Estimating the Discount Rate of S&P 500 Portfolio With Cointegration Analysis

Kai Chen
SUNY Oneonta

Richard D. Marcus
University of Wisconsin-Milwaukee

Using cointegration analysis, this paper examines the evolution of the discount rate of S&P 500 portfolio from 1926 to 2019. By estimating on a 30-year time window moving over time, we find that the discount rate has gradually become significantly smaller. The results suggest that capital cost in the U.S. stock market, represented by the discount rate of S&P 500 portfolio, has been declining as time goes by, which implies that the U.S. stock market has become more informative and efficient, since the risk of a stock, which determines its capital cost, is associated with the stock's asymmetric information.

Keywords: discount rate, capital cost, market efficiency, cointegration

INTRODUCTION

The present value model is widely applied in valuation for a variety of assets. Theoretically, the fundamental value of an asset is determined by the present value of all cash flows that the asset will generate in its future life, *i.e.*, the sum of all cash flows discounted at an appropriate risk-adjusted rate of return. Mathematically, if investors predict that a firm issuing the stock will pay dividend (d_t) per share in time (t) in its future life, then the expected price of the firm's stock per share or the stock's intrinsic value would be estimated as:

$$V = \sum_{t=1}^{\infty} \frac{d_t}{(1+r)^t}. \tag{1}$$

In the model, discount rate, r, plays a connecting role that transforms future cash flows into the present value. The level of discount rate is associated with the extent of the risk of the firm's future cash flows. The higher risk inherent in the firm's business, the higher is the discount rate that investors would require as a compensation for holding the firm's stock. In this sense, the discount rate is also viewed, from the firm's perspective, as the capital cost that the firm should pay its equity investors in exchange for using the funds contributed by the investors. Furthermore, the risk of a stock is affected by the firm's degree of asymmetric information at both macroeconomic and firm-specific level. If there were more asymmetric information between a firm and its investors, the investors facing more uncertainty in future would take on higher risk, therefore requiring a higher discount rate as a compensation.

A financial market with more asymmetric information would be running less efficiently, as information about public firms may not be fully incorporated in their current stock prices. Since the discount rate is somewhat linked to the extent of asymmetric information through the stock risk, the average discount rate of all stocks in a market may measure the overall efficiency of the market. The change in the average discount rate over time may show the trend of the market efficiency. If the average discount rate in an economy becomes smaller, it indicates that the information environment at macroeconomic level improves and the financial market is working more efficiently. For instance, if the government of a country reinforces the regulations of firms' information disclosure to public investors, we expect that the macro-environment of information would make firms more transparent. This would reduce asymmetric information and the financial market would be running more efficiently. As a result, the average discount rate would be lowered. Since the discount rate, the rate of return required by investors for yielding their funds, measures the capital cost that firms should pay for using the funds, the average discount rate of all stocks would represent the average opportunity cost of funds, which would influence many types of interest rate as well as other important macroeconomic parameters. Thus, estimating the average discount rate and examining its evolution over time is meaningful for us to understand not only corporate finance but also financial markets and macroeconomy.

If a stock pays a dividend (*d*) per share constantly to the eternity, the equation (1) can be simplified to $V = \frac{d}{r}$. Theoretically, if the intrinsic value of a stock is reflected by its market price when the stock market reaches its equilibrium and the firm's dividend is expected to be stable in a long run, then the discount rate (*r*) is obtained by solving the simplified equation, and so is the firm's capital cost. Following the principle of the present value model, we estimate the discount rate by estimating the relationship between the stock price and its dividend, assuming this relationship is stable over the estimated period. In fact, the existing literature, pioneered by Lintner (1956), shows that dividend-smoothing behavior was widespread. Even though both the stock price and the dividend would fluctuate from time to time, the underlying relationship between the stock price and its dividend, defined by the real discount rate, would be stable if the stock market keeps reaching its equilibrium over time. Empirically, using a firm's time series data of its stock price (*STKPRC*) and dividend (*DIV*), we can run the following regression to estimate the stock's discount rate.

$$STKPRC_t = \alpha + \beta \times DIV_t + \varepsilon_t, \qquad (2)$$

where the reciprocal of β is the firm's discount rate, or equivalently from the firm's perspective, capital cost over the sample period, *i.e.*, $r=1/\beta$.

