Is stock return predictability of sectoral labor reallocation shocks time-varying?
Statement of Originality

This document is written by Student Sofia Frascali who declares to take full responsibility for the contents of this document.

I declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it.

The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.
Abstract

This paper investigates the variation in the predictability of sectoral labor reallocation for stock returns over time. The main finding of this research is that the predictability of sectoral labor reallocation and other variables, which are considered by the academic literature to be good predictors, change over time. The results indicate that predictability is unstable over time and changes according to the business cycle. Particularly, when the industry production growth rate increases, the future excess return decreases even more, when sectoral labor reallocation shocks increase at long horizons. Additionally, an increase in the unemployment rate changes the effect of sectoral labor reallocation shocks for future stock returns from negative to positive.
# Table of contents

1. **Introduction** ............................................. 4

2. **Literature review** ....................................... 7
   2.1 Conceptualization of sectoral labor reallocation shocks and unemployment growth .............................................. 7
   2.2 Other labor market related return predictors ............................................. 9
   2.3 Time-varying stock return predictability for non-labor market related variables ............................................. 11
      2.3.1 Time-varying predictability and the business cycle ............................................. 12
      2.3.2 Stock return predictability and structural break ............................................. 13
   2.4 The predictive power of CSV over time ............................................. 14

3. **Data** ..................................................... 16

4. **Methodology** ............................................ 17
   4.1 The CSV measure ............................................. 17
   4.2 Stock returns predictability of CSV ............................................. 18
   4.3 The instability of the predictive power of CSV over time ............................................. 18
   4.4 Predictive regression with industry production growth ............................................. 19
   4.5 Predictive regression with unemployment growth ............................................. 19

5. **Empirical results** ....................................... 21
   5.1 Predictability of CSV and other stock return predictors ............................................. 21
   5.2 IS and OOS performance over time ............................................. 24
   5.3 Predictability of CSV and industry production growth ............................................. 32
   5.4 CSV stock return predictability and unemployment change ............................................. 34

6. **Robustness Checks** ..................................... 37

7. **Conclusion** ............................................... 44

References ...................................................... 46
1. Introduction

The existence of stock return predictability for variables, such as the dividend yield and the term spread, is one of the most debated issues in the empirical literature (McMillan, 2014). In particular, previous findings show a strong link between labor conditions and stock return predictability. Five main labor market related variables that predict the aggregate stock market return have been identified. These variables include the payroll growth, the net hiring rate, the net job creation rate, sectoral labor reallocation shocks and announcements of unemployment (Chen & Zhang, 2011; Eiling, Kan & Sharifkhani, 2016). These factors have been studied empirically and most of them are statistically significant. However, there are only a few studies that analyze predictability in a time-series context (McMillan, 2014). More importantly, there is no conclusive evidence on whether stock return predictability changes over time, which motivates additional research.

This paper will mainly focus on sectoral labor reallocation shocks, because this variable is superior to other labor market related return predictors (Eiling, Kan and Sharifkhani, 2016). It predicts future stock market excess returns through labor adjustment costs, which lower future returns. Indeed, search and training costs are higher, when workers are hired from different industries, as each industry has a certain degree of specialized labor (Davis, 1987). Thus, when there is a higher need for sectoral labor reallocation, the labor adjustment costs will rise. Higher labor adjustment costs in turn reduce the returns to hiring, which results in lower future returns (Eiling, Kan and Sharifkhani, 2016).

The expectation is that the predictive power of sectoral labor reallocation shocks is time-varying and that this time-variation depends on the business cycle. In other words, the predictive power of this variable changes depending on whether there is an economic expansion or recession. This is because part of the excess stock returns is determined by unpredictable random shocks that fluctuate with the business cycle. Indeed, the state of the market and the economy may have a different effect on the predictive power of sectoral labor reallocation shocks during recessions than in good times (McMillan, 2014; Devpura, Narayan & Sharma, 2018). Therefore, the expectation is that the predictive power of sectoral labor reallocation shocks changes over time and that this change depends on specific economic and market conditions, namely industry production and unemployment. This paper investigates whether the predictive power of sectoral labor reallocation shocks changes depending on the level of industrial production growth, which
can be regarded as a proxy for the business cycle. Furthermore, the effect of the unemployment rate on the predictive power of sectoral labor reallocation will be studied.

In the empirical literature, a positive and statistically significant link between sectoral labor reallocation shocks and the unemployment rate has been detected (Lilien, 1982). Usually, high performing industries increase their employment level and low performing industries decrease their employment level. Nevertheless, when there is a high need for sectoral labor reallocation, high performing industries do not increase their employment level right away, which leads to a rise in aggregate unemployment. Hence, sectoral labor reallocation shocks lead to an increase in the unemployment rate (Davis, 1987). The relationship between unemployment and sectoral labor reallocation shocks depends on the opportunity costs of employment, which in turn depends on the business cycle. In point of fact, sectoral reallocation of specialized labor is often followed by an increase in asymmetric labor adjustment costs, which inhibit the immediate move of workers from contracting to expanding industries. This process rises the unemployment levels. In turn, these costs depend on the business cycle, as they rise during economic expansions and decrease during recessions (Davis, 1987; Chodorow-Reich & Karabarbounis, 2016). The impact of unemployment growth on sectoral labor reallocation shocks will be studied, because the predictability of sectoral labor reallocation shocks for stock returns might depend on it. Also, economic expansions are often characterized by drops in unemployment and economic recessions by a rise in unemployment (Alexopoulos, 2004; Lawrence & Trabandt, 2010). Hence, unemployment is an indirect indicator of the business cycle, as high levels of unemployment denote a recession, whereas low levels of unemployment imply an economic expansion.

This leads to the following research question: “Does the predictive power of sectoral labor reallocation shocks for stock return change according to the business cycle?” If the results indeed show that the predictability of the factors mentioned above change over time, then it needs to be established whether it changes with the business cycle or indirectly with the unemployment rate. Once this is established, then the effect of an economic expansions versus economic contractions will be analyzed.

This paper will proceed as follows. First, it will be examined whether stock return predictability changes over time. Once this is established, the determinants of variation in predictability will be investigated. Two main factors that are related to the business cycle and that might affect the predictability of stock return will be investigated, namely industry
production growth and the unemployment rate. One would expect that variations in the predictability of sectoral labor reallocation shocks coincide with changes in the unemployment rate and in the industry production growth.

The motivation behind this thesis for investigating the time-variation in the predictive content of sectoral reallocation shocks for stock returns follows from an extensive empirical literature reporting evident correlations between stock market returns and labor market related variables (Bazdreh, Belo, & Lin, 2014). It will shed light on the debate whether return predictability is time-varying and also on the nature of the relationship between stock return predictors and macro-economic variables. The investigation of the predictive power for sectoral reallocation costs over time is of interest for both investors and policymakers because it is important for them to understand how strong the predictive power is of the variables they use to forecast stock returns. The results of this paper will further our knowledge of stock price predictability.

The main finding of this paper is that the predictive power of sectoral labor reallocation shocks is time-varying. In particular, the performance of this variable seems to vary around specific recession dates, such as the crisis of 1990-1991, the market decline of 1999-2001 and the financial crisis in 2008. Furthermore, the results show that for long horizons, industry production growth, which proxies the business cycle, affects the predictability of sectoral labor reallocation shocks over time. That is, higher values of output growth strengthen the negative effect of sectoral labor reallocation shocks on future excess returns. In addition, the predictive power of sectoral labor reallocation shocks, changes dramatically depending on the unemployment rate, which fluctuates with the business cycle. Notably, the negative effect of sectoral labor reallocation shocks on future excess returns becomes positive. In other words, sectoral labor reallocation shocks increase future stock returns, when unemployment growth is high. In sum, by linking sectoral labor reallocation shocks to business cycle related variables, it can be confirmed that its predictive power changes over time.
2. Literature review

Many researchers have studied stock return predictability in a time-series setting in many ways. This has led to inconclusive and sometimes ambiguous results (Timmerman, 2008; Van Nieuwerburgh & Lettau, 2007). The failure to provide conclusive evidence of predictability has led to different approaches in the empirical finance. Some scholars developed new econometric methods (Stambaugh, 1999; Lewellen, 2004), while others extended the time-series regression to panel data or to time-varying predictive regression models (Westerlund & Narayan, 2014). In this section the existing literature that studies time-varying predictability will be reviewed. First, the conceptualization of sectoral labor reallocation will be reviewed, after which other labor market related variables that predict stock returns will be discussed. The chapter ends with the analyses of different approaches undertaken to examine the time-variability of the predictive power of many predictive variables for stock return.

2.1 Conceptualization of sectoral labor reallocation shocks and unemployment growth

This section illustrates sectoral labor reallocation shocks as stock return predictor and its link with labor adjustment costs and unemployment.

Macroeconomic shocks affect labor demand and supply across different sectors, which changes the ideal allocation of human capital across industries. Indeed, the need for labor reallocation changes over time, as there are periods where labor demand increases unequally across sectors. Also, during recessions the opportunity costs of changing jobs are lower and the asymmetry between resources across industries is higher making it easier to transfer workers from one industry to another. Thus, labor reallocation across sectors is higher during economic contractions and lower in good times. (Black, 1995; Lilien, 1982).

Sectoral reallocation shocks predict future stock returns through labor adjustment costs (Eiling, Sharifkani, & Kan, 2016). These costs are generated, because the reallocation of specialized labor from declining industries to expanding industries is time-consuming and costly (Davis, Fluctuations in the pace of labor reallocation, 1987). Hiring workers from other industries is costlier and slow, when workers are tied to the industry (Black, 1995). Hence, when the necessity for sectoral labor reallocation is high, labor adjustment costs will increase. In other
words, sectoral reallocation shocks reduce future stock returns, through labor adjustment costs, because it leads to lower returns to hiring (Eiling, Sharifkani, & Kan, 2016).

Lilien (1982) uses the cross-sectional dispersion in the employment index as a proxy for sectoral labor reallocation shocks. However, Abraham and Katz (1986) have questioned this measure, because a rise in the employment dispersion does not always reflect an increase in labor reallocation (Loungani & Trehan, 1997). Therefore, many scholars use the cross-sectional dispersion in industry stock returns as a proxy for sectoral labor reallocation shocks. For instance, when new information is disclosed about industry-specific profitability, the stock price dispersion increases. This process leads to a change in labor demand from one sector to another, followed by the reallocation of labor across sectors. Hence, there will be a greater need for sectoral labor reallocation, when there is a higher return dispersion across industries. (Loungani, Rush, & Tave, 1990; Loungani & Trehan, 1997). In line with these findings, Eiling et al. (2016) use cross-sectional volatility (CSV) as a proxy for sectoral labor reallocation shocks and find that it has a highly significant and robust predictive power for stock returns. This paper will follow the Eiling et al. (2016), by using the CSV as proxy for sectoral labor reallocation shocks. CSV is related to labor adjustment costs, as the latter are determined by dispersion in productivity shocks across sectors, which can be quantified by CSV.

The sectoral shift hypothesis suggests that sectoral labor reallocation shocks and unemployment growth are interrelated (Davis, 1987; Lilien , 1982). Indeed, sectoral labor reallocation shocks, are followed by the necessity to reallocate human capital from declining to expanding industries. Sectoral shifts affect the unemployment rate, because labor allocation is time-consuming, as workers that are laid off do not find a new job right away. This means that the level of unemployment depends on the time it takes for workers to be reemployed. In addition, when workers are attached to the industry, because of specific skills, they will be less eager to find new jobs in other sectors (Black, 1995). Therefore, after sectoral labor reallocation shocks, the adjustment process is slow and usually leads to temporary unemployment (Lilien, 1982). Thus, sectoral labor reallocation shocks increase unemployment. This is in line with the findings by Eiling et al. (2016), who find that cross-sectional volatility (CSV) predicts higher aggregate unemployment growth. They state that expanding sectors usually increase employment, while declining sectors lower the unemployment level. However, when necessity for labor reallocation is high, expanding industries do not increase employment instantaneously.
leading to an increase in unemployment. They show that CSV is a statistically significant and a robust predictor of unemployment growth.

At the same time, the expectation is that unemployment affects the predictive power of CSV, because it is an indirect proxy for the business cycle. Many scholars associate higher levels of unemployment growth with economic recessions and low unemployment growth levels with economic expansions (Alexopoulos, 2004; Lawrence & Trabandt, 2010). Since it is expected that the predictive power will be stronger during recessions, it must be the case that higher levels of unemployment weaken the predictive power of CSV.

