Critical Finance Review, 2015, 4: 45–115

Seasonal Variation in Treasury Returns

Mark J. Kamstra, York University
Lisa A. Kramer, University of Toronto
Maurice D. Levi, University of British Columbia

ABSTRACT

We document an annual cycle in U.S. Treasuries, with variation in mean monthly returns of over 80 basis points from peak to trough. This seasonal Treasury return pattern does not arise due to macroeconomic seasonalities, seasonal variation in risk, cross-hedging between equity and Treasury markets, conventional measures of investor sentiment, the weather, seasonalities in the Treasury market auction schedule, seasonalities in the Treasury debt supply, seasonalities in the Federal Open Market Committee (FOMC) cycle, or peculiarities of the sample period considered. Rather, it is correlated with a proxy for variation in risk aversion linked to seasonal mood changes. Such a model can explain more than sixty percent of the average seasonal variation in monthly Treasury returns. The White (2000) reality test suggests this is not data snooping.

Keywords: Treasury bond returns, Treasury note returns, Market seasonality, Time-varying risk aversion.

JEL Codes: G11, G12

We thank Ivo Welch and three anonymous referees for helpful suggestions. We have benefited from conversations with Geert Bekaert, Hank Bessembinder, Michael Brennan, Hyung-Suk Choi, Ramon DeGennaro, Alex Edmans, Mark Fisher, Michael Fleming, Scott Frame, Ken Garbade, David Goldreich, Rob Heinkel, Shimon Kogan, Alan Kraus, and Monika Piazzesi. We thank seminar participants at the FRB of Atlanta, Michigan State, UBC, Maryland, Utah, and conference participants at the WFA meetings, EFA meetings, NFA meetings, and the CIPRÉE Applied Financial Time Series Workshop. We gratefully acknowledge support of the Social Sciences and Humanities Research Council of Canada. Kramer thanks the Canadian Securities Institute Research Foundation for support.

ISSN 2164-5744; DOI 10.1561/104.00000021

In this paper we establish the presence of a striking and previously unrecognized anomalous seasonal pattern in U.S. Treasury security returns. This seasonal pattern is strongly statistically and economically significant, with holders of Treasuries earning a monthly return that peaks in autumn, declines monotonically through to spring, and is on average 80 basis points higher in October than it is in April. Our focus is first to identify and document the previously unknown seasonality in Treasury returns and to show that it is both economically and statistically significant. Next we attempt to determine the exact source of the seasonal patterns in Treasury returns, and while we find support for many of the existing hypotheses on bond return movements, we demonstrate that none of these can account for the particular seasonal patterns we find. Exploring an alternative possibility, we find the seasonal cycle in Treasury returns is significantly correlated with a proxy for the timing of seasonal variation in investor risk aversion.

A large literature has explored return patterns of risky assets and the factors that explain them (see Cochrane (2005) for a comprehensive review of the asset pricing literature), however, much less attention has been devoted to the risk-free rate of return. Several papers have shown seasonalities in returns of various classes and maturities of bonds. These include Athanassakos (2008) who finds Canadian government bonds perform better in the May to October period than in the November to April period (describing this as "opposite" to the pattern in risky securities); Schneeweis and Woolridge (1979) who demonstrate the presence of autocorrelation in bond index returns; Jordan and Jordan (1991) who find no evidence of a day-of-the-week effect in corporate bonds over the past few decades but do find evidence of a January seasonal effect, a week-of-the-month effect, and a turn-of-the-year effect; Chang and Huang (1990) and Wilson and Jones (1990) who demonstrate the presence of a January seasonal effect in various U.S. corporate bond returns; and Fridson (2000) who documents seasonality in the spread between high-yield corporate bonds and Treasury bonds. Other papers have attempted to explain time-varying bond returns based on time-varying risk. For example, Boudoukh (1993) considers macroeconomic factors like consumption growth and inflation. Connolly *et al.* (2005) find that Treasury and stock markets can move in opposite directions for short periods, perhaps due to cross-market hedging. De Bondt and Bange (1992) and Brandt and Wang (2003) suggest that predictable, time-varying term premia on government bonds could arise due to unexpected inflation. Still other studies have explored the possibility of time-varying risk premia having an influence on government bond returns. For instance, Ilmanen (1995) examines long-term government bond returns in six countries and finds evidence of risk premia that depend on aggregate relative wealth measures. There is a closely related literature on bond yields that demonstrates time-varying risk premia on nominal bonds. See, for instance, Ang and Piazzesi (2003), Cochrane and Piazzesi (2005), and Murfin and Petersen (2014) for some recent evidence, and the classic work of Fama and Bliss (1987) and Campbell and Shiller (1991). Research on yields strongly supports bond return predictability based on yield spreads and macroeconomic factors. Collectively, these studies suggest possible alternative sources of seasonality in Treasury returns, and each is explored below. There are also behavioral explanations that potentially underlie the seasonality we demonstrate, for instance investor sentiment.[1] Baker and Wurgler (2006) find investor sentiment can impact security returns, and so we utilize their measure, as well as the Michigan Consumer Sentiment Index, to explore whether these sentiment measures help explain the seasonal Treasury return pattern we have identified.

Our findings contribute to the body of evidence that even in markets dominated by professional market participants, behavioral considerations can play a role. For example, Jin and Scherbina (2011) find that mutual fund managers exhibit the disposition effect, a behavioral tendency to hold onto loser stocks too long. Fleming *et al.* (2005) show that Treasury bill auction participants frequently place what the authors call “inefficient bids,” which result in them transacting a quantity lower than would be optimal. Further, Fleming and Garbade (2007) find that U.S. Treasury market dealers operating in noncompetitive auctions routinely forgo arbitrage opportunities and overpay to borrow securities. Our paper also joins the growing body of literature that explores the possible influence of affect (emotions) on financial markets. See Kamstra *et al.* (2001) and Kamstra *et al.* (2003), Statman *et al.* (2008), Kaplanski and Levy (2008), and Bassi *et al.* (2013) for instance.

[1]Note that common usage of the term “sentiment” typically refers to investor mistakes. For example, Shefrin (2008, p. 213), observes “in finance, sentiment is synonymous with error . . . errors of individual investors, particularly representativeness and overconfidence, combine to produce market sentiment.” Sentiment could encompass the seasonally varying risk aversion we investigate (e.g. seasonal variation in investor sentiment *toward taking risk*), but for simplicity in this paper we refer to sentiment and seasonally varying risk aversion as distinct concepts.

After documenting a statistically significant and economically large seasonal cycle in Treasury returns, we consider a broad range of possible explanations, including macroeconomic shocks, cross-hedging (whereby periods of stock market uncertainty may induce effects in Treasury returns), investor sentiment, the Fama-French and momentum risk factors, and several factors related to activities of the U.S. Treasury and Federal Reserve. Possible Treasury and Federal Reserve influences that we consider include the management of the supply of Treasury debt, the Federal Reserve Board's annual cycle of rate-setting meetings, and a significant change to the Treasury auction announcement policy that was introduced in the late 1970s to facilitate liquidity in the Treasury market. We find that none of these alternatives are capable of fully explaining the seasonal pattern in Treasury returns. Only models that include a proxy for time-varying investor risk aversion linked to seasonal depression appear able to explain seasonally varying Treasury security returns.

Application of the White (2000) reality test shows that the relation between seasonal depression and Treasury returns is unlikely to be a result of data mining. We report on a variety of sub-sample analyses to investigate the stability of the seasonal pattern in Treasury returns and we find that evidence of the seasonal pattern did not appear until after the Treasury introduced auctions for the sale of notes and bonds in the 1970s. Before this market-driven price-setting mechanism was in place, there was very little seasonal variation in Treasury note and bond returns. However, after auctions were introduced and Treasury issuances began following a regular, predictable schedule in the early 1980s, we demonstrate that seasonal variation became a stable feature of the Treasury returns.

Our findings are robust to alternate estimation techniques, the consideration of different parts of the Treasury maturity spectrum, different ways of measuring many of the variables (such as using currently available, updated macroeconomic data or vintage series that were available at the time investors made decisions), different proxies for seasonally varying investor risk aversion, different ways of modeling time-series characteristics of the data, the inclusion of raw or seasonally adjusted weather data in models, and other factors.

In terms of economic magnitude, seasonally varying investor risk aversion is able to explain a large portion of seasonal variation in Treasury returns. Of the 80 basis point swing in average Treasury returns from October through April, seasonally varying investor risk aversion appears to be responsible for 50 basis points of the movement, which is more than 60 percent of the variation. This finding bears importantly on our understanding of financial markets.

1 Treasury Returns

In this section we document seasonal patterns in Treasury returns, based on both nominal returns and returns in excess of the 30-day T-bill rate (which we generically refer to as "excess" returns). We consider monthly returns to holding the medium-to-long end of Treasury market securities, specifically 20-year, 10-year, 7-year, and 5-year Treasury bond and note returns, where the returns include interest and capital gains/losses. We consider data from 1952 onward, consistent with the Campbell (1990) observation that interest rates were almost constant in the United States until 1951, after which an accord between the Federal Reserve Board and the U.S. Treasury permitted interest rates to respond more freely to market forces. We limit our primary focus to the medium-to-long end because rate movements in the short end do not respond freely to market forces, even following the accord between the U.S. Treasury and the Federal Reserve Board. Gibson (1970), for instance, notes in reference to the short end that an "aim of the Federal Reserve System is to accommodate seasonal swings in the financial needs of trade, and the System tries to do this by removing seasonal fluctuations from interest rates" (p. 442). In Appendix C we do, however, provide results for the short end of the Treasury market and we discuss related institutional details. These results show evidence of seasonality in the short end, though weaker than found in the longer-term Treasuries.[2] (Note that Appendices A and B appear at the end of the paper; all other appendices are available on the journal's web site and at http://www.markkamstra.com.)

Table 1 contains summary statistics on the Treasury and equity return series we consider in this study, and Appendix A contains details about the sources of all data used in this paper. The top portion of Table 1 reports summary statistics for nominal Treasury return series, the middle portion reports on Treasury returns in excess of the 30-day T-bill rate, and the bottom portion reports on the 30-day Treasury return data used in constructing the excess returns, as well as the U.S. stock index return data

[2]Gibson (1970) also notes weak seasonal patterns in 90-day T-bill rates.

Series	Mean	Std	Beta	Seasonality test: Asymptotic p-values			
				Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
20-year Treasury: Nominal	0.54	2.64	0.11	0.332	**0.077**	0.169	**0.007**
10-year Treasury: Nominal	0.52	2.13	0.08	0.365	0.198	0.113	**0.002**
7-year Treasury: Nominal	0.55	1.78	0.06	0.257	**0.034**	**0.027**	**0.001**
5-year Treasury: Nominal	0.53	1.49	0.04	0.453	**0.090**	**0.023**	**0.001**
Average Treasury: Nominal	0.53	1.94	0.07	0.467	**0.080**	**0.064**	**0.001**
20-year Treasury: Excess	0.13	2.64	0.11	0.300	**0.078**	0.172	**0.007**
10-year Treasury: Excess	0.11	2.11	0.08	0.322	0.203	0.121	**0.002**
7-year Treasury: Excess	0.14	1.76	0.06	0.238	**0.033**	**0.032**	**0.001**
5-year Treasury: Excess	0.12	1.47	0.04	0.414	**0.096**	**0.027**	**0.001**
Average Treasury: Excess	0.13	1.92	0.07	0.421	**0.081**	**0.071**	**0.001**
30-day Treasury: Nominal	0.41	0.24	—	0.166	0.485	0.892	0.600
Equity: Nominal	0.98	4.19	—	**0.014**	0.293	**0.054**	0.744
Equity: Excess	0.57	4.21	—	**0.017**	0.320	**0.055**	0.751

Table 1: Summary statistics—Treasury and equity return series (nominal and excess).

Description: This table contains summary statistics on monthly data. See Appendix A for data sources. For each series we present the mean (Mean) and standard deviation (Std). For the Treasury bond and note return series, we also report the CAPM beta. We present asymptotic p-values associated with four tests for seasonality: nonspecific monthly, fall vs. winter, September vs. March, and October vs. April. P-values below 10 percent are indicated in bold. See Section 1 for estimation details on regressions that are used to perform the seasonality tests. Appendix D contains a broader set of summary statistics, including minimum, maximum, skewness, and kurtosis. The sample period for the Treasury and equity series is 01/1952–12/2007 ($N = 672$).

(including dividends), for reference. Table 1 also contains the results of seasonality tests. The seasonality tests we consider are motivated with reference to Kamstra *et al.* (2003), following from their observation that "on balance the seasonally asymmetric effects of SAD [a form of seasonal depression] are shifting [stock] returns from the fall to the winter" (p. 336). Our hypothesis is that seasonally varying investor risk aversion due to seasonally varying investor mood also drives seasonal patterns in Treasury returns, shifting those returns from the winter to the fall.[3] We elaborate on the specifics of this hypothesis in the next section, but first we document the annual seasonality that we find in the Treasury return data.