However, we are not interested in a specific firm's capital cost. Rather, we are interested in estimating the average capital cost of all firms in an economy that would help us understand the macro-information environment and the efficiency of a financial market. To this end, we construct a portfolio that includes all stocks on the market and then apply the regression model (2) to the time series data of the market portfolio's price and dividend. Thus, the reciprocal of the estimated β represents the average capital cost in an economy over the sample period.

An econometric issue arises at this moment: time series variables might present non-stationary processes. For example, they could be random walks, changing their variances or covariances as time goes by. If the ordinary least squares regression (OLS) is used to the model (2), the estimators might not normally distribute, resulting in a spurious regression. Therefore, the results could be misleading because the relationship between two variables estimated from non-stationary time series data might not truly exist.

A traditional approach to this issue is to difference time series variables. If a test fails to reject the hypothesis of non-stationarity of time series variables, we difference the variables in question before applying them in OLS regressions. Though this approach is acceptable, differencing may result in a loss of low-frequency (long-run) information that might be important to the long-run relationship between two variables (Miller, 1991).

Engle and Granger (1987) develop the cointegration analysis that provides a solution to this problem. Sometimes two variables are non-stationary, but a linear combination of those variables is stationary. When this is true, these two variables are said to be cointegrated. We can construct a relation model of these two cointegrated variables and estimate the model using the ordinary least squares regression. Engle and Granger (1987) prove that if the cointegration of two variables exists, the estimation of the parameters by OLS would be consistent. Furthermore, the residual of this regression then can be used to test whether the two variables are indeed cointegrated or not.

Applying cointegration analysis to the time series data of S&P 500 Index, which proxies for the U.S. stock market portfolio, we estimate the average discount rate in the U.S. stock market during a period from 1926 to 2019. We first do it for the full period to have an overall look at the average discount rate throughout this period. Then we estimate the average discount rate across a 30-year subperiod that moves along quarter by quarter to examine the evolution of the discount rate throughout the overall period from 1926 to 2019. A historical perspective of the average discount rate would show us not only the dynamics of U.S. firms' capital cost, but also the evolution of the efficiency of the U.S. stock market.

Our finding is straightforward; the discount rate has been generally declining throughout the past 94 years, though there are ups and downs in between. It indicates that the average capital cost of U.S. corporations has become smaller and that the U.S. stock market has become more transparent and more efficient over time. The conventional financial economics hypothesizes market efficiency in three forms: weak, semi-strong, and strong. The weak form suggests that today's stock prices reflect the information of all past prices, the semi-strong form contends that current stock prices integrate all publicly available information, and the strong form states that all information, whether public or private, is completely accounted for in current stock prices (Fama, 1970; Fama, 1998; Malkiel, 2003). However, the empirical literature has mixed evidence about how efficient the stock market is. Recent research has studied capital market efficiency from different perspectives. For instance, Hendershott and Moulton (2011) and Corwin and Schultz (2012) document narrowing of the bid-ask spreads of stocks. Chen *et al.* (2020) and Breugem and Buss (2019) study how investors' information acquisition affects market efficiency. This study contributes to the literature by looking at market efficiency from a historical and macroeconomic perspective; we use the average discount rate in a stock market to proxy for the market efficiency that evolves in history. Our result provides new evidence that the U.S. financial markets have become more informative and efficient in a long run.

The remainder of this article is organized as follows. The next section describes the data that we use to perform our analysis. The procedure of testing for cointegration and our estimates for the discount rates over time are illustrated in the empirical results. The last section concludes the study.

DATA DESCRIPTION

This study uses the dataset, *U.S. Stock Markets 1871-Present and CAPE Ratio*, which is downloaded from the Home Page of Robert J. Shiller.[1] Starting January 1871 and updated ongoing, this dataset contains monthly stock prices and dividends of a broad stock index portfolio, the Standard and Poor's (S&P) 500 Index. The stock price is monthly average of daily closing prices. The dividend data before 1926 are from Cowles and associates (1939), interpolated from annual data to monthly basis. Since 1926, the monthly dividend data are computed from the S&P four-quarter totals for each quarter, with linear interpolation to monthly frequency (Shiller, n.d.).