### 2.2 Other labor market related variables as stock return predictors

In this section, other stock return predictors that are related to the labor market will be illustrated. These variables include the payroll growth, the hiring rate, the net job creation rate, the announcement of rising unemployment and labor income.

Recently, researchers have investigated the relationship between stock return and labor market related variables. Chen and Zhang (2011) investigate three labor market variables, namely payroll growth, the hiring rate and the net job creation rate. They argue that payroll growth and the net job creation rate forecast future market returns. However, their results indicate that the net hiring rate is not a statistically significant predictor. In contrast, Bazdrech et al. (2014) find a significant and robust link between the hiring rate and stock returns. They show that firms with higher labor hiring rates have lower future stock returns. The intuition behind this predictive variable, is that the hiring and the investment decisions of a firm determine its profits, which includes rents from capital and labor. Firms that have higher hiring rates, incur higher labor adjustment costs and will take advantage of macroeconomic shocks that lower labor adjustment costs. Therefore, firms with higher hiring rates are relatively less risky and have lower expected future returns.

Boyd et al. (2005) provide empirical support showing that news of rising unemployment increases stock returns during economic expansions and reduces stock returns during economic contractions. They argue that news of rising unemployment discloses information about two variables that affect future stock returns, namely the future interest rate and future profits and dividends (expected growth). An increase in unemployment indicates a decrease in interest rates, which rises the stock price, and a decline in expected growth, which lowers the stock price. Thus,
at the announcement of rising unemployment, these two effects move in opposite directions. Depending on the stage of the business cycle, the weight of the two effects changes over time. During economic expansions, the effect on expected growth following the news is weaker than the good news effect of lower expected interest rates. Therefore, stock prices react positively to the news of increasing unemployment. During contractions, the opposite happens, and the relevance of the two effects is reversed. The explanation for this phenomenon is the following. Boyd et al. (2005) find that during economic contractions only stock prices decrease at the announcement of rising unemployment but bond prices do not change significantly. This indicates that during economic contractions the announcement of increasing unemployment does not disclose relevant information about future interest rates, but only about future expected growth. Kreuger and Fortson (2003) investigate the market reaction to the availability of more relevant information regarding unemployment. In line with the finding by Boyd et al. (2005), they find that announcements of rising unemployment have a significant effect on market prices.

Many scholars find a relationship between labor income and stock return. For instance, Campbell (1996) adds human capital to the CAPM model, as he considers it a principal component of wealth. He states that by omitting human capital the CAPM risk estimation for stock investments is too large and the estimation of the coefficient of risk aversion is too small. They find that with the inclusion of human capital to the model, 28% of the variation in returns can be explained by the CAPM model. Jagannathan and Wang (1996) state that the stock-index return factor does not entirely capture labor-income risk. Therefore, they develop an asset pricing model with multiple betas, where one beta is the sensitivity of an asset’s return to growth rate per capita of labor income. Their study finds a negative correlation between stock return and the per capita labor income growth rate. Finally, Santos and Veronesi (2005) show with a general equilibrium model that labor income predicts stock returns through consumption. The intuition behind this finding is the following. Since financial assets are only a small fraction of consumption and most of it is determined by labor income, it must be the case that the correlation between financial assets and consumption is very small. Therefore, investors demand a small premium for holding financial assets. Consequently, stock returns can be predicted by the labor income to consumption ratio. They find that the labor income to consumption ratio significantly and positively predicts stock returns, especially at long horizons. The reason is that
the higher the fraction of income that is determined by labor income, the higher the premium that investors require and thus the higher the returns.

2.3 Time-varying stock return predictability for non-labor market related variables

This section presents previous studies that investigated the change in the predictive power of variables that forecast stock returns. In addition, different models employed to study time-varying predictability will be illustrated. As already mentioned, there are several studies in the empirical literature that focus on the change in the predictive power of variables over time. However, none of these variables are related to the labor market. Some scholars show evidence of time-varying predictability by studying the predictive power for different subperiods of their sample. Chen (2009) reports evidence of time-varying predictability by studying the predictability of the dividend growth rate over time. He argues that the predictability of the dividend growth is often insignificant, because its predictability changes dramatically over time. He finds that for most variables there is no predictability in the first seven decades between 1872-2005 but become predictable in the postwar period. The predictability of dividend growth shows opposite patterns over time. Ang and Bekaert (2007) test for time-variation in the predictive power of their variables, by dividing the sample period in different subperiods. They find statistically significant evidence for a time-varying pattern in the predictability of their variables. For instance, they show that excess return predictability by dividend yield is not statistically significant at longer horizons and its coefficient is twice as large if the 1990’s are excluded from the sample period.

Other researchers investigate the variation in the predictive power over time by using new econometric methods. For instance, Bannigidadmath and Narayan (2016) extend a generalized least squared estimator (GLS) predictive regression to a time-varying model to examine predictability over time. They find that the factors that change predictability over time are financial ratio shocks, both expected and unexpected. They regress the predictive variables on expected and unexpected financial ratio shocks to determine the factors that change predictability over time. The results imply that expected risk predicts stock return in most sectors, while for unexpected risk this is not always the case. In other words, expected risk is the main determinant of time-variation in the predictability of return, while for unexpected risk this is not always the case. Their model confirms the existence of time-variation in the predictive
power of stock return forecasters. Timmerman (2008) uses two models to determine whether predictability is time-varying, namely the factor augmented autoregressive model and the two-layers neural net model. Time-variability is detected through rolling window estimates of the out-of-sample $R^2$. His results imply that there is indeed time-variation in the predictability of stock returns. The out-of-sample RMSE values oscillate around zero and are negative most of the time, which indicates little evidence of predictability most of the time. However, since for some periods the out-of-sample RMSE reports positive values, it must be the case that the predictive power changes over time. Park (2010) uses an adapted version of a residual-based ratio test to attest whether the time-variation in the predictive power of the dividend-price ratio could be due to changes in its persistence. He finds that the persistence of the dividend-price ratio and its performance are correlated.

2.3.1 Time-varying predictability and the business cycle

In this section the time-variation in the predictability of stock return and its relationship with the business cycle will be discussed.

Some scholars have addressed the issue of time-variation in stock return predictability by investigating the changes in the business cycle. Dangl and Halling (2012) associate the change in the predictability of stock returns with the business cycle, by using predictive regressions that allow for the time-variation of the coefficients. Moreover, they find that predictive models with time-varying coefficients are stronger than models with constant coefficients. Guidolin et al. (2012) also provide empirical support for this view by studying the predictability of the dividend yield, using monthly data for US sector portfolios. They find a negative relationship between predictability and output growth. That is, the predictive power of the dividend yield is stronger during economic contractions than during economic expansions. Analogously, Henkel et al. (2011) investigate the time-variation of stock returns predictability by using a regime-switching vector autoregressive model (RSVAR). They find a significant and robust relationship between returns predictability and the business cycle in all the G7 countries except for Germany. They show that stock return predictors related to the term structure are only effective during recessions but not during economic expansions. These results are consistent with the literature mentioned above, as they show that predictability is the strongest and more volatile during economic contractions. Furthermore, they find higher volatility during recessions. They also find that
predictors show a smoother behavior during expansions, meaning that they are more persistent and less volatile than during recessions. Finally, McMillan (2015) confirms the existence of time-variation in stock return predictability, by using a space-modelling approach. In particular, he revisits the relationship between the stock return predictability of the dividend yield and macroeconomic factors. He finds that the time-varying predictive parameters are determined in part by output growth, the volatility of output growth and the correlation between stock return. His results imply a higher predictive power, when there is more macroeconomic risk.

**2.3.2 Stock return predictability and structural breaks**

In this section different studies will be discussed that have investigated the stability of the predictive power of stock return predictors. This is often referred to as structural breaks.

Recently, scholars have questioned the stability of stock return predictability models. Some scholars associate time-changing predictability of stock return predictors with structural breaks Lettau and Ludvigson (2001) report evidence of instability in the 1990s, by examining the predictive behavior of the dividend yield and the earnings yield. Likewise, Goyal and Welch (2007) show that variables that have been used extensively to predict equity premium are unstable over time. They discover instability in the predictive regression based on the dividend yield when the sample size is extended with data from the 90s. Moreover, signs of instability of financial forecasting models have also been detected by Ang and Bekaert (2004), who show weaker evidence of predictability in the US in the 90s.

These studies are important because they attest instability in predictive return models. However, they do not indicate the specific time in which the forecasting models change, nor do they consider the likelihood of earlier structural breaks. Moreover, it is also important to determine the economic significance of instability, as this can only be evaluated if it can be determined how broad such instability is, both geographically as well as over time. Only then, the extent to which it has an effect on predictability can be ascertained. Paye and Timmerman (2006) investigate the existence of structural breaks, using a dataset of returns for ten OECD countries. Differently from the studies mentioned above, they examine the timing and the characteristics of the structural breaks. They find evidence of structural breaks and show that a break affects the predictive power of stock return forecasters. That is, stock return predictability is not persistent over time and is concentrated in certain time periods. Their empirical results
indicate the presence of a common break during the oil crisis in the 70s and in some European countries in the second half of the 70s, which can be associated with the construction of European Monetary system. Likewise, Viceira (1997) reported evidence of instability of predictive variables that change over time. He finds support for the presence of structural breaks in the behavior of stock returns in the 50’s, after the agreement on interest rates by the Federal Reserve and during the oil and monetary crisis in the 70’s. Lettau and van Nieuwenburgh (2007) argue that shifts in the mean of financial ratio change the predictability of returns forecasters and affects the predictive regressions. That is, shifts in the steady-state of expected returns and growth rate of fundamentals determine the instability of the relationship between stock returns and the predictors. Pastor and Stambaugh (2009) also present evidence on the instability of return predictors and the time-varying nature of these variables. They argue that these predictive variables might be affected by structural breaks, such as institutional, legislative and technological change, large macroeconomic shocks and changes in monetary policies.

2.4 The predictive power of CSV over time

This section describes different approaches undertaken in the empirical literature to test time-variation of the predictive power of stock returns predictors over time. This paper differs from the existing literature, because a different methodology is used and a different variable is analyzed.

As already mentioned, sectoral labor reallocation shocks, change the equilibrium allocation of workers across sectors. The main expectation is that the predictive power of CSV changes over time, because it will be affected by random shocks that occur during the business cycle, causing instability over time (Goyal & Welch, 2007; Guidolin et al., 2012). Since CSV predicts stock returns through labor adjustment costs, it must be the case that the predictive power changes with random shocks that are not predictable. As suggested by Henkel et al. (2011), the predictive power of CSV will be stronger during recessions, as it will be more volatile. As most variables studied in the empirical literature, the predictability of CSV will be countercyclical. That is, its predictive power will be stronger during recessions and weaker during economic expansions.

This paper proceeds as follows. Following Goyal and Welch (2007), the first part of the paper tests whether the predictive performance of CSV changes over time. A graphical analysis will
help to determine when these changes occur. In order to assess whether the predictive power changes depending on the stage in the business cycle, the effect of the industrial production growth (ΔIP1) and the unemployment rate (un_ch) will be tested.

Considering the empirical literature, the main expectation is that ΔIP1, which proxies for the business cycle, will have a negative effect on the predictability of CSV. Indeed, the impact of sectoral labor reallocation shocks during economic expansions, will be more negative, as the foregone production will be higher. That is, following an increase in ΔIP1 indicating an expansion phase, CSV will decrease stock prices even more, because of higher labor adjustment costs. The opposite is true for a decrease in ΔIP1. For unemployment growth the expectation is the following. Higher levels of unemployment are more likely during a recession. When the economy is in a contractionary phase, the effect of CSV on future excess returns will be less negative, as the foregone costs of production due to lost labor time will be lower.
3. Data

Following Eiling, Kan and Sharifkhani (2016), cross-sectional volatility (CSV) is used as a proxy for sectoral labor reallocation shocks. This measure is constructed by using monthly excess returns on 47 industry portfolios. These will be retrieved from Kenneth French’s website. The CSV measure is based on industry specific shocks because the effect of total shocks on sectoral labor reallocation is zero. Data for the stock market return and industrial production growth will be downloaded from CRSP.

The CSV measure will be compared to three other stock return predictors. The first predictor is the cyclically adjusted log price-earnings ratio (logPE), which will be downloaded from Robert Shiller’s website. The second alternative predictor is the inflation rate (INFL), calculated as the log of the Consumer Price Index growth rate. The last alternative variable is the term spread (TERM), computed as the difference between the yield of a ten-year US government bond and the yield on a three-month T-bills. Both the term spread and the inflation rate are downloaded from the Federal Reserve Bank website.

