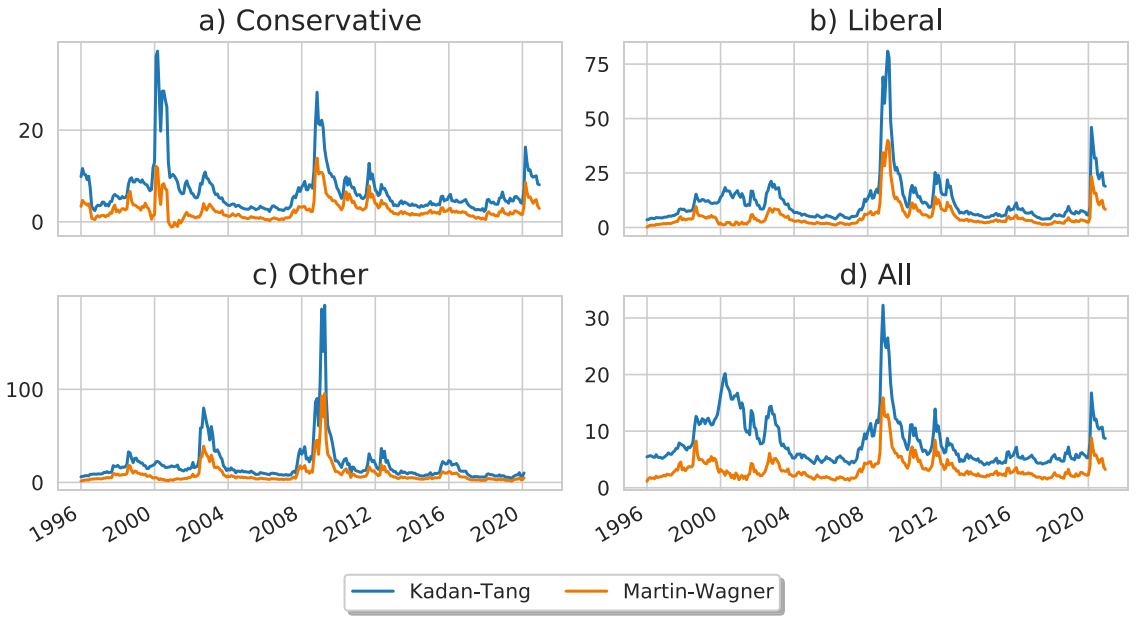

Fig. 3.4. Stock bounds by δ.
The median Kadan–Tang and Martin–Wagner bounds at the 6-month horizon are shown (in % per year) in each δ group and for the full sample.

Tang's Very Conservative and Conservative groups), Liberal ($3 \leq \delta \leq 7$, which is the union of Kadan and Tang's Moderate, Liberal, and Very Liberal groups) and Other ($\delta > 7$). Of the approximately 150,000 stock-months in the panel, about 40% fall in the Conservative group, 45% belong to the Liberal group, and the last 15% are in the Other group, as reported in the table. The table shows that the average and median levels of both bounds are increasing across δ groups. There are also clear patterns for subsequent excess returns as a function of δ: average and median excess returns are inversely related to δ, and the standard deviation of excess returns is directly related to δ. For the Conservative group, the average slackness of the Martin–Wagner bound is about 9–11%, while it is 5–6% for the Kadan–Tang bound. For the highest δ group, the average Kadan–Tang bound is 11–14% *greater* than the average realized excess return, consistent with the Kadan–Tang bound being an upper bound for higher levels of δ. For the full sample and all horizons, the Martin–Wagner bound is on average about half of the realized excess return, suggesting it may be a slack lower bound for risk premia. Without conditioning on δ, the Kadan–Tang bound is on average larger than the subsequent realized excess return.

Fig. 3.4 plots the time series of the median 6-month Kadan–Tang and Martin–Wagner bounds for the full sample and within each of the three δ groups. The plots for the other horizons look similar. The figure shows that the median Kadan–Tang bound is higher than the median Martin–Wagner bound in each δ group at each point in time. This is true at the 1, 3, and 12-month horizons also. As discussed in Section 1, the Kadan–Tang bound is twice the Martin–Wagner bound plus a number that varies over time but is constant across stocks at each date. Fig. 3.5 shows how this number varies over time. It is usually positive, implying that the Kadan–Tang bound is more than twice the Martin–Wagner bound for every stock.

Fig. 3.6 shows average excess stock returns in percentiles of the Martin–Wagner bound. Most points lie above the 45° line, consistent with the bound being a lower bound, particularly for horizons greater than 1 month. Fig. 3.7 shows average excess stock returns in percentiles of the Kadan–Tang bound for the Conservative group of stocks. Here also, most points lie above the 45° line, consistent with the bound being a lower bound for this group of stocks. Fig. 3.8 is the same plot for the Liberal group. For this group, most of the points are below the 45° line, which is inconsistent with it being a lower bound. The plot for the Other group, which is not presented, is a more extreme version of the plot for the Liberal group: almost all of the points lie below the 45° line. These figures are consistent with Kadan and Tang's result that the bound should be a lower bound for low δ and an upper bound for high δ. They are also consistent with empirical results presented in Section 4.2.

4. Validity and tightnesss

We report the tests described in Section 2 for the market bounds and for the stock-level bounds. For the conditioning variables, we use positive versions of variables from Welch and Goyal (2008).[4] Our general conclusion is that we can reject tightness but not validity, though the

[4] Our results are qualitatively unchanged if we include the bounds themselves as a conditioning variable in addition to the positive versions of variables from Welch and Goyal (2008). These results are reported in the internet appendix (Tables IA.1 and IA.2 for the market and stock-level bounds, respectively).

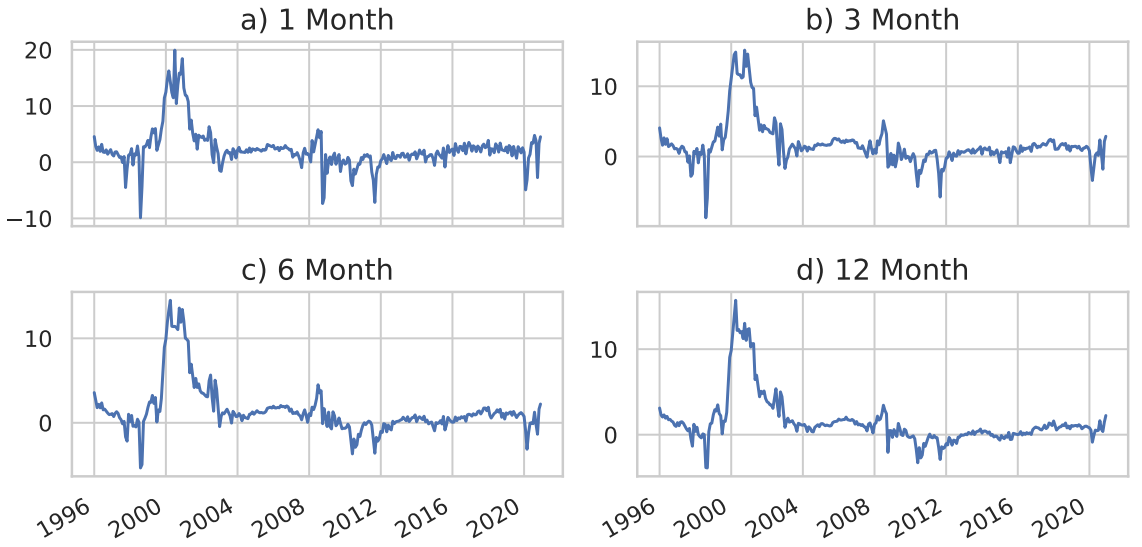

Fig. 3.5. Comparison of Kadan–Tang and Martin–Wagner bounds.
For each horizon and at each point in time, the Kadan–Tang bound minus twice the Martin–Wagner bound is constant across stocks. The plots are of the time series KT $-$ 2 $\times$ MW in percent per year.

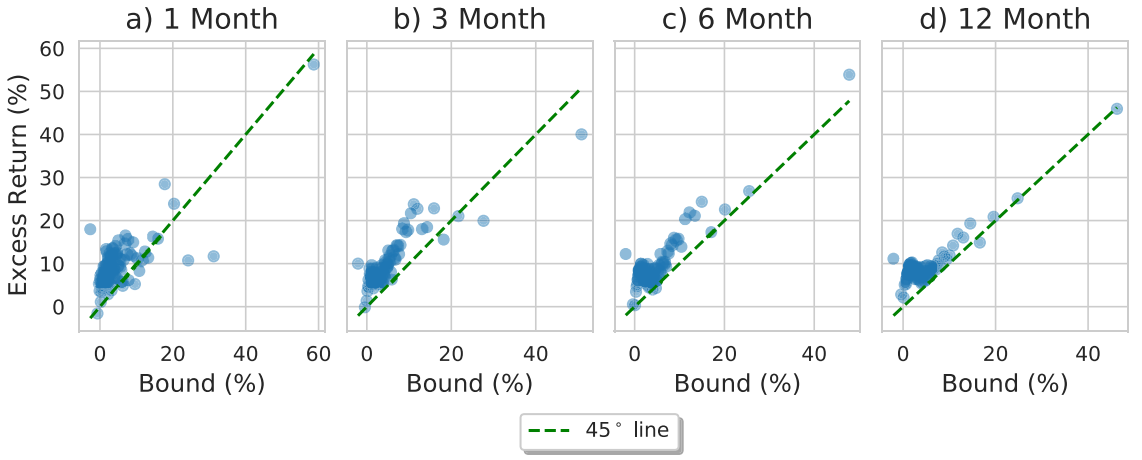

Fig. 3.6. Martin–Wagner bound and subsequent stock returns.
The Martin–Wagner bound for each horizon is sorted into percentiles, and the mean bound and mean subsequent excess return are computed within each percentile. The green dashed line is the 45° line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

conclusion varies by δ group for the Kadan–Tang bound. We show in Section 4.3 that the results for validity are driven more by average realized slackness than by the predictive power of the bounds.

4.1. Validity and tightnesss of market bounds

Table 3 reports the tests for the market bounds for the original sample periods of Martin and Chabi-Yo and Loudis and also for our extended 30-year sample. We report asymptotic p-values and also finite sample p-values based on the simulation analysis described in Section 2. We conduct all tests with daily data and overlapping return periods, using the Hansen and Hodrick (1980) covariance matrix estimator with lags equal to the number of days in the horizon.[5]

In the extended sample, the sample moments are all positive, so the statistic D_1 described in Section 2 is zero, and the null of validity is certainly not rejected. In the original samples, the statistic is positive but small, well below the lower bound on the critical value for a 10% test size of 1.64 from Kodde and Palm (1986), so we do not

[5] We use the Newey and West (1987) estimator with lags equal to 1.5 times the number of days in the horizon when the Hansen and Hodrick (1980) covariance matrix is not positive semidefinite. This occurs for the moments of slackness interacted with conditioning variables—used in our tests of validity and tightness—at the 12-month horizon in our sample and at the 6 and 12 month horizons in the original samples.

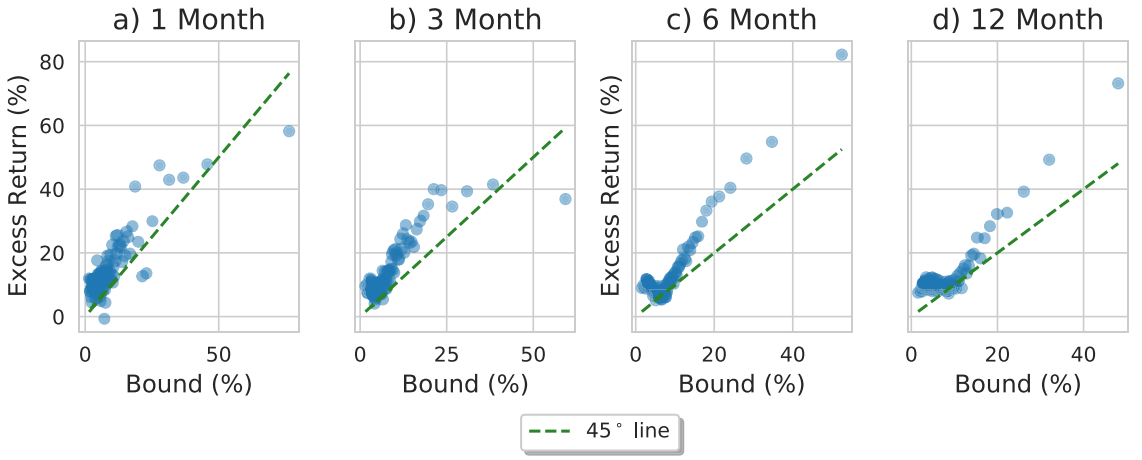

Fig. 3.7. Kadan–Tang bound and subsequent conservative stock returns.
Within the Conservative group, the Kadan–Tang bound for each horizon is sorted into percentiles, and the mean bound and mean subsequent excess return are computed within each percentile. The green dashed line is the 45° line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

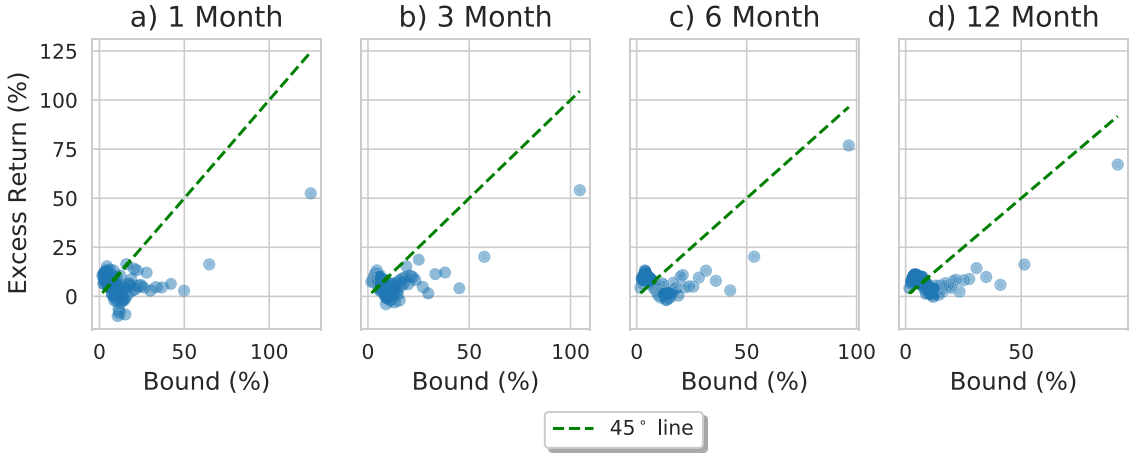

Fig. 3.8. Kadan–Tang bound and subsequent liberal stock returns.
Within the Liberal group, the Kadan–Tang bound for each horizon is sorted into percentiles, and the mean bound and mean subsequent excess return are computed within each percentile. The green dashed line is the 45° line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reject validity in the subsamples either. Inference is unchanged using finite-sample p-values.