To avoid the interpolation of quarterly dividends between months, our analysis is performed on a quarterly basis so that the frequency in our analysis is exactly consistent with the most dividend issuance in the U.S. stock market. We choose the time series period that begins with 1926Q1, the first quarter when the dividend data without interpolation are available, and ends with 2019Q4, when the latest data are updated on the Home Page of Robert J. Shiller. The quarterly stock price is obtained by averaging the monthly prices across the quarter. The CPI-U (Consumer Price Index-All Urban Consumers) published by the U.S. Bureau of Labor Statistics is used to convert the stock prices and dividends to their real terms. We use the real stock prices and dividends to estimate the real discount rate.

EMPIRICAL RESULTS

The Engle-Granger cointegration analysis follows two steps. First, we test for the stationarity of the time series under consideration. If they are not stationary, we difference each series successively until stationary series emerge. Usually, one-order difference would be a stationary series. Second, we estimate cointegration regressions with OLS using variables with the same order of integration, and test for stationarity of residuals from cointegration regressions (Engle and Granger, 1987; Miller, 1991). Our analysis proceeds in this order.

Testing for Stationarity of a Time Series Variable

Dickey and Fuller (1979) develop a procedure for testing whether a variable follows a unit root process or, equivalently, a random walk. The null hypothesis is that the variable has a unit root, and the alternative hypothesis is that the variable is generated by a stationary process. We follow this methodology to test the stationarity of a time-series variable. If the null hypothesis is rejected, then we claim that the time series is stationary over time.

We begin with testing for the stationarity of two key variables in our study. The Standard and Poor's 500 Index is used to measure the price of the S&P 500 portfolio. The real price, named as *SPPRC*, is obtained as follows. In Robert Shiller's dataset, the monthly averages of daily closing indexes have been adjusted by the Consumer Price Index into the real term.[2] We further average the monthly prices across the quarter to represent the quarterly real price of S&P 500 portfolio. The term *DIV* stands for the dividends paid by S&P 500 firms in the past four-quarters adjusted by CPI. For each time series, we calculate three *t*-statistics for zero mean, single mean, and trend, and two *F*-statistics in the single mean case and the trend case to test the unit-root hypothesis. In general, given a time series variable denoted by *y*, the three *t*-statistics for zero mean, single mean, and trend are assumed to be computed from the following models, respectively.

$$y_t = \alpha_1 y_{t-1} + e_t \tag{3}$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + e_t \tag{4}$$

$$y_t = \alpha_0 + \gamma t + \alpha_1 y_{t-1} + e_t \tag{5}$$

Let $\hat{\alpha}_1$ be the estimated regression coefficient for the first lag of the series and let $se_{\hat{\alpha}_1}$ be the standard error of $\hat{\alpha}_1$. Then the *t*-statistic for the null hypothesis that the variable has a unit root is:

$$t = \frac{\hat{\alpha}_1 - 1}{se_{\hat{\alpha}_1}}. \tag{6}$$

In empirical analysis, we carry out the Dickey-Fuller unit root test by estimating the unrestricted regression:

$$y_t - y_{t-1} = \alpha_0 + \gamma t + (\alpha_1 - 1) y_{t-1} + e_t. \tag{7}$$

Then the test boils down to computing the *t*-statistics for H_0: $(\alpha_1 - 1) = 0$ in three cases: in the case of zero mean, we set both α_0 and γ equal to zero; in the case of simple mean, we set γ equal to zero; and in the case of trend, we assume that both α_0 and γ are not equal to zero. In addition, we compute the *F*-statistic to test H_0: α_0=0 and $(\alpha_1 - 1) = 0$ in the simple mean case and the *F*-statistic to test H_0: γ=0 and

$(\alpha_1 - 1) = 0$ in the trend case. The *F*-statistic provides supplementary evidence to support the null hypothesis if it is not significant at the traditional level.

