

Beliefs and Stock Market Fluctuations: New Evidence from the Past Seven Decades

Online Appendix

David Thesmar Emil Verner

May 26, 2026

- Appendix A: Derivation of the model described in Section 2.
- Appendix B: Additional figures and tables.
- Appendix C: Details on Stambaugh bias correction.
- Appendix D: Data appendix.

*Thesmar: MIT Sloan School of Management, CEPR, and NBER, thesmar@mit.edu. Verner: MIT Sloan School of Management and NBER, everner@mit.edu.

A Model Derivation

The market-clearing condition is given by:

$$1 = \alpha \frac{F_t^N R_{t+1}}{\gamma \sigma^2} + (1 - \alpha) \frac{F_t^S R_{t+1}}{\gamma \sigma^2}$$

Return expectations are assumed to follow:

$$\begin{aligned} F_t^S R_{t+1} &= E_t R_{t+1} \\ F_t^N R_{t+1} &= E_t R_{t+1} + \mu R_t. \end{aligned}$$

which leads to the equilibrium condition:

$$\gamma \sigma^2 = E_t R_{t+1} + \alpha \mu R_t \tag{A.1}$$

We are looking for a linear Rational Expectations Equilibrium so that

$$P_t = a + b D_t + c G_t + d R_t.$$

In equilibrium, rational expectations are given by:

$$\begin{aligned} E_t R_{t+1} &= a + (1 + b) E_t D_{t+1} + c E_t G_{t+1} + d E_t R_{t+1} \\ &\quad - (1 + r) (a + b D_t + c G_t + d R_t) \\ &= \frac{1}{1 - d} (-ra + (1 - rb) D_t + (\rho(1 + b) - c(1 + r - \rho)) G_t - (1 + r) d R_t) \end{aligned}$$

Substituting this expression into the equilibrium equation (A.1) yields the equilibrium price:

$$P_t = -\frac{\gamma \sigma^2}{r} \cdot \frac{1 + r}{1 + r + \alpha \mu} + \frac{1}{r} D_t + \frac{1 + r}{r} \frac{\rho}{1 + r - \rho} G_t + \frac{\alpha \mu}{1 + r + \alpha \mu} R_t$$

Prediction 1 The first bullet point follows from:

$$\begin{aligned} F_t^S R_{t+1} &= E_t R_{t+1} = \gamma \sigma^2 - \alpha \mu R_t \\ F_t^N R_{t+1} &= E_t R_{t+1} + \mu R_t = \gamma \sigma^2 + (1 - \alpha) \mu R_t \end{aligned}$$

where the first equation comes from the market clearing condition (A.1) and the second one follows by combining with the definition of naive trader forecasts. Since these two are affine functions of R_t with opposite signs, the two forecasts are negatively correlated.

The second bullet point is related to volume. Volume is the change in position of each agent from one period to the next, i.e.:

$$\text{Volume}_t = \frac{|\Delta F_t^S R_{t+1}|}{\gamma\sigma^2} = \frac{|\Delta F_t^N R_{t+1}|}{\gamma\sigma^2}.$$

From the market clearing condition (A.1), we know that $\Delta F_t^S R_{t+1} = -\Delta F_t^N R_{t+1}$ so that:

$$\frac{|\Delta F_t^S R_{t+1} - \Delta F_t^N R_{t+1}|}{\gamma\sigma^2} = \frac{2|\Delta F_t^S R_{t+1}|}{\gamma\sigma^2} = 2\text{Volume}_t$$

Prediction 2 This follows directly from the expressions of forecasts above.

Prediction 3 We rewrite the price equation as:

$$\underbrace{P_t - \frac{1}{r}D_t}_{\text{valuation}} = \text{constant} + \underbrace{\frac{1+r}{r} \frac{\rho}{1+r-\rho} G_t}_{\text{cash flow shocks}} - \underbrace{\frac{1}{1+r+\alpha\mu} F_t^S R_{t+1}}_{\text{sophisticated expected returns}}$$

which can alternatively be written as a function of naive expected returns:

$$\underbrace{P_t - \frac{1}{r}D_t}_{\text{valuation}} = \text{constant} + \underbrace{\frac{1+r}{r} \frac{\rho}{1+r-\rho} G_t}_{\text{cash flow shocks}} + \underbrace{\frac{\alpha}{1-\alpha} \frac{1}{1+r+\alpha\mu} F_t^N R_{t+1}}_{\text{naive expected returns}}.$$

A comparison between these two expressions shows that valuations are (1) more sensitive to sophisticated expectations when G_t has low persistence and (2) more sensitive to sophisticated expectations than naive ones when α is low.

B Additional Figures and Tables

Table B.1: Correlates of Stock Returns Expectations: Literature Review

| | Data | Returns forecasters | Period | Corr. with valuation | Corr. with past returns | Predicts fut. returns | Analyzes flow cash forecasts |
|-------------------------------|------------------------|---------------------|----------------|----------------------|-------------------------|-----------------------|------------------------------|
| Bacchetta et al. (2009) | UBS-Gallup | Retail/Institutions | 1998-2003 | > 0 | | | No |
| Greenwood and Shleifer (2014) | AII/II/Gallup | Retail/Institutions | 1996-2012 | > 0 | > 0 | | No |
| Adam et al. (2017) | UBS-Gallup | Retail/Institutions | 1998-2008 | > 0 | | | No |
| Wang (2020) | Bloomberg | Institutions | 2002-2019 | | < 0 | | No |
| De La O and Myers (2021) | IBES/CFO survey | Analysts/CFOs | 2003-2015 | > 0 | | | Yes, mix |
| Andonov and Rauh (2021) | Disclosure | Pension funds | 2014-2017 | | > 0 | | No |
| Nagel and Xu (2022) | Various+imputation | Individuals | 1987/2000-2021 | = 0 | > 0 | | No |
| Nagel and Xu (2023) | CFO survey | CFOs | 2000-2021 | = 0 | = 0 | | No |
| | Livingston | Forecasters | 1952-2021 | = 0 | < 0 | | No |
| Couts et al. (2024) | Institutions' websites | Institutions | 1987-2022 | | | | No |
| Dahlquist and Ibert (2024) | Institutions' websites | Institutions | 2000-2024 | < 0 | | | No |
| Büsing and Mohrschladt (2024) | IBES | Analysts, PTG+Rec | 2002-2024 | < 0 | < 0 | > 0 | No |
| Bastianello (2025) | IBES | Analysts PTG | 2002-2024 | = 0 | < 0 | > 0 | No |
| This paper | Value Line | Value Line | 1956-2024 | < 0 | < 0 | > 0 | Yes |

Notes: This table summarizes papers that explicitly study expectations about returns. Nagel and Xu (2023) data has six observations from 1972 to 1977, but consistent coverage without gaps starts in 1987.

Table B.2: *Subjective Expected Returns and the Earnings-Price Ratio*

| | E/P _t | | | | | | | | |
|--|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-----------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| F ^{VL} r _{t→t+4} | 0.260 (4.75) | | | | | | | | |
| F ^{IBES} r _{t→t+1} | | -0.091 (-2.21) | | | | | | | |
| F ^{Sh-inst} r _{t→t+1} | | | -0.038 (-0.81) | | | | | | |
| F ^{Liv} r _{t→t+1} | | | | -0.040 (-0.34) | | | | | |
| F ^{Sh-indiv} r _{t→t+1} | | | | | -0.114 (-0.87) | | | | |
| F ^{SPF} r _{t→t+10} | | | | | | -0.001 (-0.01) | | | |
| F ^{GH} r _{t→t+10} | | | | | | | -0.673 (-2.24) | | |
| F ^{NX} r _{t→t+1} | | | | | | | | 0.156 (0.84) | |
| F ^{GH} r _{t→t+1} | | | | | | | | | 0.149 (0.57) |
| R ² | 0.593 | 0.184 | 0.009 | 0.005 | 0.038 | 0.000 | 0.199 | 0.045 | 0.027 |
| N | 69 | 22 | 36 | 73 | 27 | 34 | 23 | 36 | 23 |

Notes: This table reports results from annual time series regressions of the current earnings-price ratio on expected returns from various surveys. The earnings-price ratio measures earnings over year $t - 1$ relative to the market cap at the end of year $t - 1$. Expected returns are as of Q1 of year t . t -statistics in parentheses based on Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.3: *Inverse-CAPE and Subjective Expected Excess Returns*

| | $\frac{1}{\text{CAPE}_t}$ | | | | | | | | |
|---|---------------------------|-----------------|-----------------|-------------------|-----------------|-----------------|-----------------|-------------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $F^{\text{VL}}_{\text{xr}_{t \rightarrow t+4}}$ | 0.275 (3.16) | | | | | | | | |
| $F^{\text{IBES}}_{\text{xr}_{t \rightarrow t+1}}$ | | 0.036 (0.93) | | | | | | | |
| $F^{\text{Sh-inst}}_{\text{xr}_{t \rightarrow t+1}}$ | | | 0.029 (0.54) | | | | | | |
| $F^{\text{Liv}}_{\text{xr}_{t \rightarrow t+1}}$ | | | | -0.163 (-2.83) | | | | | |
| $F^{\text{Sh-indiv}}_{\text{xr}_{t \rightarrow t+1}}$ | | | | | 0.035 (0.53) | | | | |
| $F^{\text{SPF}}_{\text{xr}_{t \rightarrow t+10}}$ | | | | | | 0.246 (2.81) | | | |
| $F^{\text{GH}}_{\text{xr}_{t \rightarrow t+10}}$ | | | | | | | 0.055 (0.34) | | |
| $F^{\text{NX}}_{\text{xr}_{t \rightarrow t+1}}$ | | | | | | | | -0.268 (-2.52) | |
| $F^{\text{GH}}_{\text{xr}_{t \rightarrow t+1}}$ | | | | | | | | | 0.013 (0.27) |
| R^2 | 0.401 | 0.067 | 0.022 | 0.128 | 0.008 | 0.209 | 0.010 | 0.196 | 0.001 |
| N | 69 | 22 | 35 | 72 | 26 | 33 | 23 | 36 | 23 |