The statistic D_2 provides a test of tightness against the alternative that the bound is valid but not tight. In all but one case, the statistic exceeds the upper bound on the critical value for a 0.1% test size of 27.13 from Kodde and Palm (1986). Thus, the asymptotic test strongly rejects tightness. Inference is generally unchanged using finite-sample p-values, with the exception of the Martin bound at the 1-month horizon and the Chabi-Yo and Loudis bound at the 12-month horizon in the original samples. In the extended sample, we can reject tightness at the 10% level for both bounds at all horizons using finite sample inference.

Table 3 also reports p-values from F-tests of the null hypothesis that $\alpha_h = 0$ and $\beta_h = 1$ in predictive regression (18). In both the extended data and the original samples, we cannot reject the null hypothesis that the bounds are tight using this regression-based test. As mentioned earlier, Martin (2017) and Chabi-Yo and Loudis (2020) conduct the same test and reach the same conclusion. The rejection of tightness using conditional tests when the unconditional F-test fails to reject is consistent with the former having greater power.

4.2. Validity and tightnesss of stock bounds

We follow the same procedure that we use for the market risk premium but averaging across stocks and using monthly rather than daily data. We compute the excess return of each stock and the excess return interacted with the Goyal–Welch variables and compute the sample means, averaging over dates and stocks for each variable. That is, we calculate

$$\bar{\lambda}_h = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}(R^e_{i,t,t+h} - b_{i,t,t+h})z_t$$

Table 3
Validity and tightness of market bounds.
The table reports tests statistics for bound validity and bound tightness for the Martin (2017) and Chabi-Yo and Loudis (2020) bounds in Panels A and B, respectively. Each panel reports results for the indicated original sample period as well as for our extended sample. The extended sample contains daily observations on the bounds from January 1990 through December 2020, and data on subsequent realized market excess returns runs from January 1990 through March 2021. The test statistics D_1 and D_2 are defined in Section 2. The sample moments are means of daily realized slackness and daily realized slackness interacted with each of the following variables from Welch and Goyal (2008): Dividend Price Ratio (defined as the difference between the log of dividends and the log of the S&P 500 index price level, plus 5 to ensure a positive conditioning variable); Earnings Price Ratio (defined as the difference between the log of earnings and the log of the S&P 500 index price level, plus 5 to ensure a positive conditioning variable); Book-to-Market Ratio (the ratio of book value to market value for the Dow Jones Industrial Average); T-bill Rate (the 3-month Treasury bill rate); 1 + Term Spread (defined as the difference between the long-term yield from Ibbotson's and the 3-month T-bill rate, plus 1 to ensure a positive conditioning variable); Credit Spread (defined as difference between BAA and AAA-rated corporate bond yields); Stock Variance (defined as the sum of squared daily returns on the S&P 500); 1 + Net Equity Issuance (the ratio of 12-month moving sums of net issues by NYSE-listed stocks to the total end-of-year market capitalizations of NYSE stocks, plus 1 to ensure a positive conditioning variable); 1 + Inflation (the Consumer Price Index, plus 1 to ensure a positive conditioning variable). The covariance matrix of the moments is calculated using the Hansen–Hodrick estimator with the number of lags equal to the number of days in the horizon (with a month defined as 21 days), unless the matrix is not positive-definite, in which case the Newey–West estimator with the number of lags equal to 1.5 times the number of days in the horizon is used. Asymptotic p-values (in percent notation) are based on the methodology described in Appendix A. Finite-sample p-values (in percent notation) are based on test statistics simulated under the null of a tight bound as detailed in Appendix B. The table also reports the p-value from an F-test of the null hypothesis that $\alpha_h = 0$ and $\beta_h = 1$ in predictive regression (18).

	Panel A. Martin Bound			
Original Sample (Jan 1996–Jan 2012)				
	1	3	6	12
p(F test of $\alpha = 0, \beta = 1$)	96.5	99.5	38.0	74.3
D_1 (Validity)	0.61	0.32	0.00	0.01
p(D_1) - Asymptotic	39.4	48.5	65.4	61.1
p(D_1) - Finite Sample	33.9	46.8	62.7	61.3
D_2 (Tightness)	24.06	39.11	38.71	69.30
p(D_2) - Asymptotic	0.5	0.0	0.0	0.0
p(D_2) - Finite Sample	10.7	0.4	7.5	8.6
Extended Sample (Jan 1990–Dec 2020)				
p(F test of $\alpha = 0, \beta = 1$)	17.9	16.9	23.8	23.4
D_1 (Validity)	0.00	0.00	0.00	0.00
p(D_1) - Asymptotic	100.0	100.0	100.0	100.0
p(D_1) - Finite Sample	100.0	100.0	100.0	100.0
D_2 (Tightness)	29.93	39.51	49.28	71.21
p(D_2) - Asymptotic	0.0	0.0	0.0	0.0
p(D_2) - Finite Sample	3.6	0.4	2.6	8.3
Panel B. Chabi-Yo/Loudis Bound				
Original Sample (Jan 1996–Aug 2015)				
p(F test of $\alpha = 0, \beta = 1$)	55.7	81.1	17.7	68.3
D_1 (Validity)	0.52	0.42	0.04	0.01
p(D_1) - Asymptotic	42.4	45.5	59.6	62.8
p(D_1) - Finite Sample	37.2	43.7	57.3	63.4
D_2 (Tightness)	31.15	77.27	40.32	40.21
p(D_2) - Asymptotic	0.0	0.0	0.0	0.0
p(D_2) - Finite Sample	2.9	0.0	6.5	18.7
Extended Sample (Jan 1990–Dec 2020)				
p(F test of $\alpha = 0, \beta = 1$)	23.8	29.0	44.1	47.4
D_1 (Validity)	0.00	0.00	0.00	0.00
p(D_1) - Asymptotic	100.0	100.0	100.0	100.0
p(D_1) - Finite Sample	100.0	100.0	100.0	100.0
D_2 (Tightness)	28.72	36.78	44.12	61.82
p(D_2) - Asymptotic	0.1	0.0	0.0	0.0
p(D_2) - Finite Sample	4.6	1.5	4.4	9.9

for each horizon h. We test the validity of the bounds by testing the null that the vector of population means is nonnegative against an unrestricted alternative. When we cannot reject validity of a bound, we test its tightness by testing the null that the vector of means is zero, against the alternative that it is nonnegative. We calculate the covariance matrix of the moments following a block bootstrap approach similar to that of Martin and Wagner (2019). See Appendix A for details.

Table 4 reports the results of the validity and tightness tests. We cannot reject validity of the Martin–Wagner bound, but we do reject tightness. We reach the same conclusion for the Kadan–Tang bound for the Conservative group. As a lower bound, the Kadan–Tang bound is rejected for the Other group and rejected or nearly rejected for the Liberal group, depending on the horizon. On the other hand, as an upper bound, the Kadan–Tang bound is accepted as valid and rejected as tight for the Liberal

Table 4
Validity and tightness of stock bounds.
Bound validity and tightness are tested for the Martin and Wagner (2019) bounds for the full sample and for the Kadan and Tang (2020) bounds for the indicated subsamples based on δ. The test statistics D_1 and D_2 for testing lower bound validity and tightness are defined in Section 2, and the test statistics D_3 and D_4 for testing upper bound validity and tightness are defined in Section 4. The table reports p-values for each statistic in percent notation. p-values are based on the methodology described in Appendix A. The sample moments are means of monthly realized slackness and monthly realized slackness interacted with each of the Welch and Goyal (2008) variables described in Table 3. The covariance matrix of the moments is calculated following the block bootstrap procedure described in Appendix A. For the Kadan/Tang bound, p-values are presented for stocks binned into δ ranges (δ is the ratio of beta to R^2 from the stock's market-model regression). The δ groups are Conservative ($\delta \leq 3$), Liberal ($3 < \delta \leq 7$) and Other ($\delta > 7$). The statistics are based on monthly observations for S&P 500 constituent stocks with options passing the filters described in Appendix C from January 1996 through December 2020.

	1	3	6	12
Panel A. Lower Bound Validity Tests				
Martin–Wagner (All)	100.0	100.0	100.0	100.0
Kadan–Tang				
Conservative	63.8	57.7	30.5	25.6
Liberal	6.0	11.8	13.2	12.5
Other	0.2	0.1	0.2	1.2
Panel B. Upper Bound Validity Tests				
Martin–Wagner (All)	14.2	6.2	4.0	3.3
Kadan–Tang				
Conservative	20.5	11.1	6.9	3.8
Liberal	100.0	100.0	100.0	100.0
Other	100.0	100.0	100.0	100.0
Panel C. Lower Bound Tightness Tests				
Martin–Wagner (All)	0.2	0.0	0.0	0.0
Kadan–Tang (Conservative)	0.0	0.0	0.0	0.0
Panel D. Upper Bound Tightness Tests				
Kadan–Tang (Liberal)	2.3	0.1	0.0	0.2
Kadan–Tang (Other)	4.9	1.6	3.9	1.2

and Other groups.[6] So, for the Liberal and Other groups, the results are consistent with the Kadan–Tang bound being a slack upper bound. These results for the Kadan–Tang bound for the different δ groups are consistent with Kadan and Tang's analysis, which shows that the bound should be a lower bound for low δ and an upper bound for high δ. They are also consistent with Figs. 3.7 and 3.8 discussed in Section 3.2. In the remainder of the paper, we analyze the Kadan–Tang bound exclusively for the Conservative group.

4.3. What drives the validity results?

Our conditional tests do not reject validity of the lower bounds on risk premia (with the exception of the Kadan–Tang bound for high δ groups). The conditions (19) show that validity depends on covariances of realized slackness with conditioning variables not being 'too negative,' relative to unconditional mean slackness. It is an interesting question whether this occurs in the data because the bounds are good predictors of excess returns, so covariances of the predictive variables with slackness are small in absolute value (for example, the covariances would be zero if the bounds were the best available predictors), or whether it occurs simply because unconditional mean slackness is large. Our conclusion is that the latter explanation is the correct one. We determine this by running the same tests but replacing each bound by its time-series mean, so it has zero predictive power for subsequent returns.[7] We cannot reject validity of the Martin bound, the Chabi-Yo/Loudis bound, the Martin–Wagner bound, or the Kadan–Tang bound (for the Conservative group) when we replace the bound by its time-series mean. The results are tabulated in the internet appendix (Tables IA.3 and IA.4). Thus, we conclude that the failure to reject validity is due to high average slackness rather than to any predictive power of the bounds. This does not mean that the bounds lack predictive power; it simply means that any predictive power is irrelevant for the validity results. The remainder of the paper analyzes the predictive power of the bounds.

5. Full-sample estimation

The previous section shows that the bounds appear to be valid but slack. This, of course, raises the question of how much information the bounds contain for forecasting returns. This section reports in-sample analyses of this question. We find that the market bounds are correlated with subsequent returns, at least when we control for other standard predictors. For the stock bounds, the correlation with subsequent returns is primarily time-series correlation rather than cross-sectional correlation.

5.1. Full-sample estimation for market bounds

Table 5 reports predictive regressions of the market excess return on the Martin and Chabi-Yo/Loudis bounds. We employ the augmented regression method of Amihud and Hurvich (2004) to adjust point estimates and standard errors for potential small-sample bias (Stambaugh, 1999). We also report p-values bootstrapped under the null of no-predictability as detailed in Appendix B.3. The point estimate of the slope coefficient is positive for all bounds and horizons, but the standard errors are relatively large. This is consistent with evidence presented by Martin and by Chabi-Yo and Loudis. Those authors emphasize that we cannot reject that the slope coefficient is 1 and the intercept is 0, which is true in our data too. However, we also cannot reject that the slope coefficient is 0 in five of the eight regressions using Amihud–Hurvich adjusted standard errors. Inference is similar using using bootstrapped

[6] Extending the analysis presented in Section 1, we define D_3 as the squared distance in the norm defined by the inverse covariance matrix of the sample moment vector from the *nonpositive* orthant, and we define $D_4 = D_0 - D_3$, which is the squared distance from the origin minus the squared distance from the nonpositive orthant. We use D_3 and D_4 in Kodde–Palm tests to test validity and tightness of each bound as an upper bound on stock risk premia.

[7] An essentially equivalent exercise would be to reshuffle the bound in time so that it becomes uncorrelated with the conditioning variables, i.e., converting the bound to pure noise with the same mean. Both exercises eliminate $\text{cov}(b_{t,T}, z_{jt})$ on the left-hand side of (19b) and leave the other terms in (19) unchanged. We thank a referee for suggesting these exercises.

Table 5
Regressions of market excess return on bounds.
Realized excess market returns (in percent) are regressed on a constant and the bound following the Amihud and Hurvich (2004) augmented regression method. Panel A shows predictions of market excess returns using the Martin (2017) bounds; Panel B uses the Chabi-Yo and Loudis (2020) bounds. The regressions use daily observations from January 1990 through December 2020. Standard errors (in parentheses) are calculated as in Amihud and Hurvich (2004), except for the use of Hansen-Hodrick standard errors (with the number of lags equal to the number of days in the return horizon (with a month defined as 21 days)) in the Amihud-Hurvich augmented regression. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. The bottom row of each panel reports the bootstrapped p-value for the bound coefficient. Data is simulated under the null of no predictability, and univariate predictive regressions (without the Amihud/Hurvich augmentation) are run for each simulated time-series. The bootstrapped p-value is the fraction of simulations with t-values that exceed the t-value for the bound coefficient (also in a non-augmented regression) in the data. See Appendix B.3 for details.