Following Loungani et al. (1990) aggregate unemployment rates for the US are from the Current Population Survey, retrieved from the Bureau of Labor Statistics (BLS). This variable is calculated through the log transformation where \( \text{un} = \log(\text{UN}/(1 - \text{UN})) \). Finally, data for the industrial production growth \( \Delta IP1 \), calculated as the growth rate of the Industry Production Index, is downloaded website of the Federal Reserve Bank in St. Louis.

The sample will cover the period from January 1952 to December 2017. The choice of the large sample period makes this study more informative and allows for robustness tests.
4. Methodology
Following Eiling et al. (2016) and Brainard and Cutler (1993), cross-sectional volatility of industry-specific returns will be used as a proxy for sectoral labor reallocation shocks. The intuition is as follows. Industry-specific returns reflect uneven shocks that affect each industry separately, leading to an increase in cross-sectional dispersion of industry returns. A rise in CSV is a consequence of shift in the equilibrium of consumer preferences and technology across sectors, which creates the necessity for the reallocation of labor from one industry to another (Eiling, Sharifkani, & Kan, 2016).

The use of monthly stock returns instead of employment growth rates is preferred, because stock returns enable to separate the allocation effects from cyclical effects (Brainard & Cutler, 1993).

4.1 The CSV Measure
The CSV measure is derived from industry-specific returns of 47 different industries. The cross-sectional volatility is computed by using the estimated results of the following regression, which is computed with data from the past 36 months:

\[ R_{i,s} = \alpha_i + \beta_i R_{M,s} + \varepsilon_{i,s}, \quad s = t - 35, ..., t \]  

(1)

where \( R_{i,s} \) is the industry excess return and \( R_{M,s} \) is the market portfolio excess returns. The industry-specific return for industry \( i \) in period \( s \) is estimated as the abnormal return from the CAPM:

\[ \eta_{i,s} = \alpha_i + \hat{\varepsilon}_{i,s} \]

where \( \hat{\alpha}_i \) is the predicted value estimated by OLS and \( \hat{\varepsilon}_{i,s} \) are the residuals estimated in Eq. (1). The CSV measure is defined as the cross-sectional standard deviation over the last 12 months according to the following formula:

\[ \text{CSV}_i = \left[ \frac{1}{48} \sum_{i=1}^{49} (\eta_{i,t-11:t} - \bar{\eta}_{t-11:t})^2 \right]^{1/2} \]

(2)

where

\[ \eta_{i,t-11:t} = \prod_{s=t-11}^{t} (1 + \eta_{i,s}) - 1, \]

(3)
\[ \bar{\eta}_{t-11:t} = \frac{1}{49} \sum_{i=1}^{49} \eta_{t-11:t} \]

### 4.2 Stock return predictability of CSV

The predictability of CSV and other stock return predictors will be estimated with the following linear regression model:

\[
\text{EXRET}_{t:t+h} = \alpha_h + \beta_h X_t + u_{t+h} \tag{4}
\]

Where \( \text{EXRET}_{t:t+h} \) is the continuously compounded h-month excess return from month t to month t+h and \( X_t \) is the stock return predictor at month t. \( \alpha_h \) is a constant, \( \beta_h \) is the coefficient of sectoral labor reallocation shocks and \( u_{t+h} \) is the error term. The estimators in Eq. (4) will be used to test for in sample predictability.

Following Goyal and Welch (2007), the out-of-sample (OOS) will be computed with the following formula:

\[
R^2_{OOS} = 1 - \frac{MSE_A}{MSE_N} \tag{5}
\]

where \( MSE_A \) is the mean squared error based on rolling errors from the historical mean model and \( MSE_N \) is the mean squared error from the rolling errors estimated with the OLS model.

### 4.3 The instability of the predictive power of CSV over time

Following Lettau and van Nieuwerburgh (2007), the instability of the forecasting relationship will be determined by analyzing the change in the predictive power of the estimated coefficient over time. The betas will be extracted from a 30-year rolling window regression. This means that the first estimates are calculated over the period 1952:1-1981:12. Then the sample period is rolled one month forward to 1952:2-1982:1 and the next parameter estimates are calculated. This
mechanism continuous through the whole sample period. For this purpose, Eq. (4) is used for the estimation, where $h=48$ for the 4-year rolling window.

Following Goyal & Welch (2007), the performance of the CSV and other variables over time will be determined both in-sample (IS) and out-of-sample (OOS). The IS and OOS performance will be plotted over time. The IS performance graph is calculated as the difference between the cumulative squared demeaned excess return and the cumulative squared regression residuals. The OOS performance is the difference between the cumulative sum of the squared prediction errors of the prevailing mean and the cumulative sum of the squared prediction error of the regressor from the linear IS regression.

### 4.4 Predictive regression with industry production growth

Following Guidolin et al. (2013), Eq. (4) will be extended with $\Delta IP_1$, which is a proxy for economic output growth, since the assumption in this paper is that the predictability of CSV varies with the business cycle. Therefore, the following equation will be estimated:

$$EXRET_{t:t+h} = \alpha_h + \beta_1 CSV_t + \beta_2 \Delta IP_1_t + \beta_3 \Delta IP_1_t CSV_t + u_{t+h}$$

(6)

If $\beta_3$ in Eq. (6) is statistically significant, then this indicates that the predictability of the CSV measure varies according to the level of output growth and thus varies depending on the state of the economy over time. Eq. (6) makes it possible to directly analyze the effect of the state of the economy on the predictability of the CSV measure. The expectation is that $\Delta IP_1_t$ increases the absolute effect of CSV on excess returns. That is, future excess returns decrease even more when output growth is positive and hence, the economy is in an expansionary phase. Therefore, if the predictability of CSV varies over time depending on the state of the economy, then I would expect $\beta_3$ in Eq. (6) to be negative and statistically significant.

### 4.5 Predictive regression with unemployment growth

In order to determine the impact of unemployment growth on the predictability of CSV, Eq. (4) will be extended with the unemployment growth variable $un\_ch$:

$$RET_{t:t+h} = \alpha_h + \beta_1 CSV_t + \beta_2 un\_ch_t + \beta_3 un\_ch_t * CSV_t + u_{t+h}$$

(7)
If $\beta_3$ in Eq. (7) is statistically significant, then the predictability of the CSV measure changes depending on the values of un_ch, which is an indirect proxy for the business cycle. Since unemployment is related to the business cycle, and unemployment growth and CSV are correlated, I would expect that the predictability CSV varies accordingly. Since the labor adjustment costs and the costs of foregone production are lower during recessions, it must be the case that the slope parameters $\beta_3$ plus $\beta_2$ have a total effect that is smaller than the effect of $\beta_2$ in Eq. (4).
5. Empirical Results
In this section the regression results will be analyzed and discussed. First the main variables will be described and their predictability will be tested using a simple OLS regression over different horizons. Second the performance over time of CSV and other well-known stock predictors will be investigated and interpreted. Finally, the relationship between the predictability of CSV and unemployment change will be studied.

5.1 Predictability of CSV and other stock return predictors
Table 1 presents the summary statistics of CSV, three other stock return predictors, ΔIP1 and un_ch. The statistics indicate that CSV is time-varying, as the mean is 0.139 and the standard deviation is 0.043. The autocorrelation coefficients of the first and the second order for CSV are 0.90 and 0.79. These values are lower than for TERM and logPE, but larger those for INFL. The last column reports the correlation between the other predictive variables and CSV, which range from -0.1433 to 0.2620. Most correlations are significant but their magnitudes are small, this could be an indication that CSV is has an independent impact on stock returns, which has not been captured by other alternative variables.

Table 1. Summary statistics
This table shows the summary statistics of three stock return predictors, the industrial production growth and unemployment growth. CSV is computed as the cross-sectional volatility of 47 industry returns, which are calculated from the past 12-month returns. The other predictive variables include the term spread (TERM), which is the difference between the 10-year government bond yield and the three-month treasury bill rate, the inflation rate (INFL), calculated from the consumer price index, the log of the price-earnings ratio (logPE), the industry production growth (ΔIP1) and unemployment growth (un_ch). The variables are at monthly frequencies and the sample runs from January 1952 to December 2017. Only TERM starts in January 1954. The table includes the mean, median, standard deviation, the minimum and the maximum of each predictive variable, the first order autocorrelation coefficient, the second order autocorrelation coefficient and the correlation of every variable with CSV.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Std.Dev.</th>
<th>min</th>
<th>max</th>
<th>Obs</th>
<th>AR(1)</th>
<th>AR(2)</th>
<th>Corr. CSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSV</td>
<td>0.1393</td>
<td>0.1300</td>
<td>0.0431</td>
<td>0.0684</td>
<td>0.4270</td>
<td>780</td>
<td>0.9002</td>
<td>0.7899</td>
<td>1.0000</td>
</tr>
<tr>
<td>un_ch</td>
<td>0.0005</td>
<td>0.0000</td>
<td>0.0364</td>
<td>-0.1123</td>
<td>0.2617</td>
<td>780</td>
<td>0.1539</td>
<td>0.2703</td>
<td>0.1279</td>
</tr>
<tr>
<td>ΔIP1</td>
<td>0.0022</td>
<td>0.0026</td>
<td>0.0090</td>
<td>-0.0433</td>
<td>0.0618</td>
<td>780</td>
<td>0.3851</td>
<td>0.2546</td>
<td>-0.1329</td>
</tr>
<tr>
<td>TERM</td>
<td>1.55%</td>
<td>1.58%</td>
<td>1.21%</td>
<td>-2.65%</td>
<td>4.42%</td>
<td>698</td>
<td>0.9572</td>
<td>0.8943</td>
<td>-0.1433</td>
</tr>
<tr>
<td>logPE</td>
<td>2.8822</td>
<td>2.9352</td>
<td>0.4083</td>
<td>1.8929</td>
<td>3.7887</td>
<td>793</td>
<td>0.9935</td>
<td>0.9853</td>
<td>0.2620</td>
</tr>
<tr>
<td>INFL</td>
<td>0.28%</td>
<td>0.25%</td>
<td>0.31%</td>
<td>-1.77%</td>
<td>1.18%</td>
<td>793</td>
<td>0.6171</td>
<td>0.4845</td>
<td>0.0806</td>
</tr>
</tbody>
</table>
Table 2 reports the regression results of Eq. (4), where the CSV measure is the independent variable. It presents the estimated regression coefficient $\hat{\beta}$ and the Newey-West t-ratio with h-1 lags to correct for the standard errors. The table shows the results for h=1,3,12,24 and 36 months. The table reports negative regression coefficients of CSV for all h, which confirms a negative relationship between the lagged CSV measure and excess returns. These findings confirm the intuition that labor adjustments costs decrease future expected returns, through labor adjustment costs. For h=1, the $\hat{\beta}$ is significant at a 5% level, while for h>1, $\hat{\beta}$ is significant at a 1% level. As h increases the in-sample (IS) $R^2$ increases drastically from 0.78% to 19.53% and decreases again for h=36. The OOS $R^2$s show the same pattern. They increase as the horizon h becomes larger. However, the increased $R^2$ level might not be due to a higher performance level. Indeed, this could be caused by an upward bias, which might result from increasing h.

The CSV results are compared to three other stock return predictors. The significance levels of alternative predictors are lower for short horizons and sometimes not significant. The coefficient estimate of the log price-earnings ratio (logPE) is not significant for h=1,3. For h=36, the term spread (TERM) has higher IS and OOS $R^2$s than CSV. Comparing the remaining variables to the CSV, the CSV clearly outperforms the other predictors. For instance, at h=3, CSV has an OOS $R^2$ of 5.13%, while the OOS $R^2$s of the other stock return predictors go from -0.23% to -0.10%. Moreover, the IS $R^2$ of CSV is equal to 4.51%, whereas the IS $R^2$s of the other predictors are between 0.27% and 2.46% for h=3.

Table 2. Predicting Stock Market Returns Using CSV and Alternative Predictors

The table shows the results of the following predictive regression:

$$EXRET_{t:t+h} = \alpha_h + \beta_h X_t + u_{t+h}$$

Where $EXRET_{t:t+h}$ is the continuously compounded excess return for h-months from month t to month t+h. The predictive regressions are estimated with the following predictive variables ($X_t$): sectoral labor reallocation (CSV), the term spread (TERM), the inflation rate (INFL) and the log dividend price ratio (logDP). The sample starts in January 1952 and ends in December 2017. Only the TERM spread is available for a shorter period from January 1960 until December 2017. The table presents the estimation of the coefficient $\hat{\beta}$, the Newey-West (1987) t-ratio with h-1 lags and both the IS and OOS $R^2$s. The panels report the regression results for h = 1, 3, 12, 24 and 36.
months. Significance levels of 10%, 5% and 1% are denoted by one, two and three stars.