Notes: This table reports results from annual time series regressions of the current inverse CAPE ratio on expected excess returns from various surveys. Expected returns are as of Q1 in year t . Expected excess return is calculated by subtracting the 1-year bond yield as of December of year $t - 1$. The inverse-CAPE is as of December of year $t - 1$. t -statistics in parentheses based on Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.4: *Earnings-Price Ratio and Subjective Expected Excess Returns*

| | E/P _t | | | | | | | | |
|--------------------------------------|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $F^{VL}xr_{t \rightarrow t+4}$ | 0.256 (2.37) | | | | | | | | |
| $F^{IBES}xr_{t \rightarrow t+1}$ | | -0.096 (-2.36) | | | | | | | |
| $F^{Sh-inst}xr_{t \rightarrow t+1}$ | | | -0.047 (-1.02) | | | | | | |
| $F^{Liv}xr_{t \rightarrow t+1}$ | | | | -0.176 (-3.97) | | | | | |
| $F^{Sh-indiv}xr_{t \rightarrow t+1}$ | | | | | -0.179 (-1.26) | | | | |
| $F^{SPF}xr_{t \rightarrow t+10}$ | | | | | | -0.026 (-0.15) | | | |
| $F^{GH}xr_{t \rightarrow t+10}$ | | | | | | | -0.334 (-2.73) | | |
| $F^{NX}xr_{t \rightarrow t+1}$ | | | | | | | | -0.192 (-1.17) | |
| $F^{GH}xr_{t \rightarrow t+1}$ | | | | | | | | | 0.047 (0.26) |
| R^2 | 0.364 | 0.208 | 0.033 | 0.159 | 0.087 | 0.001 | 0.123 | 0.066 | 0.005 |
| N | 69 | 22 | 35 | 72 | 26 | 33 | 23 | 36 | 23 |

Notes: This table reports results from annual time series regressions of the current earnings-price ratio on expected excess returns from various surveys. Expected returns are as of Q1 in year t . Expected excess return is calculated by subtracting the 1-year bond yield as of December of year $t - 1$. t -statistics in parentheses based on Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.5: *Earnings-Price Ratio and Value Line Expected Returns and Earnings Growth*

| | E/P _t | | | | | |
|--|------------------|-----------------|-------------------|-------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| F ^{VL} _{r_t→t+4} | 0.260 (4.75) | | 0.295 (6.86) | 0.183 (2.45) | 0.216 (4.51) | 0.156 (2.61) |
| F ^{VL} _{g_t→t+4} | | 0.204 (1.48) | -0.204 (-1.88) | -0.146 (-1.35) | -0.377 (-3.17) | -0.306 (-3.07) |
| Inflation | | | | 0.359 (2.00) | | 0.239 (1.89) |
| 10-year bond yield | | | | | 0.422 (3.04) | 0.343 (2.94) |
| R ² | 0.593 | 0.045 | 0.627 | 0.685 | 0.712 | 0.735 |
| N | 69 | 69 | 69 | 69 | 69 | 69 |

Notes: This table reports results from regressions of current earnings-price ratio on VL expected returns and expected EPS growth. The earnings-price ratio measures earnings over $t - 1$ to the market cap at the end of $t - 1$. Expectation variables are as of Q1 of year t . Inflation and the 10-year bond yield are as of December of year $t - 1$. Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.6: Subjective Expected Returns and Cyclically Adjusted Earnings Growth

| Panel A: Expected returns and cyclically adjusted earnings growth | | | | | | | | | |
|--|----------------------------------|------------------------------------|---------------------------------------|-----------------------------------|--|------------------------------------|-----------------------------------|----------------------------------|----------------------------------|
| | $F^{VL}_{r_{t \rightarrow t+4}}$ | $F^{IBES}_{r_{t \rightarrow t+1}}$ | $F^{Sh-inst}_{r_{t \rightarrow t+1}}$ | $F^{Liv}_{r_{t \rightarrow t+1}}$ | $F^{Sh-indiv}_{r_{t \rightarrow t+1}}$ | $F^{SPF}_{r_{t \rightarrow t+10}}$ | $F^{GH}_{r_{t \rightarrow t+10}}$ | $F^{NX}_{r_{t \rightarrow t+1}}$ | $F^{GH}_{r_{t \rightarrow t+1}}$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $e_{t-1} - cae_{t-6}$ | 0.043 (0.60) | -0.047 (-1.69) | -0.041 (-2.81) | -0.035 (-1.40) | 0.003 (0.24) | -0.000 (-0.09) | 0.000 (0.11) | 0.026 (2.75) | 0.010 (0.96) |
| R^2 | 0.023 | 0.091 | 0.128 | 0.038 | 0.001 | 0.000 | 0.000 | 0.175 | 0.065 |
| N | 69 | 22 | 36 | 73 | 27 | 34 | 23 | 36 | 23 |
| Panel B: Controlling for past returns | | | | | | | | | |
| | $F^{VL}_{r_{t \rightarrow t+4}}$ | $F^{IBES}_{r_{t \rightarrow t+1}}$ | $F^{Sh-inst}_{r_{t \rightarrow t+1}}$ | $F^{Liv}_{r_{t \rightarrow t+1}}$ | $F^{Sh-indiv}_{r_{t \rightarrow t+1}}$ | $F^{SPF}_{r_{t \rightarrow t+10}}$ | $F^{GH}_{r_{t \rightarrow t+10}}$ | $F^{NX}_{r_{t \rightarrow t+1}}$ | $F^{GH}_{r_{t \rightarrow t+1}}$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| $e_{t-1} - cae_{t-6}$ | 0.063 (0.95) | -0.011 (-0.38) | -0.021 (-2.11) | -0.033 (-1.32) | 0.003 (0.21) | -0.004 (-0.53) | 0.000 (0.10) | 0.013 (1.79) | -0.002 (-0.14) |
| $r_{t-2 \rightarrow t-1}$ | -0.115 (-3.63) | -0.136 (-3.84) | -0.084 (-2.28) | -0.009 (-0.42) | -0.003 (-0.11) | 0.016 (0.92) | -0.000 (-0.02) | 0.053 (4.60) | 0.044 (3.49) |
| R^2 | 0.085 | 0.218 | 0.241 | 0.039 | 0.002 | 0.035 | 0.000 | 0.333 | 0.245 |
| N | 69 | 22 | 36 | 73 | 27 | 34 | 23 | 36 | 23 |

Notes: This table reports results from annual time series regressions of subjective expected equity returns as of Q1 of year t on log earnings in December $t - 1$ relative to the cyclically adjusted log earnings in December $t - 6$ (panel A). Panel B adds past returns from December $t - 2$ to December $t - 1$ as a second explanatory variable. t -statistics in parentheses based on Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.7: *VL Expectation Errors: Returns and the Inverse CAPE*

| Panel A: 1956-2024 | | | |
|---------------------------|---|-------------------------------------|---|
| | $\frac{F^{VL}r_{t \rightarrow t+4}}{(1)}$ | $\frac{r_{t \rightarrow t+4}}{(2)}$ | $\frac{r_{t \rightarrow t+4} - F^{VL}r_{t \rightarrow t+4}}{(3)}$ |
| $\frac{1}{CAPE_t}$ | 2.369 (6.21) | 1.164 (3.92) | -1.205 (-2.09) |
| R^2 | 0.648 | 0.192 | 0.144 |
| N | 66 | 66 | 66 |
| Panel B: 1956-1987 | | | |
| | $\frac{F^{VL}r_{t \rightarrow t+4}}{(1)}$ | $\frac{r_{t \rightarrow t+4}}{(2)}$ | $\frac{r_{t \rightarrow t+4} - F^{VL}r_{t \rightarrow t+4}}{(3)}$ |
| $\frac{1}{CAPE_t}$ | 3.517 (9.14) | 1.511 (4.27) | -2.005 (-2.89) |
| R^2 | 0.836 | 0.466 | 0.368 |
| N | 32 | 32 | 32 |
| Panel B: 1988-2020 | | | |
| | $\frac{F^{VL}r_{t \rightarrow t+4}}{(1)}$ | $\frac{r_{t \rightarrow t+4}}{(2)}$ | $\frac{r_{t \rightarrow t+4} - F^{VL}r_{t \rightarrow t+4}}{(3)}$ |
| $\frac{1}{CAPE_t}$ | 1.609 (5.48) | 3.438 (2.38) | 1.829 (1.18) |
| R^2 | 0.428 | 0.293 | 0.092 |
| N | 34 | 34 | 34 |

Notes: This table reports results from regressions of realized return, Value Line's expected returns, and Value Line's return forecast errors on inverse of the Shiller CAPE ratio. t -statistics in parentheses based on Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.8: *VL Expectation Errors: Returns and Past Returns*

| | $\frac{F_t^{VL} r_{t \rightarrow t+4}}{(1)}$ | $\frac{r_{t \rightarrow t+4}}{(2)}$ | $\frac{r_{t \rightarrow t+4} - F_t^{VL} r_{t \rightarrow t+4}}{(3)}$ |
|---------------------------|--|-------------------------------------|--|
| $r_{t-2 \rightarrow t-1}$ | -0.089 (-3.27) | -0.039 (-1.23) | 0.050 (1.18) |
| R^2 | 0.038 | 0.009 | 0.010 |
| N | 66 | 66 | 66 |

Notes: This table reports results from regressions of VL expected returns, realized returns, and VL return forecast errors on past 1-year market returns. Expectations are as of Q1 of year t . t -statistics in parentheses based on Newey-West standard errors.