Panel A. Martin Bound				
	1	3	6	12
Constant	3.46	3.34	1.83	5.01
	(4.35)	(5.08)	(3.18)	(4.17)
Bound	1.36	1.39	1.78**	1.04
	(1.23)	(1.47)	(0.85)	(0.88)
N	7808	7806	7743	7617
Adj. R^2	0.011	0.022	0.047	0.020
Bootstrap p-value	0.178	0.402	0.113	0.258
Panel B. Chabi-Yo/Loudis Bound				
Constant	3.94	3.51	1.77	4.33
	(4.13)	(4.74)	(3.04)	(3.93)
Bound	1.07	1.09	1.36***	0.86**
	(0.99)	(1.09)	(0.53)	(0.43)
N	7808	7806	7743	7617
Adj. R^2	0.011	0.026	0.062	0.037
Bootstrap p-value	0.170	0.330	0.054	0.056

p-values—we reject that the slope coefficient is 0 in only two of the eight regressions (the Chabi-Yo/Loudis bound at 6 and 12 month horizons).

Table 5 shows that the correlations of the market bounds with subsequent returns are generally insignificant, but we obtain a different result when we control for other standard predictors. Table 6 reports regressions of market excess returns on the Martin bound and Goyal–Welch variables. The results are virtually identical for the Chabi-Yo/Loudis bound, due to the high correlation between the bounds; those results are provided in the internet appendix (Table IA.5). With controls for the Goyal–Welch variables, the bound is significant at all horizons using Amihud–Hurvich adjusted standard errors and at all but the 1-month horizon using bootstrapped p-values. The economic magnitude of the bound coefficient is large. A one-standard-deviation increase in the 12-month bound predicts an increase in the 12-month market excess return of over 6%. The magnitudes are even larger for shorter horizons (in annualized terms). Note that the significance of some of the Goyal–Welch variables in these regressions is further evidence that the bounds are not tight: if they were tight, then slackness would be unpredictable.

5.2. Full-sample estimation for stock bounds

We first look at cross-sectional predictability of returns by running Fama–MacBeth regressions. Panel A of Table 7 reports Fama–MacBeth regressions of excess returns on the Martin–Wagner bound. The mean slope coefficient for the full panel of stocks is negative and insignificant for each horizon. Regressions on the Kadan–Tang bound would produce the same results, except for a factor of 1/2, due to the perfect correlation between the bounds in each cross-section. When we consider only the Conservative group of stocks, the point estimate of the slope is positive at each horizon but still insignificant. We conclude from the Fama–MacBeth regressions that the bounds have little cross-sectional information about future excess returns.

To isolate time-series predictability, we run panel regressions of stock excess returns on the bounds with stock fixed effects and bootstrapped standard errors.[8] We consider the full panel of stocks for the Martin–Wagner bound and the Conservative group for the Kadan–Tang bound. These regressions, reported in Panel B of Table 7, show that there is time-series predictability. The slope estimates are uniformly positive and are strongly significant at the 6 and 12-month horizons.

To see if we can detect information in the bounds when time-series and cross-sectional variation are combined, we run panel regressions without fixed effects (Panel C). When we drop the fixed effects, the slope coefficient remains significant for the Kadan–Tang bound in the Conservative group at the longer horizons but is no longer significant for the Martin–Wagner bound. We conclude that the stock bounds have predictive power, and the predictive power comes from the time series rather than the cross section. In the internet appendix (Table IA.6), we present a version of Table 7 in which we include standard stock characteristics in the Fama–MacBeth and panel regressions: size, book-to-market, asset growth, operating profitability, and momentum.[9] The estimates and statistical significance of the bound coefficient are virtually unchanged when those characteristics are included.

6. Out-of-sample analysis

To see if the predictive power of the bounds can be detected out of sample, we calculate out-of-sample R^2s and conduct Diebold and Mariano (1995) tests, using the post-1926 expanding window market mean as the benchmark. We use the same benchmark for both market and stock-level forecasts. We have also examined using zero as a

[8] Martin and Wagner (2019) employ panel regressions of stock excess returns on a constant, $SVIX^2_{i,t,T}$, and $SVIX^2_{i,t,T} - \overline{SVIX^2}_{t,T}$, testing if the intercept equals 0 and the two slope coefficients equal 1 and 0.5, respectively, as in Eq. (5), among other tests. They report these tests using regressions both with and without firm fixed effects in their Tables IV and V. They cannot reject these null hypotheses. Our panel regressions in Panels B and C of Table 7 are similar to their specifications except that we estimate the predictive coefficient on the bound (5) rather than separately estimating coefficients for its components $SVIX^2_{i,t,T}$ and $SVIX^2_{i,t,T} - \overline{SVIX^2}_{t,T}$.

[9] We follow the definitions of these characteristics in Green et al. (2017) and produce them using the SAS code helpfully provided by Jeremiah Green on his website.

Table 6
Regressions of market excess return on Martin bound and predictive variables.
Realized excess market returns (in percent) are regressed on the Martin bound and variables from Welch and Goyal (2008) defined in Table 3. The regression follows the Amihud and Hurvich (2004) multipredictor augmented regression method, assuming AR(1) processes for each predictor variable. The sample consists of monthly observations from January 1990 through December 2020. The predictor variables (including the bound) are standardized to have zero means and unit variances. Standard errors (in parentheses) are calculated as in Amihud and Hurvich (2004), except for the use of Hansen–Hodrick standard errors (with the number of lags equal to the number of months in the horizon) in the Amihud–Hurvich augmented regression. Statistical significance is represented by * $p < 0.10$, *** $p < 0.05$, and *** $p < 0.01$. The bottom row reports the bootstrapped p-value for the bound coefficient. Data is simulated under the null of no predictability, and multivariate predictive regressions (without the Amihud/Hurvich augmentation) are run for each simulated time-series. The bootstrapped p-value is the fraction of simulations with t-values that exceed the t-value for the bound coefficient (also in a non-augmented regression) in the data. See Appendix B.3 for details.

	1	3	6	12
Constant	9.7***	9.4***	9.1***	9.4***
	(2.5)	(2.1)	(2.0)	(1.9)
Bound	36.4***	25.1***	15.0***	6.6***
	(9.3)	(4.4)	(3.2)	(2.4)
Div Price Ratio	18.7***	15.2***	12.6***	12.2***
	(5.4)	(4.5)	(4.6)	(3.7)
Earnings Price Ratio	3.8	3.3	1.9	5.1*
	(4.7)	(3.4)	(3.4)	(2.8)
Book-to-Market Ratio	−6.3	−1.9	−0.1	−4.1
	(5.4)	(5.0)	(4.9)	(4.1)
T-bill Rate	−17.4***	−13.0***	−12.0***	−9.7***
	(4.6)	(4.0)	(3.6)	(2.7)
Term Spread	−15.5***	−14.1***	−13.7***	−10.1***
	(5.5)	(5.1)	(4.9)	(3.3)
Credit Spread	−8.7	−8.4**	−4.9	1.2
	(6.2)	(4.2)	(4.0)	(2.9)
Stock Variance	−28.2*	−13.0	−4.2	0.8
	(14.6)	(10.5)	(3.5)	(3.4)
Net Eq Issuance	9.2*	9.2*	10.6**	9.6**
	(5.5)	(5.3)	(5.2)	(4.2)
Inflation	9.2	−1.5	−3.0	−1.2
	(6.7)	(3.8)	(3.0)	(1.5)
N	371	371	368	362
Adj. R^2	0.178	0.325	0.421	0.505
Bootstrap p-value	0.207	0.000	0.002	0.095

Table 7
Regressions of excess stock returns on bounds.
Realized stock returns in excess of the risk-free rate (in percent) are regressed on a constant and the indicated bound. The regressions use monthly observations from January 1996 through December 2020. Panels A reports Fama–MacBeth regressions using the Martin/Wagner bounds using all stocks and the Conservative subsample ($\delta \leq 3$). (We do not report Fama–MacBeth regressions for the Kadan/Tang bound because the Fama–MacBeth coefficient on the Kadan/Tang bound is half that of the Martin/Wagner bound by definition.) Standard errors of the time-series average cross-sectional regression coefficients are Hansen-Hodrick standard errors with the number of lags equal to the number of months in the return horizon. Panels B and C report panel regressions with and without stock fixed effects, respectively. For the Martin/Wagner bounds, the panel regressions contain all stocks; for the Kadan/Tang bound, the panel regression uses the Conservative subsample. Standard errors are computed based on the bootstrap procedure described in Appendix A. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Panel A. Fama–MacBeth Regressions				
	1	3	6	12
Martin–Wagner (All)	−0.17	−0.41	−0.41	−0.25
	(0.35)	(0.4)	(0.43)	(0.4)
Martin–Wagner (Conservative)	0.35	0.68	0.17	0.34
	(0.52)	(0.86)	(1.22)	(1.12)
Panel B. Panel Regressions with Stock Fixed Effects				
Martin–Wagner (All)	0.88	1.02*	1.31***	1.08***
	(0.66)	(0.61)	(0.5)	(0.42)
Kadan–Tang (Conservative)	0.67	0.65	1.24***	1.04***
	(0.66)	(0.74)	(0.44)	(0.35)
Panel C. Panel Regressions without Fixed Effects				
Martin–Wagner (All)	0.64	0.75	1.04	0.85
	(0.72)	(0.67)	(0.64)	(0.59)
Kadan–Tang (Conservative)	0.86	0.92	1.53***	1.27**
	(0.76)	(0.86)	(0.55)	(0.52)

benchmark for the stock forecasts, as in Gu et al. (2020) for example, but we find that in our sample the market mean is a more difficult benchmark to beat. We find that some models outperform the benchmark, but we do not find statistical significance.

6.1. Out-of-sample analysis of market bounds

The out-of-sample R^2 of a bound obviously depends on its bias (mean slackness). Even if the bias-adjusted bound were the best possible predictor of the excess return, the out-of-sample R^2 of the bound as a forecast will be negative if the mean slackness exceeds the standard deviation of the bound. To see this, let s denote mean slackness, and set $u_{t,T} = R^e_{t,T} - b_{t,T} - s$, which we assume is zero-mean and unpredictable. Denote the mean of the bound $b_{t,T}$ by b. Then, $\mathrm{E}[R^e_{t,T}] = b + s$, so, asymptotically, the benchmark forecast is approximately $b + s$, and its forecast error is approximately

$$R^e_{t,T} - (b+s) = (b_{t,T} + s + u_{t,T}) - (b+s) = b_{t,T} - b + u_{t,T}\,.$$

On the other hand, the forecast error of the bound is

$$R^e_{t,T} - b_{t,T} = (b_{t,T} + s + u_{t,T}) - b_{t,T} = s + u_{t,T}\,.$$

So, in this circumstance, the out-of-sample R^2 of the bound as a forecast is asymptotically positive if and only if the standard deviation of the bound is greater than its mean slackness.

In this model, adding past mean slackness to the bound produces a forecast with asymptotic forecast error equal to $u_{t,T}$, so the 'bound + mean slackness' forecast has a positive out-of-sample R^2 asymptotically. However, this forecast is likely to fail to beat the benchmark in our sample due to the brevity of the 30-year sample for estimating mean slackness and the much longer time series we use to estimate the mean market excess return. Section 6.2 presents simulations that show we would likely need 150 years of data before the 'bound + mean slackness' forecast would have a better than even chance of achieving a significantly positive out-of-sample R^2, given the 65 year head start that we have given the market mean.

Table 8 reports the out-of-sample R^2s for the market return forecasts. We forecast at the end of each month for the subsequent 1, 3, 6 and 12-month periods.[10] Row 0 of Panel A shows that the out-of-sample R^2 of the Martin bound is negative at every horizon.[11] Table 1 shows that the mean slackness of the Martin bound exceeds its standard deviation at every horizon, so, as discussed above, we would expect negative out-of-sample R^2s for the bound as a forecast even if the bias-adjusted bound were the best available predictor. Row 1 of Panel A shows that adding past mean slackness to the Martin bound produces a better forecast, which has a positive but insignificant out-of-sample R^2 at 3 and 6 month horizons. Fig. 6.1 shows the 3 and 12 month 'bound plus mean slackness' forecasts and their cumulative squared errors compared to the benchmark (as in Goyal and Welch, 2003). The plots for 1 and 6 months are similar. Positive slopes in Panels (b) and (d) indicate that the forecast is outperforming the benchmark, and negative slopes indicate underperformance. For both horizons, the forecast beat the benchmark in the late 1990s before underperforming around the turn of the century. Post 2002, the forecast and benchmark have performed similarly, aside from two notable episodes for the 3-month horizon: (1) the 3-month forecast substantially underperformed and then partially recovered in the financial crisis, and (2) the forecast substantially beat the benchmark in predicting the recovery from coronavirus-related market declines in 2020. The outperformance during the recovery from the coronavirus crisis resulted in the overall R^2 being positive at the 3-month horizon, as shown in Panel (b) of Fig. 6.1 and in Row 1 of Panel A of Table 8.

Row 2 of Panel A of Table 8 reports the performance of a forecast created by running OLS on the Martin bound in expanding windows. This forecast performs worse than using the bound as the forecast. Fig. 6.2 shows its performance at all horizons. The forecast has essentially no predictive power, and in fact the regression line of subsequent excess returns on the forecast has a negative slope in three of the four cases (regressing mean excess returns on the mean forecast in 50 bins). The poor performance of forecasts derived from running OLS on the bound is unsurprising, given the high standard errors of the bound coefficient in full sample regressions.

Rapach et al. (2010) show that an effective way to combine multiple predictors is to form individual forecasts based on the individual predictors and then average them. This is called a combination forecast. We create individual forecasts by running univariate regressions on the Goyal–Welch predictors and then average them (Row 3 of Panel A of Table 8). We also average (50–50) this combination forecast with the 'bound plus past mean slackness' forecast, thereby producing another combination forecast (Row 4 of Panel A of Table 8).[12] As Panel A shows, these forecasts based upon the Martin bound do not beat the benchmark. It is interesting, however, that the combination forecast is improved at every horizon by averaging it with the bound-based forecast, though the improvement is statistically insignificant in Diebold–Mariano tests (Row 5 of Panel A of Table 8).

Panel B of Table 8 repeats Panel A but using the Chabi-Yo/Loudis bound instead of the Martin bound. The results are very similar with the notable exception that using the

[10] We have overlapping returns and want to avoid look-ahead bias, so when we calculate past average slackness or run predictive regressions in expanding windows, we allow h months to elapse after the end of the window before using the results to form a forecast, where h months is the length of the return period.

[11] Martin (2017) reports positive out-of-sample R^2s in his sample period (1996–2012) but does not test significance. Our results differ primarily due to our extended sample period. In particular, Fig. 6.3 below shows that the benchmark forecast has generally beaten the bound as a forecast from 2012 to 2020. Our expanding market mean benchmark differs as well due to Martin's use of data going back to 1871.