<table>
<thead>
<tr>
<th>Panel A: h=1</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>t-ratioNW</td>
<td>$R^2_{fs}$</td>
<td>$R^2_{gos}$</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>-0.0887</td>
<td>-2.23**</td>
<td>1.55%</td>
<td>1.26%</td>
</tr>
<tr>
<td>INFL</td>
<td>-1.1393</td>
<td>-1.93*</td>
<td>1.52%</td>
<td>1.35%</td>
</tr>
<tr>
<td>logPE</td>
<td>-0.0051</td>
<td>-1.25</td>
<td>0.16%</td>
<td>0.14%</td>
</tr>
<tr>
<td>TERM</td>
<td>0.3263</td>
<td>2.15**</td>
<td>1.49%</td>
<td>-0.07%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: h=3</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>t-ratioNW</td>
<td>$R^2_{fs}$</td>
<td>$R^2_{gos}$</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>-0.4380</td>
<td>-4.52***</td>
<td>4.51%</td>
<td>5.13%</td>
</tr>
<tr>
<td>INFL</td>
<td>-3.9967</td>
<td>-2.18**</td>
<td>1.34%</td>
<td>-0.10%</td>
</tr>
<tr>
<td>logPE</td>
<td>-0.01512</td>
<td>-1.43</td>
<td>0.27%</td>
<td>-0.20%</td>
</tr>
<tr>
<td>TERM</td>
<td>1.1523</td>
<td>3.01***</td>
<td>2.46%</td>
<td>-0.23%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: h=12</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>t-ratioNW</td>
<td>$R^2_{fs}$</td>
<td>$R^2_{gos}$</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>-1.1430</td>
<td>-5.89***</td>
<td>8.71%</td>
<td>11.96%</td>
</tr>
<tr>
<td>INFL</td>
<td>-11.2157</td>
<td>-4.30**</td>
<td>3.56%</td>
<td>3.68%</td>
</tr>
<tr>
<td>logPE</td>
<td>-0.7146</td>
<td>-3.47***</td>
<td>2.00%</td>
<td>-1.09%</td>
</tr>
<tr>
<td>TERM</td>
<td>3.6276</td>
<td>4.86***</td>
<td>7.32%</td>
<td>7.72%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel D: h=24</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>t-ratioNW</td>
<td>$R^2_{fs}$</td>
<td>$R^2_{gos}$</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>-1.7103</td>
<td>-8.01***</td>
<td>11.40%</td>
<td>12.76%</td>
</tr>
<tr>
<td>INFL</td>
<td>-12.8639</td>
<td>-4.95***</td>
<td>2.11%</td>
<td>-0.72%</td>
</tr>
<tr>
<td>logPE</td>
<td>-0.1361</td>
<td>-5.06***</td>
<td>3.88%</td>
<td>-2.21%</td>
</tr>
<tr>
<td>TERM</td>
<td>5.5974</td>
<td>6.89***</td>
<td>7.32%</td>
<td>11.44%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel E: h=36</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>t-ratioNW</td>
<td>$R^2_{fs}$</td>
<td>$R^2_{gos}$</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>-1.8315</td>
<td>-6.47***</td>
<td>8.71%</td>
<td>5.36%</td>
</tr>
<tr>
<td>INFL</td>
<td>-14.0917</td>
<td>-5.42***</td>
<td>1.67%</td>
<td>-4.89%</td>
</tr>
<tr>
<td>logPE</td>
<td>-0.18013</td>
<td>-6.06***</td>
<td>5.36%</td>
<td>-1.35%</td>
</tr>
<tr>
<td>TERM</td>
<td>5.5974</td>
<td>6.69***</td>
<td>16.00%</td>
<td>14.85%</td>
</tr>
</tbody>
</table>
Figure 2 plots the estimated rolling beta coefficients over time, obtained from a 48-month rolling window regression. From the graph, it is evident that the estimate of beta is affected by time-variation. The beta shows a cyclical pattern and there are both large positive and negative values. The changing beta might be an indicator of time-varying predictability. The line shows very large positive values in the 80’s and around the financial crisis in 2008. This could be an indication that the predictability changes during recessions, while it remains stable during economic expansions. In order to determine whether predictability of the CSV measure depends on the state of the economy, Eq. (6) and Eq. (7) need to be estimated.

5.2 IS and OOS Performance CSV over time

Figure 1 plots the performance of the predictive variables presented in Table 4 at a monthly frequency. The figure shows the IS and OOS performance of variables in Table 4. The performance is computed as the difference between the cumulative squared demeaned equity premium and the cumulative squared regression residual for the IS regression. For the OOS regression, this is the difference between the cumulative squared prediction errors of the
prevailing mean and the cumulative squared prediction error of the predictive variable from the linear regression. The two performance lines can be interpreted as follows: an increase in the line means that the variable in Figure 1 performs better than the historical average. Thus, an increasing line indicates superior performance of the predictive regression model, while a decreasing line suggests a better performance of the NULL. The time-series pattern allows to analyze periods of good and bad performance, even if the units are not intuitive. As expected, the lines of the variables plotted in Figure 1 change drastically around some recession dates. These findings might indicate that the predictability of these variable varies according to the business cycle.

Following Goyal and Welch (2007), this paper uses two basic principles to establish whether a variable can be defined as a stable predictor:

1. It has a significant IS predictability over the whole sample period. That, is the IS $R^2$ needs to be positive and the IS t-statistic needs to be significant;
2. There is a good OOS performance over the whole sample period, which means that the OOS $R^2$ needs to be positive throughout the whole sample.

The first principle is not true for all stock return predictors as the values reported in Table 2, show that logPE is not statistically significant for $h=1, 3$. Nevertheless, as previously proven, this variable is significant for $h>3$. Additionally, the inconclusive results regarding the predictability of these variables, can be explained by differences in the estimation period (Goyal & Welch, 2007). For this reason, the predictive performance of this predictors will also be investigated.

Goyal and Welch (2007) use the past three decades, because they want to study the effect of the Oil shock recession of 1973-1975 on the predictive performance. This thesis will be focused on the performance of the last two decades, which include the “bubble” period of 1999-2001 and the financial crisis of 2008. The expectation is that these recession periods will affect and change the performance of predictive of CSV. The entire sample period will also be analyzed and compared with the performance of the past 20 years. Differently from their paper, this research will not study the performance after the oil shock, but the performance after 1997, as the plots in Figure 1 show large changes in the performance at that point in time.

CSV in Figure 1 depicts monthly data and is statistically significant at a 5% level. The performance plot displays three different periods, both IS as OOS. CSV shows evidence of underperformance from 1980 to 1985, neither good or bad performance from 1985 to 2000 and
good performance thereafter. Furthermore, the graph indicates that the IS performance was constantly superior to the OOS performance. The difference between the IS and OOS performance decreases after 2001. Table 3 shows that over the past two decades (1997 to 2017) the R²s for CSV are large and positive both IS and OOS, which is an indicator for good performance. Over the entire sample period, the R² is still positive but the R² is smaller for IS and OOS for all horizons. These findings might be an indication that the predictive power of CSV varies over time.

Table 2 shows that the term spread (TERM) is significant for all the time horizons. Unlike CSV, there is no radical change in performance as a result of the market decline of 2000-2002. In the last two decades there is evidence of underperformance, especially around the financial crises there is a large drop in performance. Figure 1 indicates that the overall performance of TERM is better in the first half of the sample period. This can be in part confirmed by the R²'s values in Table 3, as the OOS R²'s in the last 20 years are more negative than for the whole sample periods for h=1, 3, 12. Also the IS R²'s are smaller in the last two decades than in the entire sample period, except for h=36. The differences in R²'s between the whole sample period and the subsample including the past 20 years, might indicate the change in performance of TERM that is shown in Figure 1.

The estimation results in Table 2 show that inflation rate (INFL) is significant for all time horizons. The plot in Figure 1 shows a gradual increase in performance from the 1980s to 2007, followed by a large and sudden drop in performance at the start of the financial crisis in 2008. Its performance pattern is different from the CSV variable, as after 1987, there a large decrease in the predictive performance instead of an increase both IS and OOS. Compared to the performance of the CSV variable, this variable shows opposite patterns, as it underperforms in the last decade. Therefore, the performance of INFL is smaller in the last two decades both IS, while the OOS performance in the last two decades is negative, which can be confirmed by the IS and OOS R²'s presented in Table 3. Indeed, for most horizons, both the IS R²'s are smaller and the IS R²'s are more negative in the past 20 years than in the whole sample period. Only for h=24, 36 the IS R²'s are larger for the subsample period.

The log price-earnings ratio (logPE) is not statistically significant for the small horizons (h= 1, 3) as reported in Table 2. Nevertheless, for h>3, logPE is significant at a 1% level. Its OOS performance pattern is more volatile than its IS performance pattern. The overall
performance is good most of the time. However, in the 1980s and around 1997 there is a large drop in performance, followed by large peaks thereafter. Since these large drops did not occur in the last 20 years, the plot in Figure 1 suggests superior performance of logPE in the last two decades compared to the entire sample period. These assumptions are confirmed by the $R^2$ values in Table 3. Indeed, for all the horizons, both IS and OOS $R^2$s are larger for the subsample period including the past two decades, than for the whole sample period.

Concluding, periods with large stock market changes, such as the “bubble period” from 1999–2001, followed by its collapse and the sovereign debt crisis of 2008 affect the forecasting performance of different stock return predictors both IS and OOS. Most models except for the CSV, do not perform well in the last two decades, which indicates that these stock market changes affect the predictability of stock return. Differently from all the other variables, which decrease after such a shock occurs, the CSV performance seems to increases. Hence, these plots might indicate time-varying predictive power of CSV. Moreover, the changes in the predictive performance occur around specific recession dates, which might imply that these changes occur around recession dates.

Although CSV satisfies the two criteria for a stable predictor (IS significance, OOS performance), there is evidence of variation in the predictive performance of CSV. In the next section, it will be determined to what extend the predictive power of CSV varies with the business cycle.
Table 3

**OOS and IS R² over the entire sample and the past 20 years.**

This table reports both the IS and OOS R²s for CSV and alternative stock return predictors for different time horizons. The R²s are computed from the predictive regression, which includes the following variables. The cross-sectional volatility (CSV), the inflation rate (INFL), the term spread (TERM) and the log price-earnings ratio. The OOS R² is computed according to the formula in Eq. (5). The column “Past 20 Years” presents the IS and OOS R²s estimated in the period between January 1998 and December 2017. The column “All time” shows the IS and OOS R²s for the whole sample period that runs from January 1952 to December 2017.
### Panel A: $h=1$

<table>
<thead>
<tr>
<th></th>
<th>Past 20 Years</th>
<th>All time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_S$</td>
<td>$R^2_{OOS}$</td>
</tr>
<tr>
<td>CSV</td>
<td>5.24%</td>
<td>7.21%</td>
</tr>
<tr>
<td>INFL</td>
<td>0.00%</td>
<td>-10.79%</td>
</tr>
<tr>
<td>logPE</td>
<td>1.12%</td>
<td>2.70%</td>
</tr>
<tr>
<td>TERM</td>
<td>0.22%</td>
<td>-2.84%</td>
</tr>
</tbody>
</table>

### Panel B: $h=3$

<table>
<thead>
<tr>
<th></th>
<th>Past 20 Years</th>
<th>All time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_S$</td>
<td>$R^2_{OOS}$</td>
</tr>
<tr>
<td>CSV</td>
<td>13.85%</td>
<td>18.76%</td>
</tr>
<tr>
<td>INFL</td>
<td>0.11%</td>
<td>-14.04%</td>
</tr>
<tr>
<td>logPE</td>
<td>2.50%</td>
<td>8.26%</td>
</tr>
<tr>
<td>TERM</td>
<td>0.22%</td>
<td>-5.16%</td>
</tr>
</tbody>
</table>

### Panel C: $h=12$

<table>
<thead>
<tr>
<th></th>
<th>Past 20 Years</th>
<th>All time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_S$</td>
<td>$R^2_{OOS}$</td>
</tr>
<tr>
<td>CSV</td>
<td>30.56%</td>
<td>33.00%</td>
</tr>
<tr>
<td>INFL</td>
<td>4.96%</td>
<td>-12.08%</td>
</tr>
<tr>
<td>logPE</td>
<td>16.91%</td>
<td>31.28%</td>
</tr>
<tr>
<td>TERM</td>
<td>5.28%</td>
<td>-0.97%</td>
</tr>
</tbody>
</table>