Table B.9: *VL Expectation Errors: Price-Earnings Multiple*

| | $\frac{F_t^{VL} P_{t+4}}{F_t^{VL} EPS_{t+4}} (1)$ | $\frac{P_{t+4}}{F_{t+4}^{VL} EPS_{t+4}} (2)$ | $\frac{P_{t+4}}{F_{t+4}^{VL} EPS_{t+4}} - \frac{F_t^{VL} P_{t+4}}{F_t^{VL} EPS_{t+4}} (3)$ |
|------------------------------|---|--|--|
| $\frac{P_t}{F_t^{VL} EPS_t}$ | 0.504 (5.28) | 0.623 (3.67) | 0.119 (0.48) |
| R^2 | 0.509 | 0.292 | 0.011 |
| N | 65 | 65 | 65 |

Notes: This table reports results from regressions of the VL expected long-term price-earnings multiple, the realized long-term price-earnings multiple, and VL long-term price-earnings multiple forecast error on the current price-earnings multiple. Expectations are as of Q1 of year t . t -statistics in parentheses based on Newey-West standard errors.

Table B.10: *VL Expectation Errors: EPS Growth*

| Panel A: Past EPS Growth from $t - 2$ to $t - 1$ | | | |
|---|-------------------------------|-------------------------|---|
| | $F^{VL}g_{t \rightarrow t+4}$ | $g_{t \rightarrow t+4}$ | $g_{t \rightarrow t+4} - F^{VL}g_{t \rightarrow t+4}$ |
| | (1) | (2) | (3) |
| $g_{t-2 \rightarrow t-1}$ | -0.052 (-3.46) | -0.038 (-1.63) | 0.015 (0.50) |
| R^2 | 0.124 | 0.029 | 0.003 |
| N | 65 | 65 | 65 |

| Panel B: Earnings Relative to Cyclically Adjusted EPS Growth | | | |
|---|-------------------------------|-------------------------|---|
| | $F^{VL}g_{t \rightarrow t+4}$ | $g_{t \rightarrow t+4}$ | $g_{t \rightarrow t+4} - F^{VL}g_{t \rightarrow t+4}$ |
| | (1) | (2) | (3) |
| $e_{t-1} - cae_{t-6}$ | -0.012 (-1.04) | -0.065 (-3.99) | -0.054 (-2.40) |
| R^2 | 0.013 | 0.189 | 0.102 |
| N | 65 | 65 | 65 |

Notes: This table reports results from regressions of expected EPS growth, realized EPS growth, and forecast errors on past EPS growth. Expectations are as of Q1 of year t . Expectations of EPS growth are from Value Line. In Panel A, past EPS growth is from year $t - 2$ to $t - 1$. In Panel B, past EPS growth is the log of earnings for the S&P 500 in year t relative to cyclically adjusted earnings in year $t - 5$ as in Bordalo et al. (2024). t -statistics in parentheses based on Newey-West standard errors.

Table B.11: *Predicting Future Returns with Subjective Expected Returns from Other Surveys: Excess Returns*

| | $xr_{t \rightarrow t+1}$ (1) | $xr_{t \rightarrow t+2}$ (2) | $xr_{t \rightarrow t+3}$ (3) | $xr_{t \rightarrow t+4}$ (4) | $xr_{t \rightarrow t+5}$ (5) | $xr_{t \rightarrow t+6}$ (6) |
|---------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| $F^{IBES} xr_{t \rightarrow t+1}$ | -0.427 | 0.394 | 1.023 | 0.672 | 1.126 | 1.086 |
| {Bias-adj Coef} | {-0.402} | {0.065} | {0.683} | {0.380} | {0.572} | {1.618} |
| (p-value) | (0.209) | (0.691) | (0.298) | (0.665) | (0.519) | (0.612) |
| R^2 | 0.021 | 0.010 | 0.051 | 0.012 | 0.021 | 0.018 |
| N | 22 | 21 | 20 | 19 | 18 | 17 |
| $F^{Sh-inst} xr_{t \rightarrow t+1}$ | 0.457 | 0.840 | 1.184 | 1.835 | 2.037 | 1.143 |
| {Bias-adj Coef} | {0.470} | {0.935} | {1.330} | {1.751} | {1.816} | {0.954} |
| (p-value) | (0.447) | (0.419) | (0.501) | (0.374) | (0.449) | (0.729) |
| R^2 | 0.020 | 0.031 | 0.035 | 0.050 | 0.041 | 0.010 |
| N | 35 | 34 | 33 | 32 | 31 | 30 |
| $F^{Liv} xr_{t \rightarrow t+1}$ | 0.057 | -0.210 | -0.260 | 0.008 | -0.203 | -0.630 |
| {Bias-adj Coef} | {0.006} | {-0.251} | {-0.227} | {-0.142} | {-0.353} | {-1.019} |
| (p-value) | (0.883) | (0.733) | (0.782) | (0.992) | (0.880) | (0.685) |
| R^2 | 0.000 | 0.003 | 0.003 | 0.000 | 0.001 | 0.004 |
| N | 69 | 68 | 67 | 66 | 65 | 64 |
| $F^{Sh-indiv} xr_{t \rightarrow t+1}$ | 0.795 | -0.608 | 0.711 | -0.078 | -6.086 | -13.408 |
| {Bias-adj Coef} | {-0.075} | {-1.236} | {-1.950} | {-3.838} | {-10.479} | {-16.160} |
| (p-value) | (0.747) | (0.865) | (0.920) | (0.987) | (0.451) | (0.086) |
| R^2 | 0.010 | 0.002 | 0.001 | 0.000 | 0.050 | 0.215 |
| N | 26 | 25 | 24 | 23 | 22 | 21 |
| $F^{SPF} xr_{t \rightarrow t+10}$ | 2.022 | 5.651 | 7.518 | 11.104 | 9.866 | 15.456 |
| {Bias-adj Coef} | {2.037} | {5.721} | {7.390} | {10.199} | {9.786} | {16.035} |
| (p-value) | (0.366) | (0.270) | (0.528) | (0.416) | (0.479) | (0.139) |
| R^2 | 0.039 | 0.110 | 0.097 | 0.107 | 0.047 | 0.084 |
| N | 33 | 32 | 31 | 30 | 29 | 28 |
| $F^{GH} xr_{t \rightarrow t+10}$ | 2.551 | 6.011 | 10.350 | 14.003 | 24.858 | 40.877 |
| {Bias-adj Coef} | {2.683} | {5.928} | {9.529} | {13.256} | {25.651} | {43.897} |
| (p-value) | (0.425) | (0.477) | (0.430) | (0.353) | (0.104) | (0.010) |
| R^2 | 0.046 | 0.131 | 0.277 | 0.293 | 0.335 | 0.524 |
| N | 23 | 22 | 21 | 20 | 19 | 18 |
| $F^{NX} xr_{t \rightarrow t+1}$ | -0.588 | -1.348 | -2.780 | -3.770 | -7.492 | -5.861 |
| {Bias-adj Coef} | {-0.346} | {-1.226} | {-2.474} | {-3.094} | {-5.641} | {-5.457} |
| (p-value) | (0.672) | (0.546) | (0.597) | (0.600) | (0.434) | (0.523) |
| R^2 | 0.005 | 0.011 | 0.024 | 0.026 | 0.062 | 0.029 |
| N | 36 | 36 | 35 | 34 | 33 | 32 |
| $F^{GH} xr_{t \rightarrow t+1}$ | 0.298 | 0.034 | 2.900 | 3.929 | -2.882 | 2.103 |
| {Bias-adj Coef} | {0.643} | {0.743} | {3.490} | {4.984} | {-0.679} | {3.564} |
| (p-value) | (0.879) | (0.993) | (0.485) | (0.290) | (0.574) | (0.603) |
| R^2 | 0.001 | 0.000 | 0.036 | 0.042 | 0.012 | 0.006 |
| N | 23 | 22 | 21 | 20 | 19 | 18 |

Notes: This table reports results from estimating equation (8) with future cumulative realized excess returns on the left-hand-side and expected excess returns from various surveys on the right-hand-side. The realized excess return is the return on the market from April of year t to April of year $t + h$ minus the h -year yield to maturity from December of year $t - 1$. Expectations are based on surveys from Q1 of year t . We address Stambaugh (1999) bias using the bootstrap approach in Nagel and Xu (2023). For each independent variable, OLS coefficients are reported; bootstrap bias-adjusted coefficients are reported in braces; bootstrapped p-values are reported in parentheses.