The average of each of the monthly nominal (excess) Treasury index return series is roughly 50 (10) basis points. The standard deviations of the Treasury index returns are well below that of the equity index over the same period, increasing monotonically with maturity. The stock index has a mean nominal return close to one percent per month and a standard deviation exceeding 4 versus 1.47 to 2.64 for the U.S. Treasury bond and note indices. Exposure to market risk is a traditional measure of systemic risk, thus we also report the capital asset pricing model (CAPM) beta for each of the individual Treasury bond and note series. Beta is measured by regressing the Treasury excess returns on the equity index excess returns. The beta of all the Treasury classes is virtually zero.[4] Appendix D contains a broader set of summary statistics for the series shown in Table 1. Those statistics show that the minimum and maximum observed for each Treasury series generally span a smaller range as maturity shortens. Additionally, the Treasury series are leptokurtotic and skewed toward positive returns.

Figures 1 and 2 contain plots of the monthly average Treasury excess return series, starting with September and ending with August. Results are qualitatively identical for nominal returns. Figure 1 depicts monthly Treasury excess returns averaged across the 5-, 7-, 10-, and 20-year maturities, represented with a thick solid line. Dotted lines depict a 90 percent

---

[3] Kamstra *et al.* (2014) explore an asset pricing model with a representative agent who experiences seasonally varying risk preferences. They find plausible values of risk-preference parameters are capable of generating the empirically observed seasonal patterns in equity and Treasury returns.

[4] It is a commonly held belief that short- and long-term Treasury securities represent a safe haven from risk. For example, during the 2008/2009 financial crisis, and even more recently in August 2011, all Treasury maturities were in high demand. Press coverage on this matter includes *Wall Street Journal* articles by Lauricella *et al.* (2011) and Zeng (2011).

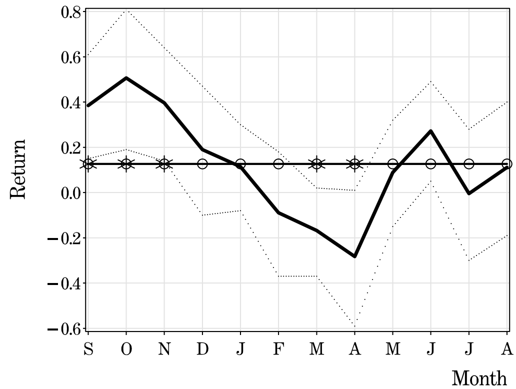


Figure 1: Average monthly excess Treasury returns, averaged across maturities.

Description: This figure contains plots of monthly Treasury excess returns, running from September through to August. Monthly returns are averaged across the 5-, 7-, 10-, and 20-year maturities. The thick solid line is the monthly mean residuals and dotted lines represent a 90 percent confidence interval around the monthly means. The average annual return is the thin solid line with circles (and an X in cases where the average return falls outside the confidence interval). The data span 01/1952 to 12/2007.

Interpretation: We see economically and statistically significant monthly seasonal variation in Treasury returns. Monthly returns averaged across the series decline more than 80 basis points from their peak in October to their low in April.

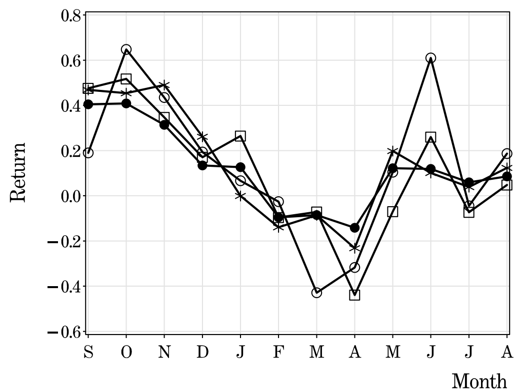


Figure 2: Average monthly excess Treasury returns, individual maturities.

Description: This figure contains plots of monthly Treasury excess returns, running from September through August. The 5-, 7-, 10-, and 20-year series are represented by lines with solid circles, asterisks, hollow squares, and hollow circles, respectively. The data span 01/1952 to 12/2007.

Interpretation: The monthly seasonal variation in the average Treasury returns shown in Figure 1 is also evident in the returns for individual maturities.

confidence interval around the monthly means.[5,6] The thin solid line with circles represents the average annual return, and an X appears over the circle in months where the average return falls outside of the confidence interval. Monthly average Treasury excess returns are high and above the annual average (of approximately 0.13 percent) through the fall months and are below average in the winter months. In April, the monthly average excess return reaches its lowest point of the year. The decline in returns is monotonic from the annual peak in October to the annual trough in April. Further, the magnitude of the decline in average monthly returns from October through to April is striking: the difference is about 80 basis points. The decline from October to April is also statistically significant, and five months of the year (September, October, November, March, and April) are significantly different from the annual mean. Figure 2 contains plots of each of the four individual average monthly Treasury excess return series, which all show very similar seasonal variation.

Formal tests also support the notion that Treasury returns are in effect shifted between the fall and winter seasons. We consider three tests. First, we use an indicator variable equal to one in the fall (October, November, and December), equal to minus one in the winter (January, February, and

---

[5]There are several approaches one could adopt to calculate the confidence interval around the mean monthly returns. The simplest is to use the standard deviation of the monthly mean returns directly. However, this would ignore information about the cross-sectional variability of returns across the four Treasury series. Instead, we form a system of equations with the four series and estimate a fixed-effects model with twelve dummy variables (one for each month) and no intercept. Consistent with the typical implementation of a fixed-effects model, we allow each series to have a different mean, while estimating one set of parameter values for the variables each series has in common, in this case the monthly dummy variables. From this regression we obtain standard errors on the monthly dummies to form the confidence intervals around the monthly mean returns.

[6]We use generalized method of moments (GMM) as in Hansen (1982) and Newey and West (1987) heteroskedasticity and autocorrelation consistent standard errors, and following Newey and West (1994) we use the Bartlett kernel and an automatic bandwidth parameter (autocovariance lags) equal to the integer value of $4 \cdot (T/100)^{2/9}$. The moment conditions we use include orthogonality between a small set of instruments and the errors. For instruments we use the constant, a lag of the Center for Research in Security Prices (CRSP) value-weighted return (entire U.S. market return, including dividends), the contemporaneous 30-day T-bill rate as suggested by Ferson and Foerster (1994), and the 12 monthly dummy variables. The confidence intervals are similar if we use full information maximum likelihood and MacKinnon and White (1985) bootstrap heteroskedasticity-consistent standard errors. For a detailed discussion on the use of GMM, see Cochrane (2005), Chapters 10 and 11.

March), and equal to zero otherwise ($\text{DM}_{t,\text{fall/winter}}$). Second, we employ an indicator variable equal to one in September, equal to minus one in March, and equal to zero otherwise ($\text{DM}_{t,\text{Sep/Mar}}$). Third, we use an indicator variable equal to one in October, equal to minus one in April, and equal to zero otherwise ($\text{DM}_{t,\text{Oct/Apr}}$). The October/April and September/March indicator variable specifications come closest to matching the typical timing of the onset of individuals' seasonal decline in mood and their ultimate recovery in mood, as documented in clinical studies of individuals who suffer from seasonal mood variation (we elaborate on these studies below), and the fall/winter indicator variable should pick up the average impact across the full fall and winter seasons. Our null hypothesis is that there is no seasonal difference in returns, i.e., that the coefficient on a given indicator variable is zero, against the alternative of returns being shifted from winter into fall. Our alternative hypothesis implies that these indicator variables should have positive coefficients when applied to Treasury returns.

For instance, to test whether a Treasury return series has the same mean value in the fall and winter versus the alternative that the fall and winter means deviate from the annual average by an equal and opposite amount, we estimate the following model:

$$r_{i,t} = \alpha_i + \beta_{i,\text{fall/winter}} \cdot \text{DM}_{t,\text{fall/winter}} + \epsilon_{i,t}.$$

The dependent variable is the Treasury return series, where i indexes 5-, 7-, 10-, or 20-year maturity. We estimate alternate versions of this model to produce the various seasonality tests, replacing $\text{DM}_{t,\text{fall/winter}}$ with either $\text{DM}_{t,\text{Sep/Mar}}$ or $\text{DM}_{t,\text{Oct/Apr}}$. A given seasonality test is a two-sided t-test on the indicator variable coefficient to differ from zero.

We also perform a fourth test for seasonal variation. This one tests for seasonal variation of nonspecific form, involving a regression of the return series on a constant and monthly dummy variables, excluding January. We test whether the monthly dummy variables jointly differ from zero, an eleven degree of freedom χ^2 test. Each seasonality test is performed by estimating the model using the Hansen (1982) generalized method of moments (GMM);[7] tests are performed using Newey and West (1987) and Newey and West (1994) heteroskedasticity and autocorrelation consistent

[7] Hansen (1982), Staiger and Stock (1997), and Stock and Wright (2000) detail conditions sufficient for consistency and asymptotic normality of GMM estimators. The moment conditions we use include orthogonality between a small set of instruments and the errors. For instruments we use the constant, a lag of the CRSP value-weighted return (entire U.S.

(HAC) standard errors. See footnote 6 for details. The HAC standard errors control for well-known heteroskedasticity and autocorrelation effects in returns. Note that we employ GMM and Newey and West (1987) and Newey and West (1994) standard errors for all estimations reported in this paper. In the tables of summary statistics provided in Appendix D we report a second set of p-values based on bootstrapping the distribution of the seasonality test statistics.[8,9] Results for all four sets of seasonality tests are virtually identical based on both sets of standard errors. In untabulated results, we find the seasonality test results are also very similar if we use full information maximum likelihood estimation and MacKinnon and White (1985) bootstrap heteroskedasticity-consistent standard errors, and/or if

---

market return, including dividends), the contemporaneous 30-day T-bill rate as suggested by Ferson and Foerster (1994), and the indicator variables used for the regression slightly modified as follows: For the fall versus winter seasonality test, dummies for each of the fourth and first quarters are included in the instrument list, for the September versus March seasonality test, dummies for each of September and March are included, and for the October versus April seasonality test, dummies for each of October and April are included.

[8]Ferson and Foerster (1994) note that in cases where there are too many over-identifying restrictions relative to the sample size, the asymptotic distribution of test statistics can be a poor approximation of the finite-sample test distribution.

[9] We employ block bootstrap resampling to allow for data dependence, as detailed by Politis and Romano (1994) and employed by White (2000). Politis and Romano (1994) show that this technique produces valid bootstrap approximations for means of alpha-mixing processes, so long as the block length increases with sample size. Results of Gonçalves and White (2002) and Gonçalves and White (2005) establish the consistency of the bootstrap variance estimator of Politis and Romano (1994) for the sample mean in the presence of heteroskedasticity and dependence of unknown form. Politis and Romano (1994) use blocks of data of random length, distributed according to the geometric distribution with mean block length b. The parameter b is chosen so that block length is data-dependent, with Politis and Romano (1994) recommending a scaling proportional to $N^{1/3}$, where N=sample size. The setting $b = N^{1/3}$ would lead to a mean block length of approximately 9 observations in our sample, which is a fairly long block length for monthly return data. White (2000) remarks that a mean block length of 10 for daily data is appropriate given the weak autocorrelation of returns. This would translate to the minimum mean block length of 2 for our monthly data. We set the block length to 5 but find our results are virtually identical for block lengths between 2 and 10. We use 1,000 resamples, which we find produces stable results. White (2000) suggests 500 or 1,000 resamples and uses 500 in his empirical application on S&P 500 stock returns. Although the tests reported in Table 1 are all one series at a time, we perform much of the subsequent analysis on all four Treasury series with system-of-equation estimation. Rilstone and Veall (1996) show substantially better inference can result using the bootstrap in a system-of-equations estimation context. Palm *et al.* (2011) show asymptotic validity of block bootstrap tests in the context of panel data with cross-sectional dependence.

we include a sufficient number of lags of the dependent variable to directly control for return autocorrelation.

The last four columns of Table 1 contain the results of the seasonality tests on the Treasury return series. In each cell, we provide the asymptotic p-value. Cases significant at the 10 percent level or better are indicated in bold. We consider first the three tests for seasonality of a specific form. All of the Treasury bond and note return series exhibit strong seasonality, with each exhibiting p-values below 0.1 percent for the October/April test, all but the 10-year returns exhibiting p-values below 10 percent for the fall/winter test, and all but the 10- and 20-year returns exhibiting p-values below 10 percent for the September/March test. Considering the nominal and excess average returns across the series, we reject the null at the 10 percent level or better for all three sets of seasonality tests of specific form. Analysis based on the bootstrapped distributions of these test statistics (reported in Appendix D) verifies the robustness of the finding of seasonality. The test for nonspecific monthly seasonality, based on regressing returns on a constant and a dummy variable for each month except January, is insignificant for all of the Treasury series. Note that the test for nonspecific monthly seasonality is a weak test for a specific form of seasonal variation such as that which may arise due to the seasonally varying risk aversion hypothesis we investigate. Because there is legitimate concern for data mining, in later analysis we perform the White (2000) reality test to investigate whether the results in support of the seasonally varying risk aversion hypothesis are an artifact of data snooping.

For comparison, the bottom portion of Table 1 contains summary statistics and seasonality tests based on the U.S. CRSP value-weighted equity index return series (including dividends). These returns show a statistically significant September/March seasonality and significant evidence of nonspecific monthly seasonality; the latter is likely primarily a consequence of the well-known January seasonal effect in small-capitalization stocks.