### Panel D: $h=24$

<table>
<thead>
<tr>
<th></th>
<th>Past 20 Years</th>
<th>All time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_S$</td>
<td>$R^2_{OOS}$</td>
</tr>
<tr>
<td>CSV</td>
<td>21.01%</td>
<td>22.29%</td>
</tr>
<tr>
<td>INFL</td>
<td>22.85%</td>
<td>-9.93%</td>
</tr>
<tr>
<td>logPE</td>
<td>41.53%</td>
<td>49.48%</td>
</tr>
<tr>
<td>TERM</td>
<td>5.28%</td>
<td>20.04%</td>
</tr>
</tbody>
</table>

### Panel E: $h=36$

<table>
<thead>
<tr>
<th></th>
<th>Past 20 Years</th>
<th>All time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2_S$</td>
<td>$R^2_{OOS}$</td>
</tr>
<tr>
<td>CSV</td>
<td>30.56%</td>
<td>17.65%</td>
</tr>
<tr>
<td>INFL</td>
<td>3.06%</td>
<td>-9.95%</td>
</tr>
<tr>
<td>logPE</td>
<td>60.89%</td>
<td>58.22%</td>
</tr>
<tr>
<td>TERM</td>
<td>51.05%</td>
<td>21.75%</td>
</tr>
</tbody>
</table>
Figure 1
Monthly performance of stock return predictors.
These graphs plot the performance of monthly predictive regressions both in-sample (IS) and out-of-sample (OOS). In particular, this is the difference between the cumulative squared prediction errors of the NULL and the cumulative squared prediction error of the ALTERNATIVE. The ALTERNATIVE is a model that is based on the forecasting variables that are reported in each graph. For the IS graph, the NULL is equity premium mean for the whole sample period, while for the OOS graph this is the prevailing equity premium average. The relative IS performance is the dotted red line, the OOS relative performance is denoted by the blue line. An increase in a line suggests better performance of the model based on the predictive variables. A decrease in the line means is an indication of superior performance of the NULL model.
5.3 Predictability of CSV and industry production growth

Table 4 reports the estimation results for Eq. (6). Compared to the regression without the industrial production growth measure ($\Delta IP_1$), there are some differences in the magnitude of estimated $\hat{\beta}_1$ coefficient for $h=12$ and $h=24$. For all the horizons smaller than 36 months ($h=36$) the estimated coefficient $\hat{\beta}_3$ is not significant, indicating that for small horizons the level of output growth has no effect on the predictive power of CSV. However, for $h=36$, $\hat{\beta}_3$ is significant at a 5% level. The estimated coefficient on the interaction variable is -85.42, which means that a given value of CSV is associated with a decrease in future excess returns of approximately 0.85, when there is positive output growth. The total effect is the following. A 1% increase of the CSV reduces future returns by 0.87 units, when there is positive output growth.

The effect is CSV on future stock returns in Table 2 is already negative, by adding $\Delta IP_1$ and its interaction variable to the regression, the effect of CSV on future excess stock returns becomes more negative. Thus, when $\Delta IP_1$ is high, the predictive power of CSV becomes even higher, as the negative affect that CSV has on future excess stock returns is enhanced.
The estimation results presented in Table 4, show that for 36-month horizons, the predictive power of CSV is higher, when the economy is in an expansionary phase. These findings confirm the expectations in the previous sections. During an economic expansion, the foregone production costs that follow from sectoral labor reallocation are higher, which means that that sectoral allocation shocks reduce future stock returns even more (Davis, 1987). Thus, the predictive power of CSV varies over time and depends on the business cycle, as CSV has a different effect on future excess stock returns, when output growth is positive.

Table 4. Predicting Stock Market Returns Using CSV and IP1

This table presents the regression results of Eq. (6), where the dependent variable is the continuously compounded excess return of h months from month t to month t+h. The dependent variables are the CSV measure and the industry production growth rate (IP1). Additionally, INT_CSV_IP1 is the interaction variable between the CSV and IP1. The sample period is from 1952 to 2017. The table reports the regression coefficient estimates, the Newey-West (1987) t-ratio with k − 1 lags, and as the in-sample and out-of-sample R²'s. Significance levels of 1%, 5% and 10% are denoted respectively by three, two and one stars.

<table>
<thead>
<tr>
<th>h</th>
<th>CSV</th>
<th>IP1</th>
<th>INT_CSV_IP1</th>
<th>R²IS</th>
<th>R²OOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.1857</td>
<td>0.1324</td>
<td>-2.3842</td>
<td>1.64%</td>
<td>2.27%</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td></td>
<td>( t - ratio_{NW} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.4638</td>
<td>-0.0340</td>
<td>-6.8936</td>
<td>5.30%</td>
<td>4.45%</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td></td>
<td>( t - ratio_{NW} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-1.2226</td>
<td>3.3980</td>
<td>23.7696</td>
<td>12.64%</td>
<td>15.62%</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td></td>
<td>( t - ratio_{NW} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-1.7666</td>
<td>-3.8486</td>
<td>-7.8717</td>
<td>14.73%</td>
<td>15.28%</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td></td>
<td>( t - ratio_{NW} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>-1.7588</td>
<td>7.2658</td>
<td>-85.4162</td>
<td>13.82%</td>
<td>9.62%</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td></td>
<td>( t - ratio_{NW} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.4 Stock return predictability CSV and unemployment change

In this section the relationship between CSV and unemployment growth will be analyzed. In particular, the effect of unemployment growth on the predictability of CSV to predict excess returns will be investigated. The expectation is that the effect of un_ch will weaken the predictive power of CSV. Indeed, higher unemployment growth levels occur during recessions, when the costs of foregone production induced by sectoral labor reallocation are lower.

Table 5. Predicting Stock Market Returns Using CSV and un_ch

This table presents the regression results of Eq. (7), where the dependent variable is the continuously compounded excess return of h months from month t to month t+h. The dependent variables are the CSV measure and unemployment change (un_ch). Additionally, INTER is the interaction variable between the CSV and un_ch. The sample period is from 1952 to 2017. The table reports the regression coefficient estimates, the corresponding Newey-West (1987) t-ratio with k − 1 lags, as well as the in-sample and out-of-sample R2s. Significance levels of 1%, 5% and 10% are denoted respectively by three, two and one stars.

<table>
<thead>
<tr>
<th>h</th>
<th>CSV</th>
<th>un_ch</th>
<th>INTER</th>
<th>$R_{IS}^2$</th>
<th>$R_{OOS}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\hat{\beta}$</td>
<td>0.4503</td>
<td>0.0653</td>
<td>0.2012</td>
<td>3.34%</td>
</tr>
<tr>
<td></td>
<td>$t - \text{ratio}_{NW}$</td>
<td>1.80*</td>
<td>0.69*</td>
<td>2.65***</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\hat{\beta}$</td>
<td>0.72514</td>
<td>0.0546</td>
<td>0.3671</td>
<td>7.21%</td>
</tr>
<tr>
<td></td>
<td>$t - \text{ratio}_{NW}$</td>
<td>1.81*</td>
<td>0.38</td>
<td>3.11***</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>$\hat{\beta}$</td>
<td>1.6412</td>
<td>0.0915</td>
<td>0.8886</td>
<td>16.86%</td>
</tr>
<tr>
<td></td>
<td>$t - \text{ratio}_{NW}$</td>
<td>2.30**</td>
<td>3.28***</td>
<td>5.34***</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>$\hat{\beta}$</td>
<td>3.8879</td>
<td>0.582</td>
<td>1.7584</td>
<td>23.74%</td>
</tr>
<tr>
<td></td>
<td>$t - \text{ratio}_{NW}$</td>
<td>5.43***</td>
<td>2.46**</td>
<td>7.85***</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>$\hat{\beta}$</td>
<td>6.7214</td>
<td>0.3137</td>
<td>2.6398</td>
<td>29.67%</td>
</tr>
<tr>
<td></td>
<td>$t - \text{ratio}_{NW}$</td>
<td>8.65***</td>
<td>1.44</td>
<td>11.45***</td>
<td></td>
</tr>
</tbody>
</table>
The regression results in Table 5 include unemployment growth (un_ch) and the interaction term between un_ch and the CSV measure (INTER). These results are in sharp contrast with the results reported in Table 2. The coefficient on CSV has become positive instead of negative. Also, the significance level of the estimated coefficient of CSV has dropped compared to the regression without un_ch. Moreover, the coefficient on INTER is positive and statistically significant for all horizons. Since both the estimated coefficient on INTER and the coefficient on CSV are positive, it must be the case that when un_ch is high, CSV increases future stock returns. That is, un_ch does not weaken the predictability of CSV. Instead, it changes the effect of CSV on future stock returns from negative to positive. In other words, CSV predicts higher future excess stock returns, when un_ch is high. For instance, for h=1, a 1% increase in CSV is associated with an increase in future excess returns of 0.0065 units, when there is a positive growth in unemployment. For h=36, a rise in CSV is associated with an increase in future excess returns of approximately 0.095 units, when unemployment growth is positive. For all horizons, both the IS and the OOS R^2s have increased, compared to the regression results in Table 2, which might suggest better predictability. Since unemployment changes cyclically over time, and the predictability of CSV varies with the levels of un_ch, it must be the case that the predictive power of CSV varies over time depending on the stage of the business cycle.

These findings can be confirmed by Figure 3, which plots the CSV and the unemployment rate time series. The graph shows that both variables fluctuate more during specific recession dates that have been highlighted in the previous sections, such as the market decline in 2001 and the financial crisis in 2008. In particular, Figure 3 shows that the CSV peaks correspond with the unemployment peaks in the last two decades and less in the earlier years.

**Figure 3. Time series of CSV and aggregate unemployment.**
The blue line is the CSV monthly time series, calculated as the most recent 12 months of industry specific returns. The blue line shows aggregate unemployment rates in levels over time. The dashed line shows the aggregate unemployment rates (in levels) at time t. The yellow areas correspond to the most important crisis dates, namely the oil crisis of 1973-1975, the crisis of 1990-1991, the market decline of 2001 and the financial crisis of 2008.
6. Robustness Checks

In this section robustness checks are performed over time. The sample will be divided in three subsample periods to determine the robustness of the results. In order to compare these results to existing studies, the first subsample period starts in 1973 instead of 1952. Panel A in Table 6 shows an overall higher $R^2$s both IS and OS for most variables, which indicates that there is a stronger relationship between future excess returns and CSV in this specific sample period compared to the whole sample period. This is also an indication that the predictive performance of the variables reported in Table 5 is better over the subsample period than over the entire sample period. For CSV the regression coefficients do not change significantly compared to the whole sample period. However, the IS and the OOS $R^2$s are for most horizons larger than in the whole sample period except for the 12-month horizon. This confirms the previous findings in this paper that show that the performance of CSV improves over time.