Table B.12: Predicting Returns with Value Line Expected Returns: Robustness using VL Universe of Firms

| Panel A: Nominal Returns | | | | | | |
|---------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | $r_{t \rightarrow t+1}$ | $r_{t \rightarrow t+2}$ | $r_{t \rightarrow t+3}$ | $r_{t \rightarrow t+4}$ | $r_{t \rightarrow t+5}$ | $r_{t \rightarrow t+6}$ |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $F^{VL}r_{t \rightarrow t+4}$ | 0.499 | 0.763 | 1.167 | 1.875 | 2.711 | 3.603 |
| {Bias-adj Coef ₁ } | {0.231} | {0.560} | {1.053} | {1.787} | {2.570} | {3.743} |
| {Bias-adj Coef ₂ } | {0.343} | {0.463} | {0.732} | {1.313} | {2.028} | {2.806} |
| (p-value) | (0.045) | (0.022) | (0.013) | (0.010) | (0.007) | (0.001) |
| R^2 | 0.058 | 0.080 | 0.115 | 0.174 | 0.237 | 0.299 |
| N | 69 | 68 | 67 | 66 | 65 | 64 |

| Panel B: Excess Returns | | | | | | |
|--------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | $xr_{t \rightarrow t+1}$ | $xr_{t \rightarrow t+2}$ | $xr_{t \rightarrow t+3}$ | $xr_{t \rightarrow t+4}$ | $xr_{t \rightarrow t+5}$ | $xr_{t \rightarrow t+6}$ |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $F^{VL}xr_{t \rightarrow t+4}$ | 0.486 | 0.553 | 0.757 | 1.294 | 1.944 | 2.455 |
| {Bias-adj Coef ₁ } | {0.187} | {0.299} | {0.587} | {1.051} | {1.609} | {2.338} |
| {Bias-adj Coef ₂ } | {0.320} | {0.238} | {0.304} | {0.724} | {1.267} | {1.670} |
| (p-value) | (0.246) | (0.372) | (0.277) | (0.157) | (0.103) | (0.035) |
| R^2 | 0.035 | 0.027 | 0.031 | 0.057 | 0.089 | 0.108 |
| N | 69 | 68 | 67 | 66 | 65 | 64 |

Notes: This table reports results from regressions of future realized returns on expected returns (panel A) and of future realized excess returns on expected excess returns (panel B). Realized returns are calculated for the universe of firms in VL at time t . Expected excess returns are calculated by subtracting the horizon-matched Fama-Bliss/Liu-Wu bond yield from cumulative nominal returns. The realized excess return from t to $t+h$ is computed as the cumulative nominal return over that horizon minus the time t yield to maturity on a Treasury bond with maturity of h years. We address Stambaugh (1999) bias using two complementary corrections: a bootstrap approach following Nagel and Xu (2023) (method 1) and an alternative correction following Boudoukh et al. (2022) (method 2). For each independent variable, OLS coefficients are reported; bootstrap bias-adjusted coefficients are reported in braces; bootstrapped p-values are reported in parentheses.

Table B.13: *Price-Earnings Variance Decomposition: Robustness to Aggregation Methodology*

| | CF | ER | LT |
|--------------|--------|--------|--------|
| | (1) | (2) | (3) |
| $\ln(P/E_t)$ | 0.110 | 0.597 | 0.287 |
| | (1.57) | (6.01) | (4.04) |

Notes: This table reports results from regressions of each term in the Campbell-Shiller decomposition on the current log price-earnings ratio. Relative to Table 5, we calculate each term in the Campbell-Shiller decomposition by first aggregating all variables in the VL expectations data and then calculating the expected return based directly on the aggregated series, instead of calculating the expected return as the market-cap-weighted firm-level expected return. Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.14: Implications of Belief Disagreement: Change in Beliefs**Panel A: Belief Disagreement and Trading Volume**

| | Volume | | | | |
|--|-----------------|-----------------|-----------------|-----------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) |
| $ \Delta F^{NX}_r - \Delta F^{VL}_r $ | 3.261 (2.73) | | | | |
| $ \Delta F^{Sh-indiv}_r - \Delta F^{VL}_r $ | | 1.144 (1.98) | | | |
| $ \Delta F^{GH}_r - \Delta F^{VL}_r $ | | | 2.185 (1.84) | | |
| $ \Delta F^{Indiv-avg-all}_r - \Delta F^{VL}_r $ | | | | 2.314 (1.68) | |
| $ \Delta F^{Indiv-avg-unbal}_r - \Delta F^{VL}_r $ | | | | | 1.602 (1.31) |
| R^2 | 0.055 | 0.015 | 0.037 | 0.030 | 0.015 |
| N | 145 | 100 | 86 | 83 | 148 |

Panel B: Belief Disagreement and Return Volatility

| | Volatility | | | | |
|--|-----------------|-----------------|-----------------|-----------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) |
| $ \Delta F^{NX}_r - \Delta F^{VL}_r $ | 0.237 (7.75) | | | | |
| $ \Delta F^{Sh-indiv}_r - \Delta F^{VL}_r $ | | 0.083 (3.05) | | | |
| $ \Delta F^{GH}_r - \Delta F^{VL}_r $ | | | 0.143 (6.77) | | |
| $ \Delta F^{Indiv-avg-all}_r - \Delta F^{VL}_r $ | | | | 0.160 (5.11) | |
| $ \Delta F^{Indiv-avg-unbal}_r - \Delta F^{VL}_r $ | | | | | 0.169 (5.42) |
| R^2 | 0.422 | 0.076 | 0.185 | 0.169 | 0.243 |
| N | 145 | 100 | 86 | 83 | 148 |

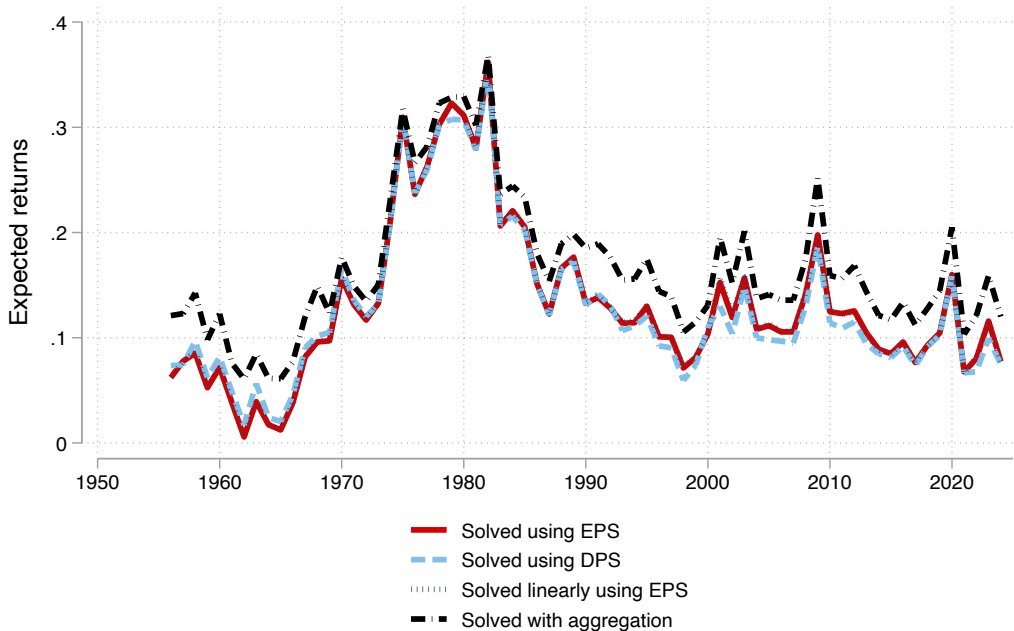
Notes: This table reports results from regressions of trading volume and standard deviation of daily market returns in quarter t on the absolute difference between the change in expected returns of individuals (from the Shiller survey, Nagel-Xu, Graham-Harvey survey) and the change in expected returns of Value Line in the same quarter t . $|\Delta F^{Indiv-avg-avail}_r - \Delta F^{VL}_r|$ uses the average of the individual expected return series available in a given quarter to maximize the sample length (if only one series is available, the average is just the value from that one series). $|\Delta F^{Indiv-avg-all}_r - \Delta F^{VL}_r|$ uses the average of all three individual expected return series, so it requires that all three are available in a given quarter. t-statistics in parentheses. Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