[12] We also computed forecasts by running multivariate regressions on the Goyal–Welch variables and on the Goyal–Welch variables and the Martin (or Chabi-Yo/Loudis) bound, and we computed forecasts from those variables using partial least squares (Kelly and Pruitt, 2013). None of those forecasts ever beat the benchmark, so we did not tabulate the results.

Table 8
Out-of-sample tests of market forecasts.
Out-of-sample R^2s are computed for forecasts of the market excess return. Out-of-sample R^2 is defined as

$$R^2_{OOS} = 1 - \frac{\sum_t \varepsilon_t^2}{\sum_t \nu_t^2},$$

where ε_t is the error when a particular forecast is used and ν_t is the error when the benchmark forecast is used. The benchmark forecast for rows (0) to (4) is the expanding window average market excess return (using the Fama–French market excess return series starting in 1926). The benchmark forecast in the last row is the combination forecast (4) that uses only the Welch and Goyal (2008) predictors. Forecast (0) is simply the Martin (2017) or Chabi-Yo and Loudis (2020) bound. Forecast (1) is the bound + past average slackness. Forecast (2) is a linear models of excess returns regressed on a constant and the bound. Forecast (3) is a combination forecast of linear models; each month's forecast is the equal-weighted average of linear models using a single Welch and Goyal (2008) predictor. Forecast (4) is an equal-weighted average of forecast (3) and (1). Panels A and B report R^2s for these forecasts using the Martin (2017) and Chabi-Yo and Loudis (2020) bound, respectively. Panels C and D report R^2s using forecasts that are truncated below at either the bound (forecasts (2) and (4)) or at zero (forecast (3)). The expanding window regressions use monthly observations from January 1990 through December 2020. To ensure our results are not contaminated by look-ahead bias, we allow h months to elapse after the end of the window before using the results to form a forecast, for return horizon h. The initial estimation window uses the first 60 months of bound observations and the first 60+h months of realized returns. For each model, p-values for Diebold–Mariano tests are calculated using Hansen–Hodrick standard errors with the number of lags equal to the number of months in the return horizon and are reported in the internet appendix. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Positive R^2 values are shaded gray.

Panel A. Martin

	1	3	6	12
(0) Tight Bound	-0.012	-0.012	-0.002	-0.030
(1) Bound + Avg Slackness	-0.007	0.003	0.005	-0.071
(2) OLS on Bound	-0.032	-0.079	-0.037	-0.182
(3) Combination (GW)	-0.006	-0.036	-0.073	-0.121
(4) Combination (GW + Bound)	-0.003	-0.009	-0.023	-0.083
(5) (4) vs. (3)	0.004	0.026	0.047	0.034

Panel B. Chabi-Yo/Loudis

	1	3	6	12
(0) Tight Bound	-0.012	0.003	0.043	0.036
(1) Bound + Avg Slackness	-0.009	0.007	0.026	-0.058
(2) OLS on Bound	-0.034	-0.093	-0.015	-0.188
(3) Combination (GW)	-0.006	-0.036	-0.073	-0.121
(4) Combination (GW + Bound)	-0.002	-0.004	-0.006	-0.066
(5) (4) vs. (3)	0.004	0.031	0.062	0.050

Panel C. Martin with Truncation

	1	3	6	12
(0) Tight Bound	-0.012	-0.012	-0.002	-0.030
(1) Bound + Avg Slackness	-0.007	0.003	0.005	-0.071
(2) max(OLS on Bound,Bound)	-0.019	-0.046	-0.010	-0.066
(3) max(Combination (GW),Zero)	-0.006	-0.016	-0.025	-0.099
(4) max(Combination (GW + Bound),Bound)	-0.006	-0.000	0.014	-0.053
(5) (4) vs. (3)	-0.000	0.015	0.038	0.042

Panel D. Chabi-Yo/Loudis with Truncation

	1	3	6	12
(0) Tight Bound	-0.012	0.003	0.043	0.036
(1) Bound + Avg Slackness	-0.009	0.007	0.026	-0.058
(2) max(OLS on Bound,Bound)	-0.018	-0.049	0.017	-0.030
(3) max(Combination (GW),Zero)	-0.006	-0.016	-0.025	-0.099
(4) max(Combination (GW + Bound),Bound)	-0.008	0.004	0.043	-0.024
(5) (4) vs. (3)	-0.002	0.020	0.066	0.069

Chabi-Yo/Loudis bound as the forecast (Row 0 of Panel B of Table 8) beats the benchmark at all horizons other than 1 month (though the outperformance is insignificant).[13] The difference between the two bounds in this regard is not unexpected, given that Table 1 shows that the mean slackness of the Chabi-Yo/Loudis bound is less than that of the Martin bound at every horizon. Table 1 also

[13] Chabi-Yo and Loudis report positive out-of-sample R^2s for all horizons in their sample of 1996–2015 (their Table 7B) but do not test significance.

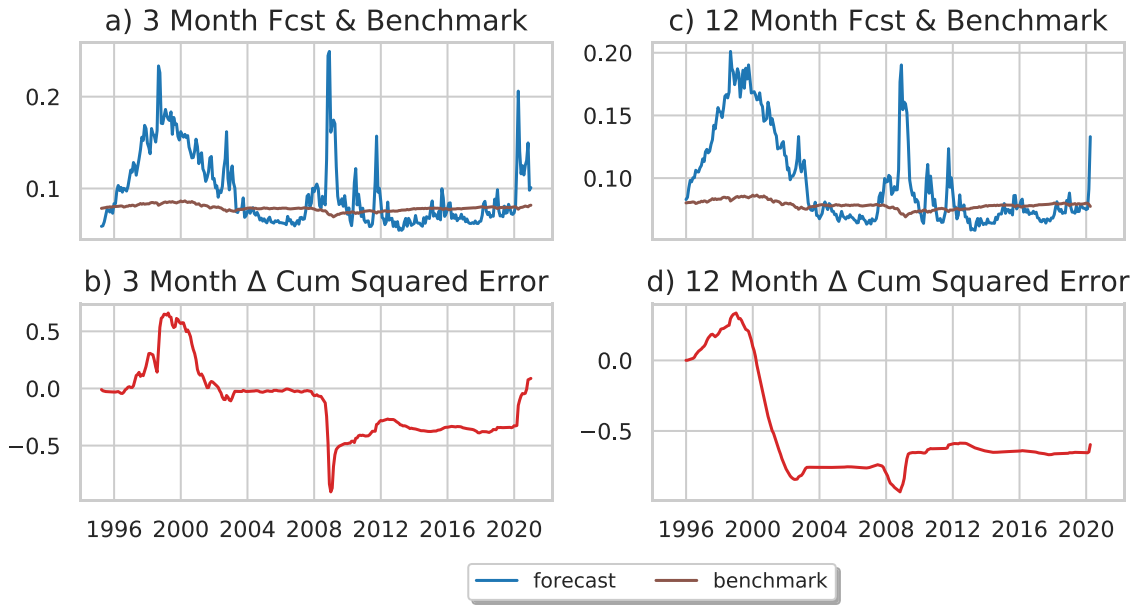

Fig. 6.1. Forecast and cumulative squared errors.
The market excess return at different horizons is forecast monthly as the Martin bound plus expanding-window mean slackness. Panels (a) and (c) show the forecast and the benchmark, which is the post-1926 expanding window market mean, at 3 and 12 month horizons. Panels (b) and (d) show the cumulative squared error of the benchmark minus the cumulative squared error of the bound-based forecast at 3 and 12 month horizons.

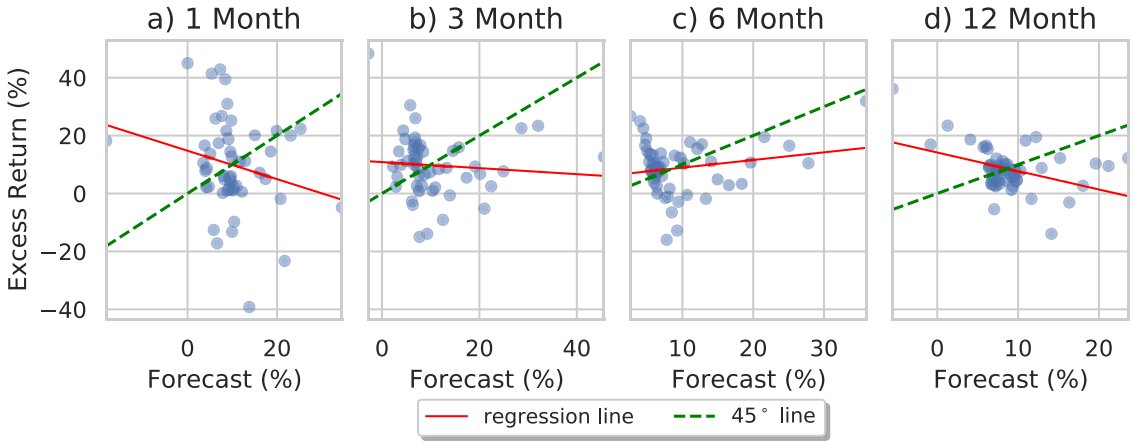

Fig. 6.2. OLS forecast from Martin bound and subsequent returns.
Forecasts of market excess returns at different horizons are computed monthly by running OLS on the Martin bound in expanding windows. The monthly forecasts are sorted into 50 groups, and the mean forecast and mean subsequent excess return are computed within each group. The red solid line is the regression line through the 50 points. The green dashed line is the 45° line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

shows that the volatility of the Chabi-Yo/Loudis bound is greater than or roughly equal to its mean slackness at every horizon, so, as discussed before, we would expect positive out-of-sample R^2s asymptotically for the bound as a forecast if the bias-adjusted bound were the best possible predictor. The difference in performance of the bounds as forecasts is shown in Fig. 6.3. The better performance of the Chabi-Yo/Loudis bounds at the longer horizons has been concentrated in the post-financial crisis period.

Campbell and Thompson (2008) show that market excess return forecasts can be improved by constraining the forecasts to be nonnegative. It is likewise reasonable to constrain forecasts based on a lower bound to be at least as large as the lower bound. Panels C and D of Table 8 repeat Panels A and B but using truncated forecasts. The combination forecast that uses only the Goyal–Welch variables is truncated at zero, and the other forecasts are truncated at the bound. Truncation improves the forecasts in almost every case. With truncation, the combination fore-

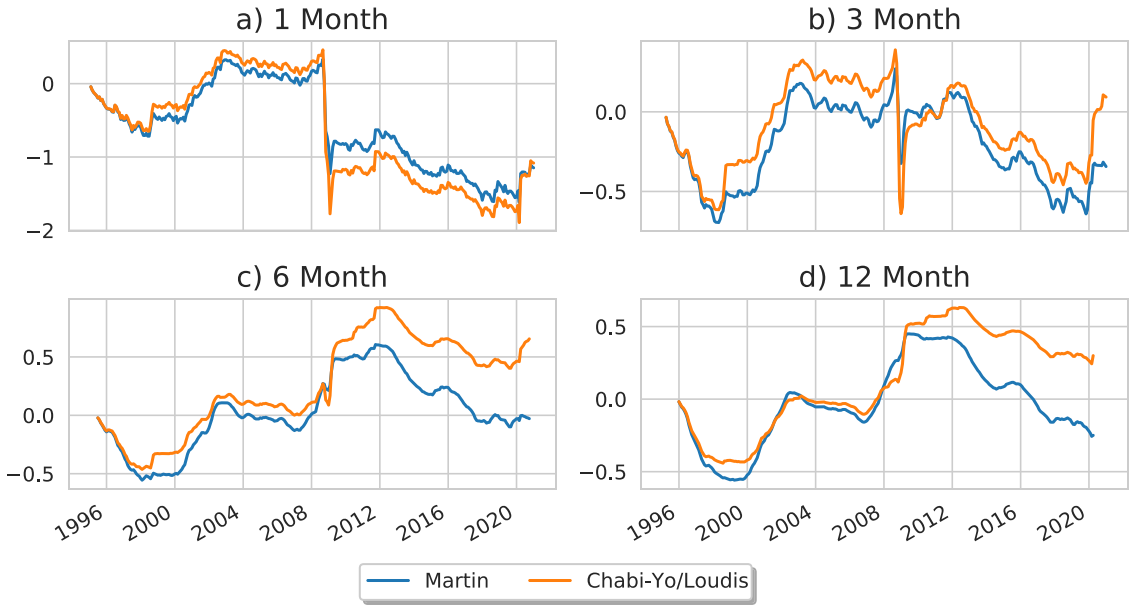

Fig. 6.3. Cumulative squared errors of market bounds.
For each horizon, the panels show the cumulative sum of squared errors of the benchmark minus the cumulative squared error of using the indicated bound as the forecast. The benchmark is the post-1926 expanding window market mean.

cast that averages the bound with the combination forecast from the Goyal–Welch variables (Row 4) outperforms the market at the 6-month horizon for both bounds and at the 3-month horizon for the Chabi-Yo/Loudis bound, but the improvement is insignificant.

An interesting fact shown in Table 8 is that none of the forecasts beats the benchmark at the 1-month horizon. Furthermore, only one of the forecasts (the Chabi-Yo/Loudis bound) beats the benchmark at the 12-month horizon. Predictability peaks at 6 months, with the Chabi-Yo/Loudis bound forecast reaching an out-of-sample R^2 of 4.3%.

6.2. How much data is needed to detect a market bound's forecasting power?

In this section, we use simulations as described in Appendix B.4 to determine how long a sample we would need to reliably detect that a bound or a bound plus past average slackness outperforms the expanding-window market mean when the bound is tight or when the bound is valid but slack. We look at both out-of-sample R^2s and p-values of the Diebold–Mariano test.

Fig. 6.4 plots distributions of simulated out-of-sample R^2s as a function of the sample length. The simulation is calibrated to the Martin bound at the 12-month horizon. As in our empirical setting, we use an expanding-window mean excess return as a benchmark forecast, and we assume that there are 65 years of realized returns prior to the first bound observation. The volatility of the bound in the simulation is the same as in the data (2.1% as shown in Table 1). When we simulate with a slack bound, we take slackness to be 5%. Panels (a) and (b) show that, with a 30-year sample, we expect (with a probability near 75%) to obtain a positive out-of-sample R^2 using the bound as the forecast if the bound is tight and (again with a probability near 75%) a negative out-of-sample R^2 if the bound is slack. Our empirical results of positive slackness and a negative out-of-sample R^2 are consistent with this. As the horizon lengthens, the simulation results in Panels (a) and (b) approach the asymptotic results that the bound is a better forecast than the benchmark if the bound is tight and a worse forecast if the bound is slack (given that slackness in the calibration is higher than the bound volatility).