The sample period is also divided into two equal sub periods. The first runs from January 1952 to December 1985 and the second runs from January 1984 to December 2017. The values in Panel B of Table 6 show lower $R^2$s for CSV both IS and OOS, which is in line with the findings in Figure 1. Indeed, by excluding the most recent decades from the sample period, the overall performance and the $R^2$ decreases, which in what can be seen in Panel B. In Panel C of Table 5, there is an increase in $R^2$s for CSV both IS and OOS compared to the results over the entire sample in Table 2. Again, this confirms the main findings in this paper according to which there is an increase in predictability of CSV in the last decades. The IS $R^2$ is higher for longer horizons and slightly lower for smaller horizons. This is also the case for TERM, which reports an overall increase in the subsample OOS $R^2$ compared to the whole sample. This is in line with the results in section 5.0, where evidence is shown for an increased performance of TERM, especially in the last two decades. Moreover, logPE has higher $R^2$ both OOS and IS and INFL has different $R^2$ values both higher and lower. Overall, by comparing the values in Table with the values in Table 2, there is clear evidence of time variation the performance of the variables. Furthermore, these results indicate that the second subsample period has the strongest predictive power for excess stock market returns both IS and OOS. Particularly, for the CSV this could be due to the fact that the second sub sample period includes 1991-1992 and 1999-2001, during which sectoral shifts were frequent and large in magnitude (Groshen & Potter, 2003).
Table 6. Predicting Stock Market Returns Using CSV and Alternative Predictors

The table shows the results of the predictive regression reported in Eq. (2), where the dependent variable is the continuously compounded h-month excess returns from month t to month t+h. The independent variables are the following stock return predictors: sectoral labor reallocation (CSV), the term spread (TERM), the inflation rate (INFL) and the log dividend price ratio (logDP). The sample starts in January 1973 and ends in December 2017. Only the TERM spread is available for a shorter period from January 1960 until December 2017. The table reports the regression coefficient estimate $\hat{\beta}$, the corresponding Newey-West (1987) t-ratio with h-1 lags and the $R^2$. The five panels show results for $h = 1, 3, 12, 24$ and 36 months. Three, two and one stars indicate statistical significance levels of 1%, 5% and 10% levels, respectively. In Panel A, the sample period starts in January 1973 and ends in December 2017. For Panel B, the sample period runs from January 1952 to December 1985. The sample period in Panel C runs from January 1986 to December 2017.
Panel A: Predicting stock return from January 1973 to December 2017

<table>
<thead>
<tr>
<th>h</th>
<th>β</th>
<th>t-ratio</th>
<th>$R^2_{IS}$</th>
<th>$R^2_{OOS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CSV</td>
<td>-0.1688</td>
<td>-2.22***</td>
<td>1.46%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-1.951</td>
<td>-1.46</td>
<td>1.00%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.0018</td>
<td>-0.23</td>
<td>0.13%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>0.4949</td>
<td>1.65*</td>
<td>0.88%</td>
</tr>
<tr>
<td>3</td>
<td>CSV</td>
<td>-0.4271</td>
<td>-3.78***</td>
<td>4.67%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-2.1358</td>
<td>-0.97</td>
<td>6.00%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.0093</td>
<td>-0.79</td>
<td>0.23%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>0.8893</td>
<td>2.04**</td>
<td>1.43%</td>
</tr>
<tr>
<td>12</td>
<td>CSV</td>
<td>-1.1695</td>
<td>-5.22***</td>
<td>11.77%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-7.1383</td>
<td>-2.66***</td>
<td>2.25%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.0503</td>
<td>-2.27**</td>
<td>2.20%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>2.6500</td>
<td>3.33***</td>
<td>4.26%</td>
</tr>
<tr>
<td>24</td>
<td>CSV</td>
<td>-1.6716</td>
<td>-7.09***</td>
<td>14.84%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-7.6798</td>
<td>-2.79***</td>
<td>1.61%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.0928</td>
<td>-3.22***</td>
<td>4.62%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>5.1931</td>
<td>5.25***</td>
<td>4.26%</td>
</tr>
<tr>
<td>36</td>
<td>CSV</td>
<td>-1.8196</td>
<td>-5.66***</td>
<td>11.77%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-9.0722</td>
<td>-3.09***</td>
<td>1.67%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.1302</td>
<td>-4.04***</td>
<td>6.79%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>8.1187</td>
<td>-8.41***</td>
<td>18.43%</td>
</tr>
</tbody>
</table>
Panel B: Predicting stock return from January 1952 to December 1985

<table>
<thead>
<tr>
<th>h</th>
<th>$\hat{\beta}$</th>
<th>t-ratio</th>
<th>$R^2_{fs}$</th>
<th>$R^2_{oos}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CSV</td>
<td>-0.0624</td>
<td>-0.57</td>
<td>0.14%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-3.8276</td>
<td>-3.46***</td>
<td>0.56%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.0097</td>
<td>-0.87</td>
<td>0.55%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>1.2140</td>
<td>3.50***</td>
<td>5.90%</td>
</tr>
<tr>
<td>3</td>
<td>CSV</td>
<td>-0.2222</td>
<td>-1.30</td>
<td>0.80%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-6.1262</td>
<td>-3.76***</td>
<td>5.69%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.0284</td>
<td>-1.65</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>2.0904</td>
<td>4.14***</td>
<td>8.60%</td>
</tr>
<tr>
<td>12</td>
<td>CSV</td>
<td>-0.1687</td>
<td>-0.50</td>
<td>0.13%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-11.0329</td>
<td>-3.06***</td>
<td>5.06%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.1239</td>
<td>-4.04***</td>
<td>6.50%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>3.6932</td>
<td>3.35***</td>
<td>8.19%</td>
</tr>
<tr>
<td>24</td>
<td>CSV</td>
<td>-0.9145</td>
<td>-2.50**</td>
<td>2.52%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-10.9062</td>
<td>-3.06***</td>
<td>3.38%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.1792</td>
<td>-4.20***</td>
<td>9.28%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>0.4158</td>
<td>0.45</td>
<td>8.19%</td>
</tr>
<tr>
<td>36</td>
<td>CSV</td>
<td>-0.6795</td>
<td>-1.86*</td>
<td>0.13%</td>
</tr>
<tr>
<td></td>
<td>INFL</td>
<td>-11.3994</td>
<td>-3.32***</td>
<td>3.37%</td>
</tr>
<tr>
<td></td>
<td>logPE</td>
<td>-0.1788</td>
<td>-5.04***</td>
<td>8.43%</td>
</tr>
<tr>
<td></td>
<td>TERM</td>
<td>0.2640</td>
<td>0.29</td>
<td>0.33%</td>
</tr>
</tbody>
</table>
Table 6 shows the regression results for measuring the predictability of CSV by adding unemployment growth for the three different subsample periods mentioned above. Both in Panel A and B of Table 6, there is a large change in the magnitude of the CSV regression coefficient, when compared to coefficients over the whole sample in Table 4. Most coefficient remain
positive, except for the longer horizons in Panel B (which have negative values). Additionally, there is no change in the significance levels of the corresponding Newey West t-ratio’s, confirming that the predictability of the CSV variable changes with unemployment growth. In Panel C of Table 6 the sign of the regression coefficients does not change, but there both the IS and OOS R²’s are larger when compared to the entire sample period. These findings might suggest a higher level of predictive performance of the variables in this subsample. However, in Panel C the interaction variable INTER is not significant for small horizons, while it is significant for all the other sample periods.

Table 6. Predicting excess stock returns with CSV by adding unemployment growth
The Table reports the estimation results of Eq. (5) in which the dependent variable is the continuously compounded excess return for h months from month to month t+h. The independent variables are the cross-sectional volatility (CSV) and unemployment growth (un_ch). The variable INTER is the interaction between CSV and un_ch. In Panel A, the sample runs from January 1973 to December 2017. Panel B, includes the period from January 1952 to December 2017 and the sample period in Panel C runs from January 1986 to December 2017. The table includes the regression coefficient estimates, the corresponding Newey-West (1987) t-ratio with k-1 lags and both the IS and OOS R²’s.

| Panel A: Predicting stock return with CSV by adding un_ch from January 1973 to December 2017 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| h               | CSV             | un_ch           | INTER           | R² IS           | R² OOS          |
| 1               | 0.3961          | 0.0931          | 0.1856          | 2.93%           | 5.40%           |
|                 | t - ratio₃₅     | 1.35            | 0.73            | 2.01**          |                 |
| 3               | 0.6357          | 0.0172          | 0.3414          | 6.65%           | 15.23%          |
|                 | t - ratio₃₅     | 1.37            | 0.09            | 2.41**          |                 |
| 12              | 1.7736          | 0.5230          | 0.9685          | 18.71%          | 31.42%          |
|                 | t - ratio₃₅     | 2.69**          | 2.18**          | 4.50***         |                 |
| 24              | 3.6477          | 0.502           | 1.7289          | 26.53%          | 42.94%          |
|                 | t - ratio₃₅     | 3.85***         | 1.64            | 5.75***         |                 |
| 36              | 6.6059          | 0.384           | 2.7186          | 33.20%          | 48.09%          |
|                 | t - ratio₃₅     | 6.27***         | 1.36            | 8.41***         |                 |
### Panel B: Predicting stock return with CSV by adding un_ch from January 1952 to December 1985

<table>
<thead>
<tr>
<th>h</th>
<th>$\hat{\beta}$ CSV</th>
<th>un_ch</th>
<th>INTER</th>
<th>$R^2_{IS}$</th>
<th>$R^2_{OOS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6746</td>
<td>0.1227</td>
<td>0.2522</td>
<td>4.58%</td>
<td>4.54%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>2.45**</td>
<td>1.65*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.1024</td>
<td>0.1914</td>
<td>0.4516</td>
<td>7.31%</td>
<td>8.76%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>2.56**</td>
<td>1.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.7144</td>
<td>1.0993</td>
<td>0.6655</td>
<td>11.91%</td>
<td>11.83%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>2.32**</td>
<td>5.49***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-0.71692</td>
<td>1.2899</td>
<td>0.1033</td>
<td>10.08%</td>
<td>3.71%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>-0.76</td>
<td>4.64***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>-0.5307</td>
<td>1.1869</td>
<td>0.0839</td>
<td>7.09%</td>
<td>4.40%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>-0.46</td>
<td>4.20***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Panel C: Predicting stock return with CSV by adding un_ch from January 1986 to December 2017

<table>
<thead>
<tr>
<th>h</th>
<th>CSV</th>
<th>un_ch</th>
<th>INTER</th>
<th>$R^2_{IS}$</th>
<th>$R^2_{OOS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1138</td>
<td>-0.1419</td>
<td>0.104</td>
<td>4.06%</td>
<td>7.78%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>0.31</td>
<td>-0.93</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0445</td>
<td>-0.2309</td>
<td>0.1772</td>
<td>10.56%</td>
<td>16.16%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>0.08</td>
<td>-1.13</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.8912</td>
<td>0.0477</td>
<td>0.8066</td>
<td>31.31%</td>
<td>34.41%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>1.16</td>
<td>0.17</td>
<td>3.40***</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>6.331</td>
<td>0.1402</td>
<td>2.7452</td>
<td>40.32%</td>
<td>37.62%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>5.02***</td>
<td>0.40</td>
<td>6.89***</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>10.0467</td>
<td>0.2835</td>
<td>3.8855</td>
<td>45.60%</td>
<td>67.47%</td>
</tr>
<tr>
<td></td>
<td>$t - ratio_{NW}$</td>
<td>7.92***</td>
<td>0.81</td>
<td>10.36***</td>
<td></td>
</tr>
</tbody>
</table>
7. Conclusion

This paper studies the in-sample (IS) and out-of-sample (OOS) time-variation of the predictive power of stock returns predictors that have been used in the academic literature. In particular, this paper analyzes the time-variation and thus, the stability of the predictability of CSV over time. The main conclusion is that the predictability is time-varying and depends on shocks that occur during the business cycle.

Despite the robustness of the results, there is a substantial change in the predictive performance of CSV depending on the selection of the sub sample. The different plots help diagnose when these variables perform well or poorly, both IS and OOS. They shed light on the most relevant sub periods that occurred especially in the last two decades. These shocks include market changes such as the crisis of 1990-1991, the “bubble period” from 1999–2001, followed by its collapse and the sovereign debt crisis of 2008. The plots show when CSV performance is good and when the performance is poor, both IS and OOS. Indeed, by excluding or adding these specific subperiods, CSV has different predictive performance levels. In the last decades, there is evidence of better performance of stock return predictors both IS and OOS. The graphs and the different R²s for different subsample periods might suggest a change in the predictive power of CSV over time.

To confirm the assumptions mentioned above, the effect industrial production growth ΔIP1 and unemployment growth un_ch on the predictability of CSV is studied, as these variables are well-known indicators of business cycle phases. High values of ΔIP1 indicate that the economy is in an expansionary phase, as increased output levels are often associated with an economic expansion. High values of un_ch suggest that the economy is in a recession, as high unemployment levels are often associated with a contractionary phase of the economy. Therefore, ΔIP1 should strengthen the effect of CSV on future stock returns, whereas un_ch should weaken the predictability of CSV on future excess stock returns. These expectations are confirmed by the statistical results in this paper. Indeed, ΔIP1 strengthens the negative effect that CSV has on future excess returns. The effect of un_ch on the predictability of CSV does not only weaken the predictive power of CSV, but also changes the sign of the coefficient. This means that CSV decreases future excess stock returns more, when ΔIP1 is high and increases future excess stock returns when un_ch is high. These empirical results show that the predictive power of CSV varies over time and that these variations in predictability depend on the business cycle.
Hence, the predictive power of CSV is stronger during economic expansions and weaker during recessions.