2 Seasonally Varying Mood and Risk Aversion

A seasonal connection between investor mood, investor risk aversion, and asset returns was first proposed by Kamstra *et al.* (2003), who found evidence consistent with the hypothesis that seasonally varying risk aversion impacts equity index returns, with shorter days leading first to declining daily returns in the fall and then higher daily returns as the days lengthen (and consequently higher expected returns for investors who hold equities over the fall and into the winter season). The connections between the seasons, investor mood, investor risk aversion, and financial markets rest on two key building blocks. First, investor mood varies systematically by season. Second, seasonal variation in investor mood leads to seasonal variation in investor risk aversion, which in turn impacts asset returns.

The first of these building blocks, the connection between mood and season, is supported by extensive medical evidence which establishes that up to 10 percent of the population experiences severe depression during the fall and winter seasons, a condition known as seasonal affective disorder (SAD).[10] Additionally, recent studies suggest most people experience some degree of seasonal mood variation. For instance, Harmatz *et al.* (2000) and Kramer and Weber (2012) find even healthy individuals, i.e., those who do not meet the medical criteria for a diagnosis of severe seasonal depression, are more depressed on average in the fall and winter, a phenomenon commonly described as 'winter blues.' Young *et al.* (1997) and Lam (1998) document the clinical onset of depressive symptoms and recovery from depressive symptoms among North Americans known to experience seasonal depression. These data indicate that depression symptoms typically begin in early-to-mid fall and fully dissipate by early spring, though exact timing varies by individual.[11,12]

[10] Precise estimates of the prevalence of SAD vary depending on latitude and diagnostic method. See Kamstra *et al.* (2012) for details.

[11] Young *et al.* (1997) study 190 Chicago residents who experience seasonal depression and find that 74 percent of them first experience depression symptoms between mid-September and early November. Lam (1998) studies 454 Vancouver residents who experience seasonal depression and also finds that the peak timing of onset is in early fall. Lam further establishes that the timing of clinical remission peaks in April, closely followed by March. Onset and recovery are typically separated by several months.

[12] September and October are the months during which the highest proportion of individuals experience the onset of depression. If investors begin rearranging their portfolios when they first become risk averse, then September and October should be the approximate time of year when we observe the largest positive impact on Treasury returns due to SAD. Although some individuals begin recovering in January, the peak time for recovery is March/April. Thus we should see mood improvement impacting returns as early as January, but the peak effects should occur roughly in March/April. In short, security returns, which are an income flow, should respond to the flow of depression-affected investors, not the stock of investors actively suffering from depression.

Regarding the second link, connecting seasonality in mood with seasonality in risk aversion, Kramer and Weber (2012) study hundreds of individuals' risk preferences across seasons in a survey/experimental context, including individuals who meet the diagnostic criteria for SAD and a comparison group of individuals who do not. They develop a task with real financial consequences, called the safe asset versus risky (SAVR) task, and find those who suffer from SAD exhibit greater financial risk aversion year-round than those who do not. Further, they find both groups are significantly more likely to choose a safe asset than a risky asset in winter, especially the SAD group. More generally, Pietromonaco and Rook (1987), Carton *et al.* (1992), Carton *et al.* (1995), and Smoski *et al.* (2008) show that (not necessarily seasonally) depressed individuals are more averse to risk, including risk of a financial nature.

We explore the possibility that seasonally varying risk aversion may help to explain the seasonal cycle in Treasury returns. Specifically, if investors experience a dampening of mood and hence an increase in risk aversion in the fall, the price of Treasuries should rise, resulting in higher-than-average realized Treasury returns in the fall. Then when investors' mood rebounds and their risk aversion diminishes in the spring, Treasury prices fall, resulting in lower-than-average realized returns. This seasonal pattern is the converse of the pattern of returns we would expect for risky assets.

2.1 Measuring Seasonally Varying Mood and Risk Aversion

Medical research clearly shows seasonal depression primarily arises as a consequence of seasonally reduced exposure to daylight and not other environmental factors; see Molin *et al.* (1996) and Young *et al.* (1997). In studying the connection between seasonally varying investor risk aversion and equity returns, Kamstra *et al.* (2003) therefore use a proxy for seasonally varying investor risk aversion based on seasonal variation in daylight. We consider in place of their measure an alternative measure linked more directly to the clinically observed timing of seasonal depression symptoms in individuals, constructed using data from Lam (1998) and yielding the best currently available proxy of the timing of seasonally varying risk aversion in the general population.[13] While individuals in the Lam sample have

[13]There exist other clinical studies that document the timing of seasonal depression symptoms, including Young *et al.* (1997). We base our measure on data from the Lam (1998) study because, unlike other clinical studies, his study details the timing of both

been clinically diagnosed as suffering from seasonal depression, results from studies including Harmatz *et al.* (2000) and Kramer and Weber (2012) support the view that marked seasonal variation in mood occurs in healthy individuals as well.

Details on the construction of our measure of seasonally varying investor risk aversion are as follows. First, we form an 'incidence' variable which reflects the monthly proportion of seasonal-depression-sufferers who are actively experiencing symptoms in a given month. The incidence variable is calculated by cumulating, monthly, the proportion of subjects who have experienced the *onset* of their seasonal depression symptoms (cumulated starting in late summer, when a small proportion of subjects are first diagnosed with onset) and then deducting the cumulative proportion of people who have experienced full *recovery*. The resulting monthly incidence variable takes on values between zero percent, in summer, and close to 100 percent, in winter. This measure of incidence is based on *estimates* of onset and recovery in the broader population of all North Americans who suffer from seasonal depression, hence incidence is measured with error. To avoid an error-in-variables bias (see Levi (1973)), we construct an instrumented version of the incidence variable.[14] Finally, the *monthly change* in this instrumented incidence variable yields the onset/recovery used in our tests, which we denote $\hat{\mathrm{OR}}_t$ (short for onset/recovery, with the hat indicating the variable is the fitted value from a regression). More specifically, the monthly variable $\hat{\mathrm{OR}}_t$ is calculated as the value of the daily instrumented incidence value on the 15th day of a given month minus the value of the daily instrumented incidence value on the 15th day of the

onset of and recovery from seasonal depression symptoms. Our measure and findings are qualitatively identical if we combine data from the Lam and Young et al. studies.

[14]To produce the instrumented version of incidence, first we smoothly interpolate the monthly incidence variable to daily frequency using a spline function. We need to produce an instrumented value of incidence that is strictly positive but no more than 100 percent, so we run a logistic regression of the daily incidence on our chosen instrument, the length of day. (The nonlinear model is $1/(1+e^{\alpha+\beta\cdot \mathrm{day}_t})$, where day_t is the length of day t in hours in New York and t ranges from 1 to 365. The $\hat{\beta}$ coefficient estimate is 1.18 with a standard error of 0.021, the intercept estimate is −13.98 with a standard error of .246, and the regression R^2 is 94.9 percent.) The fitted value from this regression is the instrumented measure of incidence. Employing additional instruments, such as change in the length of the day, makes no substantial difference to the fit of the regression or the subsequent results using this fitted value.

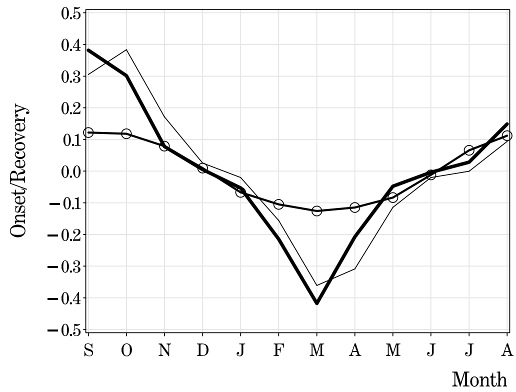


Figure 3: Onset/recovery and change in length of night.

Description: The thick plain line depicts the onset/recovery variable ($\hat{OR}_t$), reflecting the change in the instrumented proportion of seasonal-depression-affected individuals actively experiencing symptoms. The thin plain line represents the observed onset/recovery data (OR_t) based on the Lam (1998) study. The line with circles is the change in the length of night, normalized by dividing by 12 (the average annual length of night). These monthly series are all calibrated to the 15th day of each month and are plotted starting with September and ending with August.

Interpretation: This figure depicts monthly values of the following: the onset/recovery variable, the clinical incidence of onset of and recovery from seasonal depression symptoms in a sample of people who suffer from SAD, and the change in length of night in New York City.

previous month.[15,16]

$\hat{OR}_t$ reflects the change in the proportion of individuals actively experiencing depression symptoms. The monthly values of $\hat{OR}_t$ are plotted as a thick plain line in Figure 3, again starting with September and ending with August, together with the corresponding values of OR_t (thin plain line) and the change in the length of night divided by 12 (thin line with circles). Notice that all measures are positive in summer and fall and negative in

[15] The values of $\hat{OR}_t$ by month, rounded to the nearest integer and starting in July, are: 3, 15, 38, 30, 8, 1, −5, −21, −42, −21, −5, 0. These values represent the instrumented *change* in incidence of symptoms. The correlation of the instrumented fitted value with the realized onset/recovery is 0.96 and the correlation of the fitted value with the change in length of night is 0.91.

[16] We find qualitatively identical results when we perform our analysis replacing $\hat{OR}_t$ with either OR_t or the change in the length of night. See Appendix E.

winter and spring. The values peak near the fall equinox and reach a trough near the spring equinox: the equinoxes are the points of inflection in the annual daylight cycle.

2.2 Does Seasonally Varying Risk Aversion Help Explain the Treasury Return Annual Cycle?

We turn now to testing whether the onset/recovery variable helps explain the seasonal patterns in Treasury returns evident in Table 1 and Figures 1 and 2. Excess returns are required for some of the alternative models we consider, thus we focus on excess returns in our regression analysis. Results are virtually identical when using nominal returns. We regress excess Treasury returns on $\hat{\mathrm{OR}}_t$:

$$r_{i,t} = \mu_i + \mu_{i,\hat{\mathrm{OR}}} \cdot \hat{\mathrm{OR}}_t + \epsilon_{i,t}. \tag{1}$$

We call this Model 1. The sample period used for estimation is 01/1952–12/2007.

Panel A of Table 2 contains the system-of-equations estimation results for Model 1, using the Hansen (1982) GMM and Newey and West (1987) and Newey and West (1994) standard errors, accounting for cross-equation covariance between the four return series, and heteroskedasticity and autocorrelation in returns.[17,18,19]

---

[17] To calculate the standard errors we follow Newey and West (1994), and use the Bartlett kernel and an automatic bandwidth parameter (autocovariance lags) equal to the integer value of $4 \cdot (T/100)^{2/9}$.

[18] The instruments we use in all of our regressions to form the GMM moment conditions, unless noted otherwise, are a constant, the explanatory variables (in Equation (1) this is $\hat{\mathrm{OR}}_t$), 30-day T-bill returns, and the lagged CRSP value-weighted equity index returns including dividends. See footnote 6 for further estimation details.

[19] Throughout the paper, regression results are very similar if we use full information maximum likelihood (FIML) or a seemingly unrelated regression rather than GMM, and/or if we include a sufficient number of lags of the dependent variable to directly control for return autocorrelation, and/or if we introduce small changes in the number of instruments used to identify model parameters and window width smoothing parameters employed in GMM estimation. (See Appendix F for results based on FIML.) In general, the more instruments used to identify model parameters, the more significant are the parameter estimates, consistent with the intuition that the more over-identifying information used, the better we are able to estimate parameters of the system. The small-sample properties of our tests degrade with excessive numbers of moment conditions, however. Ferson and Foerster (1994) consider the use of GMM and HAC standard errors in the context

Panel A: Regression estimates and AR/ARCH test statistics

	Treasury Excess Returns			
Parameter or Statistic	20-Year Coeff. (Std Err)	10-Year Coeff. (Std Err)	7-Year Coeff. (Std Err)	5-Year Coeff. (Std Err)
μ	0.153*	0.106	0.135**	0.116**
	(0.088)	(0.076)	(0.062)	(0.051)
$\mu_{\hat{OR}}$	1.103**	1.027***	0.949***	0.776***
	(0.454)	(0.362)	(0.294)	(0.243)
R^2	0.0072	0.0087	0.0115	0.0110
AR(12)	16.85	11.26	9.27	13.09
ARCH(12)	90.11***	106.68***	95.53***	122.26***

Panel B: Model statistics

Parameter or Statistic	Value
Proportion of Monthly Seasonal Variation in Returns Explained	0.63
GMM Test of Overidentification Restrictions	10.49
MMSC-BIC of Full Model/Constant-Only Model	−41.59/−40.82
MMSC-HQIC of Full Model/Constant-Only Model	−20.98/−20.21
Number of Parameters	8
Number of Moment Conditions	16

Table 2: Model 1—Regression estimates and model statistics.

Description: Panel A contains coefficient estimates, standard errors, R^2, and tests for autocorrelation (AR) and ARCH based on estimating Equation (1) as a system of equations. See Section 2.2 for estimation details. One, two, and three asterisks denote significance at the 10, 5, and 1 percent level respectively, based on two-sided tests. Panel B reports the proportion of monthly seasonal variation in returns that the model is able to explain, stated in percent returns, two information criteria (defined so that we wish to minimize them), the number of parameters, and the number of moment conditions. The sample period is 01/1952–12/2007 ($N = 672$).