Further, we perform the Augmented Dickey-Fuller tests by expanding the regression above to include four lagged changes $\sum_{j=1}^{4} \lambda_j \Delta y_{t-1}$ on the right-hand side of the equation. The *t*-statistics and *F*-statistic are defined in the same way.

Table 1 reports the results of the Dickey-Fuller (DF) and the Augmented Dickey Fuller (ADF) tests for the stationarity of *SPPRC* and *DIV* and their first differences. The *p*-value is reported in parenthesis below each statistic.[3] In Panel A where the results of the level regressions are displayed, we find that no *t*-statistics in DF and ADF tests rejects the null hypothesis of non-stationarity for both *SPPRC* and *DIV* at 1% significance level. The *F*-statistics also provides the supplementary evidence for the non-stationarity in almost all tests, except for *SPPRC* in the case of single mean and for *DIV* in the cases of single mean and trend. Generally, the results do not reject the null hypothesis of non-stationary for the time series *SPPRC* and *DIV*.

TABLE 1
TEST FOR STATIONARITY

	t-statistic in zero mean	*t*-statistic in single mean	*F*-statistic in single mean	*t*-statistic in trend	*F*-statistic in trend
Panel A: level variable					
Part 1: DF Test					
SPRRC	3.9039	2.7738	7.64***	0.7477	4.31
	(1.00)	(1.00)	(<0.01)	(1.00)	(0.31)
DIV	6.8293	6.0307	29.71***	4.1912	18.82***
	(1.00)	(1.00)	(<0.01)	(1.00)	(<0.01)
Part 2: ADF Test					
SPRRC	2.0925	1.2182	2.21	-0.6493	2.04
	(0.99)	(1.00)	(0.51)	(0.98)	(0.77)
DIV	2.113	1.5045	2.46	-0.2947	2.00
	(0.99)	(1.00)	(0.44)	(0.99)	(0.78)
Panel B: 1st difference variable					
Part 1: DF Test					
SPRRC	-13.0783***	-13.2644***	87.98***	-13.4696***	90.73***
	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)
DIV	-9.5057***	-9.834***	48.36***	-10.1845***	51.89***
	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)
Part 2: ADF Test					
SPRRC	-6.6952***	-6.9011***	23.83***	-7.1596***	25.67***
	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)
DIV	-5.1281***	-5.405***	14.62***	-5.7593***	16.66***
	(<0.01)	(<0.01)	(<0.01)	(<0.01)	(<0.01)

This table shows the results of the tests for stationarity of two time-series variables, *SPRRC* and *DIV*, from 1926Q1 to 2019Q4, inclusive. Panel A shows the results of the tests for level variables and Panel B shows the results of the tests for first difference variable. *p*-values are put in the parentheses below the statistics. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

In Panel B, the results of first order differencing show that the *t*- and *F*-statistics in both DF and ADF tests reject the null hypothesis of non-stationarity at 1% significance level. These results suggest that *SPPRC* and *DIV* are stationary series in first differences. In sum, we conclude that using the level variables of *SPPRC* and *DIV* in OLS regressions might provide misleading statistical evidence of a linear relationship between *SPPRC* and *DIV*. We should apply the cointegration analysis to avoid the spurious regressions.

TABLE 2
ANALYSIS OF COINTEGRATION REGRESSIONS

Time Period	α	β	Discount Rate	$\bar{R}^2$	DW	DF	ADF
Panel A: Overall Period							
1926Q1-2019Q4	-521.7137***	63.3702***	1.58%	0.8491	0.0566	-2.3335**	-2.9935***
	(<0.01)	(<0.01)				(0.02)	(<0.01)
Panel B: Subperiod							
1926Q1-1955Q4	36.9218	16.1223***	6.2%	0.321	0.2047	-1.9661**	-3.3948***
	(0.15)	(<0.01)				(0.05)	(<0.01)
1936Q1-1965Q4	-250.9988***	42.2049***	2.37%	0.7874	0.164	-2.4766***	-2.4302**
	(<0.01)	(<0.01)				(0.01)	(0.02)
1940Q1-1969Q4	-272.9575	43.819	2.28%	0.8794	0.1279	-1.9619**	-2.2806**
	(<0.01)	(<0.01)				(0.05)	(0.02)
1959Q4-1989Q3	-457.908***	51.6345***	1.94%	0.4398	0.1318	-2.2675**	-2.5692***
	(<0.01)	(<0.01)				(0.02)	(0.01)
1969Q1-1998Q4	-1730.92***	117.0567***	0.85%	0.7236	0.0827	-0.0363	-1.9992**
	(<0.01)	(<0.01)				(0.67)	(0.04)
1975Q4-2005Q3	-2706.5***	168.342***	0.59%	0.7308	0.0574	-1.1652	-2.1958**
	(<0.01)	(<0.01)				(0.22)	(0.03)
1980Q1-2009Q4	-1149.58***	94.6798***	1.06%	0.5335	0.0578	-1.3924	-1.7814*
	(<0.01)	(<0.01)				(0.15)	(0.07)
1990Q1-2019Q4	79.3076	48.9855	2.04%	0.6779	0.0701	-1.8142*	-2.1938**
	(0.79)	(<0.01)				(0.07)	(0.02)