If stock return predictability of CSV and other variables changes over time, then the models used so far are unstable and alternative methods should be employed. For instance, more alternative stock return predictors, such as the dividend yield and the default spread should be included in the tests performed in this paper to compare the CSV with other variables. Also, simulation models could be used to test the time-variation of the predictability of CSV. This paper has implications for future research, as only variables that have a robust and meaningful predictive power, both IS and OOS, should be used to forecast excess stock returns.
References


Appendix

//alternative set data set as time series
clear
tostring Date, replace
gen newdate = date(Date,"YM")
format newdate %tm
g mydate=date(Date, "YM")
format mydate %td
g dm=mofd(mydate)
format dm %tm

rename TERM TERM1
gen TERM=TERM1/100
destring Date, replace

//computation of the unemployment change variable
gen UN_=UN/100
gen un=log(UN_/(1-UN_))
gen un_ch=un[_n]-un[_n-1]

foreach i of varlist Agric Food Beer Smoke Toys Fun Books Hshld Clths MedEq Drugs Chems Txtls BldMt Cnstr Steel Mach ElcEq Autos Aero Ships Mines Coal Oil Util Telem PerSv BusSv Hardw Chips LabEq Boxes Trans Whlsl Rtail Meals Banks Insur RlEst Fin Other MktRF{
    replace `i'=`i'/100
}

//loop 2.0
local Y Agric Food Beer Smoke Toys Fun Books Hshld Clths MedEq Drugs Chems Txtls BldMt Cnstr Steel Mach ElcEq Autos Aero Ships Mines Coal Oil Util Telem PerSv BusSv Hardw Chips LabEq Boxes Trans Whlsl Rtail Meals Banks Insur RlEst Fin Other
foreach y of local Y  {
    rangestat (reg) `y' MktRF, int(dm 0 36)
gen `y'_residual = `y' - b_cons - b_MktRF * MktRF
rename (reg_* b_* se_*) (`y'_=)
}

//creation of N variable
foreach i of varlist Agric Food Beer Smoke Toys Fun Books Hshld Clths MedEq Drugs Chems Txtls BldMt Cnstr Steel Mach ElcEq Autos Aero Ships Mines Coal Oil Util Telem PerSv BusSv Hardw Chips LabEq Boxes Trans Whlsl Rtail Meals Banks Insur RlEst Fin Other{
gen N_`i'=`i'_b_cons+`i'_residual
//creation N12
tset dm
    gen double running_product_`k' = (`k'+1) if _n == 1
    replace running_product_`k' = running_product_`k'[_n-1] * (`k'+1) if _n > 1
    gen double rolling_product_`k' = running_product_`k'/L11.running_product_`k'
    gen N12`k' = rolling_product_`k' - 1
}

//generate the mean variable

//generate difference for each industry
    gen diff`l'=(N12`l'-mean)^2
}
gen CSV1 =CSV_sqrt^0.5

//gen continuously compounded returns
gen ret=exp(MktRF)-1
rename CSV1 CSV
drop if CSV==.
gen logPE=ln(P_Eratio)

**drop if TERM==.
destring Date, replace

//drop all unneeded variables
drop Agric_reg_nobs-CSV_sqrt

//regression k=1
forvalues j=0/1{
gen ret`j' = f`j' .ret
}
egen ret_av1=rowtotal(ret0-ret1)
newey ret_av1 CSV, lag(1)
predict ret_CSV
corr ret_av1 ret_CSV
di r(rho)^2
newey ret_av1 INFL, lag(1)
predict ret_INFL
corr ret_av1 ret_INFL
di r(rho)^2
newey ret_av1 logPE, lag(1)
predict ret_logPE
corr ret1 ret_logPE
di r(rho)^2
newey ret_av1 TERM, lag(1), if Date>195912
predict ret_TERM
corr ret_av1 ret_TERM
di r(rho)^2
drop ret1 ret0

//regression k=3
forvalues j=0/3{
gen ret`j' = f`j' .ret
}
egen ret_av3=rowtotal(ret0-ret3)
newey ret_av3 CSV, lag(1)
predict ret_CSV3
corr ret_av3 ret_CSV3
di r(rho)^2
newey ret_av3 INFL, lag(1)
predict ret_INFL3
corr ret_av3 ret_INFL3
di r(rho)^2
newey ret_av3 logPE, lag(1)
predict ret_logPE3
corr ret_av3 ret_logPE3
di r(rho)^2
newey ret_av3 TERM, lag(1), if Date>195912
predict retTERM3
corr ret_av3 ret_TERM3
di r(rho)^2
drop ret0-ret3

//regression k=12
forvalues j=0/12{
gen ret'j' = f'j'.ret
}  
egen ret_av12=rowtotal(ret0-ret12)
newey ret_av12 CSV, lag(1)
predict ret_CSV12
corr ret_av12 ret_CSV12
di r(rho)^2
newey ret_av12 INFL, lag(1)
predict ret_INFL12
corr ret_av12 ret_INFL12
di r(rho)^2
newey ret_av12 logPE, lag(1)
predict ret_logPE12
corr ret_av12 ret_logPE12
di r(rho)^2
newey ret_av12 TERM, lag(1), if Date>195912
predict ret_TERM12
corr ret_av12 ret_TERM12
di r(rho)^2
drop ret0-ret12

//regression k=24
forvalues j=0/24{
gen ret'j' = f'j'.ret
}  
egen ret_av24=rowtotal(ret0-ret24)
newey ret_av24 CSV, lag(1)
predict ret_CSV24
corr ret_av24 ret_CSV24
di r(rho)^2
newey ret_av24 CSV IP1, lag(1)
predict ret_IP24
corr ret_av24 ret_IP24
di (r(rho))^2
newey ret_av24 INFL, lag(1)
predict ret_INFL24
corr ret_av24 ret_INFL24
di (r(rho))^2
newey ret_av24 logPE, lag(1)
predict ret_logPE24
corr ret_av24 ret_logPE24
di (r(rho))^2
newey ret_av24 TERM, lag(1), if Date>195912
predict ret_TERM24
corr ret_av12 ret_TERM24
di (r(rho))^2
drop ret0-ret24

// regression k=36
forvalues j=0/36{
gen ret'j' = f'j'.ret
}
egen ret_av36=rowtotal(ret0-ret36)
newey ret_av36 CSV, lag(1)
predict ret_CSV35
corr ret_av36 ret_CSV12
di (r(rho))^2
newey ret_av36 CSV IP1, lag(1)
predict ret_IP36
corr ret_av36 ret_IP36
di (r(rho))^2
newey ret_av36 INFL, lag(1)
predict ret_INFL36
corr ret_av36 ret_INFL36
di (r(rho))^2
newey ret_av36 logPE, lag(1)
predict ret_logPE36
corr ret_av36 ret_logPE36
di (r(rho))^2
newey ret_av36 TERM, lag(1), if Date>195912
predict ret_TERM36
corr ret_av36 ret_TERM36
di (r(rho))^2
drop ret0-ret36

// establish the performance of the predictors
//plotting beta over time
gen ret1 = f1.ret
rangestat (reg) ret1 CSV, int(dm 0 48)
tsline b_CSV

//creating the OOS-R-squared h=1
**the CSV will be replaced with TERM, INFL, logPE in the second loop to calculate the other estimates**
//OOS line CSV
forvalues t=240/780 {
    newey ret_av1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

default t=240/780 {
    newey ret_av1 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//OOS r-squared h=3
forvalues t=240/780 {
    newey ret_av3 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT
forvalues t=240/780 {
newey ret_av3 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//OOS r-squared h=12
forvalues t=240/780 {
newey ret_av12 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//OOS r-squared h=24
forvalues t=240/780 {
newey ret_av24 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}

egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
egen cum_sum2=sum(res_squared)
gen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av24 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

//OOS r-squared h=36
forvalues t=240/780 {
    newey ret_av36 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

drop cum_sum3 MSEn cum_sum2 MSEa r_squared1
//regression with IP variable h=1
gen INT_CSV_IP1= CSV*IP1
ewey ret_av1 CSV IP1 INT_CSV_IP1, lag(1)
predict ret_IP1
corr ret_av1 ret_IP1
di r(rho)^2

//regression with IP variable h=3
newey ret_av3 CSV IP1 INT_CSV_IP1, lag(1)
predict ret_IP13
corr ret_av3 ret_IP13
di r(rho)^2

//regression with IP variable h=12
newey ret_av12 CSV IP1 INT_CSV_IP1, lag(1)
predict ret_IP12
corr ret_av12 ret_IP12
di r(rho)^2

//regression with IP variable h=24
newey ret_av24 CSV IP1 INT_CSV_IP1, lag(1)
predict ret_IP24
corr ret_av24 ret_IP24
di r(rho)^2

//regression with IP variable h=36
newey ret_av36 CSV IP1 INT_CSV_IP1, lag(1)
predict ret_IP36
corr ret_av24 ret_IP36
di r(rho)^2

//OOS
//creating the OOS-R-squared h=1 IP1
gen INT_CSV_IP1= CSV*IP1

forvalues t=240/780 {
    newey ret_av1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}  
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT
forvalues t=240/780 {
    newey ret_av1 CSV IP1 INT_CSV_IP1 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//creating the OOS-R-squared h=3 IP1
forvalues t=240/780 {
    newey ret_av3 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av3 CSV IP1 INT_CSV_IP1 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//creating the OOS-R-squared h=12 IP1
forvalues t=240/780 {
    newey ret_av12 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)

drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av12 CSV IP1 INT_CSV_IP1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)

drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//creating the OOS-R-squared h=24 IP1
forvalues t=240/780 {
    newey ret_av24 CSV IP1 INT_CSV_IP1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)

drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av24 CSV IP1 INT_CSV_IP1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)

drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//creating the OOS-R-squared h=36 IP1
forvalues t=240/780 {
    newey ret_av36 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av36 CSV IP1 INT_CSV_IP1 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression with h=1 unemployment growth
//IS
gen INTER=CSV*un
newey ret_av1 CSV un_ch INTER, lag(1)
predict ret_CSV_un
corr ret_av1 ret_CSV_un
di r(rho)^2
//OOS
forvalues t=240/780 {
    newey ret_av1 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT
forvalues t=240/780 {
    newey ret_av1 CSV un_ch INTER if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression with h=3
//IS
newey ret_av3 CSV un_ch INTER, lag(1)
predict ret_CSV_un3
corr ret_av3 ret_CSV_un3
di r(rho)^2
//OOS
forvalues t=240/780 {
    newey ret_av3 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av3 CSV un_ch INTER if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression with h=12
newey  ret_av12 CSV un_ch INTER, lag(1)
predict ret_CSV_un12
corr ret_av12 ret_CSV_un12
di r(rho)^2
//OOS
forvalues t=240/780 {
    newey ret_av12 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} 
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT
genvs t=240/780 {
    newey ret_av12 CSV un_ch IP1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} 
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression with un_ch h=24
newey  ret_av24 CSV un_ch INTER, lag(1)
predict ret_CSV_un24
corr ret_av24 ret_CSV_un24
di r(rho)^2
//OOS
forvalues t=240/780 {
    newey ret_av24 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} 
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av24 CSV un_ch INTER if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[^`t'+1], dm[^`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression with un_ch h=36
newey ret_av36 CSV un_ch INTER, lag(1)
predict ret_CSV_un36

corr ret_av36 ret_CSV_un36
di r(rho)^2

//OOS
forvalues t=240/780 {
    newey ret_av36 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[^`t'+1], dm[^`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

forvalues t=240/780 {
    newey ret_av36 CSV un_ch INTER if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[^`t'+1], dm[^`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

**Establish the performance of the predictors**

//IS line CSV
newey ret_av1 CSV, lag(1)
predict retHAT
gen residuals=ret_av1-retHAT
gen re_sq= residuals^2
gen cum_sum=sum(re_sq) if dm>240
egen mean_ret=mean(ret_av1)
gen demean_ret= ret_av1-mean_ret
gen demean_ret_sq= demean_ret^2
gen cum_demean= sum(demean_ret_sq) if dm>240
gen performance_IS= cum_demean-cum_sum
drop cum_demean demean_ret_sq demean_ret mean_ret cum_sum re_sq residuals retHAT