Interpretation: The onset/recovery variable is statistically significant, consistent with the hypothesis that seasonally varying risk aversion is associated with seasonality in Treasury returns.

We see that the $\hat{OR}$ coefficients on all four Treasury series are positive and significant. The onset/recovery variable itself is positive in the fall and negative in the winter, thus the positive coefficients imply above-average Treasury returns in the fall and below-average Treasury returns in the winter, consistent with the seasonally varying risk aversion hypothesis.

We also present R^2, a Wald χ^2 test statistic for the presence of up to 12 lags of autocorrelation (AR), and a Wald χ^2 test statistic for the presence of up to 12 lags of ARCH. The test for ARCH is a standard LM test of order 12 (see Engle (1982)). To perform the test for autocorrelation, we augment our regression with 12 lags of the residuals, estimate MacKinnon and White (1985) bootstrap heteroskedasticity-consistent standard errors with OLS and test for the joint significance of these terms.

In Panel B of Table 2 we provide additional estimation details. The first row of Panel B contains an indicator of economic significance, calculated as follows. We compute the average monthly return across the four Treasury series, and we calculate the average monthly predicted return based on Model 1. We regress the actual monthly averages on the predicted monthly averages, suppressing the intercept, and the resulting R^2 is the proportion of seasonal variation explained by the model. Based on this indicator, we see that Model 1 explains 63 percent of the average annual seasonal variation in Treasury returns.

In the remaining rows of Panel B we report a GMM test of overidentifying restrictions, two information criteria (labeled MMSC-BIC and MMSC-

---

of a system-equation estimation with monthly U.S. Treasury and stock returns. They perform Monte Carlo experiments to evaluate the small-sample performance of GMM and HAC standard errors with system-equation estimation and testing in the presence of autocorrelation and autoregressive conditional heteroskedasticity (ARCH). Their case of sample size $N = 720$, with 5 equations and a small instrument set of 3 variables, is closest to most of our model estimations, in particular our onset/recovery model. The GMM approach shows poor performance when very small data sets are used, 60 observations, and even with moderately sized data sets with many instruments (numbering 8 relative to 1 or 3 parameters per equation). But with a large number of observations, such as we employ, even use of a large number of instruments does not compromise performance markedly. We work with over-identified models, though none are as heavily over-identified as the extreme case Ferson and Foerster explore. Finally, their main experiments do not incorporate conditional heteroskedasticity, but their main results are robust to the impact of conditional heteroskedasticity, as they describe in Section 5.4 of their paper. Ferson and Harvey (1992) provide a review of the literature on simulation studies of GMM estimation in small samples that also strongly supports the use of GMM methods in samples as large as ours, spanning about 50 years of monthly data.

HQIC) specifically designed by Andrews and Lu (2001) for application to GMM estimation in a dynamic panel setting, the number of model parameters, and the number of moment conditions. Lower values of the information criteria identify better model performance. For each information criterion, we present two values. One is for the model that includes onset/recovery and the other is for a model that includes only a constant. On the basis of both criteria, we see that the onset/recovery model performs better than a constant return model. The test of overidentifying restrictions, χ^2 with 8 degrees of freedom, does not reject the null of no misspecification.

Panel A of Table 3 contains the economic magnitudes of seasonal variations in returns, by series, calculated based on both realized returns and fitted returns arising from estimating Equation (1). We consider seasonal variation in returns from fall to winter, September to March, and October to April. In each case the variations for the "realized" series are positive, ranging in magnitude from a low around 30 basis points to a high over 90 basis points. The variations for the "fitted" series reveal that Model 1 is accurately capturing seasonal variability in Treasury returns, both in terms of positive sign and on the basis of rough magnitudes.

Panel B of Table 3 contains additional seasonality tests based on estimation of Equation (1). The first two lines report p-values associated with joint tests on the onset/recovery coefficients across the four Treasury series, the first testing whether the estimates are jointly zero and the second testing whether they are jointly equal (but not necessarily zero). We present asymptotic p-values and find virtually identical results based on bootstrapped p-values (see Table G.1 in Appendix G, which reports a more detailed set of results for Model 1). We reject the null that the onset/recovery coefficients are jointly zero, with a p-value below 2 percent. Overall, the results are consistent with investors shunning risk in the fall, resulting in higher Treasury prices (and higher realized Treasury returns) in the fall than would otherwise be the case. Similarly, the results are consistent with investors resuming their previous level of risk aversion as daylight becomes more plentiful through the winter season, resulting in lower Treasury prices (and lower realized Treasury returns) than would otherwise be the case.

The remaining lines in Panel B of Table 3 contain p-values associated with the tests for residual seasonality across all of the return series. These tests are analogous to the seasonality tests performed in Section 1 on

Panel A: Economic magnitudes of seasonal differences in returns, stated in percent

	Treasury Excess Returns			
	20-Year	10-Year	7-Year	5-Year
Fall-Winter				
Fitted	0.39	0.37	0.34	0.28
Realized	0.55	0.31	0.48	0.30
September-March				
Fitted	0.88	0.82	0.76	0.62
Realized	0.62	0.55	0.56	0.49
October-April				
Fitted	0.56	0.52	0.48	0.39
Realized	0.97	0.96	0.69	0.55

Panel B: Statistical significance of joint tests and seasonality tests

	Asymptotic p-value
Onset/Recovery Coefficients Jointly 0:	**0.016**
Onset/Recovery Coefficients Jointly Equal:	0.153
Nonspecific Monthly Seasonality:	0.902
Fall vs. Winter Seasonality:	0.592
September vs. March Seasonality:	0.396
October vs. April Seasonality:	0.292

Table 3: Model 1—Seasonality tests, economic magnitudes of seasonal differences in returns, and joint tests.

Description: Panel A reports seasonal differences in realized and fitted returns from estimating Equation (1). Panel B reports p-values associated with joint tests on $\hat{OR}$ values arising from estimating Equation (1) and p-values associated with four seasonality tests described in Section 2.2.

Interpretation: Seasonality tests indicate that there remains no significant evidence of seasonality in the residuals of Model 1, and the fitted values based on Model 1 are large relative to realized Treasury returns, explaining an economically large portion of returns.

the raw and excess Treasury returns, one series at-a-time, but now we explore whether there exists *joint* seasonality across the four series after having controlled for onset/recovery. For instance, the test for nonspecific monthly seasonal variation involves a regression of the return series on a constant, $\hat{\text{OR}}$, and 11 monthly dummy variables, restricting coefficients on the dummy variables to have the same value across series. That is,

$$r_{i,t} = \alpha_i + \mu_{i,\hat{\text{OR}}} \cdot \hat{\text{OR}}_t + \sum_{j=2}^{12} \beta_j \cdot \text{DM}_{j,t} + \epsilon_{i,t},$$

where $\text{DM}_{j,t}$ is a dummy variable equal to 1 if the month of the year for observation t equals j (with February designated month 2, March month 3, and so on). We test whether the monthly dummy coefficients each equal zero, an eleven degree of freedom χ^2 test. To test whether a Treasury return series, controlling for onset/recovery, has the same mean value in the fall and winter versus the alternative that the fall and winter means deviate from the annual average by an equal and opposite amount, we estimate the following model:

$$r_{i,t} = \alpha_i + \mu_{i,\hat{\text{OR}}} \cdot \hat{\text{OR}}_t + \beta_{\text{fall/winter}} \cdot \text{DM}_{t,\text{fall/winter}} + \epsilon_{i,t}.$$

Note again that the coefficient on the seasonality test variable, $\beta_{\text{fall/winter}}$, is restricted to have the same value across series. We estimate alternate versions of this model to produce the alternate seasonality tests, replacing $\text{DM}_{t,\text{fall/winter}}$ with either $\text{DM}_{t,\text{Sep/Mar}}$ or $\text{DM}_{t,\text{Oct/Apr}}$. A given test for seasonality is a two-sided t-test on the indicator variable coefficients to each equal zero.[20]

We see in Panel B of Table 3 that all four test statistics (associated with the test for nonspecific monthly seasonality and the three tests for seasonal-depression-related seasonality) are insignificant. That is, there is no significant evidence of seasonal variation in the returns once we control for onset/recovery.

In Figure 4, the panel labeled “Model 1” contains a plot of the monthly mean Treasury return residuals from estimating Equation (1); we consider

[20]In untabulated analysis, we found that the restriction of constant coefficients on the seasonality variables $\text{DM}_{t,\text{fall/winter}}$, $\text{DM}_{t,\text{Sep/Mar}}$, and $\text{DM}_{t,\text{Oct/Apr}}$, relative to the unrestricted case with the coefficients allowed to vary across Treasury series i, had little qualitative effect on test results.

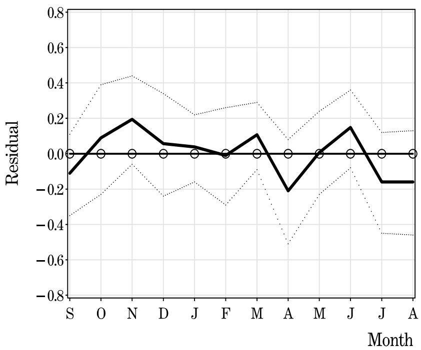
Models 2–12 later. As before, the plot starts with September and ends with August. Observe that the seasonal pattern in the Equation (1) residual series is largely purged. All of the monthly mean bond residuals lie within the confidence interval around the expected value of zero. As we show in the next section, the magnitude of the deviations around zero is smaller than that achieved by other models and the confidence intervals are no wider than those produced by other models. That is, the lack of statistically significant seasonality in the monthly residuals is not an artifact of a relatively noisy regression error.

2.3 Do Treasury Returns Vary Seasonally Due to Weather?

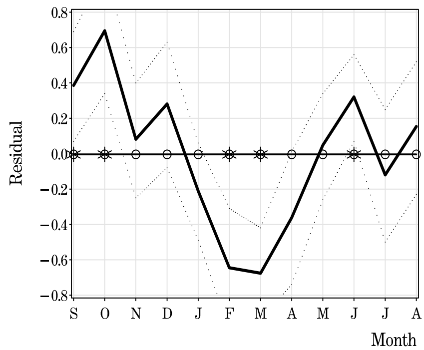
Hirshleifer and Shumway (2003) study index returns for 26 international stock exchanges and find a significant relationship between index returns and the amount of morning sunshine (which depends on the amount of cloud cover) in a given exchange's location. They explain this finding in terms of misattribution: an investor who is in a good or bad mood on account of the weather may misattribute his/her feelings to investment prospects, and this could result in a link between returns and the weather.

Medical research has established that *seasonal* depression arises as a consequence of diminished exposure to daylight, not weather. Nevertheless, the weather in any given location can follow distinct seasonal patterns, and hence one might wonder whether seasonality in Treasury returns is related to seasonality in the weather. To address this concern, we estimate a series of models. The full results appear in Appendix H; we summarize them briefly below.

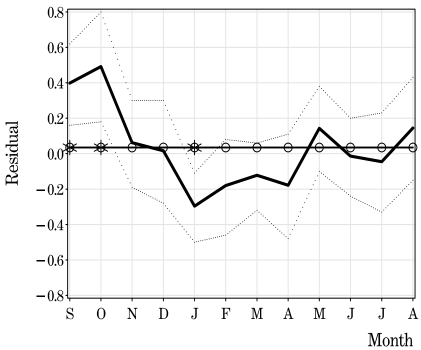
When we regress daily Treasury returns on temperature, cloud cover, and rainfall (all seasonally adjusted following the method of Hirshleifer and Shumway (2003)), the coefficient estimates are uniformly negative, with some significant estimates for cloud cover at standard levels of significance. When we include the onset/recovery variable in the model, the coefficient estimates on the weather variables remain negative (with some significant estimates for cloud cover and temperature), and the onset/recovery variable is positive and significant at the 5 percent level for each of the 5-, 7-, 10-, and 20-year series. When we employ seasonally unadjusted weather data in these models instead of demeaned weather data, the results are qualitatively identical. The coefficient estimates on the weather variables are uniformly negative and occasionally significant, and the onset/recovery variable has a