Table B.15: Long-term and Short-term Multiples

| | $\frac{F^{VL}P_{t+4}}{F^{VL}EPS_{t+4}}$ | | | | | | | | | |
|-----------------------------------|---|-----------------|-----------------|-----------------|-----------------|-----------------|-------------------|-------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| $\frac{P_t}{F^{VL}EPS_t}$ | 0.564 (7.03) | 0.297 (2.76) | 0.226 (1.91) | 0.217 (2.40) | 0.242 (2.83) | 0.255 (2.79) | 0.273 (2.98) | 0.272 (3.18) | 0.285 (3.28) | 0.311 (3.72) |
| $\frac{P_{t-1}}{F^{VL}EPS_{t-1}}$ | | 0.337 (4.77) | 0.191 (3.76) | 0.154 (3.14) | 0.142 (2.46) | 0.144 (2.57) | 0.143 (2.37) | 0.173 (4.40) | 0.168 (4.18) | 0.157 (3.66) |
| $\frac{P_{t-2}}{F^{VL}EPS_{t-2}}$ | | | 0.247 (3.17) | 0.139 (2.03) | 0.094 (1.56) | 0.102 (1.51) | 0.105 (1.46) | 0.086 (1.20) | 0.102 (1.61) | 0.097 (1.49) |
| $\frac{P_{t-3}}{F^{VL}EPS_{t-3}}$ | | | | 0.179 (4.20) | 0.105 (2.80) | 0.049 (1.16) | 0.050 (0.98) | 0.064 (1.75) | 0.056 (1.17) | 0.075 (3.09) |
| $\frac{P_{t-4}}{F^{VL}EPS_{t-4}}$ | | | | | 0.131 (2.27) | 0.023 (0.32) | -0.028 (-0.39) | -0.015 (-0.18) | -0.001 (-0.01) | -0.025 (-0.31) |
| $\frac{P_{t-5}}{F^{VL}EPS_{t-5}}$ | | | | | | 0.159 (2.06) | 0.066 (1.12) | -0.000 (-0.01) | 0.015 (0.31) | 0.051 (1.22) |
| $\frac{P_{t-6}}{F^{VL}EPS_{t-6}}$ | | | | | | | 0.140 (2.30) | 0.066 (1.18) | -0.007 (-0.15) | -0.006 (-0.12) |
| $\frac{P_{t-7}}{F^{VL}EPS_{t-7}}$ | | | | | | | | 0.119 (1.99) | 0.020 (0.42) | -0.051 (-1.13) |
| $\frac{P_{t-8}}{F^{VL}EPS_{t-8}}$ | | | | | | | | | 0.154 (2.91) | 0.089 (2.35) |
| $\frac{P_{t-9}}{F^{VL}EPS_{t-9}}$ | | | | | | | | | | 0.116 (1.61) |
| Constant | 6.901 (4.99) | 5.890 (3.71) | 5.500 (3.55) | 5.184 (3.29) | 4.867 (2.96) | 4.625 (3.01) | 4.416 (3.19) | 4.216 (3.31) | 3.856 (3.23) | 3.573 (3.11) |
| R^2 | 0.618 | 0.696 | 0.737 | 0.758 | 0.772 | 0.787 | 0.808 | 0.830 | 0.853 | 0.877 |
| N | 69 | 68 | 67 | 66 | 65 | 64 | 63 | 62 | 61 | 60 |

Notes: This table reports results from regressions of Value Line's forecast for long-term multiples on current forward multiples and its lags. Newey-West standard errors with bandwidth $S = 1.3\sqrt{T}$.

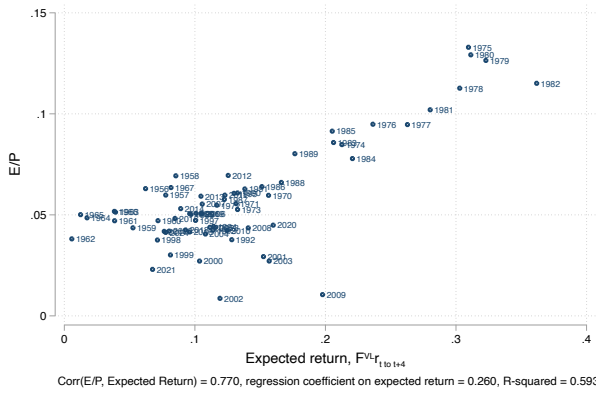
Figure B.1: Value Line Expected Returns: Alternative Calculation Approaches



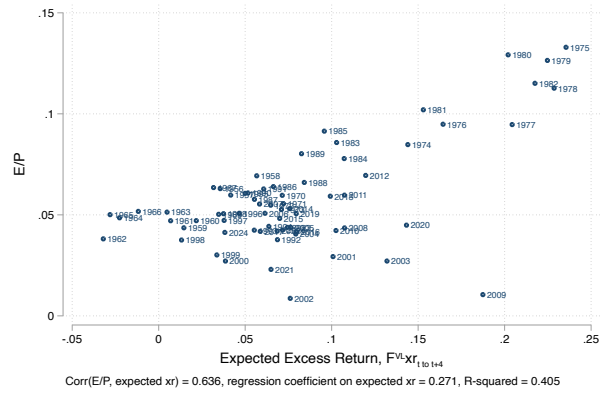
Notes: This figure plots the VL expected return using various calculation approaches. “Solved using EPS” is the baseline measure in the paper. “Solved using DPS” directly uses VL’s forecast of dividends per share. “Solved using linearly interpolated level of EPS” uses the level of EPS and linearly interpolates EPS forecasts from year t to $t + 4$, allowing for negative values of year t EPS. “Solved with aggregation” first aggregates all variables (current market cap, expected earnings in t , expected earnings in $t + 4$, and expected market cap in $t + 4$) in VL and then computes the expected return from the aggregate series.

Figure B.2: Valuation Ratios and Measures of Expected Return from Value Line

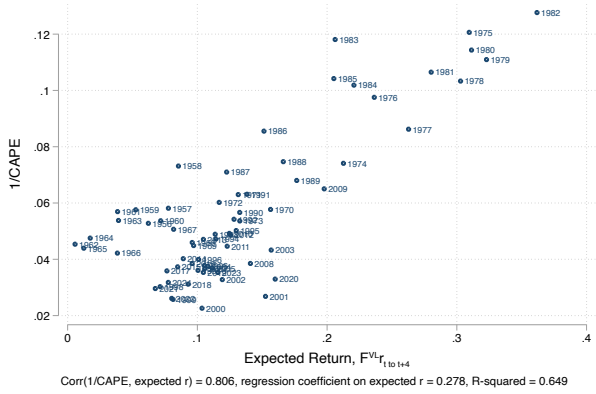
(a) Current EP Ratio and VL Expected Return



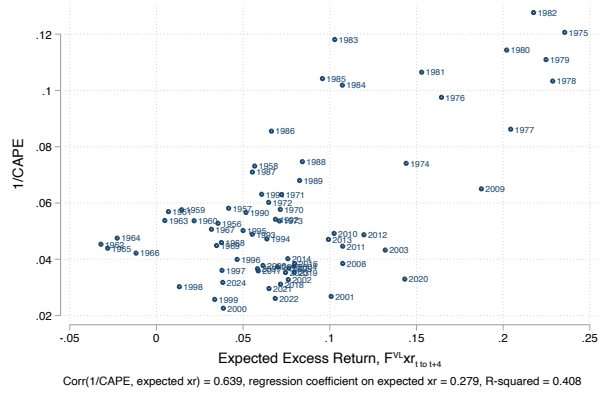
(b) Current EP Ratio and VL Expected Excess Return