Adding mean slackness to the bound to forecast will beat the market mean benchmark in the long run, under the hypotheses of our simulation. This can be seen in Panel (c). However, Panel (c) also shows that, with a 30-year sample, we expect (with a probability near 75%) to obtain a negative out-of-sample R^2 using 'bound + mean slackness' as the forecast. This is because errors in estimating mean slackness cause the 'bound + mean slackness' forecast to be inferior to the market mean (which has a 65 year head start) for shorter samples, even though it is superior in the long run. Panel (c) shows that we would need 150 years of data before the 'bound + mean slackness' forecast would generate a positive out-of-sample R^2 with 50% probability.

We would need even more than 150 years before the outperformance of the 'bound + mean slackness' forecast would be statistically significant. Fig. 6.5 plots distributions of p-values from Diebold–Mariano tests for the simulations. Panel (c) shows that we would need around 500 years of data before reaching a 50% probability of finding statistically significant outperformance by the 'bound + mean slackness' forecast. Even if the bound is tight and we use the bound as the forecast, we would need 150 years of data before reaching a 50% chance of finding statistically significant forecasting ability (Panel (a)).

The results differ somewhat if we calibrate the simulations to the Chabi-Yo/Loudis bound. Figs. 6.6 and 6.7

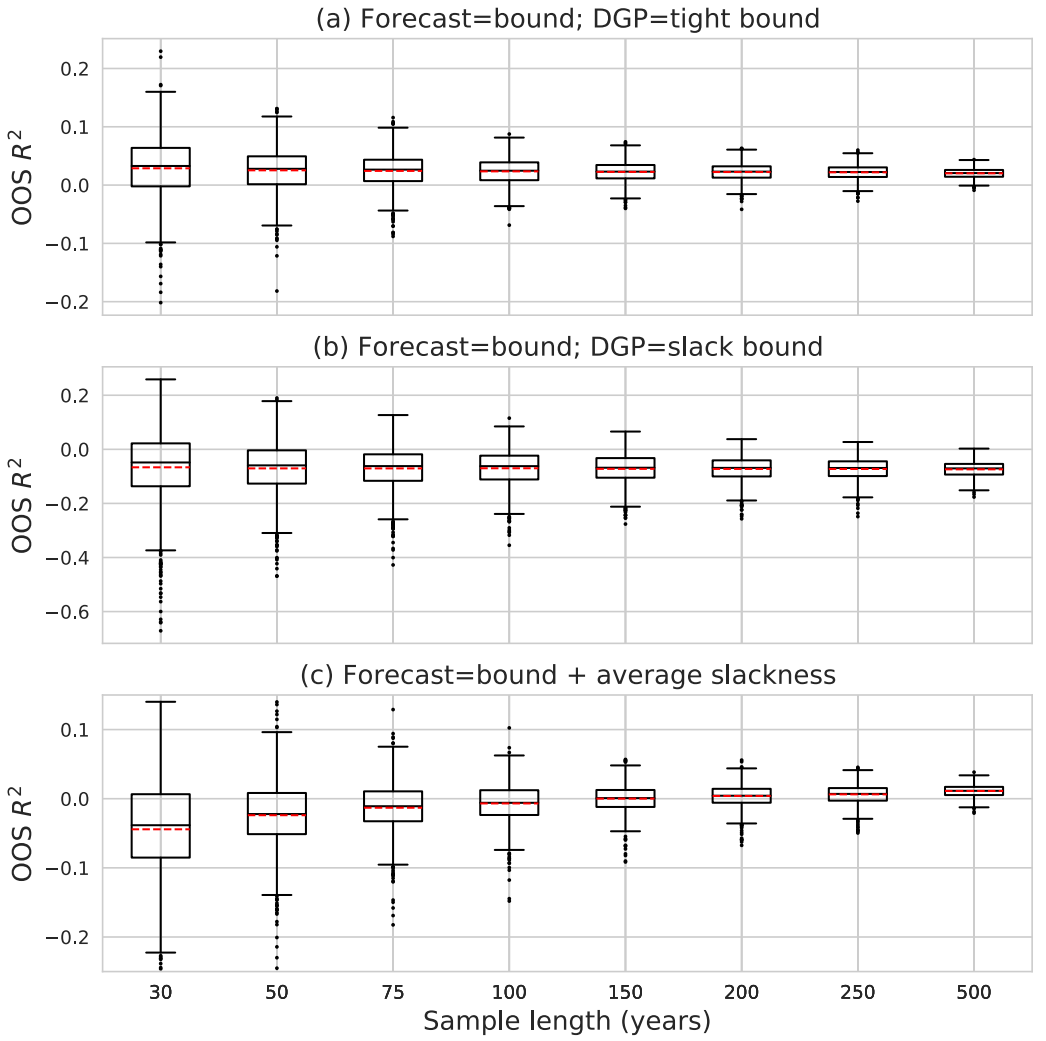

Fig. 6.4. Simulated out-of-sample R^2s: Martin calibration.
Each panel plots distributions of simulated out-of-sample R^2s for various sample lengths. The forecasts are either a tight bound (panels (a) and (b)) or the bound plus an expanding window average realized slackness (Panel (c)). The initial estimation window for average slackness is 60 months. The benchmark forecast is an expanding window mean excess return with 65 years of realized returns available prior to the first bound observation, as in our empirical setting. The simulations are calibrated using the 12-month Martin bound. For Panel (a), returns are simulated assuming the bound is tight. For Panel (b), returns are simulated assuming the bound is slack (5% per year). The results in Panel (c) are the same under either data-generating process, because any non-zero mean slackness contributes the same to the forecasts as it does to the returns. Some outliers in Panel (c) are suppressed for legibility. Each panel is based on 1000 simulations. The dashed red line represents the average out-of-sample R^2 across simulations. The solid black line within each box represents the median. The box represents the interquartile range, and the whiskers extend no more than 1.5 times the interquartile range from the edge of the box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plot distributions of simulated out-of-sample R^2s and p-values from Diebold–Mariano tests, respectively, for simulations calibrated to the 12-month Chabi-Yo/Loudis bound. The simulated bound has the same volatility as the 12-month Chabi-Yo/Loudis bound (3.47% as shown in Table 1). When we simulate with a slack bound, we take slackness to be 4%.

With this calibration and a slack bound, the out-of-sample R^2 of the bound is positive with about 50% probability with 30 years of data, though the probability decreases as the sample period increases (Panel b). The out-of-sample R^2 of the 'bound + mean slackness' forecast is also positive with about 50% probability with 30 years of data, and the probability increases as the sample period increases (Panel c). However, we would still need a fairly long sample before the outperformance of the 'bound + mean slackness' forecast would be statistically significant. Panel (c) of Fig. 6.7 shows that we would need between 150 and 200 years before reaching a 50% probability of finding statistically significant outperformance.

6.3. Out-of-sample analysis of stock bounds

As with the market bounds, we consider the stock bounds as forecasts and also examine OLS and combina-

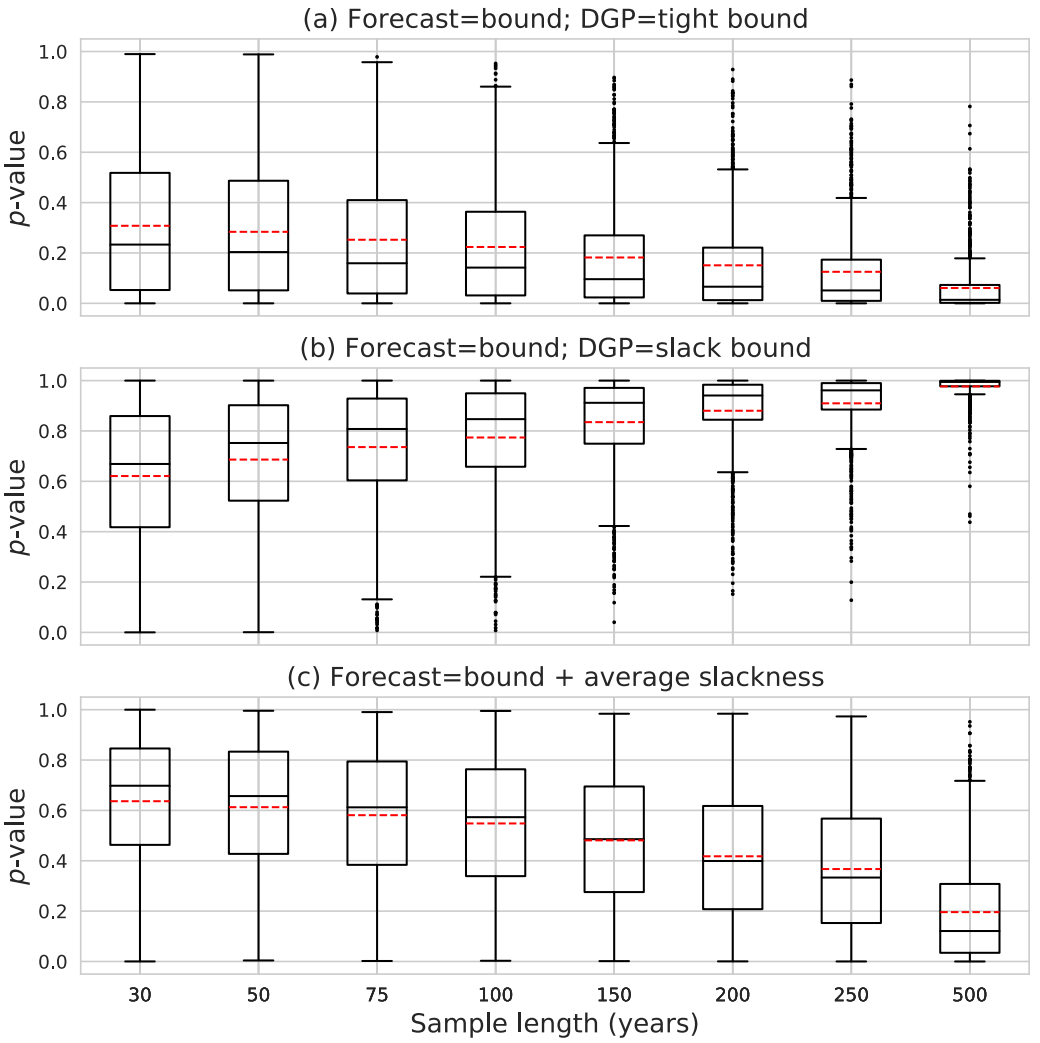

Fig. 6.5. Simulated Diebold–Mariano *p*-values: Martin calibration.
Each panel plots distributions of simulated Diebold–Mariano *p*-values for various sample lengths. The forecasts are either a tight bound (panels (a) and (b)) or the bound plus an expanding window average realized slackness (Panel (c)). The initial estimation window for average slackness is 60 months. The benchmark forecast is an expanding window mean excess return with 65 years of realized returns available prior to the first bound observation, as in our empirical setting. The simulations are calibrated using the 12-month Martin bound. For Panel (a), returns are simulated assuming the bound is tight. For Panel (b), returns are simulated assuming the bound is slack (5% per year). The results in Panel (c) are the same under either data-generating process, because any non-zero mean slackness contributes the same to the forecasts as it does to the returns. Each panel is based on 1000 simulations. The dashed red line represents the average *p*-value across simulations. The solid black line within each box represents the median. The box represents the interquartile range, and the whiskers extend no more than 1.5 times the interquartile range from the edge of the box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tion forecasts based on the bounds.[14] Motivated by the results in Section 5.2, we want to isolate time-series variation and cross-sectional variation. Let f_{it} denote the forecasted excess return of stock i at date t and, for the sake of brevity, let r_{it} denote the realized excess return denoted by $R^e_{i.t.T}$ before. Let $\bar{f}_t$ and $\bar{r}_t$ denote the cross-sectional means at date t. Given T time periods and N_t stocks in the cross section at date t, the mean squared forecast error for the panel is

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}(r_{it}-f_{it})^2\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}(r_{it}-\bar{r}_t+\bar{r}_t-\bar{f}_t+\bar{f}_t-f_{it})^2\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}(\bar{r}_t-\bar{f}_t)^2+\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i=1}^{N_t}\left[(r_{it}-\bar{r}_t)-(f_{it}-\bar{f}_t)\right]^2.
\end{aligned} \tag{20}$$

[14] We also investigated forecasting models based on multivariate regressions on stock characteristics, the Goyal–Welch variables, and the bounds, and we evaluated training a random forest on polynomials of those variables. However, the multivariate and random forest models never beat the benchmark.