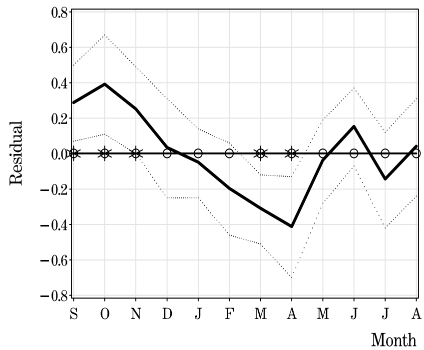
Model 1



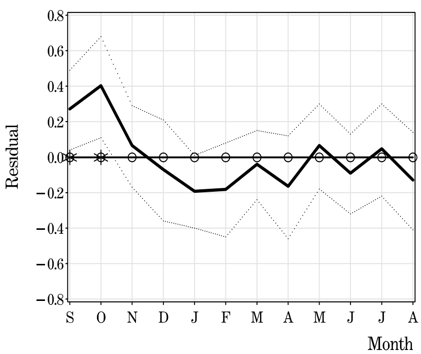
Model 2



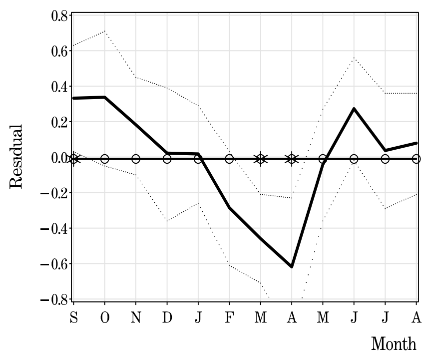
Model 4



Model 3

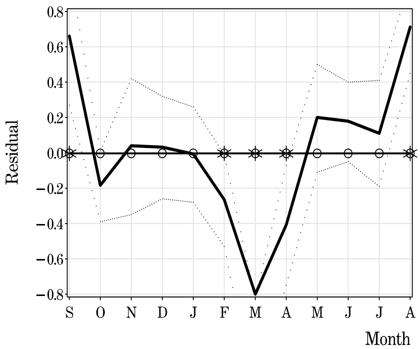


Model 5

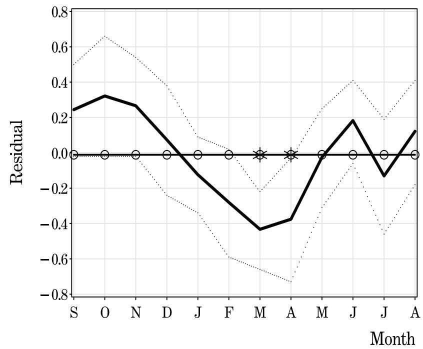


Model 6

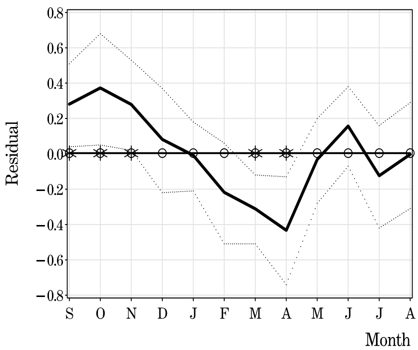
Figure 4: Monthly average residuals for Models 1–12.



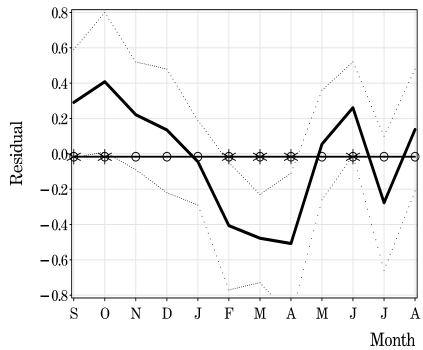
Model 7



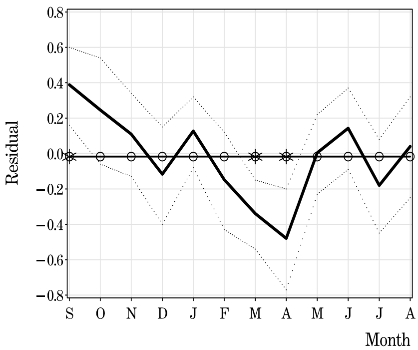
Model 8



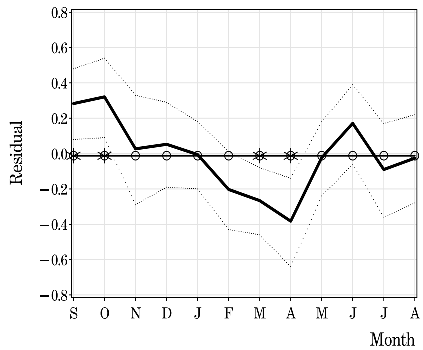
Model 10



Model 9



Model 11



Model 12

Figure 4: *(Continued)*

Figure 4: *(Continued)*

Description: This figure contains monthly average Treasury return residuals from regressing the Treasury returns on explanatory variables associated with Models 1–12 (i.e., Equations (1)–(12)). A thick solid line represents the monthly mean residuals and dotted lines represent a 90 percent confidence interval around the monthly means, calculated as described in reference to Figure 1. The average monthly expected value of the residuals (zero) is plotted with a thin solid line with circles (and X in cases where the expected value falls outside the confidence interval). The sample period varies by model; see Table 4 for details.

Interpretation: We see economically and statistically significant monthly seasonal variation in Treasury residuals from each model with the single exception of Model 1. The seasonality in Models 2–12 is very similar to that found unconditionally in Treasury returns, suggesting that these models are not capable of accounting for this seasonality.

uniformly positive coefficient estimate, which is significant at the 5 percent level. Thus the observed seasonal variation in Treasury returns does not appear to arise due to investors reacting to weather.

In untabulated robustness checks, we considered the use of weather data over different time periods, using seemingly unrelated regressions with MacKinnon and White standard errors instead of GMM with a system of equations and Newey and West standard errors, and using monthly instead of daily data. Results are similar, although the coefficients on the weather variables are insignificant based on monthly data, consistent with the reduction in power that can arise from using monthly averages instead of daily values for weather variables that can vary greatly over the course of a month.

3 Alternative Models

In this section we consider several potential alternative explanations for the seasonal cycle in Treasury returns, including cross-hedging, Baker and Wurgler (2006) and Baker and Wurgler (2007) sentiment, consumer sentiment as measured by the Michigan Consumer Sentiment Index, Fama-French risk factors, momentum, a broad range of macroeconomic shocks, and several factors related to activities of the U.S. Treasury, including the supply of Treasury debt, the Federal Reserve Board's annual cycle of rate-setting meetings, and a significant change to the Treasury auction announcement policy that was introduced in the late 1970s to reduce shocks and facilitate liquidity in the Treasury market. We also consider a conditional capital asset pricing model that permits a seasonally varying price of risk. We introduce each of the various possible explanations immediately below. For simplicity, we postpone discussion of the detailed results from estimating each of the models until Section 4. Data sources (and where appropriate, data construction methods) for each model's variables are summarized in Appendix A. Summary statistics for each model's variables appear in the tables in Appendix B. For ease of reference, we summarize the model numbers, names, and sample periods for each of the models in Table 4.

3.1 Model 2: The FOMC Meeting Cycle, Treasury Auctions, and Treasury Debt Supply

The first alternative we consider is the possibility that the seasonal cycles in Treasury returns that we have identified can be explained by activities of the U.S. Department of the Treasury or the Federal Open Market Committee (FOMC). Throughout most of our sample, mid-quarterly Treasury auctions of notes and bonds have been held in February, May, August, and November. In the early part of our sample, however, the maturity and supply of securities offered at these auctions was typically determined by surveying buyers of the Treasury issues, and then making adjustments in a "tactical" fashion. Thus the selection and quantity of Treasuries offered for sale did not follow a predictable pattern, an occurrence that occasionally disrupted the market by catching investors off guard.[21] During the mid-1970s U.S. Treasury officials, concerned about growing financing demands due to fiscal deficits, began to regularize Treasury offerings of notes and bonds. Quarterly and mid-quarterly auction schedules were put in place for most maturities of notes and bonds by 1980, and by 1982 the choice and supply of offered maturities was announced well in advance of auctions. The posted dates are tentative and can change, but changes are rare.[22] The U.S. Treasury currently sells bills, notes, bonds, and Treasury inflation-protected securities (TIPS) at more than 150 auctions held throughout the year. See Dupont and Sack (1999) for an overview of the operations of the Treasury securities market.

[21] See Garbade (2007) for further details.

[22] See Garbade (2007) for details.

Model Number	Model Name	Sample Period
1	Onset/Recovery	01/1952–12/2007
2	FOMC, Treasury, & Debt Supply Factors	01/1980–11/2007
3	CRR Macro Factors	01/1952–12/2007
4	Seasonally Unadjusted Macro Factors	01/1952–12/2006
5	CRR Macro Factors & Seasonally Unadjusted Macro Factors	01/1952–12/2006
6	Real-Time Macro Factors	12/1965–12/2003
7	Cross Hedging & Treasury Liquidity	08/1960–12/2007
8	Cross Hedging, Treasury Liquidity, & Treasury Volatility	01/1994–12/2007
9	Baker-Wurgler Sentiment	03/1966–12/2005
10	Michigan Consumer Sentiment	02/1953–12/2007
11	Fama-French Factors	01/1952–12/2007
12	Conditional CAPM	02/1952–12/2007

Table 4: Models 1–12.

Description: We provide the model numbers, names, and sample periods for each of Models 1–12. Section 3 contains a detailed description of each model. Appendix A lists data sources.

We seek to control for features of the Treasury auction arrangements that might help explain the seasonal cycle in Treasury returns. The first variable we introduce for this purpose is a dummy variable for the auction-announcement months (IsAuction_t). This variable helps us determine whether the mid-quarterly auction schedule, which is a more prominent feature of the post-1980 period, induces a seasonal pattern in returns. Second, because the supply of debt has been shown to impact the Treasury market,[23] we control for Treasury debt supply changes. We measure

[23] For instance, Krishnamurthy (2002) and Krishnamurthy and Vissing-Jorgensen (2012).

the impact of Treasury debt supply, following Krishnamurthy and Vissing-Jorgensen (2012), by forming the ratio of Treasury debt to gross domestic product (DebtToGDP_t).[24] Finally, the Federal Reserve conducts open market operations, the sale or purchase of Treasury debt, as a tool to implement monetary policy. The explicit intent of these efforts is to manage the money supply, short-term interest rates, and seasonal movements of funds. The decision to conduct open market operations is based on directives from the FOMC, which meets only six to eight times a year. Further, the preparation for and follow-up to the FOMC meetings generates a vast amount of microeconomic and macroeconomic information, some of it released shortly before the meetings (e.g. the Beige Book), some of it released on the meeting date (rate changes, statement of bias, etc.), and some released shortly after (e.g. minutes of the meeting). Long-term rates can react strongly to statements made by the FOMC, even if the FOMC announces no immediate rate change and makes no recommendation for open market operations. It is thus interesting to control for FOMC meeting dates, which we accomplish using a dummy variable (FOMC_t) set equal to one in months when the FOMC has a meeting.[25]

Although DebtToGDP_t displays no evidence of seasonality itself (see Table B1 in Appendix B) and thus may be unlikely to account for the seasonal variation we document in Treasury returns, both FOMC_t and IsAuction_t are highly seasonal. The model we estimate is:

$$\begin{aligned} r_{i,t} &= \mu_i + \mu_{i,\text{Auction}} \cdot \text{IsAuction}_t + \mu_{i,\text{DebtToGDP}} \cdot \text{DebtToGDP}_t \\ &\quad + \mu_{i,\text{FOMC}} \cdot \text{FOMC}_t + \epsilon_{i,t}. \end{aligned} \tag{2}$$

[24] Note that our debt supply change variable, DebtToGDP, is measured at time t, contemporaneous with returns. In a set of untabulated robustness checks, we replace the DebtToGDP measure of supply with (sequentially) the contemporaneous change in the amount of Treasury debt, the contemporaneous net federal U.S. government saving, and a three-month lead of *each* of the three measures of Treasury supply. (Using the *lead* of these variables allows for the fact that these measures are based on quarterly data and thus the information they contain may have been at least partially anticipated by market participants.) Each of these checks produces results that are qualitatively identical.

[25] FOMC meetings are typically held during the months of January/February, March, May, June, August, October/November and December, though the schedule varies enough from year to year that every month of the year has involved a FOMC meeting one year or another.

3.2 Macroeconomic Risks

Ang and Piazzesi (2003), among others, have shown that bond prices embed macroeconomic information. Thus it is plausible that the seasonal variation we observe in bond prices arises as a simple consequence of macroeconomic seasonality. We test for the possibility that the cycle in Treasury returns is explained by any of several types of macroeconomic variables, including the macroeconomic data typically investigated in the asset pricing literature, seasonally unadjusted macroeconomic data, and real-time macroeconomic data.[26] We discuss each in turn below. Note that the macroeconomic variables we consider are intended to capture news that would have been available to market participants at the time prices were being formed, allowing us to identify comovements of returns and macroeconomic news. This means that many of these variables are measured at time t, contemporaneous with returns.

3.2.1 Model 3: Chen, Roll, and Ross Macroeconomic Risks

Chen *et al.* (1986), (henceforth CRR) found the following factors to be significantly priced in the stock market: an interest rate variable measured by the spread in the return to holding a long bond and a short bill (Term),[27] expected and unexpected inflation (Inf and InfSurp, respectively), the growth in industrial production (IP), and the spread between high- and low-grade bonds (Default). See Appendix A for data sources and series construction details. Here we consider whether they explain the observed

[26]We also explored a model predicting returns with the yield structure of Treasury notes and bonds, specifically with the 1-, 3-, 5-, 10-, and 20-year note and bond yields. These yields show very little seasonality, and this model was not successful at capturing the seasonality we document.