This table shows the results of the tests for the cointegration regression. We first run the OLS model,

$$SPPRC_t = \alpha + \beta \times DIV_t + \varepsilon_t, \quad (8)$$

The error from the regression is recovered to perform the test for the cointegration between *SPPRC* and *DIV*. The Augmented Dickey-Fuller (ADF) non-stationarity test is performed with the following model:

$$d\mu_t = \sigma_1 \mu_{t-1} + \sum_{i=1}^{4} \sigma_{1=i} d\mu_{t-i} + e_t, \quad (9)$$

where μ_t is the error term from the model (8) and e_t is the error term in the model (9). The null hypothesis of non-stationarity is rejected when σ_1 is significantly different from zero. The Dickey-Fuller (DF) test for non-stationarity is conducted by removing the summation. We also report the adjusted coefficients of determination ($\bar{R}^2$) and the Durbin-Watson statistics (DW) from the cointegration regressions. *p*-values are put in the parentheses below the statistics. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Testing for Cointegration

After verifying the stationarity of the first difference of *SPPRC* and *DIV*, we regress model (2) discussed in the introduction. If cointegration exists between *SPPRC* and *DIV*, then the OLS estimator is normally distributed and the estimated relationship is not spurious. Two time series variables are cointegrated if the residual from a linear regression between the two variables is stationary. Therefore, testing for cointegration boils down to testing for stationarity of the residual. We run model (2) and then use Dickey-Fuller approach to check the cointegration between *SPPRC* and *DIV*. Since it is very likely for two time series not to be cointegrated when the adjusted coefficient of determination ($\bar{R}^2$) exceeds the Durbin-Watson (DW) statistic (Granger and Newbold 1974; Plosser and Schwert 1978), we also report DW statistic along with $\bar{R}^2$ for reference.

We first examine the cointegration between *SPPRC* and *DIV* over the full period from 1926Q1 to 2019Q2 and report the results in Panel A, Table 2. Then, we apply the model to a series of 30-year subperiods (moving time windows) to look at the change of the cointegration over the full period. Specifically, we move a window of 30 years, quarter by quarter, along the full period and test the cointegration on each time window. This process produces 257 consecutive windows, with the first window spanning from 1926Q1 to 1955Q4 and the last window covering from 1990Q1 to 2019Q4. In Panel B, Table 2, we present the results of eight windows chronologically, two of which are the first and last windows, with the rest randomly chosen from each decade between 1930s and 1980s.

The results in Panel A, Table 2, show that both DF and ADF tests reject the null hypothesis of a non-stationary residual at the 5% or less significance levels for the full period. Though $\bar{R}^2$ exceeds DW statistic by a wide margin, the strong results from the DF and ADF tests indicate that *STKPRC* is cointegrated with *DIV*, and the regression over the full time series sample is not spurious.

For the subperiods shown in Panel B, Table 2, we find that all ADF tests and five out of eight DF tests reject the non-stationarity of the residuals at the 10% or less significance levels. Considering the large differences between $\bar{R}^2$ and DW statistic for these eight randomly chosen subperiods, we should look at the cointegrations for all 257 windows over the full period.