(c) 1/CAPE and VL Expected Return

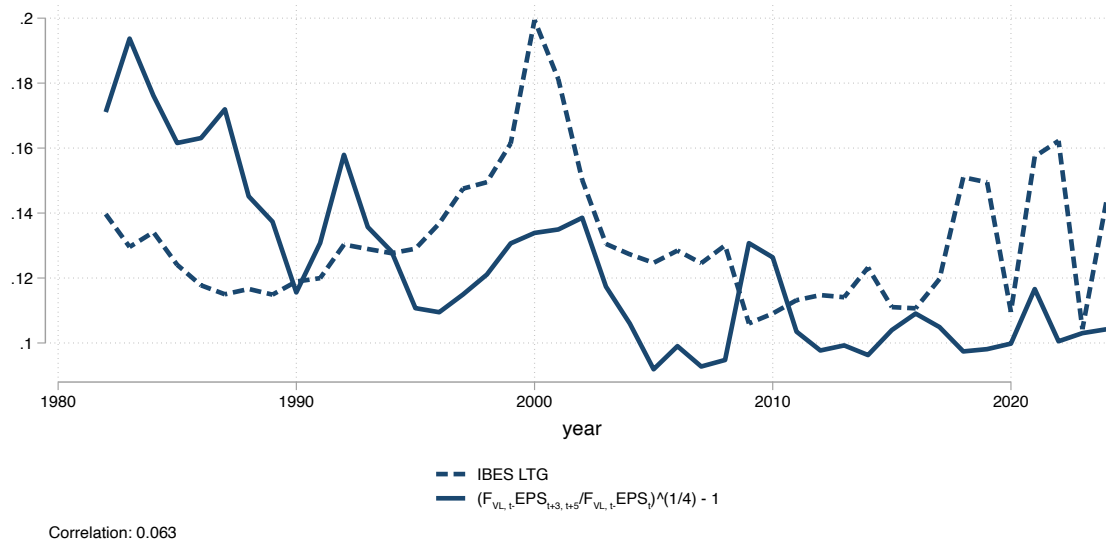


(d) 1/CAPE and VL Expected Excess Return



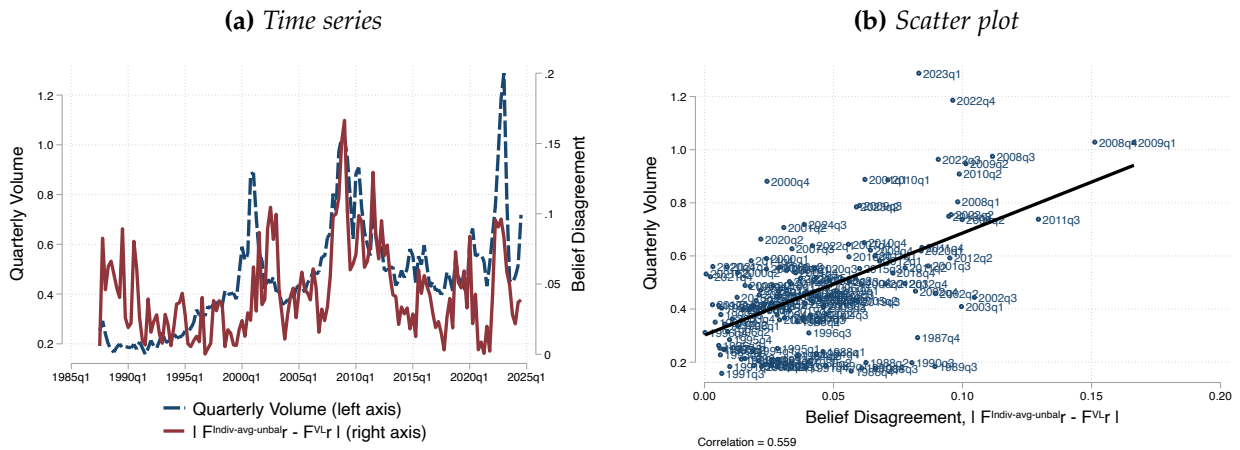
Notes: This figure plots measures of the earnings-price ratio against measures of expected returns. VL expected returns are as of Q1 of year t . The VL expected excess return is calculated by subtracting the horizon-matched bond yield (four years) as of December of year $t - 1$.

Figure B.3: Subjective Expected Long-Term EPS Growth: Value Line and IBES LTG



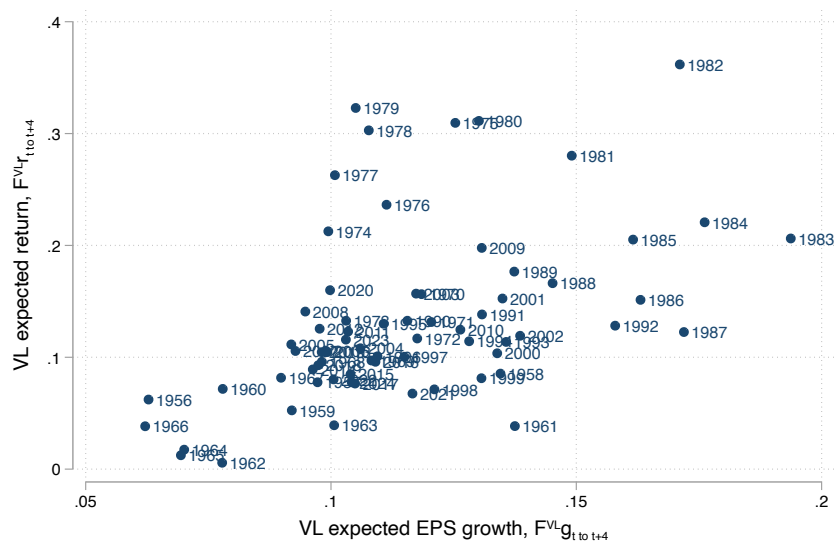
Notes: This figure shows subjective expected EPS growth from IBES and Value Line. Subjective expected EPS Growth from IBES is LTG forecast as of Q1 in year t , defined as the expected annual increase in operating earnings over the company's next full business cycle, a period ranging from three to five years. Expected EPS growth from VL is as of Q1 of year t and measures the annualized expected growth in EPS from t to $t + 4$.

Figure B.4: Disagreement about Expected Returns and Trading Volume: Robustness using Unbalanced Average



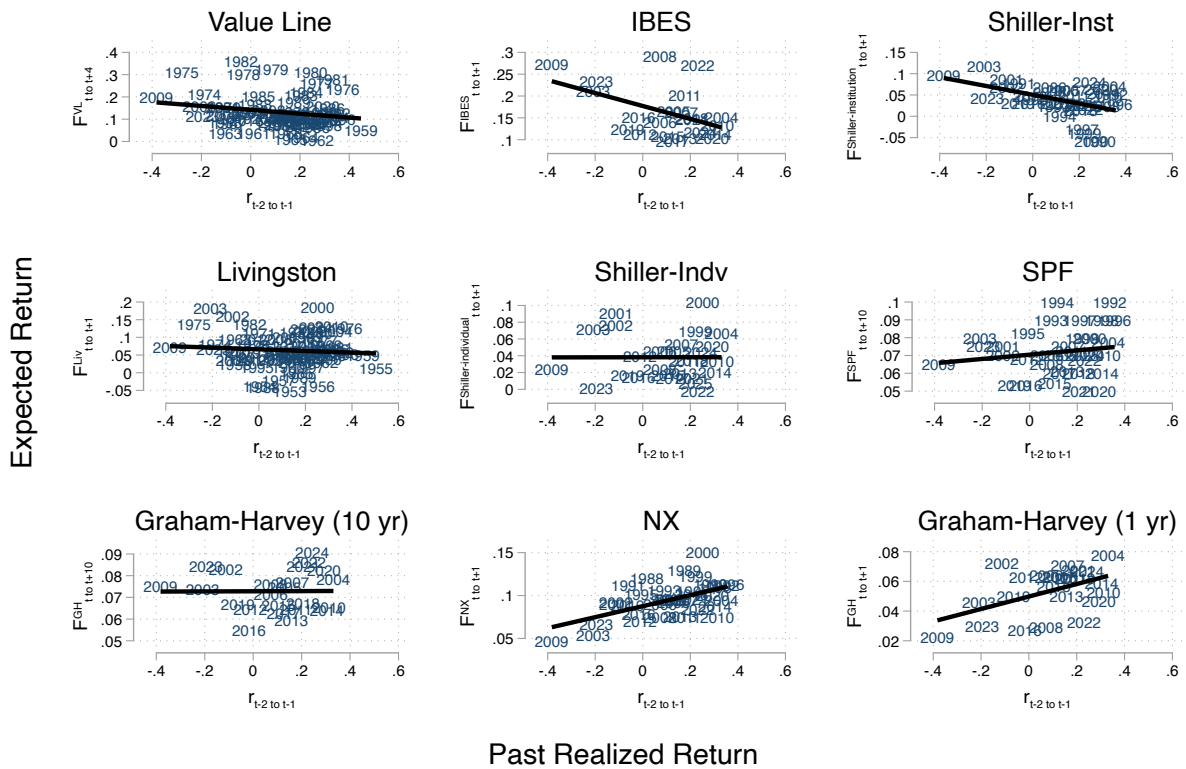
Notes: This figure is similar to Figure 7, but computes the individual expected return in the belief disagreement measure by taking the average of any available series (an unbalanced average). For example, in early years when Nagel-Xu is the only available series, we just use this series. When additional series become available, we take the within-quarter average across available series.

Figure B.5: VL Expected Cash Flow and Expected Return



Notes: The figure plots VL's expected return from t to $t + 4$ against their expected EPS growth from t to $t + 4$. Both measures are annualized. VL expectations for year t are formed in Q1.

Figure B.6: Survey Expected Return and Past Realized Return



Notes: This figure shows subjective expected returns from various surveys against past realized returns. Subjective expected returns are as of Q1 in year t . Past realized returns are returns on the market from December of year $t - 2$ to December of year $t - 1$.

C Details on Stambaugh Bias Correction

This appendix provides details on the two approaches we use to correct for Stambaugh (1999) bias.