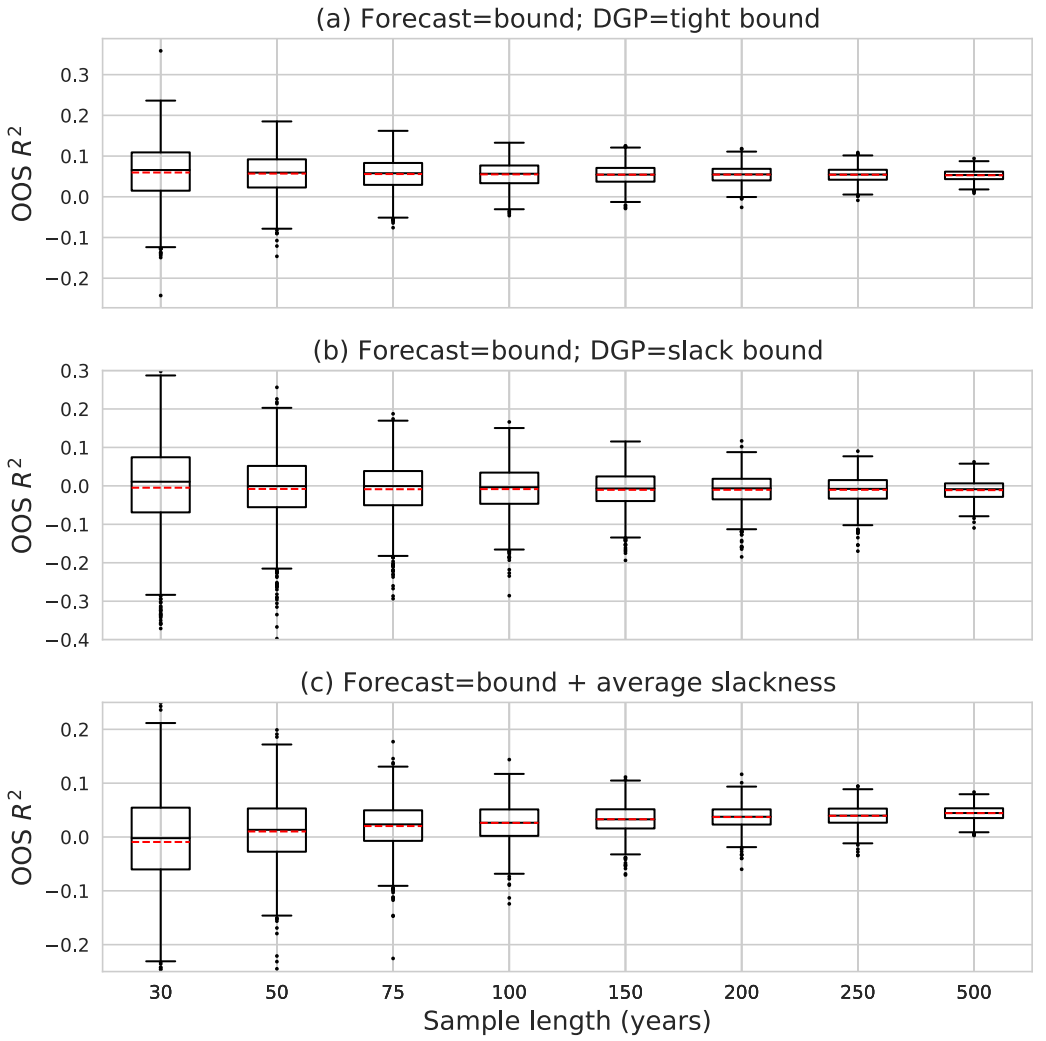

Fig. 6.6. Simulated out-of-sample R^2: Chabi-Yo/Loudis calibration.
Each panel plots distributions of simulated out-of-sample R^2s for various sample lengths. The forecasts are either a tight bound (panels (a) and (b)) or the bound plus an expanding window average realized slackness (Panel (c)). The initial estimation window for average slackness is 60 months. The benchmark forecast is an expanding window mean excess return with 65 years of realized returns available prior to the first bound observation, as in our empirical setting. The simulations are calibrated using the 12-month Chabi-Yo/Loudis bound. For Panel (a), returns are simulated assuming the bound is tight. For Panel (b), returns are simulated assuming the bound is slack (4% per year). The results in Panel (c) are the same under either data-generating process, because any non-zero mean slackness contributes the same to the forecasts as it does to the returns. Some outliers in Panels (b) and (c) are suppressed for legibility. Each panel is based on 1000 simulations. The dashed red line represents the average out-of-sample R^2 across simulations. The solid black line within each box represents the median. The box represents the interquartile range, and the whiskers extend no more than 1.5 times the interquartile range from the edge of the box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The first term on the right-hand side of (20) is a time-series mean squared error: it measures how well the average forecast each period predicts the average return each period. The second term measures how well the deviation of a forecast from the mean forecast predicts the deviation of a return from the mean return. Thus, the first term measures the time series information in the forecast, and the second term measures the cross-sectional information in the forecast. We compute out-of-sample R^2s for each of the two terms by comparing each term to the corresponding calculation based on using the benchmark as the forecast. At the end of this section, we also look at portfolio returns from sorting stocks on the bounds. Those results provide additional information about the predictive power of the bounds in the cross section.

Row 0 of Panel A of Table 9 shows that the Martin–Wagner bound outperforms the benchmark in both the cross section and the time series, except for the 3-month horizon for the cross section. However, the outperformance in the cross section is an order of magnitude smaller than the outperformance in the time series. The total R^2 is a blend of the time-series and cross-section R^2s, with more weight on the cross-section because most of the squared errors for the benchmark come from the

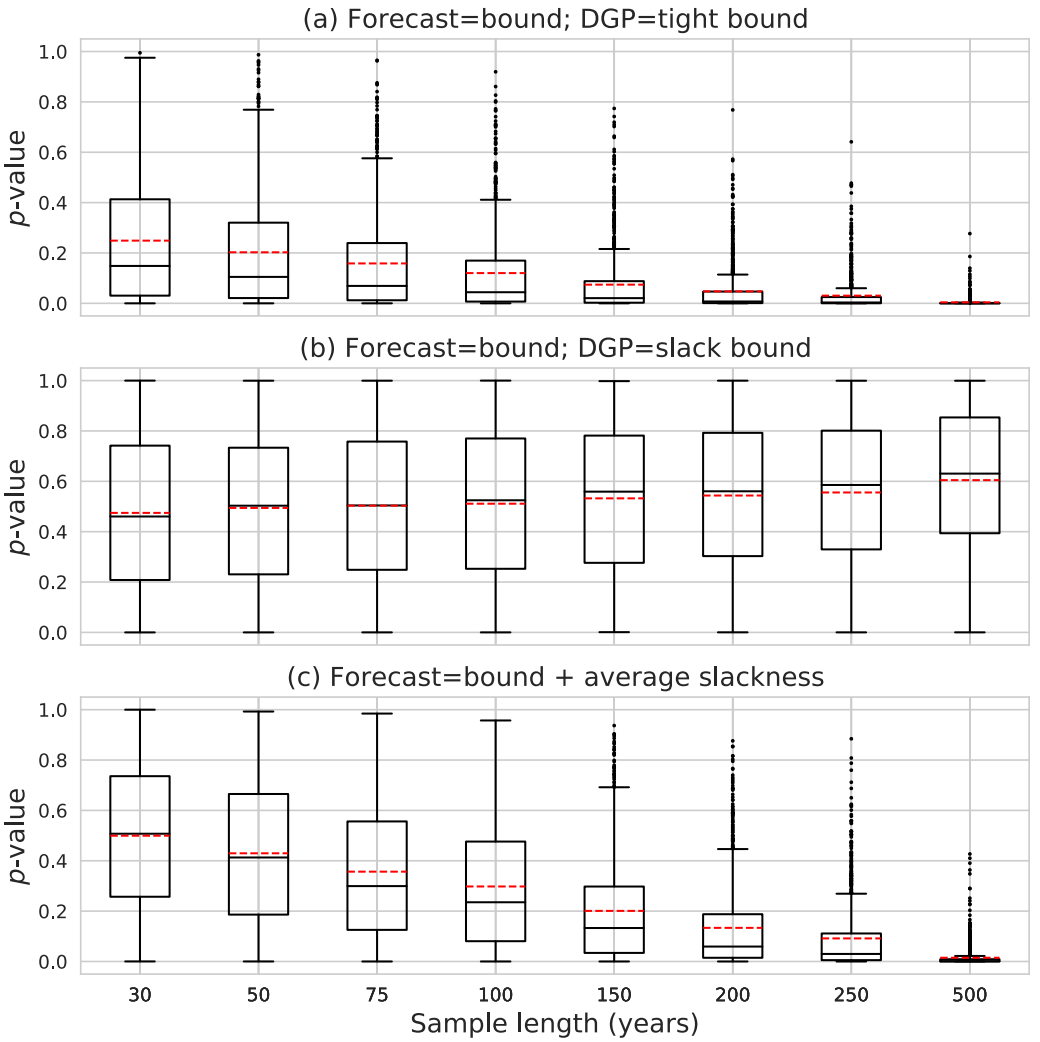

Fig. 6.7. Simulated Diebold–Mariano *p*-values Chabi-Yo/Loudis calibration.
Each panel plots distributions of simulated Diebold–Mariano *p*-values for various sample lengths. The forecasts are either a tight bound (Panels (a) and (b)) or the bound plus an expanding window average realized slackness (Panel (c)). The initial estimation window for average slackness is 60 months. The benchmark forecast is an expanding window mean excess return with 65 years of realized returns available prior to the first bound observation, as in our empirical setting. The simulations are calibrated using the 12-month Chabi-Yo/Loudis bound. For Panel (a), returns are simulated assuming the bound is tight. For Panel (b), returns are simulated assuming the bound is slack (4% per year). The results in Panel (c) are the same under either data-generating process, because any non-zero mean slackness contributes the same to the forecasts as it does to the returns. Each panel is based on 1000 simulations. The dashed red line represents the average *p*-value across simulations. The solid black line within each box represents the median. The box represents the interquartile range, and the whiskers extend no more than 1.5 times the interquartile range from the edge of the box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cross section.[15] It is positive at every horizon but never significant.[16]

The OLS forecast in Row 1 of Panel A of Table 9 is derived from an expanding-window panel regression on the Martin–Wagner bound. The forecast underperforms the benchmark at every horizon. This mirrors the results for the market bounds. The combination forecast in Row 2 of Panel A of Table 9 is an average of forecasts derived from expanding-window panel regressions on individual stock

[15] Let MSE_b denote total mean-squared error for the benchmark (left-hand side of (20) for the benchmark forecast) and MSE_b^{TS} and MSE_b^{CS} denote its time-series and cross-sectional components (right-hand side of last line of (20) for the benchmark forecast). Total R^2 is a weighted average of the time-series and cross-sectional R^2s:

$$R^2 = \left(1 - \frac{\text{MSE}_b^{\text{CS}}}{\text{MSE}_b}\right) R^2_{\text{TS}} + \frac{\text{MSE}_b^{\text{CS}}}{\text{MSE}_b} R^2_{\text{CS}}.$$

[16] Martin and Wagner report out-of-sample R^2s relative to a number of benchmarks, including historical average excess returns of the S&P 500, a CRSP value-weighted index and a constant 6% return per year, for their sample (their Table IX). In their sample (1996–2014), out-of-sample R^2s are positive at all horizons except the 1 month horizon for certain benchmarks. They do not test whether any positive out-of-sample R^2s are significant.

Table 9
Out-of-sample tests of stock forecasts.
Out-of-sample R^2s are computed for monthly forecasts of stock excess returns. Out-of-sample R^2 is defined as

$$R^2_{OOS} = 1 - \frac{\sum_t MSE_{f,t}}{\sum_t MSE_{b,t}},$$

where $MSE_{f,t}$ ($MSE_{b,t}$) is the cross-sectional mean of squared forecast errors for forecast f (benchmark b) for month t. The benchmark model is the expanding window average market excess return (using the Fama-French market excess return series starting in 1926). Forecast (0) is simply the Martin–Wagner or Kadan–Tang bound. Forecast (1) uses an expanding-window panel regressions of excess returns on a constant and the bound. Forecast (2) is a combination forecast of linear models; each month's forecast is the equal-weighted average of univariate expanding-window panel regressions using a single stock characteristic (size, book-to-market, asset growth, operating profitability, or momentum). Forecast (3) is an equal-weighted average of forecast (2) and (0). For forecast (0), the table reports R^2s using total *MSE* as well as versions using only the cross-sectional and time-series components of *MSE* defined in Eq. 20. Panels A and B report R^2s for forecasts using the Martin and Wagner (2019) bounds for the full sample of stocks and the Kadan and Tang (2020) bound for the Conservative stock subsample ($\delta \leq 3$), respectively. Panels C and D report R^2s using forecasts that are truncated below at either the bound (forecasts (1) and (3)) or at zero (forecast (2)). The expanding window regressions use monthly observations from January 1996 through December 2020. To ensure our results are not contaminated by look-ahead bias, we allow h months to elapse after the end of the window before using the results to form a forecast, for return horizon h. The initial estimation window uses the first 60 months of bound observations and the first 60+h months of realized returns. For each model, Diebold–Mariano tests are performed using the time-series of differences in $MSE_{b,t}$ and $MSE_{f,t}$ as the outcome variable. p-values for these tests are calculated using Hansen-Hodrick standard errors with the number of lags equal to the number of months in the return horizon and are reported in the internet appendix. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. Positive R^2 values are shaded gray.

Panel A. Martin-Wagner (All Stocks)

	1	3	6	12
(0) Tight Bound				
Cross-Section	0.000	−0.001	0.003	0.001
Time-Series	0.001	0.008	0.032	0.027
Total	0.000	0.002	0.011	0.007
(1) OLS on Bound	−0.006	−0.012	−0.011	−0.025
(2) Combination (FF)	−0.001	−0.005	−0.009	−0.017
(3) Combination (FF + Bound)	0.001	0.005	0.014	0.021

Panel B. Kadan-Tang (Conservative Stocks)

	1	3	6	12
(0) Tight Bound				
Cross-Section	−0.003	−0.005	−0.001	−0.010
Time-Series	−0.014	−0.041	−0.037	−0.066
Total	−0.007	−0.019	−0.015	−0.032
(1) OLS on Bound	−0.012	−0.034	−0.036	−0.052
(2) Combination (FF)	−0.003	−0.009	−0.018	−0.044
(3) Combination (FF + Bound)	0.004	0.011	0.028*	0.029

Panel C. Martin-Wagner with Truncation (All Stocks)

	1	3	6	12
(1) OLS on Bound	0.000	0.006	0.027	0.038
(2) Combination (FF)	−0.001	−0.005	−0.009	−0.017
(3) Combination (FF + Bound)	0.001	0.009	0.029	0.045

Panel D. Kadan-Tang with Truncation (Conservative Stocks)

	1	3	6	12
(1) OLS on Bound	0.002	0.012	0.052	0.047
(2) Combination (FF)	−0.003	−0.009	−0.018	−0.043
(3) Combination (FF + Bound)	0.004	0.016	0.057*	0.074*

characteristics: size, book-to-market, asset growth, operating profitability, and momentum. The second combination forecast reported in the table (Row 3 of Panel A) is an average of the first forecast and the Martin–Wagner bound. Whereas the combination forecast based on the stock characteristics always underperforms the benchmark, the forecast that includes the Martin–Wagner bound always outperforms the benchmark, reaching an out-of-sample R^2 of 2.1% at the 12-month horizon. However, the overperformance is insignificant based on the Diebold–Mariano test.

Panel B of Table 9 repeats Panel A but using the Kadan–Tang bound and only stocks in the Conservative group. The only case in which a model outperforms the benchmark in Panel B is the combination forecast that averages the bound with the combination forecast from the stock characteristics (Row 3).