TABLE 3
NUMBER OF WINDOWS WITH SIGNIFICANT TESTS

Significance Level	ADF Test	DF Test	DF or ADF Test
1%	112	29	115
5%	199	112	199
10%	238	160	238

This table shows the number of subperiods that have significant DF, ADF tests, or either of the two at different significance levels.

Table 3 reports the number of windows that have significant DF, ADF tests, or either of the two at different significance levels. We find that out of 257 windows, 238 have significant ADF tests at 10% level. Combining the results for the overall period, we maintain that the cointegration between *SPPRC* and *DIV* is relatively stable as the 30-year time window is moving along.

Estimating the Discount Rate

After verifying the cointegration, we proceed to the economic interpretation of the estimated coefficient of *DIV* in the OLS model. In Table 2, computing the reciprocal of a coefficient of *DIV* yields the estimated discount rate across a certain period. For instance, by inverting 63.3702 in Panel A, we estimate the discount rate of S&P 500 portfolio throughout the full period as 1.58%.

However, we are interested in the evolution of the discount rate over time. The estimated discount rates over the moving windows, shown in Panel B of Table 2, allow us to have a look at the change in the discount rate. Generally, we find a trend that the discount rate has been declining gradually from 6.2% in 1926Q1-1955Q4 window to 2.04% in 1990Q1-2019Q4 window. To better understand this trend, we plot, in Figure 1, the estimated discount rates on all 257 windows against the timeline marked by the beginning quarter of each moving window. The solid curve in Figure 1, representing the discount rate of S&P 500 portfolio, had decreased from the beginning until it turned to a weak reversal that emerged around 1970s. The overall trend is characterized by the dotted straight line with a pronounced downward slope. We estimate the trend line with the following model:

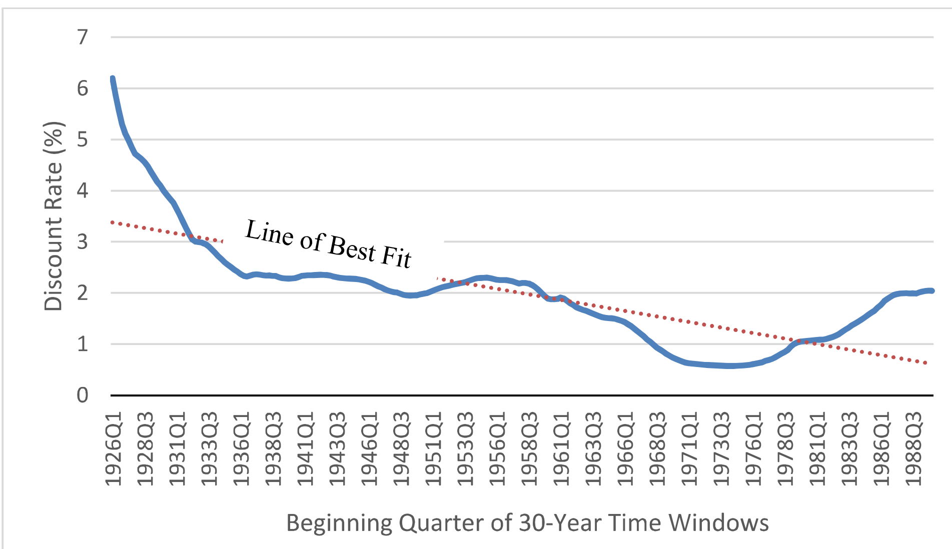
$$Rate_t = \alpha + \beta \times Quarter_t + \varepsilon_t \, , \tag{10}$$

where $Rate_t$ is the discount rate estimated from the model (8) on moving windows and $Quarter_t$ represents the beginning quarter of moving windows with its value changing from one to 257.

The results of this estimation show that the estimated value of the coefficient of $Quarter_t$ is -0.0108. Having a *p*-value less than 0.01, the coefficient of $Quarter_t$ is significant at the 1% level. The economic implication of this result is that the discount rate of S&P 500 portfolio decreases by 0.01% per quarter on average from 1926Q1-1955Q4 subperiod to 2.04% in 1990Q1-2019Q4 subperiod.