//OOS line CSV
forvalues t=240/780 {
    newey ret_av1 if dm<`t', lag(1),
predict retHAT' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

forvalues t=240/780 {
    newey ret_av1 CSV if dm<`t', lag(1),
predict retHAT' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
gen dRMSE= sqrt(MSEn)-sqrt(MSEa1)
gen MSE_F=780*(MSEn-MSEa1)/MSEa1
```stata
display r_squared1 dRMSE MSE_F

gen performance_OOS=cum_sum2-cum_sum3 if dm>240
tsline performance_OOS performance_IS if dm>240
drop performance_OOS performance_IS cum_sum3

//IS line TERM
newey ret_av1 TERM, lag(1)
predict retHAT
gen residuals=ret_av1-retHAT
gen re_sq=residuals^2
gen cum_sum=sum(re_sq) if dm>240
egen mean_ret=mean(ret_av1)
gen demean_ret=ret_av1-mean_ret
gen demean_ret_sq=demean_ret^2
gen cum_demean=sum(demean_ret_sq) if dm>240
gen performance_IS=cum_demean-cum_sum
drop cum_demean demean_ret_sq demean_ret mean_ret cum_sum re_sq residuals retHAT

//OOS line TERM
forvalues t=455/780 {
    newey ret_av1 TERM if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT455-retHAT695)
gen res_squared2=(ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared2)
egen MSEa2=mean(res_squared2)
drop retHAT455-retHAT695 res_squared retHAT

gen r_squared2=1-1-MSEa2/MSEn
gen dRMSE2=sqrt(MSEn)-sqrt(MSEa2)
gen MSE_F2=780*(MSEn-MSEa2)/MSEa2
display r_squared2 dRMSE2 MSE_F2

gen performance_OOS=cum_sum2-cum_sum3 if dm>240
tsline performance_OOS performance_IS if dm>240
drop performance_OOS performance_IS cum_sum3

//IS line un_ch
newey ret_av1 un_ch, lag(1)
predict retHAT
gen residuals=ret_av1-retHAT
gen re_sq=residuals^2
gen cum_sum1=sum(re_sq) if dm>240
```
egen mean_ret=mean(ret_av1)
gen demean_ret_sq= (ret_av1-mean_ret)^2
gen cum_demean= sum(demean_ret_sq) if dm>240
gen performance_IS= cum_demean-cum_sum1
drop cum_demean demean_ret_sq demean_ret mean_ret cum_sum1 re_sq residuals retHAT

//OOS line un_ch
forvalues t=455/780 {
    newey ret_av1 un_ch if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT455-retHAT695)
gen res_squared3= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa3= mean(res_squared)
drop retHAT455-retHAT695 retHAT res_squared

egen r_squared3=1-MSEa3/MSEn
gen dRMSE3= sqrt(MSEn)-sqrt(MSEa3)
gen MSE_F3=(MSEn-MSEa3)/MSEa3
display r_squared3 dRMSE3 MSE_F3

gen performance_OOS=cum_sum2-cum_sum3 if dm>240
tsline performance_OOS performance_IS if dm>240

drop performance_OOS performance_IS cum_sum3

//IS line INFL
newey ret_av1 INFL, lag(1)
predict retHAT
gen residuals=ret_av1-retHAT
/gen re_sq= residuals^2
/gen cum_sum=sum(re_sq) if dm>240
egen mean_ret=mean(ret_av1)
gen demean_ret= ret_av1-mean_ret
gen demean_ret_sq= demean_ret^2
gen cum_demean= sum(demean_ret_sq) if dm>240
gen performance_IS= cum_demean-cum_sum
drop cum_demean demean_ret_sq demean_ret mean_ret cum_sum re_sq residuals retHAT

//OOS line INFL
forvalues t=240/780 {
    newey ret_av1 INFL if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT695)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa4=mean(res_squared)
drop retHAT240-retHAT695 res_squared retHAT

gen r_squared4=1-MSEa4/MSEn
gen dRMSE4= sqrt(MSEn)-sqrt(MSEa4)
egen MSE_F4=(MSEn-MSEa4)/MSEa4
display r_squared4 dRMSE4 MSE_F4

gen performance_OOS=cum_sum2-cum_sum3 if dm>240
tsl ine performance_OOS performance_IS if dm>240

drop performance_OOS performance_IS cum_sum3

//IS line logPE
newey ret_av1 logPE, lag(1)
predict retHAT
gen residuals=ret_av1-retHAT
gen re_sq= residuals^2
gen cum_sum=sum(re_sq) if dm>240
egen mean_ret=mean(ret_av1)
gen demean_ret= ret_av1-mean_ret
gen demean_ret_sq= demean_ret^2
gen cum_demean= sum(demean_ret_sq) if dm>240
gen performance_IS= cum_demean-cum_sum
drop cum_demean demean_ret_sq demean_ret mean_ret cum_sum re_sq residuals retHAT

//OOS line logPE
forvalues t=240/780 {
newey ret_av1 logPE if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}e gen retHAT=rowtotal(retHAT240-retHAT695)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa5=mean(res_squared)
drop retHAT240-retHAT695 res_squared retHAT

gen r_squared5=1-MSEa5/MSEn
gen dRMSE5= sqrt(MSEn)-sqrt(MSEa5)
egen MSE_F5=(MSEn-MSEa5)/MSEa5
display r_squared5 dRMSE5 MSE_F5
gen performance_OOS=cum_sum2-cum_sum3 if dm>240
tsline performance_OOS performance_IS if dm>240
drop performance_OOS performance_IS cum_sum3

//Plotting CSV and unemployment rate over time
twoway (tsline CSV, yaxis(1) lc(blue)) (tsline UN_, yaxis(2) lc(green))

//robustness check using data from 1973
drop if Date<197301
***regression k=1
newey ret_av1 CSV, lag(1)
predict ret_CSVa
corr ret_av1 ret_CSVa
di r(rho)^2
**regression k=3
newey ret_av3 CSV, lag(1)
predict ret_CSV3a
corr ret_av3 ret_CSV3a
di r(rho)^2
**regression k=12
newey ret_av12 CSV, lag(1)
predict ret_CSV12a
corr ret_av12 ret_CSV12a
di r(rho)^2
**regression k=24
newey ret_av24 CSV, lag(1)
predict ret_CSV24a
corr ret_av24 ret_CSV24a
di r(rho)^2
**regression k=36
newey ret_av36 CSV, lag(1)
predict ret_CSV36a
corr ret_av24 ret_CSV36a
di r(rho)^2

//regression with h=1 unemployment change
newey ret_av1 CSV un_ch INTER, lag(1)
predict ret_CSV_una
corr ret_av1 ret_CSV_una
di r(rho)^2

**OOS R-squared 1973-
//regression h=1
forvalues t=240/527 {

newey ret_av1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
gen res_squared= (ret_av1-retHAT)^2 if dm>406
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

forvalues t=240/780 {
  newey ret_av1 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
gen r_squared1=1-MSEa1/MSEn
display r_squared1
}
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

//regression h=3
forvalues t=240/527 {
  newey ret_av3 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
gen res_squared= (ret_av3-retHAT)^2 if dm>406
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

forvalues t=240/527 {
  newey ret_av3 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
gen res_squared= (ret_av3-retHAT)^2 if dm>406
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av1-retHAT)^2 if dm>406
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av3-retHAT)^2 if dm>406
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=12
forvalues t=240/527 {
  newey ret_av12 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av12-retHAT)^2 if dm>406
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

forvalues t=240/527 {
  newey ret_a12 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av12-retHAT)^2 if dm>406
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=24
forvalues t=240/527 {
  newey ret_av24 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared= (ret_av24-retHAT)^2 if dm>406
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

forvalues t=240/527 {
  newey ret_av24 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[\`t'+1], dm[\`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared = (ret_av24 - retHAT)^2 if dm>406
gen cum_sum3 = sum(res_squared)
egen MSEa1 = mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

gen r_squared1 = 1 - MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=36
forvalues t=240/527 {
    newey ret_av36 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} 
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared = (ret_av36-retHAT)^2 if dm>406
gen cum_sum2 = sum(res_squared)
egen MSEn = mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

forvalues t=240/527 {
    newey ret_av36 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} 
egen retHAT=rowtotal(retHAT240-retHAT527)
gen res_squared = (ret_av36-retHAT)^2 if dm>406
gen cum_sum3 = sum(res_squared)
egen MSEa1 = mean(res_squared)
drop retHAT240-retHAT527 res_squared retHAT

gen r_squared1 = 1 - MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//ROBUSTNESS TESTS FRO JANUARY 1952 UNTIL DECEMBER 1985
drop if Date>198512
//regression h=1
forvalues t=240/394 {
    newey ret_av1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} 
egen retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared = (ret_av1-retHAT)^2 if dm>240
gen cum_sum2 = sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

forvalues t=240/394 {
    newey ret_av1 CSV if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared= (ret_av1-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=3
forvalues t=240/394 {
    newey ret_av3 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

forvalues t=240/394 {
    newey ret_av3 CSV if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared= (ret_av3-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=12
forvalues t=240/394 {

newey ret_av12 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

g forvalues t=240/394 {
newey ret_av12 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
)
gen res_squared= (ret_av12-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=24
forvalues t=240/394 {
newey ret_av24 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

forvalues t=240/394 {
newey ret_av24 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
)
gen res_squared= (ret_av24-retHAT)^2 if dm>240
gen cum_sum3=sum(res_squared)
egen MSEa=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

73
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=36
forvalues t=240/394 {
    newey ret_av36 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}  
egen retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
gen MSEn=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

geno retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen cum_sum2=sum(res_squared)
gen MSEa1=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

geno retHAT=rowtotal(retHAT240-retHAT394)
gen res_squared= (ret_av36-retHAT)^2 if dm>240
gen MSEa1=mean(res_squared)
drop retHAT240-retHAT394 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//ROBUSTNESS CHECKS FROM JANUARY 1986

geno dm5=dm-324
//h=1
forvalues t=240/371 {
    newey ret_av1 if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}  
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av1-retHAT)^2 if dm5>238
gen cum_sum2=sum(res_squared)
gen MSEn=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT

forvalues t=240/371 {
newey ret_av1 CSV if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av1-retHAT)^2 if dm5>238
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=3
forvalues t=240/371 {
newey ret_av3 if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av3-retHAT)^2 if dm5>238
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT

forvalues t=240/371 {
newey ret_av3 CSV if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av3-retHAT)^2 if dm>238
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//regression h=12
forvalues t=240/371 {
newey ret_av12 if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av12-retHAT)^2 if dm5>238
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT

forvalues t=240/371 {
    newey ret_av12 CSV if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av12-retHAT)^2 if dm5>238
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//h=24
forvalues t=240/371 {
    newey ret_av24 if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av24-retHAT)^2 if dm5>238
gen cum_sum3=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT

forvalues t=240/371 {
    newey ret_av24 CSV if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
}
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av24-retHAT)^2 if dm5>238
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//h=36
forvalues t=240/371 {
    newey ret_av36 if dm5<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av36-retHAT)^2 if dm>238
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT
forvalues t=240/371 {
  newey ret_av36 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm5, dm5[`t'+1], dm5[`t'+1])
} egen retHAT=rowtotal(retHAT240-retHAT371)
gen res_squared= (ret_av36-retHAT)^2 if dm>238
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT240-retHAT371 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//creating OOS R-squared past 20 years
//h=1
forvalues t=540/780 {
  newey ret_av1 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>453
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT
forvalues t=540/780 {
  newey ret_av1 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
} egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av1-retHAT)^2 if dm>453
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT
gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1
//h=3
forvalues t=540/780 {
    newey ret_av3 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm["t'+1], dm["t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av3-retHAT)^2 if dm>453
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//CSV h=12
forvalues t=540/780 {
    newey ret_av12 if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm["t'+1], dm["t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av12-retHAT)^2 if dm>453
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1
egen MSEa1=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//CSV h=24
forvalues t=540/780 {
    newey ret_av24 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av24-retHAT)^2 if dm>453
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//CSV h=36
forvalues t=540/780 {
    newey ret_av36 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>453
gen cum_sum2=sum(res_squared)
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

forvalues t=540/780 {
    newey ret_av36 CSV if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>453
egen MSEa1=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1

drop cum_sum3 MSEn cum_sum2 MSEa r_squared1

//CSV h=36
forvalues t=540/780 {
    newey ret_av36 if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>453
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

define t=540/780 {
    newey ret_av36 CSV if dm<`t', lag(1),
    predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>453
egen MSEn=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT
newey ret_av36 CSV if dm<`t', lag(1),
predict retHAT`t' if inrange(dm, dm[`t'+1], dm[`t'+1])
}
egen retHAT=rowtotal(retHAT540-retHAT780)
gen res_squared= (ret_av36-retHAT)^2 if dm>453
gen cum_sum3=sum(res_squared)
egen MSEa1=mean(res_squared)
drop retHAT540-retHAT780 res_squared retHAT

gen r_squared1=1-MSEa1/MSEn
display r_squared1
drop cum_sum3 MSEn cum_sum2 MSEa r_squared1