[27] We make use of the difference between the 20-year Treasury bond and the 30-day Treasury bill returns, lagged one period. It is possible that the spread itself is influenced by seasonally varying investor risk aversion, for example if, with the onset of seasonal depression, investors move to short-term Treasury securities rather than to long-term Treasury securities. Including the term spread as an explanatory variable thus may be inappropriate, if shifting assets is not uniformly distributed between the various series of Treasury securities. Our results are unaffected by excluding this variable, however, and also are unaffected if we define the term variable as the difference between the 90- and 30-day returns as Harvey (1989) suggests, the difference between the 20-year and 90-day returns, the difference between the 20-year and 1-year returns, the difference between the 20-year and 2-year returns, or the difference between the 20-year and 5-year returns. See Appendix C for details.

seasonal variation in Treasury returns. Based on the summary statistics on the Model 3 variables shown in Table B2, we see that none of these explanatory variables display evidence of seasonality (indeed several of these variables are seasonally adjusted). Still, if the seasonality we document in Treasury returns were simply an artifact of a few unusual years and otherwise these returns were well explained by the CRR model, we might observe the *statistical* evidence for this seasonality fade once we control for the CRR factors. It is difficult to rule out this possibility without formal analysis, so we estimate the following model:

$$\begin{aligned} r_{i,t} \quad = \quad & \mu_i + \mu_{i,\text{Term}} \cdot \text{Term}_{t-1} + \mu_{i,\text{Inf}} \cdot \text{Inf}_t + \mu_{i,\text{InfSurp}} \cdot \text{InfSurp}_t \\ & + \mu_{i,\text{IP}} \cdot \text{IP}_t + \mu_{i,\text{Default}} \cdot \text{Default}_t + \epsilon_{i,t}. \end{aligned} \tag{3}$$

3.2.2 *Model 4: Seasonally Unadjusted Macroeconomic Risks*

Most of the macroeconomic variables conventionally employed in the asset pricing literature to capture risk are deseasonalized; predictable seasonality is not commonly believed to influence returns. There is, however, a possibility that the seasonally predictable component of macroeconomic risk may account for the seasonal patterns we observe in Treasury returns. Although such a finding would still constitute a legitimate asset pricing puzzle, it would not necessarily be related to seasonal depression and time-variation in risk aversion. Hence we incorporate seasonally unadjusted macroeconomic data in our analysis. The seasonally unadjusted (SU) variables we consider are GDP growth rate ($\text{GDP}_{\text{SU},t}$), percent change in the producer price index ($\text{PPI}_{\text{SU},t}$), industrial production growth rate ($\text{IP}_{\text{SU},t}$), unemployment growth rate ($\text{UEG}_{\text{SU},t}$), and percent change in the consumer price index ($\text{CPI}_{\text{SU},t}$).

In Table B3 we see that *all* of these explanatory variables display evidence of a fall/winter seasonality, and most of these variables also display strong statistical evidence of September/March and October/April oscillations, much as we find in the Treasury return series. We estimate the following regression model:

$$\begin{aligned} r_{i,t} \quad = \quad & \mu_i + \mu_{i,\text{GDP}_{\text{SU}}} \cdot \text{GDP}_{\text{SU},t} + \mu_{i,\text{PPI}_{\text{SU}}} \cdot \text{PPI}_{\text{SU},t} + \mu_{i,\text{IP}_{\text{SU}}} \cdot \text{IP}_{\text{SU},t} \\ & + \mu_{i,\text{UEG}_{\text{SU}}} \cdot \text{UEG}_{\text{SU},t} + \mu_{i,\text{CPI}_{\text{SU}}} \cdot \text{CPI}_{\text{SU},t} + \epsilon_{i,t}. \end{aligned} \tag{4}$$

3.2.3 Model 5: CRR and Seasonally Unadjusted Macroeconomic Risks

Even if the set of macroeconomic factors in Models 3 and 4 are separately incapable of explaining Treasury return seasonality, there is a possibility that the combined set of seasonally adjusted and unadjusted factors may. Thus we combine both sets of factors into a single macroeconomic risk model[28]:

$$\begin{aligned} r_{i,t} \quad = \quad & \mu_i + \mu_{i,\text{Term}} \cdot \text{Term}_{t-1} + \mu_{i,\text{Inf}} \cdot \text{Inf}_t + \mu_{i,\text{InfSurp}} \cdot \text{InfSurp}_t + \mu_{i,\text{IP}} \cdot \text{IP}_t \\ & + \mu_{i,\text{Default}} \cdot \text{Default}_t + \mu_{i,\text{GDP}_{\text{SU}}} \cdot \text{GDP}_{\text{SU},t} + \mu_{i,\text{PPI}_{\text{SU}}} \cdot \text{PPI}_{\text{SU},t} \\ & + \mu_{i,\text{IP}_{\text{SU}}} \cdot \text{IP}_{\text{SU},t} + \mu_{i,\text{UEG}_{\text{SU}}} \cdot \text{UEG}_{\text{SU},t} + \mu_{i,\text{CPI}_{\text{SU}}} \cdot \text{CPI}_{\text{SU},t} + \epsilon_{i,t}. \end{aligned} \tag{5}$$

3.2.4 Model 6: Real-Time Macroeconomic Risks

We now consider a wider set of macroeconomic information that may affect Treasury returns: first, real-time data and, second, data that may affect Treasury markets differentially during economic contractions versus expansions. First, regarding real-time data, all of the macroeconomic series we have considered thus far are the most up-to-date versions of the data available, some of which have been revised since the data were first released. When we use the revised data we may be neglecting information that market participants responded to at the time the information was announced. We control for this possibility by considering real-time macroeconomic data as originally reported to the public. The real-time series we consider are the unemployment rate, industrial production growth rate, and inflation rate, from which we construct expected and surprise changes in the unemployment rate, expected and surprise industrial production growth rates, and expected and surprise inflation rates.[29] Second, we allow for some macroeconomic variables to influence Treasuries differently depending on the state of the economy, following Boyd *et al.* (2005). They find, for example, that unemployment rate surprises impact stock and bond returns symmetrically in an economic expansion but oppositely during a contraction. Boyd et al. find that in an expansion, unexpected rising

[28] In a previous version of the paper, we also explored combining all of the variables in Models 2 through 12 into one (admittedly vastly over-parameterized) large model. Results from that model are qualitatively identical to findings based on this smaller combined model.

[29] More details about the real-time ("vintage") series are provided in Appendix I.

unemployment is good news for both stocks and bonds, but in a contraction, unexpected rising unemployment is bad news for stocks and irrelevant for bonds. Constructing the surprise and expected macroeconomic series is a multi-step process, which we detail in Appendix I. To capture the probability of an expansion/contraction we use the experimental coincident recession index of Stock and Watson (1989).

Altogether we control for the influence of the expected change in the unemployment rate (UEG_t), the expected growth in industrial production (IP_t), the surprise in the industrial production growth rate (IPSurp_t), the monthly change in the spread between Baa and Aaa corporate bond rates ($\Delta\text{Default}_t$), the monthly change in the spread between 20-year and 30-day Treasury returns (ΔTerm_t), the probability of a contraction (ProbC_t), the surprise in the unemployment rate change interacted with the probability of a contraction (USurpC_t), the surprise in the unemployment rate change interacted with the probability of an economic expansion (USurpE_t), and a January dummy variable (Jan_t). De Bondt and Bange (1992) and Brandt and Wang (2003) suggest inflation surprises may lead to time-varying government bond returns, and thus we control for expected inflation (Inf_t) and inflation surprises (InfSurp_t).[30] In Table B4 we see that a few of these explanatory variables display evidence of fall/winter, September/March, or October/April seasonal oscillations. We estimate the following model:

$$
\begin{aligned}
r_{i,t} \;=\;& \mu_i + \mu_{i,\text{UEG}} \cdot \text{UEG}_t + \mu_{i,\text{IP}} \cdot \text{IP}_t + \mu_{i,\text{IPSurp}} \cdot \text{IPSurp}_t \\
&+ \mu_{i,\Delta\text{Default}} \cdot \Delta\text{Default}_t + \mu_{i,\text{Term}} \cdot \text{Term}_{t-1} + \mu_{i,\text{ProbC}} \cdot \text{ProbC}_t \\
&+ \mu_{i,\text{USurpC}} \cdot \text{USurpC}_t + \mu_{i,\text{USurpE}} \cdot \text{USurpE}_t + \mu_{i,\text{Inf}} \cdot \text{Inf}_t \\
&+ \mu_{i,\text{InfSurp}} \cdot \text{InfSurp}_t + \mu_{i,\text{Jan}} \cdot \text{Jan}_t + \epsilon_{i,t}.
\end{aligned}
\tag{6}
$$

3.3 *Model 7: Cross-Market Hedging and Treasury Market Liquidity*

Connolly *et al.* (2005) find that Treasury and stock markets can move in opposite directions during short periods such as market crashes, perhaps

[30] As we describe in Appendix I, our findings with respect to residual seasonality are virtually identical based on alternate specifications of several variables. For instance, we explore two measures of surprises, one using real-time data and the other using the most recently available data (which includes data revisions), and three alternate definitions of the interactive unemployment surprise variable.

due to cross-market hedging. They control for this possibility using a volatility measure and a turnover measure. A disproportionate share of market crashes has occurred in the early fall, leading to large negative swings in equity returns and hedging in Treasuries; such activity could lead to the seasonal patterns we consider, even though turnover and volatility variables show little or no seasonality themselves (as shown in Table B5).[31]

The first variable we control for is stock market volatility, measured using the fitted (conditional) value from a GARCH(1,1) model. We denote the conditional volatility as CondVar_t.[32] We also control for stock market turnover (Turnover_t; see Appendix A for details on the construction of this variable). Finally, we add a variable measuring bond market trading activity in month t to capture the impact of Treasury market liquidity (Liquidity_t), as this can modulate the impact of cross-market hedging. The model we estimate is:

$$\begin{aligned} r_{i,t} \quad = \quad & \mu_i + \mu_{i,\text{CondVar}} \cdot \text{CondVar}_t + \mu_{i,\text{Turnover}} \cdot \text{Turnover}_{t-1} \\ & + \mu_{i,\text{Liquidity}} \cdot \text{Liquidity}_{t-1} + \epsilon_{i,t}. \end{aligned} \tag{7}$$

3.4 *Model 8: Cross-Market Hedging, Treasury Market Liquidity, and Treasury Market Volatility*

There exists the possibility that time variation in Treasury return volatility, a proxy for risk, drives seasonal variation in Treasury returns. Andersen and Benzoni (2010) show that the realized volatility of a Treasury security of a given maturity can be derived using yields from Treasury securities with the same maturity. We utilize daily yields for the 5-year, 7-year, 10-year, and 20-year constant maturity securities. Following the procedure of Andersen and Benzoni (2010) we compute realized yield volatility. We then form a

[31] Additionally, Holland and Toma (1991) observe, "[financial] panics in pre-Fed times were more likely to occur during the autumn than in other seasons" (p. 675).

[32] We obtain similar results if instead we estimate the conditional volatility using the fitted value from an ARMA(1,2) model of realized S&P 500 stock index return volatility. The ARMA(1,2) specification is the lowest order autoregressive-moving-average (ARMA) model that removes evidence of autocorrelation from the realized volatility series. For reference to the theoretical justification for and properties of the realized volatility measure, see Andersen *et al.* (2003). Untabulated robustness checks using the conditional volatility of the CRSP value-weighted or equal-weighted return series show that our results are not sensitive to the choice of the S&P 500 volatility measure.

forecasted monthly volatility with an autoregressive moving average model of order (3,1).[33] We incorporate Treasury volatility ($\text{TreasuryVol}_{i,t}$) for series i, in the cross-hedging model:[34]

$$\begin{aligned} r_{i,t} \;=\;& \mu_i + \mu_{i,\text{CondVar}} \cdot \text{CondVar}_t + \mu_{i,\text{Turnover}} \cdot \text{Turnover}_{t-1} \\ &+ \mu_{i,\text{Liquidity}} \cdot \text{Liquidity}_{t-1} + \mu_{i,\text{TreasuryVol}} \cdot \text{TreasuryVol}_{i,t-1} + \epsilon_{i,t}. \end{aligned} \tag{8}$$

3.5 Models 9/10: Investor Sentiment

Baker and Wurgler (2006) and Baker and Wurgler (2007) suggest that investor sentiment can have an impact on security prices, with positive (negative) sentiment driving up (down) risky equities, in particular those whose valuations are highly subjective and difficult to arbitrage. They measure investor sentiment as a function of the closed-end fund discount, NYSE share turnover, the number of initial public offerings (IPOs), the average first-day IPO return, equity share (gross equity issuance divided by gross equity plus gross long-term debt issuance), and the dividend premium (the log difference of the average market-to-book ratios of dividend payers and non-payers).

The Baker-Wurgler measure of sentiment embeds data that are possibly seasonal, and other measures of sentiment, like the Michigan consumer sentiment survey, do display seasonality. We see in Table B6 that the Baker-Wurgler sentiment measure does not display significant seasonality, but the Michigan measure shows a fall/winter oscillation and unconditional seasonality. Under some conditions, say investors substituting safe assets for risky ones in negative sentiment periods and reversing in positive sentiment periods, investor sentiment plausibly causes seasonal patterns in Treasury returns. It is therefore natural to consider whether the seasonality we explore is actually a result of sentiment. We use the lag of the change in the Baker and Wurgler (2007) sentiment index (BWSentiment_{t-1}). Work by Qiu and Welch (2006) suggests the University of Michigan's Consumer Sentiment Index may be a relatively better proxy for consumer sentiment, thus we

[33] For further details on construction of the realized volatility measures from yields, see Andersen and Benzoni (2010), in particular Equation (30) in Section I, and Andersen and Benzoni (2009). This model is sufficient to capture the dependence of the realized volatility to lag length 12 (by measure of Godfrey (1978a) and Godfrey (1978b) serial correlation test) and explains roughly 70 percent of the variation of realized volatility.