In general, Figure 1 shows solid evidence that the average discount rate of U.S. stocks, represented by S&P 500 portfolio, has been declining over nearly one hundred years. The decrease in the discount rate in the long run indicates that the capital cost has become smaller and the stock market has become more transparent and efficient in the United States.

FIGURE 1
DISCOUNT RATES ESTIMATED ON MOVING WINDOWS



The solid curve represents the discount rate estimated on a 30-year subperiod that moves over the full sample period from 1926Q1 to 2019Q4. The dotted straight line is the line of best fit of the solid curve.

CONCLUSION

The discount rate contains a great deal of information. The average discount rate in a stock market is informative of opportunity cost, systematic risk, and market efficiency. A long-term discount rate can be estimated by regressing the stock price against the dividend. However, the problem of non-stationarity of time series variables might result in spurious relationships between the variables. Application of the cointegration analysis enables us to deal with this problem. In this paper, we use the cointegration analysis to examine the relationship between the stock price of S&P 500 portfolio and its dividend from 1926 to 2019. We find that the discount rate of S&P 500 portfolio has been on a declining trend during the last 94 years. With this portfolio proxying for the stock market, this result indicates that the average capital cost of U.S. public firms has become smaller and the financial market in the U.S. has become more transparent and efficient. Our study contributes to the literature of capital cost and market efficiency from a historical perspective.

ENDNOTES

1. Campbell and Shiller (1987) use the dataset to estimate the discount rate in U.S. over a period from 1871 to 1987. For detailed description of the data, see the website: http://www.econ.yale.edu/~shiller/. The same data are also published on the website: https://datahub.io/core/s-and-p-500.
2. The dataset is updated quarterly. At each update, the real terms are calculated with an estimated CPI for the last month in the last quarter. In our study, the value of the estimated CPI for June 2020 is 256.4.
3. We use a SAS procedure, PROC ARIMA, to perform the Dickey-Fuller Test. PROC ARIMA gives same test statistics as OLS regressions, but correct *p*-values due to appropriate 5% left tail critical value -2.86 (Fuller, 1996, p.642).

REFERENCES

Breugem, M., & Buss, A. (2019). Institutional investors and information acquisition: implications for asset prices and informational efficiency. *Review of Financial Studies*, *32*(6), 2260–2301.

Campbell, J. Y., & Shiller, R. J. (1987). Cointegration and tests of present value models. *Journal of Political Economy*, *95*(5), 1062-1088.

Chen, Y., Kelly, B., & Wu, W. (2020). Sophisticated investors and market efficiency: Evidence from a natural experiment. *Journal of Financial Economics*, forthcoming.

Corwin, S.A., & Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance*, *67*(2), 719-760.

Cowles, A., & Associates. (1939). *Common Stock Indexes*. Bloomington, Indiana: Principia Press.

Dickey, D.A., & Fuller, W. A. (1979). Distribution of estimates of autoregressive time Series with unit root. *Journal of the American Statistical Association*, *74*, 27-31.

Engle, R.F., & Granger, C.W.J. (1987). Co-integration and error correction: representation, estimation and, testing. *Econometrica*, *55*(2), 251-276.

Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*(2), 383–417.

Fama, E.F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306.

Fuller, W.A. (1996). *Introduction to statistical time series*. New York: John Wiley & Sons, Inc.

Granger, C.W.J., & Newbold, P. (1974). Spurious regression in econometrics. *Journal of Econometrics*, *2*, 111-120.

Hendershott, T., & Moulton, P.C. (2011). Automation, speed, and stock market quality: The NYSE's hybrid. *Journal of Financial Markets*, *14*(4), 568-604.

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review*, *46*(2), 97-113.

Malkiel, G.B. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59-82.

Miller, S.M. (1991). Monetary dynamics: An application of cointegration and error-correction modeling. *Journal of Money, Credit, and Banking*, *23*(2), 139-154.

Plosser, C.I., & Schwert, G.W. (1978). Money, income, sunspots: Measuring economic relationships and the effects of differencing. *Journal of Monetary Economics*, *4*, 637-60.

Shiller, R.J. (n.d.). *Online data - Robert Shiller*. Retrieved from http://www.econ.yale.edu/~shiller/data.htm.