Figs. 6.8 and 6.9 show the cumulative squared errors of the Martin–Wagner bound and the Kadan–Tang bound as forecasts compared to the benchmark squared errors. Fig. 6.8 is for the full panel of stocks, and Fig. 6.9 is

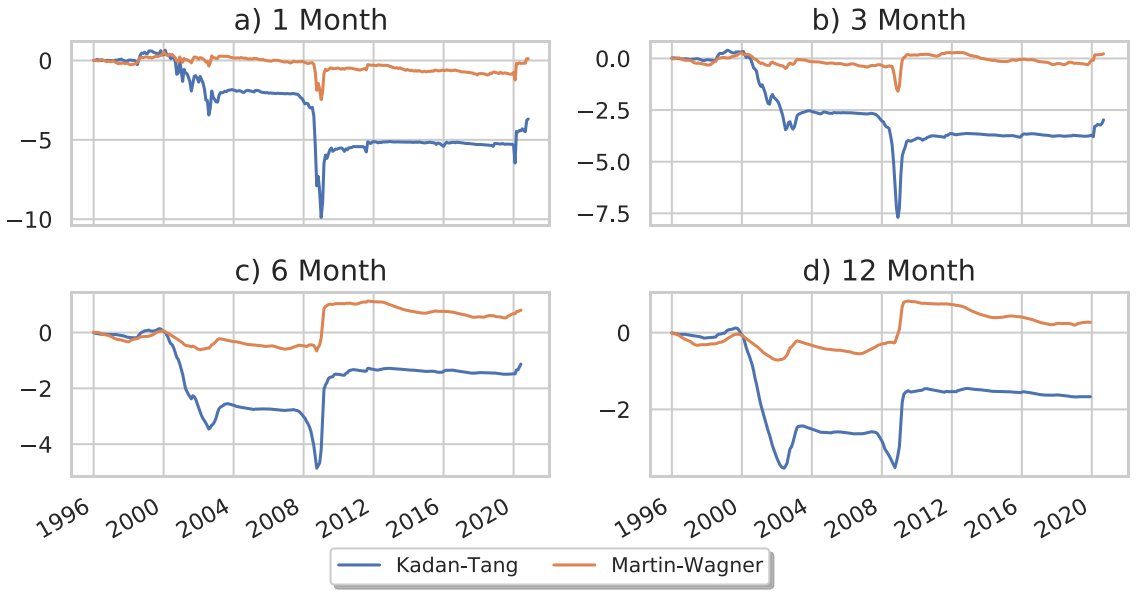

Fig. 6.8. Cumulative squared errors for all stocks.
For each horizon, the panels show the cumulative sum of squared errors of the benchmark minus the cumulative squared error of using the indicated bound as the forecast. The benchmark is the post-1926 expanding window market mean. The sample consists of all stocks in our dataset.

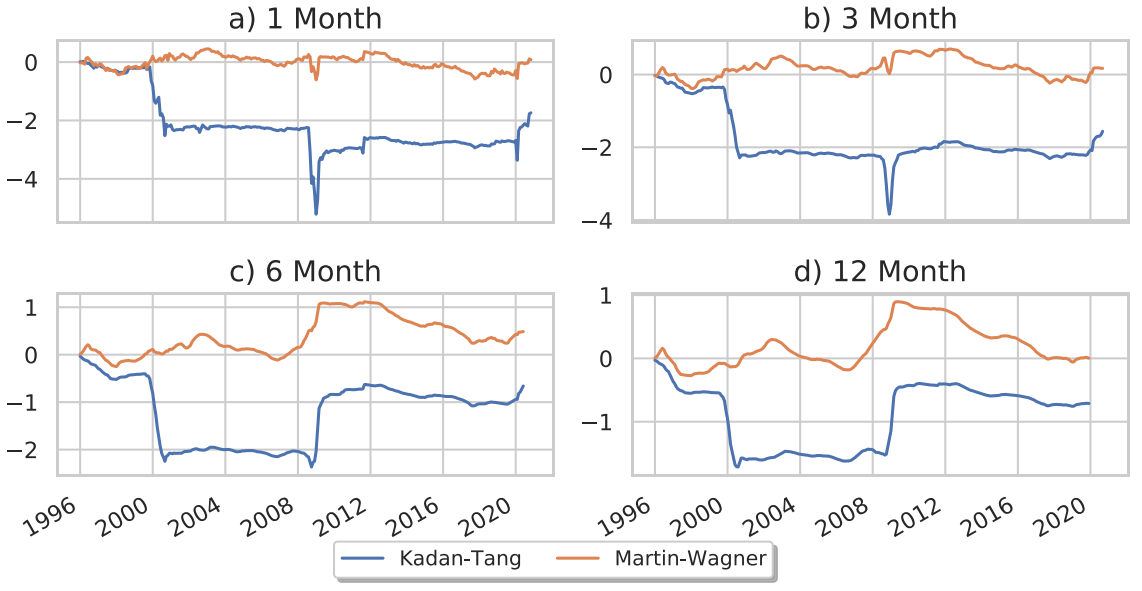

Fig. 6.9. Cumulative squared errors for conservative stocks.
For each horizon, the panels show the cumulative sum of squared errors of the benchmark minus the cumulative squared error of using the indicated bound as the forecast. The benchmark is the post-1926 expanding window market mean. The sample consists of the Conservative group of stocks.

for the Conservative group of stocks. The squared errors are averaged in each cross section and then accumulated over time. The plots end in positive territory for the Martin–Wagner bound, meaning that the out-of-sample R^2 is positive, as shown in Table 9. However, they end in negative territory for the Kadan–Tang bound. The figures show that the Martin–Wagner bound outperforms the Kadan–Tang bound in terms of mean squared errors, for both the full panel and the Conservative group.

Panels C and D of Table 9 report the results of truncating forecasts, either at zero (for the combination forecast that does not use a bound) or at the bound. As for the market forecasts, truncation of the stock forecasts improves the forecasts. For the Conservative group, the combination forecast using the Kadan–Tang bound has statistically significant outperformance at the 6 and 12 month horizons (Row 3 of Panel D).

We take a further look at cross-sectional predictability by sorting stocks on the bound and computing port-

Table 10
Portfolios from sorts on bounds.
Stocks are sorted into quintiles monthly based on the Martin–Wagner/Kadan–Tang bound. Mean annualized forward returns at horizons of 1, 3, 6, and 12 months are computed for each quintile each month. The bottom part of each panel presents statistics (mean, CAPM alpha, and five-factor Fama–French alpha) for the difference between the top quintile return and the bottom quintile return with standard errors in parentheses. The standard errors are Hansen–Hodrick standard errors with number of lags equal to the length of the horizon. Statistical significance is represented by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$. The sample in Panel A is the full sample of stocks. The sample in Panel B is the Conservative stock subsample ($\delta \leq 3$).

	Panel A. All Stocks			
	1	3	6	12
1 (Low)	7.21	7.76	7.87	7.97
2	9.62	9.34	9.27	9.17
3	10.94	10.28	9.77	9.34
4	10.49	10.41	10.31	9.58
5 (High)	10.49	9.95	9.69	9.90
5-1: Mean	3.29	2.19	1.82	1.93
	(5.22)	(5.11)	(5.30)	(4.60)
5-1: CAPM α	2.32	2.41	2.08	2.14
	(5.10)	(5.26)	(5.43)	(4.78)
5-1: Fama–French α	5.04	3.31	2.77	3.28
	(5.36)	(5.09)	(5.17)	(4.64)
	Panel B. Conservative Stocks			
	1	3	6	12
1 (Low)	6.18	6.29	6.22	6.10
2	6.19	7.25	6.91	7.34
3	8.02	7.96	7.44	8.71
4	9.28	9.15	8.29	9.39
5 (High)	12.01	9.79	8.72	7.99
5-1: Mean	5.83*	3.51	2.5	1.88
	(3.01)	(3.19)	(3.31)	(2.92)
5-1: CAPM α	5.02*	3.63	2.82	2.14
	(2.88)	(3.34)	(3.39)	(2.94)
5-1: Fama–French α	4.59	2.92	2.26	1.79
	(3.59)	(3.56)	(3.48)	(3.08)

folio returns. The Martin–Wagner and Kadan–Tang bounds produce identical results for this exercise, because they are perfectly correlated in each cross section, as discussed previously. Table 10 reports the average returns of equally-weighted quintile portfolios for the panel of all stocks and for the Conservative group. There is some monotonicity across quintiles, and the top quintile has a higher average return than the bottom quintile for both groups and all horizons. Furthermore, the point estimates of the CAPM alpha and five-factor Fama–French alpha for the 5-1 return are positive for both groups and all horizons. However, there is statistical significance only for the mean and CAPM alpha for the Conservative group at the 1-month horizon. This is consistent with our previous conclusion that the predictive power of the bounds comes primarily from the time series.

7. Conclusion

The recently developed option-based bounds are important advances in forecasting returns, making novel use of option prices to bound expected returns. Using conditional tests, we cannot reject validity, but we do reject tightness. In full-sample analyses, the market bounds are positively correlated with subsequent realized returns when we control for the Goyal–Welch variables, and the stock bounds are positively correlated with subsequent realized returns when we include stock fixed effects. Out-of-sample tests are inconclusive—bound-based forecasts sometimes beat the historical average market excess return benchmark, but the outperformance is not statistically significant in our relatively short sample period. The data show that at the market level the best out-of-sample performance comes from the Chabi-Yo/Loudis bound, at the stock level the best out-of-sample performance comes from the Martin–Wagner bound, and most of the stock performance both in sample and out of sample comes from the time series rather than the cross section. Adding past mean slackness to the bounds appears to be a promising forecasting approach, but we have a much longer historical period for estimating the mean market return than we do for estimating mean slackness of the bounds. Simulations for the market bounds show that we may need another century or more of data before the 'bound + past mean slackness' forecast can be expected to consistently generate positive out-of-sample R^2s relative to the market-mean benchmark. Additional research on the determinants of bound slackness may be useful to improve the performance of the bounds for forecasting.

Acknowledgments

Toni Whited was the editor for this article. We are grateful to an anonymous referee for especially insightful comments. We thank Ian Martin, Ohad Kadan, Xiaoxiao Tang, seminar participants at INSEAD, Baylor, and

Rice, and Alexander David (NFA discussant) for helpful comments. We thank Fousseni Chabi-Yo, Ohad Kadan, Johnathan Loudis, Ian Martin, Xiaoxiao Tang, and Christian Wagner for making their bounds data available. We thank Amit Goyal for providing predictor data on his website. A previous version of this paper was circulated under the title "Risk Premium Bounds: Slackness Tests and Return Predictions." This research did not receive any specific grant from fund agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Implementing the Kodde–Palm tests

Let $w(N, i, \Sigma)$ denote the probability that i of the N elements of $\hat{\lambda}$ are strictly positive, under the hypothesis that the population mean vector λ is zero. To test whether a bound is valid, our null hypothesis is that $\lambda \geq 0$. For any critical value c, the size of the test, conditional on the covariance matrix Σ, is defined to be

$$\sup_{\lambda \geq 0} \Pr\left(D_1 \geq c \mid \Sigma\right),$$

where D_1 is defined in Eq. (11). According to Kodde and Palm (1986), the asymptotic size is

$$\sup_{\lambda \geq 0} \Pr\left(D_1 \geq c \mid \Sigma\right) = \sum_{i=0}^{N} \Pr\left[\chi^2(N-i) \geq c\right] w(N, i, \Sigma)\,. \tag{A.1}$$

To test whether a bound is tight, our null hypothesis is that $\lambda = 0$. For any critical value c, the asymptotic size of the bound tightness test, according to Kodde and Palm (1986) is

$$\Pr\left(D_2 \geq c \mid \Sigma\right) = \sum_{i=0}^{N} \Pr\left[\chi^2(i) \geq c\right] w(N, i, \Sigma)\,. \tag{A.2}$$

For $i = 0$, the chi-square distribution function is defined as the point mass at the origin (i.e., $\Pr\left[\chi^2(0) \geq c\right] = 0$ for $c > 0$).

The weights w are complicated to calculate analytically (for additional details see Kudo, 1963; Gourieroux et al., 1982; Wolak, 1989). For this reason, Kodde and Palm (1986) provide upper and lower bounds on the critical values that circumvent the need to calculate the weights unless the test statistic falls within the bounds. In order to obtain p-values, we follow Wolak (1989) and calculate the weights through simulation. Specifically, we simulate 10,000 draws of N random variables from a multivariate normal distribution with mean zero and covariance matrix Σ. For each draw $\lambda^s \in \mathbb{R}^N$, define

$$\hat{\lambda}^s = arg\,min_{\lambda \geq 0} \left(\lambda - \lambda^s\right)' \Sigma^{-1} \left(\lambda - \lambda^s\right). \tag{A.3}$$

The weights $w(N, i, \Sigma)$ are estimated as the fraction of the 10,000 draws for which $\hat{\lambda}^s$ has exactly i elements greater than zero. Using these estimated weights, we then compute p values by evaluating Eqs. (A.1) and (A.2) at the test statistics.

For tests of stock-level bounds, we estimate the covariance matrix Σ using a block bootstrap procedure to account for time-series and cross-sectional dependencies similar to that used by Martin and Wagner (2019). We generate 1000 bootstrap samples from the original panel of bounds, forward returns, stock characteristics, and Goyal–Welch variables. Each sample consists of blocks (overlapping and circular) of h consecutive dates, including all stocks in the panel at each date, where h is the horizon.

Appendix B. Simulations

We calibrate and simulate a model for returns and bounds over an horizon consisting of T days. We variously take $T = 21$, $T = 63$, $T = 126$, and $T = 252$. Let r_t denote the excess return on day t. Let μ_t denote the conditional mean of r_{t+1} given information through day t. We assume the mean process is AR(1):

$$\mu_{t+1} = (1-a)\overline{\mu} + a\mu_t + u_{t+1}, \tag{B.1}$$

where the u_t form a mean-zero iid series. We assume that

$$r_{t+1} = \mu_t + v_{t+1}\,, \tag{B.2}$$

where the v_t also form a mean-zero iid series that are possibly contemporaneously correlated with the u_t. Ignoring compounding, define the forward excess return starting at the end of day t (which we previously denoted by $R^e_{t,T}$) as

$$R_t = \sum_{i=1}^{T} r_{t+i}\,. \tag{B.3}$$

Using iterated expectations and the AR(1) specification we can compute that the mean of R_t given information at the end of day t is $(T-\theta)\bar{\mu} + \theta\mu_t$, where

$$\theta = \frac{1-a^T}{1-a}\,. \tag{B.4}$$

We assume the bound is the mean of the forward return minus a constant slackness, namely,

$$b_t = (T-\theta)\bar{\mu} + \theta\mu_t - s\,. \tag{B.5}$$

This implies that b is also an AR(1) process:

$$b_{t+1} = (1-a)\bar{b} + ab_t + \theta u_{t+1}\,. \tag{B.6}$$

where $\bar{b} = T\bar{\mu} - s$. Furthermore, we can solve (B.5) for μ_t as

$$\mu_t = \frac{b_t - (T-\theta)\bar{\mu} + s}{\theta} = \frac{\bar{b}+s}{T} + \frac{b_t - \bar{b}}{\theta}\,. \tag{B.7}$$

It follows that the daily return is

$$r_t = \frac{\bar{b}+s}{T} + \frac{b_{t-1} - \bar{b}}{\theta} + v_t\,. \tag{B.8}$$

Assume the logs of the conditioning variables form a VAR(1) system with monthly time steps:

$$x_{m+1} = (I-A)\bar{x} + Ax_m + w_{m+1}\,, \tag{B.9}$$

where the innovation vectors w_m are mean-zero iid processes.