C.1 Method 1: Nagel and Xu (2023) Approach

We consider the following two predictive regressions for horizon H :

$$z_{t+H} = \alpha_H + \beta_H x_t + \eta_{t+H}, \quad (\text{B.1})$$

$$x_{t+1} = \kappa + \Phi x_t + \iota_{t+1}. \quad (\text{B.2})$$

1. We estimate an AR(1) model for the predictor, x_t . To correct for the bias in the AR estimates, we use the approach in Amihud et al. (2008) that relies on the analytical expressions from Nicholls and Pope (1988). This allows us to obtain the bias-adjusted coefficients and shocks: $\tilde{\kappa}$, $\tilde{\Phi}$, and $\{\tilde{\iota}_t\}$, given by

$$\begin{aligned} \tilde{\Phi} &= \hat{\Phi} + \frac{1 + 3\hat{\Phi}}{T} + \frac{3(1 + 3\hat{\Phi})}{T^2}, \\ \tilde{\kappa} &= (1 - \tilde{\Phi})\bar{x}, \\ \tilde{\iota}_t &= x_t - \tilde{\kappa} - \tilde{\Phi}x_{t-1}, \end{aligned}$$

where $\bar{x} = T^{-1} \sum_t x_t$ is the sample mean of the predictor.

2. We next estimate the predictive regression in (B.1) with OLS, obtain $\hat{\alpha}_H$, $\hat{\beta}_H$, residuals $\hat{\eta}_{t+H}$, and the corresponding t -statistic \hat{t}_H for $\hat{\beta}_H$ computed using Newey–West standard errors with bandwidth $S = 1.3\sqrt{T}$.
3. We then construct pseudo-samples by bootstrapping the time-series of residual vectors $(\tilde{\iota}_t, \hat{\eta}_t)$. We use a stationary block bootstrap, where the optimal block length is determined as in Politis and White (2004).
4. To compute the bootstrap p -value: For each bootstrap sample i and horizon H , with bootstrapped residuals, we impose the null of no predictability by generating data from

$$z_{t+H}^{null,i} = \bar{z} + \hat{\eta}_{t+H}^i, \quad (\text{B.3})$$

$$x_{t+1}^i = \tilde{\kappa} + \tilde{\Phi}x_t^i + \tilde{\iota}_{t+1}^i. \quad (\text{B.4})$$

- We then re-run the predictive regression in the sample i and record the t -statistic t_H^i .

- The $\{t_H^i\}$ are used to obtain the small-sample p -value by comparing with the sample t -statistic \hat{t}_H . The bootstrap p -value is computed as

$$\hat{p}_H = \frac{1}{B} \sum_{i=1}^B \mathbf{1}(|t_H^i| \geq |\hat{t}_H|).$$

5. To compute the bias-adjusted bootstrap coefficient: For each bootstrap sample j and horizon H , with bootstrapped residuals, we generate data under the alternative that $\beta_H = \hat{\beta}_H$ as

$$z_{t+H}^{alter,j} = \hat{\alpha}_H + \hat{\beta}_H x_t^j + \hat{\eta}_{t+H}^j, \quad (\text{B.5})$$

$$x_{t+1}^j = \tilde{\kappa} + \tilde{\Phi} x_t^j + \tilde{\nu}_{t+1}^j. \quad (\text{B.6})$$

- We then re-run the predictive regression in the sample j and record the coefficients $\hat{\beta}_H^j$.
- From the distribution of bootstrapped coefficients, we estimate the finite-sample bias as the difference between the mean of bootstrapped coefficients and the sample OLS estimate:

$$\widehat{\text{Bias}}_H = \frac{1}{B} \sum_{j=1}^B \hat{\beta}_H^j - \hat{\beta}_H.$$

- The bias-adjusted coefficient is

$$\bar{\beta}_H = \hat{\beta}_H - \widehat{\text{Bias}}_H = 2\hat{\beta}_H - \frac{1}{B} \sum_{j=1}^B \hat{\beta}_H^j.$$

where B is the number of bootstrap simulations, set to 5000.

C.2 Method 2: Boudoukh-Israel-Richardson Correction

1. Steps 1–3 are identical to Method 1.
2. To compute the bias-adjusted coefficient: For each horizon H ,
 - Estimate the AR(1) process for the predictor (eq. (B.2)), obtain the OLS coefficient $\hat{\Phi}$, residuals $\hat{\nu}_{t+1}$, and compute the variance,

$$(\sigma_\nu^2) = \text{Var}(\hat{\nu}_{t+1}).$$

- Next, estimate the one-period predictive regression (eq. (B.1) with $H=1$), obtain the OLS residuals $\hat{\eta}_{t+1}$, and compute the contemporaneous innovation covariance

$$(\sigma_{\eta\nu}) = \text{Cov}(\hat{\eta}_{t+1}, \hat{\nu}_{t+1}).$$

- Estimate the biased coefficients $\hat{\beta}_H^{\text{ols}}$ from the H -horizon regression:

$$z_{t \rightarrow t+H} = \alpha_H + \beta_H x_t + \eta_{t+H}.$$

- Compute the bias following the formula in Boudoukh et al. (2022):

$$\widehat{\text{Bias}}_H^{\text{BIR}} = -\frac{1}{T} \left[H(1 + \hat{\Phi}) + 2\hat{\Phi} \left(\frac{1 - (\hat{\Phi})^H}{1 - \hat{\Phi}} \right) \right] \frac{(\sigma_{\eta_t})}{(\sigma_t^2)}.$$

- Compute the bias-adjusted coefficient: $\bar{\beta}_H^{\text{BIR}} = (\hat{\beta}_H^{\text{ols}}) - \widehat{\text{Bias}}_H^{\text{BIR}}$.

3. To compute the bootstrap p -value: For each bootstrap sample j and horizon H , with bootstrapped residuals, impose the null of no predictability by generating data as

$$z_{t+H}^{\text{null},j} = \bar{z} + \hat{\eta}_{t+H}^j \tag{B.7}$$

$$x_{t+1}^j = \bar{\kappa} + \tilde{\Phi} x_t^j + \tilde{t}_{t+1}^j. \tag{B.8}$$

- We then re-run the predictive regression on the resulting pseudo-sample and record the t -statistic t_H^j .
- The small-sample p -value under the null is obtained by comparing the bootstrap distribution of $\{t_H^j\}$ with the sample t -statistic \hat{t}_H . Formally, the bootstrap p -value is computed as

$$\hat{p}_H = \frac{1}{B} \sum_{j=1}^B \mathbf{1}(|t_H^j| \geq |\hat{t}_H|).$$

where B denotes the number of bootstrap replications.

D Data Appendix

D.1 New Value Line Database, 1956-1989

Data source. We digitize *Value Line Investment Survey* from 1956 through 1989. We digitize the first quarter reports of each year. This dataset covers 38,696 firm-year observations, with an average of 1,138 firms each year for 34 years. The VL reports contain actual realizations of major accounting variables for approximately past 10 years. Moreover, the reports contain forecast data of the past year, current year, the next year, and a 3-5 horizon.¹

Data construction. With a team of research assistants and the assistance of the MIT Library, we scanned physical volumes and microfilm images of the VL reports published between January and March each year from 1956 to 1989. We then separated pages with industry reports from pages containing the analysis of individual firms. Figures B3 and B4 contain examples of the raw data we digitize. We then apply a few preprocessing techniques to enhance the quality of the scanned photo (enhancing the scans, rotating, and cropping the images). We then OCR the images using the Amazon Web Services Textract OCR engine to obtain the JSON file with block objects containing all the strings and coordinates that Amazon Textract was able to read. From the JSON file, we locate the firm name, ticker, and pages. We then locate the coordinates of the table we needed from the block objects and use them to calculate the slope and the skewed angle, then rotate page and all coordinates by that angle to straighten it up. We cropped the table of interest from the straightened page and uploaded to Amazon Textract again to get a precise translation of all the cells and their deskewed coordinates. The returned block objects are then reconstructed into tables. We also use Textract's query feature to extract additional information from the page, such as the recent price, dividend yield, P/E ratio, and Value Line rankings. In early years (1956-1963), the 3-5 year forecasts are reported in the text, so we extract these using AWS's query function. We then combine all the above information we obtained in one excel sheet containing: forecast and actual tables, names and dates, and query answers.

To ensure that the digitization was high quality, we applied a range of checks. We manually checked all tables for misread strings, unaligned columns, and jammed rows or cells. We pay special attention to suspicious cells and cells with a low confidence score. Further, we applied a range of accounting and financial identities. For example, we checked consistency between sales, sales per share, and common shares outstanding. We also checked for outliers in the current P/E ratio, earnings per share, and recent price.

Data cleaning and panel construction. We appended the data tables for each firm and year. We then use the CRSP database to construct a firm-year panel of firms with actuals and forecasts. To match firms across years, we cleaned the tickers and firm names. In particular, we standardize firm names by constructing a ground truth list of names and

¹For a few years, the long-term forecasts are at the 2-4 year horizon.

tickers. We then applied a fuzzy match to the CRSP database to find PERMNO code corresponding to each firm-year, applying a confidence threshold of 95 percent.

D.2 Modern Value Line Database, 1987-2024

The Value Line data we use from 1988 onwards was purchased from Value Line. Data purchased includes two products: Fundamental Datafile and Historical Estimates File.

The Value Line Fundamental Datafile only contains realizations data and includes detailed annual income statements, balance sheets, and cash flow statements for around 7,000 major public companies across 10 economic sectors and 97 industry categories.