[34] Note that daily yields on the 20-year Treasury securities are not available prior to 1994, which restricts the sample period for this model.

also employ the lag of the change in the Michigan index (MSentiment_{t-1}). To model the influence of Baker-Wurgler sentiment, we estimate:

$$r_{i,t} = \mu_i + \mu_{i,\text{BWSentiment}} \cdot \text{BWSentiment}_{t-1} + \epsilon_{i,t}, \tag{9}$$

and for the Michigan consumer sentiment measure we estimate:

$$r_{i,t} = \mu_i + \mu_{i,\text{MSentiment}} \cdot \text{MSentiment}_{t-1} + \epsilon_{i,t}. \tag{10}$$

3.6 Model 11: Fama-French Factors

Fama and French (1993) identify common risk factors in stock and bond returns, finding three equity return factors and two bond return factors. The equity return factors are the excess return on the overall market, SMB (firm size), and HML (book-to-market); the bond return factors are the term spread (long-term Treasury bond returns minus the 30-day T-bill rate) and the default spread (the difference between long-term corporate and government bond returns). Fama and French find that the shared impact of these factors—the equity return factor impact on bond returns and the bond return factor impact on stock returns—appears to come in through the excess market return, which is itself influenced by all five factors. Since bond returns have been shown to be a function of term structure factors as well as the excess market return, itself "a hodgepodge of the common factors in returns" Fama and French (1993, p. 27), we consider whether the seasonal cycle in Treasury returns arises due to seasonality in these factors.

The explanatory variables we employ are the three Fama-French equity return factors (excess return on the overall market,[35] Small Minus Big (SMB), and High Minus Low (HML)), the two bond return factors (the lagged term spread measured by long-term Treasury bond returns minus the 30-day T-bill rate for the corresponding month (Term_{t-1}), and the contemporaneous default spread, measured by the yield difference of BAA and AAA corporate bonds (Default_t). As momentum has also been shown to be an influential return factor (see Jegadeesh and Titman (1993)), we include it in our collection of factors (labeled MOM_t). Perhaps unsurprisingly, these

[35] To distinguish the roles of the bond and equity factors, we follow Fama and French (1993) and orthogonalize the excess market return with respect to all of these variables, and we use this orthogonalized variable in place of the excess return on the overall market. We label the orthogonalized excess market return MKT.

return variables show strong evidence of seasonality. See Table B7. We estimate:

$$\begin{aligned} r_{i,t} &= \mu_i + \mu_{i,\mathrm{SMB}} \cdot \mathrm{SMB}_t + \mu_{i,\mathrm{HML}} \cdot \mathrm{HML}_t + \mu_{i,\mathrm{MOM}} \cdot \mathrm{MOM}_t \\ &\quad + \mu_{i,\mathrm{Default}} \cdot \mathrm{Default}_t + \mu_{i,\mathrm{Term}} \cdot \mathrm{Term}_t + \mu_{i,\mathrm{MKT}} \cdot \mathrm{MKT}_t + \epsilon_{i,t}. \end{aligned} \tag{11}$$

3.7 Model 12: Conditional CAPM

A conditional capital asset pricing model (CCAPM) in which the reward-to-risk ratio can vary with seasonalities in risk aversion may account for Treasury return seasonalities, as Garrett *et al.* (2005) explore for equity returns. Following Harvey (1989) and Bekaert and Harvey (1995), for asset i the CCAPM is

$$E_{t-1}(\tilde{r}_{it}) = \lambda \cdot \mathrm{cov}_{t-1}(\tilde{r}_{it}, \tilde{r}_{mt}),$$

where $\tilde{r}_{it}$ is the excess return on the i^{th} asset, $\tilde{r}_{mt}$ is the excess return on the market portfolio, λ is the price of risk, and cov is the time-varying conditional covariance between excess returns on the asset and on the market portfolio. Aggregating over equities, as Bekaert and Harvey (1995) do over countries, we find

$$E_{t-1}(\tilde{r}_{mt}) = \lambda \cdot \mathrm{var}_{t-1}(\tilde{r}_{mt}),$$

where var is the time-varying conditional variance of the market. (As our proxy for $\mathrm{var}_{t-1}(r_{mt})$, we use $\mathrm{CondVar}_t$, the volatility forecast we define for the cross-hedging models above.) This CAPM formulation was first explored by Merton (1980), and he interpreted λ as the representative investor's Arrow-Pratt coefficient of relative risk aversion for wealth.

Following Bekaert and Harvey (1995) we allow the price of risk to vary over time by making it an exponential function of conditioning variables (Z_t), restricting the price of risk to be positive (Equation (12) of their paper): $\lambda_t = \exp(\delta' \cdot Z_t)$. We adopt the specification outlined in Harvey (1989), utilizing dividend yields in excess of the risk free rate (XDP_t), the excess return on the market portfolio ($\tilde{r}_{m,t}$), the junk bond premium ($\mathrm{Default}_t$), and the term premium ($\mathrm{Term90}_t$). We estimate:

$$\begin{aligned} E_{t-1}(\tilde{r}_{i,t}) &= \lambda_{t-1} \cdot \mathrm{CondVar}_t \\ \lambda_{t-1} &= \exp\big(\delta_i + \delta_{i,\mathrm{XDP}} \cdot \mathrm{XDP}_{t-1} + \delta_{i,\tilde{r}_m} \cdot \tilde{r}_{m,t-1} \\ &\quad + \delta_{i,\mathrm{Default}} \cdot \mathrm{Default}_{t-1} + \delta_{i,\mathrm{Term90}} \cdot \mathrm{Term90}_{t-1}\big). \end{aligned} \tag{12}$$

Although none of the variables in this model exhibit specific forms of seasonal oscillation (see Table B8), this model is able to capture some of the seasonal variation in Treasury returns, as we discuss below.

4 Results for Alternative Models 2–12

We now consider how well each of the alternative models introduced in the previous section explains the seasonal patterns in Treasury returns. Note that in all cases, the results are based on using excess Treasury returns as the dependent variable; our findings are qualitatively identical using raw returns.[36]

We estimate each model using system-of-equations GMM and Newey and West (1987) and Newey and West (1994) HAC standard errors; see footnote 6 for estimation details. For the interested reader, in Appendix G we provide detailed estimation results for each of Models 2–12. We summarize the primary results for the full set of models in Tables 5, 6, and 7.[37] Prior to discussing those details, we consider plots of the residuals from estimating Models 2–12, shown in Figure 4.[38] A common feature of the Model 2–12 plots is an inability to capture the above average bond returns in the early fall and/or the trough in bond returns in the winter/spring. For all of these models there remains significant evidence of residual seasonality, with months in the fall exhibiting average residuals that are significantly greater than 0 and (in all but one case) months in the winter/spring exhibiting average residuals that are significantly less than 0. That is, in contrast to the onset/recovery model, Model 1, none of Models 2–12 are able to account for the seasonality in Treasury returns.

[36]While we present results for Treasury return models only, in a previous version we included U.S. stock index returns as a fifth equation in the system, with results for the Treasury series qualitatively identical to those we discuss here.

[37]For each of Models 2–11, the instruments we use to form the GMM moment conditions are a constant, the explanatory variables, 30-day T-bill returns, and the lagged CRSP value-weighted equity index returns including dividends. For Model 12, we augment that set by including a lag of the dependent variable and the explanatory variables, to obtain identification.

[38]The peaks of the residual plots in Figure 4 shift from model to model due to the varying degree of seasonality accounted for by the various models and minor variation in the sample period. In Section 6 we discuss the robustness of our findings across various sub-samples.

Model	Nonspec. Monthly: p-value	Fall vs. Winter: p-value (Realized/ Fitted)	Sep. vs. March: p-value (Realized/ Fitted)	Oct. vs. April: p-value (Realized/ Fitted)	Proportion of Seasonality Explained: With OR_t (Without OR_t)
1	0.902	0.592 (0.41/0.34)	0.396 (0.55/0.77)	0.292 (0.79/0.49)	0.63 (0.00)
2	<**0.001**	**0.011** (0.85/−0.02)	<**0.001** (1.11/0.05)	0.226 (0.99/−0.07)	0.73 (0.00)
3	**0.002**	**0.076** (0.41/0.00)	**0.008** (0.55/−0.05)	**0.011** (0.79/−0.02)	0.65 (−0.04)
4	<**0.001**	0.136 (0.40/0.01)	**0.005** (0.55/0.03)	**0.019** (0.79/0.12)	0.54 (0.07)
5	<**0.001**	0.145 (0.40/0.12)	**0.061** (0.55/0.24)	**0.042** (0.79/0.23)	0.71 (0.37)
6	<**0.001**	0.175 (0.63/0.21)	**0.001** (0.81/0.01)	0.109 (0.94/−0.01)	0.71 (0.14)
7	**0.004**	0.113 (0.52/0.02)	**0.001** (0.66/−0.02)	0.261 (0.80/0.10)	0.73 (0.08)
8	<**0.001**	**0.028** (0.44/0.02)	<**0.001** (1.75/−0.03)	0.605 (0.58/0.36)	0.60 (0.07)
9	<**0.001**	0.141 (0.57/0.01)	**0.001** (0.70/−0.07)	**0.054** (0.94/0.02)	0.63 (−0.04)
10	**0.010**	0.163 (0.42/0.00)	**0.018** (0.59/0.00)	**0.039** (0.81/0.00)	0.63 (0.00)
11	**0.003**	0.485 (0.41/0.21)	**0.001** (0.55/−0.18)	0.118 (0.79/0.06)	0.64 (0.00)
12	**0.002**	0.135 (0.41/0.12)	**0.007** (0.55/0.00)	0.126 (0.79/0.09)	0.42 (0.20)

Table 5: Seasonality tests for Models 1–12.

Description: For each of Models 1–12 we present the p-values associated with four different seasonality tests. Below the associated p-values, we include the economic magnitude of the difference between fall and winter returns, September and March returns, and October and April returns, both for the realized series and the fitted series, averaged across the 20-, 10-, 7-, and 5-year maturity series, analogous to Panel A of Table 3. Bolded p-values are significant at the 10 percent level or better. In the last column we report the proportion of monthly seasonal variation in returns that the model is able to explain (stated in percent returns); see footnote 39 for details. The full set of estimation results appear in Appendix G.

Interpretation: There is significant evidence of seasonality in the residuals of all models except Model 1.

In Table 5 we present tests for seasonality for each of the models. Recall that the estimation periods and model names appear in Table 4; data availability limits some estimation periods. The first four columns of results contain p-values associated with seasonality tests (analogous to those that appear in Panel B of Table 3 for Model 1). In each of these cells, the asymptotic p-value appears on the top line. For each of Models 2–12 we reject the null hypothesis of no seasonality; there remains significant evidence of at least one form of seasonal-depression-related seasonality in all cases. For each of Models 2–12, there is also evidence of nonspecific monthly seasonality, though bootstrapped p-values suggest that this result is not always robust; see Appendix G.

In parentheses below those seasonality test p-values, we include the economic magnitude of the difference between fall and winter returns, September and March returns, and October and April returns, respectively, both for the realized series and the fitted series, averaged across the 5-, 7-, 10-, and 20-year maturity series. These are analogous to the economic magnitudes reported in Table 3 for Model 1, and the values are stated in terms of percent returns. The divergence in sign and/or magnitude between the realized differences and fitted differences reflects the poor ability of the alternative models to capture the seasonal oscillation we document in Treasury returns. Consider Model 2, the FOMC, Treasury, and debt supply factors model. The difference between *realized* fall and winter returns is about 85 basis points whereas the difference between the *fitted* fall and winter returns is negative and comparatively small in absolute magnitude. That is, Model 2 captures none of the 85 basis point seasonal variation in Treasury returns across the fall and winter seasons. Similarly, the difference between the realized September and March returns is over 100 basis points while the fitted difference is close to 0. (Note that the realized variability of Treasury returns changes somewhat across models due to differences in the sample periods available to us for the various series we employ.) After the onset/recovery model, Model 1, the next best fit to the seasonal oscillation comes from Model 5 (the CRR and seasonally unadjusted macro factors model), but even this model captures only about a third of the magnitude captured by Model 1. We present additional measures of economic magnitude in the last column of Table 5. The top value in each cell is the proportion of the seasonal variation in returns that is explained by a given model including the onset/recovery variable, and the bottom value (in parentheses) is the proportion of the seasonal variation in returns explained by the model that excludes the onset/recovery variable.[39] Recall that in Panel B of Table 2 we reported that for Model 1, 63 percent of the monthly seasonal variation in Treasury returns is captured by the model that includes the onset/recovery variable. For reference, this suggests Model 1 captures about 50 basis points of the 80 basis point swing from peak to trough observed on average in Treasury returns, as shown in Figure 1. Estimating Model 1 without the onset/recovery variable (which is a simple model including only a constant) captures 0 percent of the seasonal variation in Treasury returns. Looking through the other models, it is uniformly the case that the models which include the onset/recovery variable capture much more of the seasonal variation in returns than those that do not.[40]

We present information criteria for each of the models in Table 6. The first of those columns contains the Andrews and Lu (2001) model and moment selection criteria (MMSC), MMSC-BIC (Bayesian information criterion) and MMSC-HQIC (Hannan-Quinn information criterion). Of Models 1–12, only 1, 3, and 11 are estimated over identical data spans. Of these, the Fama-French model (Model 11) has the best ranked performance, with the onset/recovery model next. These two models have greatly different numbers of parameters and moment conditions, however, and the remaining models cannot be compared directly to each other as they range over different estimation periods. Because we are primarily interested in the utility of the onset/recovery variable, we report in the last column the information criteria obtained by adding the onset/recovery variable to a given model, and constraining the onset/recovery coefficient to be the same across the return series so that we only require one additional parameter to be estimated. Within a given model, we can use these information criteria to evaluate whether the addition of the onset/recovery variable improves the model performance. In every case, the information criteria in the last column are considerably smaller than the values in the middle column,

[39] This value is calculated analogously to the statistic reported in the first row of Panel B in Table 2. Specifically, we regress the actual monthly average returns on the predicted monthly average returns, suppressing the intercept. The R^2 emerging from this estimation is the proportion of seasonal variation explained by the model.