B1. Calibration

We fit the autoregression (B.6) to the bound series to estimate the parameters a and $\bar{b}$. Given a, we compute θ from (B.4) and then infer the shocks u_t to the daily mean excess return from θ and from the fitted residuals of the bound autoregression (B.6). We infer the daily return shocks v_t from (B.8), taking s to be empirical mean slackness. We fit the VAR (B.9) to the logs of the positive Goyal–Welch variables.

B2. Finite sample inference for validity and tightness tests

To conduct finite sample inference for the validity and tightness test statistics, we simulate the model by first drawing the initial bound b_0 from the stationary distribution of the AR(1) process (B.6), assuming normality. To simulate under the null that a bound is tight, we fix mean slackness s to be zero and construct the bounds b, daily excess returns r and forward excess returns R from (B.6), (B.8), and (B.3), using the fitted value of θ and bootstrapping the shocks u and v in blocks of 12 months from the fitted values. We draw x_0 from the stationary distribution of the VAR(1) process (B.9), assuming normality, and construct the x series from (B.9), bootstrapping the shocks w in blocks of 12 months from the fitted residuals of the VAR(1). We use a circular bootstrap, and we use the fitted u, v, and w series from the same block of 12 months each time to preserve contemporaneous correlations. We exponentiate the x series to obtain simulated positive conditioning variables for use in the validity and tightness tests. We run 1000 simulations to obtain finite sample distributions of the validity and tightness test statistics.

B3. Bootstrapped predictive regressions under null of no predictability

As an alternative to inference using the Amihud and Hurvich (2004) augmented regression methodology, we simulate to estimate the distribution of predictive regression coefficients under the null of no predictability. We simulate as in Appendix B.2 except that we replace (B.8) with

$$r_t = \bar{r} + v_t\,, \tag{B.10}$$

where $\bar{r}$ is the sample mean daily return, and we sample in blocks of h months where h is the length of the horizon. For the monthly predictive regressions, we sample blocks of months, using all daily fitted values of u and v within each sampled month along with the monthly fitted value of w. We use the last observation in a month of the simulated daily series of forward excess return R and bound b in the monthly regressions. We run 1000 simulations for each sample and record the t-value for the bound in univariate daily or multivariate monthly predictive regressions. The bootstrapped p-value is the fraction of simulations with t-values that exceed the t-value for the bound coefficient in the actual data.

B4. Simulation of out-of-sample R^2

To consider the out-of-sample performance of the bound relative to an expanding mean return benchmark, we simulate daily time-series of 65 + T_b years for forward excess returns R and bounds b, where T_b denotes the horizon over which the bounds are assumed available to the econometrician. Using the fitted series of u_t and v_t from the calibration described in Appendix B.1, we compute the sample covariance matrix $\widehat{\Sigma}_{uv}$ of the innovations u_t and v_t. We draw the initial bound b_0 from the stationary distribution of the AR(1) process (B.6), assuming normality. We draw the innovation vectors (u_t, v_t) independently for each day t from the normal distribution with mean vector $(0, 0)'$ and covariance matrix $\widehat{\Sigma}_{uv}$. For a given level of assumed slackness s, we construct the bounds b, daily excess returns r and forward excess returns R from (B.6), (B.8), and (B.3). Since we employ monthly frequency data in our empirical out-of-sample tests, we convert the simulated time-series to monthly observations by sampling every 21st observation.

For each simulation, we calculate out-of-sample R^2s and Diebold–Mariano p-values for two forecasts relative to an expanding mean return benchmark. As in our empirical setting, the expanding mean return uses the initial 65 years of simulated data, and we assume the econometrician can only observe the bound starting in the 66th year. We consider out-of-sample performance in the last $T_b - 5$ years of simulated data, using five years of observations of the bound and returns as an initial estimation window of bound slackness. We consider two forecast methods incorporating the bound. The first forecast is simply the simulated bound. The second forecast is the simulated bound plus an expanding window estimate of bound slackness.

Appendix C. Calculating the bounds

We follow Martin (2017); Chabi-Yo and Loudis (2020); Martin and Wagner (2019); Kadan and Tang (2020) to compute the bounds. We annualize all the bounds by dividing by $T - t$. This appendix explains the computation of the nonannualized bounds.

C1. Market-level bounds

We use option prices from Option Metrics tables opprcdYYYY, where YYYY is the year of each observation, ranging from 1996 to 2019. For 1990–1995 and 2020–2021, we use option price data from CBOE. We perform the following steps to clean the data.

1. From the CBOE data prior to 1996, keep option roots corresponding to S&P500.
2. From both the CBOE and Option Metrics data, keep only standard options and drop all others, including weekly and quarterly options.
3. Drop options with missing bid or ask prices. This includes codes 998 and 999 in the CBOE data.
4. Merge the data with S&P500 closing prices for each date.
5. Calculate option prices as the midpoint between best offers and best bids.

6. Calculate time-to-maturity.
7. For each (day, maturity), keep strike prices with both put and call options in the data.
8. For each (day, maturity, strike), keep the option (call or put) with the lowest price. This step keeps out-of-the-money options in the sample while omitting in-the-money options.
9. Drop options with the best bid of zero.
10. Drop options with less than 7 or more than 550 days to maturity.
11. Drop every (day, maturity) with less than 10 strike prices.

After cleaning the data, we perform the following steps to replicate the results of Martin (2017).

1. For each (day, maturity), integrate the price of OTM options with respect to the strike price using the method described in the appendix of Martin (2017), section B. Let $I_{t,T}$ denote this integral. $I_{t,T}$ is the numerical equivalent of
$$\int_0^{F_{t,T}} put_{t,T}(K)dK + \int_{F_{t,T}}^{\infty} call_{t,T}(K)dK.$$
2. Calculate the Martin bound as $2I_{t,T}/S_t^2$, where S_t is the closing price of the S&P500.
3. Interpolate the bound linearly to match 30, 90, 180, and 360-day maturities. We match these with realized returns over 21, 63, 126, and 252 trading days.

To calculate the Chabi-Yo and Loudis (2020) bounds, in addition to cleaning the data, we use the three-month treasury bill yields from FRED (DGS3MO) to calculate the risk-free rate at the maturity horizon of each option. Then, we proceed to perform the following steps.

1. At each (date, maturity), calculate the second, third, and fourth moments as defined in equation (B.16) of Chabi-Yo and Loudis (2020).
2. At each (date, maturity), calculate the probability of a 20% crash as defined in the internet appendix of Chabi-Yo and Loudis (2020), Section B.2. To calculate the derivative of put prices, we fit a cubic spline to the entire curve of OTM put prices and then calculate the derivative of the cubic spline. We also use the same cubic spline to find the put price at a strike of $0.8S_t$.
3. At each (date, maturity), calculate the restricted first to fifth moments at $k_0 = 0.8$ as described in equation (B.27) of Chabi-Yo and Loudis (2020).
4. For each (date, maturity) calculate the restricted lower bound as
$$\frac{Mom_2/R_{f,t,T} - Mom_3/R_{f,t,T}^2 + Mom_4/R_{f,t,T}^3}{1 - Mom_2/R_{f,t,T}^2 + Mom_3/R_{f,t,T}^3}, \tag{C.1}$$
following equation (31) in Chabi-Yo and Loudis (2020).
5. For each date and bound, interpolate the maturities to reach the desired 30, 90, 180, and 360-day maturities.

C2. Stock-level bounds

We download S&P500 constituents from CRSP and match them to Option Metrics using the linking table provided on WRDS. For these constituents, we download the volatility surface and underlying security data from Option Metrics tables vsurfdYYYY and secprdYYYY respectively, where YYYY denotes the year, ranging from 1996 to 2020. We perform the following steps to replicate the results of Martin and Wagner (2019).

1. Drop all (stock, date, maturity) triples for which some option on the stock with that maturity at that date has a missing strike or maturity.
2. Merge volatility surface data with underlying security data.
3. For each (stock, date, maturity), generate the running minimum of call prices and running maximum of put prices.
4. For each (stock, date, maturity, strike), drop the put option if the running maximum is less than or equal to the running minimum. Otherwise drop the call option.
5. For each (date, maturity), assume the series of strikes remaining in the sample are $K_1 < K_2 < \ldots < K_n$. Also, assume that the respective prices are $p_1, p_2, \ldots, p_n$. We numerically integrate the option price with respect to the strike price using the method described in the internet appendix of Kadan and Tang (2020). More specifically, we calculate the sum
$$I_{t,T} = \sum_{i=2}^{n} (K_i - K_{i-1}) \min\{p_i, p_{i-1}\}, \tag{C.2}$$
which estimates
$$\int_0^{F_{t,T}} put_{t,T}(K)dK + \int_{F_{t,T}}^{\infty} call_{t,T}(K)dK.$$
$I_{t,T}$ is always smaller than the integral above.
6. Calculate $R_{f,t,T}SVIX_{i,t,T} = 2I_{t,T}/S_{i,t}^2$, where $S_{i,t}$ is the closing price of the underlying security. This is the same quantity in the first equation of page 16 in Martin and Wagner (2019).
7. Repeat the same procedure for index options to get $R_{f,t,T}SVIX_{t,T}^2$.
8. Merge the data with daily CRSP files and calculate the market cap for each (permno, date) in the data.
9. Calculate $R_{f,t,T}\overline{SVIX}_{t,T}^2$ by taking a value-weighted average of $R_{f,t,T}SVIX_{i,t,T}^2$.
10. Calculate the bound as
$$R_{f,t,T}\left(SVIX_{t,T}^2 + \frac{1}{2}(SVIX_{i,t,T}^2 - \overline{SVIX}_{t,T}^2)\right).$$

To calculate the lower bound in Kadan and Tang (2020), we calculate $R_{f,t,T} \times SVIX_{i,t,T}^2$ as explained above. To remain consistent with the Martin and Wagner (2019) bounds, we deviate from the process in Kadan and Tang (2020) in two ways. First, we use the volatility surface tables from Option Metrics instead of individual option prices. Second, for each (stock, date), we calculate the bounds at each horizon separately as opposed to taking the average across horizons. This is possible because the volatility surface tables provide more maturities than are available from individual option prices. To calculate
$$\delta_i = \frac{\text{var}(R_i)}{\text{cov}(R_i, R_m)},$$
we download daily stock files from CRSP and calculate variances and covariances on a rolling basis with a 252-day

window. In this step, we require at least 200 observations for each observation of δ.

References

Amihud, Y., Hurvich, C.M., 2004. Predictive-regressions: a reduced-bias estimation method. J. Financ. Quant. Anal. 39, 813–841. doi:10.1017/S0022109000003227.

Bakshi, G., Crosby, J., Gao, X., Zhou, W., 2019. A new formula for the expected excess return of the market,Unpublished working paper. 10.2139/ssrn.3464298

Boudoukh, J., Richardson, M., Smith, T., 1993. Is the ex ante risk premium always positive?: a new approach to testing conditional asset pricing models. J. Financ. Econ. 34 (3), 387–408. doi:10.1016/0304-405X(93)90033-8.

Campbell, J.Y., Thompson, S.B., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud. 21 (4), 1509–1531. doi:10.1093/rfs/hhm055.

Chabi-Yo, F., Dim, C., Vilkov, G., 2022. Generalized bounds on the conditional expected excess return on individual stocks, Management Science, forthcoming. doi:10.2139/ssrn.3565130

Chabi-Yo, F., Loudis, J., 2020. The conditional expected market return. J. Financ. Econ. 137 (3), 752–786. doi:10.1016/j.jfineco.2020.03.009.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263. doi:10.1198/073500102753410444.

Gourieroux, C., Holly, A., Monfort, A., 1982. Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters. Econometrica 50, 63–80. https://www.jstor.org/stable/1912529

Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Manag. Sci. 49, 639–654. doi:10.1287/mnsc.49.5.639.15149.

Green, J., Hand, J.R.M., Zhang, X.F., 2017. The characteristics that provide independent information about average U.S. monthly stock returns. Rev. Financ. Stud. 30 (12), 4389–4436. doi:10.1093/rfs/hhx019.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273. doi:10.1093/rfs/hhaa009.

Hansen, L.P., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. J. Polit. Econ. 88 (5), 829–853. doi:10.1086/260910.

Kadan, O., Tang, X., 2020. A bound on expected stock returns. Rev. Financ. Stud. 33 (4), 1565–1617. doi:10.1093/rfs/hhz075.

Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. J. Finance 68 (5), 1721–1756. doi:10.1111/jofi.12060.

Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restrictions. Econometrica Kodde 54, 1243–1248. https://www.jstor.org/stable/1912331

Kudo, A., 1963. A multivariate analogue of the one-sided test. Biometrika 50, 403–418. doi:10.1093/biomet/50.3-4.403.

Martin, I., 2017. What is the expected return on the market? Q. J. Econ. 132 (1), 367–433. doi:10.1093/qje/qjw034.

Martin, I.W.R., Wagner, C., 2019. What is the expected return on a stock? J. Finance 74 (4), 1887–1929. doi:10.1111/jofi.12778.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Perlman, M.D., 1969. One-sided testing problems in multivariate analysis. Ann. Math. Stat. 40, 549–567. https://www.jstor.org/stable/2239474

Rapach, D.E., Strauss, J.K., Zhou, G., 2010. Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Rev. Financ. Stud. 23 (2), 821–862. doi:10.1093/rfs/hhp063.

Stambaugh, R.F., 1999. Predictive regressions. J. Financ. Econ. 54, 375–421. doi:10.1016/S0304-405X(99)00041-0.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity market prediction. Rev. Financ. Stud. 21, 1455–1508. doi:10.1093/rfs/hhm014.

Wolak, F.A., 1989. Testing inequality constraints in linear econometric models. J. Econom. 41 (2), 205–235. doi:10.1016/0304-4076(89)90094-8.