The Value Line Historical Estimates File contains both realizations and forecast data, including estimated income statement and balance sheet data for about 1,600 major public companies across 10 economic sectors and 97 industry groups, along with identifiers for over 7,000 companies. Generally, for each stock, the data provides forecasts for the current year, next year, and 3 to 5 years ahead. We use E&P data to construct forecast and realizations data for the modern period. In Historical Estimates files, Value Line uses a stock's historical eight-digit CUSIP and exchange ticker as the identifier. We merge firms to CRSP stocknames table from WRDS by CUSIP-year to assign PERMNO to each CUSIP. PERMNO is assigned by historical CUSIP and year. As Value Line sometimes backfills CUSIPs, if this merge fails, we use header CUSIP, the next historical CUSIP, or refer to Value Line's historical CUSIP changes file to assign PERMNO.

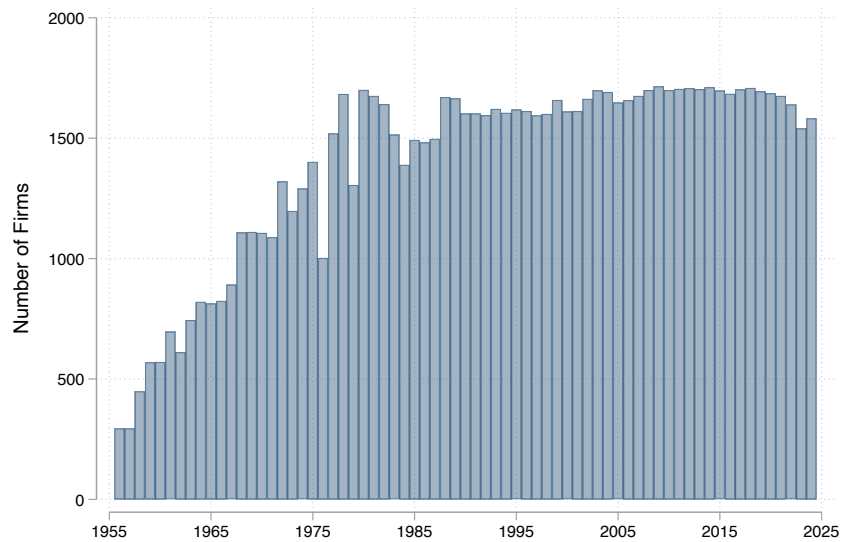
To extract actual realizations data, we take estimates which have been identified as actual data by Value Line (with actual indicator filled with an "A"). If the actual data is missing for a certain PERMNO-year, impute the data using actual values from Compustat. To be consistent with digitized historical digitized data, only keep forecasts with pricing date in quarter 1 of each year. If a CUSIP shows up multiple times in pricing quarter 1, we use the forecasts from the latest pricing date.

To combine the digitized and modern data, we simply append the datasets for non-overlapping years. For overlapping years (1988-1989), if a firm shows up in both digitized and modern data, we first use the forecast from modern data. In rare instances where a firm is not in the modern data but is in the historical data, we use the information from the historical data.

D.3 Value Line Data Coverage

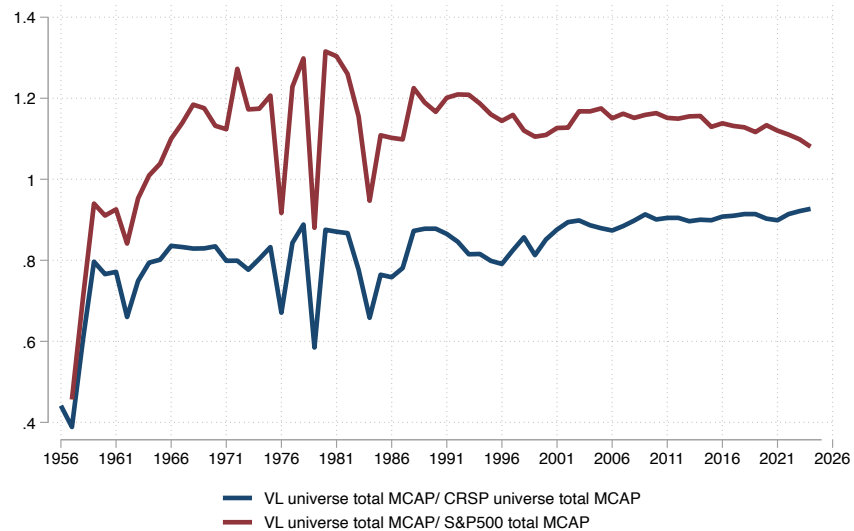
Firm coverage. Figure B1 plots the number of firms in our Value Line dataset in each year. In the early part of the sample (1950s), Value Line covers about 300-500 firms. Coverage rises to about 1000 firms in the mid-1960s and 1500-1600 by the mid-1970s. Figure B2 shows that VL firms account for 80% or more of the total CRSP market cap in most years after 1960.

Figure B1: Value Line Forecasts: Number of Firms by Year



Notes: This figure shows the number of unique firms covered in Value Line in the first quarter of each year.

Figure B2: Value Line Forecasts: Total Market Capitalization Relative to CRSP and S&P500 Universe by Year



Notes: This figure shows the total market capitalization of firms covered in Value Line in the first quarter of each year, relative to the total market capitalization of (i) firms in the CRSP universe and (ii) constituents of the S&P 500. The S&P 500 was started in 1957, so this ratio is missing for 1956.

Figure B3: Example of Value Line Original Data: General Motors, 1956

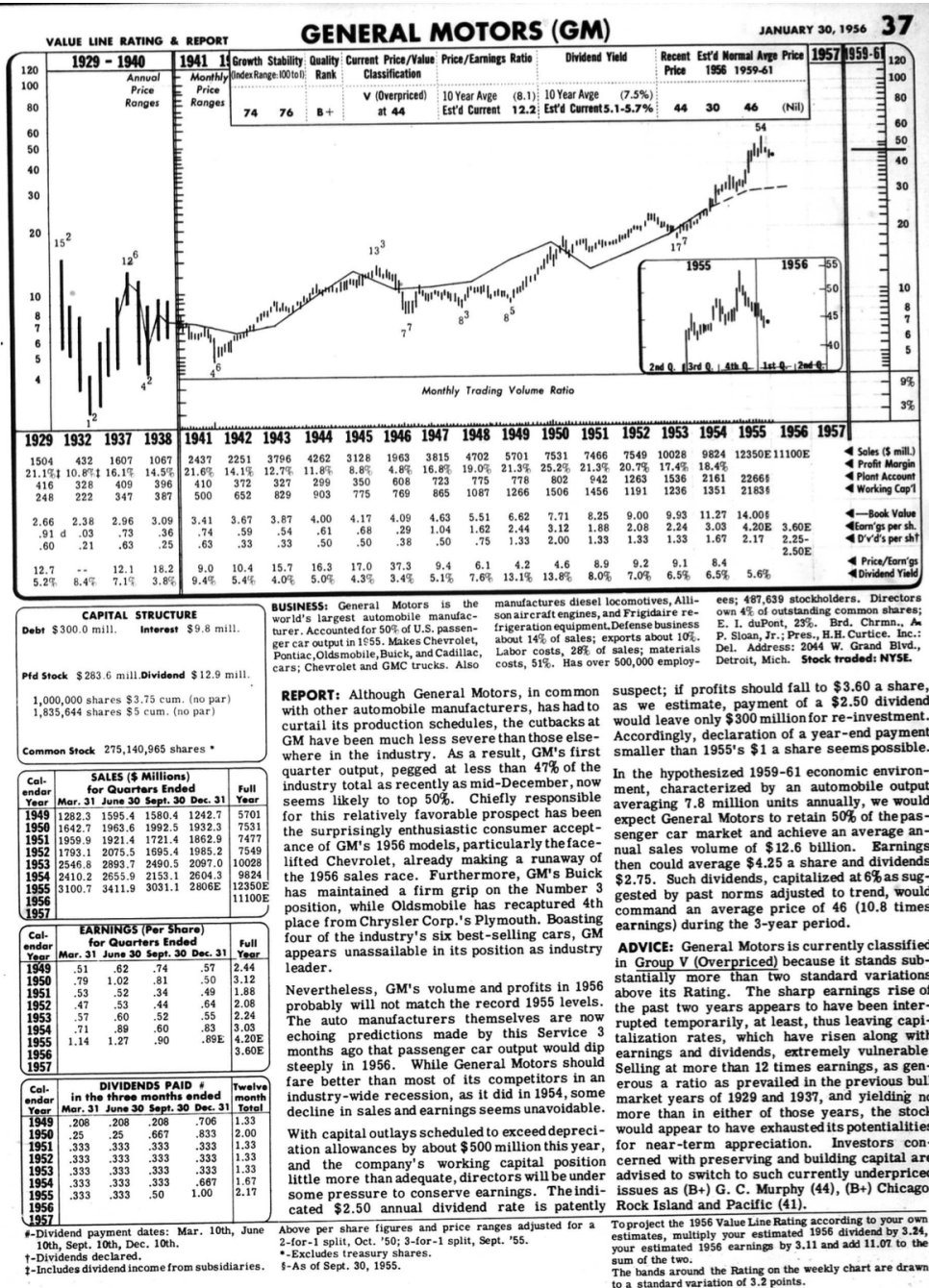


Figure B4: Example of Value Line Original Data: General Motors, 1975