[40] Note that these values are best suited for comparison of economic magnitudes within a particular model, with and without including the onset/recovery variable. In some cases, comparison *across* models is complicated by differences in sample periods and functional form (for instance Model 12 has a non-linear form).

Model (# Param.) [# Moment Cond.]	Info. Criteria For Model: MMSC-BIC (MMSC-HQIC)	Info. Criteria For Model with *Constrained* OR: MMSC-BIC (MMSC-HQIC)
1: (8) [16]	−41.59	—
	(−20.98)	—
2: (16) [24]	−37.46	−49.98
	(−20.52)	(−26.68)
3: (24) [32]	−37.65	−53.89
	(−17.05)	(−25.55)
4: (24) [32]	−38.32	−51.97
	(−17.81)	(−23.77)
5: (44) [52]	−37.45	−53.76
	(−16.94)	(−25.55)
6: (44) [52]	−37.62	−54.62
	(−19.07)	(−29.11)
7: (16) [24]	−42.68	−58.21
	(−22.96)	(−31.10)
8: (20) [40]	−82.61	−96.75
	(−48.76)	(−57.82)
9: (8) [16]	−40.00	−54.76
	(−21.21)	(−28.93)
10: (8) [16]	−40.66	−56.70
	(−20.15)	(−28.50)
11: (28) [36]	−42.57	−58.72
	(−21.96)	(−30.38)
12: (20) [60]	−222.90	−236.46
	(−119.89)	(−125.73)

Table 6: Information criteria for Models 1–12.

Description: We report information criteria for each of the models. The first column contains the number of parameters and the number of moment conditions for each model. The middle column corresponds to estimating a given model without including onset/recovery as an explanatory variable. The last column corresponds to including the onset/recovery as an explanatory variable in the model, where the onset/recovery variable coefficient estimate is constrained to be the same across the 20-, 10-, 7-, and 5-year series. The full set of estimation results, including coefficient estimates and standard errors for all variables in the models, information criteria, autocorrelation and heteroskedasticity test statistics, R^2 statistics, and other details, appear in Appendix G.

Interpretation: The information criteria (defined so that we wish to minimize them) favor models that include the onset/recovery variable over models that exclude it.

Model	20-Year Treasury Excess Returns Coeff. (Std Err)	10-Year Treasury Excess Returns Coeff. (Std Err)	7-Year Treasury Excess Returns Coeff. (Std Err)	5-Year Treasury Excess Returns Coeff. (Std Err)	Asymptotic p-values for Parameter Joint Tests: Jointly 0 (Jointly Equal)
1	1.103**	1.027***	0.949***	0.776***	**0.016**
	(0.454)	(0.362)	(0.294)	(0.243)	(0.153)
2′	2.437***	2.077***	1.704***	1.430***	**0.003**
	(0.773)	(0.570)	(0.467)	(0.383)	(**0.051**)
3′	1.270***	1.114***	1.039***	0.858***	**0.003**
	(0.428)	(0.336)	(0.271)	(0.226)	(0.126)
4′	1.624***	1.920***	1.599***	1.377***	<**0.001**
	(0.561)	(0.461)	(0.368)	(0.315)	(**0.026**)
5′	1.284**	1.620***	1.360***	1.207***	<**0.001**
	(0.593)	(0.486)	(0.386)	(0.333)	(**0.090**)
6′	1.470**	1.399***	1.240***	1.057***	**0.004**
	(0.643)	(0.479)	(0.380)	(0.314)	(0.306)
7′	1.389***	1.249***	1.140***	0.954***	**0.009**
	(0.539)	(0.421)	(0.338)	(0.284)	(0.182)
8′	2.347***	1.786***	1.386***	1.021***	0.111
	(0.809)	(0.608)	(0.489)	(0.379)	(**0.064**)
9′	1.562**	1.520***	1.346***	1.133***	**0.003**
	(0.620)	(0.481)	(0.391)	(0.327)	(0.122)
10′	1.171**	1.082***	0.987***	0.807***	**0.020**
	(0.471)	(0.375)	(0.307)	(0.256)	(0.144)
11′	1.019**	0.933***	0.818***	0.706***	**0.040**
	(0.446)	(0.345)	(0.279)	(0.231)	(0.503)
12′	8.160	11.646**	9.405***	11.874**	**0.024**
	(6.363)	(5.493)	(2.923)	(5.505)	(0.737)

Table 7: Onset/recovery coefficient estimates for Model 1 & Models 2′–12′.

Description: We report coefficient estimates and standard errors for the onset/recovery variable ($\hat{OR}_t$) based on estimating Equation (1) as well as Equations (2) through (12) that have been modified to include $\hat{OR}_t$ as an explanatory variable. One, two, and three asterisks denote significant coefficient estimates at the 10, 5, and 1 percent level respectively, based on two-sided tests. The last row contains asymptotic p-values for tests that the onset/recovery coefficient estimates jointly equal zero (or jointly equal each other, in parentheses, beneath). Bolded p-values are significant at the 10 percent level or better. Full regression results for Models 2′–12′ appear in Appendix J.

Interpretation: When the onset/recovery variable is embedded in each of Models 2–12, we find the onset/recovery coefficient estimates are statistically and economically significant, as in Model 1. This confirms that none of the alternative sets of variables drives out the significance of the onset/recovery variable.

indicating the addition of the onset/recovery variable improves model performance, model-by-model, sample-by-sample. While the specific estimates are not tabulated, we note that in each case the onset/recovery variable is also statistically significant.

We also repeat the above exercise, *without* constraining the onset/recovery coefficient across series, to determine whether onset/recovery remains significant when we simultaneously control for the competing explanations. That is, we estimate each of Models 2–12 augmenting the set of explanatory variables to include the onset/recovery variable but allowing different coefficient estimates for each of the 5-, 7-, 10-, and 20-year series within a particular model. We label these augmented model specifications Models $2'$–$12'$. In the interest of succinctly summarizing our findings, in Table 7 we report coefficient estimates and test statistics pertaining only to the onset/recovery variable arising from estimating the regressions associated with Models $2'$–$12'$. In Appendix J, we provide the full set of regression results for each of these models, including coefficient estimates and standard errors for all variables in the model, information criteria, bootstrapped standard errors, autocorrelation and heteroskedasticity tests, model statistics, R^2, and other details.

We see in Table 7 that the $\hat{OR}$ coefficient estimates are positive and statistically significant for each of Models $2'$–$12'$ and for each series, and the magnitudes of the coefficient estimates are similar to those observed for Model 1 in Table 2. (The only exception is Model $12'$, for which one of the estimates is insignificant, and for which the coefficient magnitudes cannot be compared directly with estimates arising from other models due to onset/recovery having been interacted with volatility in the CCAPM specification.) Asymptotic p-values shown in the last column of Table 7, with bold values significant at the 10 percent level or better, show that for almost all the models, the onset/recovery estimates are jointly significantly different from zero, indicating an annual Treasury return cycle of above-average returns in the fall and below-average returns in the winter (correlated with onset of and recovery from seasonal depression), even after having controlled for a range of alternative explanations.[41] Overall, the results

[41] The single exception, Model $8'$, perhaps highlights the importance of a relatively long sample with which to reliably estimate the effect of seasonal depression. Due to data constraints, this model's sample begins in 1994. Although the shorter sample does not alter the economic impact of seasonal depression, the standard errors are much larger and the joint test is insignificant.

in Table 7 confirm that none of the alternative sets of variables is able to drive out the economic and statistical significance of the onset/recovery variable.

5 Is This Just Data Snooping? The White Reality Test

When conducting inference with frequently studied data, there is a concern that statistically significant results may arise due to data snooping rather than due to any actual underlying economic phenomenon. To test for this possibility here, we employ the data-snooping test developed by White (2000), designed to account for the fact that researchers tend to report only those results that are statistically significant. To implement the procedure, a benchmark set of models must first be defined, in our case, an alternative set of patterns that would have been as remarkable to find correlated with our returns data as the onset/recovery pattern. Once the benchmark models have been defined, bootstrap resampling techniques are used to determine the data-snooping-adjusted significance of the original pattern.[42] The classic application of the White test has us first determining the most significant pattern (across all the benchmark models being considered) separately for each simulated data set, that is, the most significant test statistic finding on each simulated data set that could have been obtained by simple data mining. In a given simulation, the maximum test statistic is associated with the pattern that happens to be the most correlated with returns. Collecting these maximum test statistics across all the simulations yields a distribution of the maximum test statistic. This maximum-value test statistic does not itself have a standard distribution; the maximum t-test is not itself t-distributed, for instance. The White reality test uses simulation techniques to find the distribution of the maximum test statistic.[43] Using White's methods, we can compare statistically unusual features of our original model statistics with the bootstrap distribution, yielding a data-snooping adjusted p-value.

[42] In implementing the reality test, we follow White (2000) and use block bootstrapping to allow for return dependence. See footnote 9 for details. We use 10,000 resamples for the White reality test implementation.

[43] The essence of data mining is the reporting of the maximum test statistic found across many trials. Even if any given researcher conducts only one trial, the end result can be indistinguishable from data mining due to the tendency of journals to publish the findings of only those researchers who find significant results.

As Sullivan *et al.* (2001, p. 259) remark, the choice of benchmark models is important; considering too few benchmarks creates the potential to overlook models considered but never reported in the literature. This could lead to an over-stated level of significance because the data-mining adjustment fails to account for the full set of models from which the successful model was chosen. Considering too many models, models that were never actually considered by researchers, means the test will exhibit a loss of power. This loss of power comes about because the addition of models that are not redundant — models that incorporate return patterns that are not perfectly correlated with other models—increases the effective "span" of the universe of models and the White reality test adjustment will mechanically yield a lower data-snooping adjusted p-value, even if the researcher's model is valid. Sullivan *et al.* (2001) thus recommend using a possibly over-sampled universe of benchmark models and a smaller universe of "basic" models to estimate the data-snooping adjusted p-value.

To form our "basic" calendar anomaly benchmark models, we consider seasonal patterns that would be remarkable to observe, including lagged versions of our onset/recovery variable (lagged by 1 to 11 months), a monthly oscillation in average monthly returns (higher, then lower, then higher, then lower, etc., throughout the year) starting any month of the year, a bi-monthly oscillation in average monthly returns (higher, higher, lower, lower) starting any month of the year, quarterly oscillations in average monthly returns (higher for three months, then lower for three months) starting any month of the year. We consider trends in returns that would be striking, including a three month repeating rising (or declining) trend starting any month of the year, a three month rising then three month declining trend starting any month of the year, a six month repeating rising (or declining) trend starting any month of the year, a six month rising then six month declining trend starting any month of the year, and a twelve month rising (or declining) trend starting any month of the year. We also consider all permutations of consecutive monthly dummy variables, including the trivial (one month in a row) and two, three, four, five, six, seven, eight, nine, ten months in a row. This collection of 126 patterns, each of which would be remarkable to observe, encompasses variations such as the Sell-in-May anomaly, the turn-of-the-year effect, and the January effect model. We expand this selection to over 500 models by including non-consecutive monthly dummy variable permutations of calendar anomaly models in our set of benchmark models (an arguably over-sampled universe of benchmark models, likely to include any calendar anomaly model ever considered as well as many not considered).

Because the run of six consecutive months of declining Treasury returns evident in Figure 1 draws one's attention, we start by bootstrapping the significance of this run. We find that this pattern is indeed unusual, with a p-value less than 0.001. Continuing to our "basic" set of benchmark models, the data-snooping adjusted p-value of the correlation of the onset/recovery variable with monthly mean returns is roughly 0.025, using raw or excess returns.[44,45] When we turn to our large benchmark set of models, the data-snooping adjusted p-value is roughly 0.075. We also conduct a *joint* test, based on the data-snooping-adjusted test statistic and the run of six consecutive months of declining Treasury returns, yielding a p-value less than 0.001. While it is impossible to prove that mere chance did not generate the Treasury return patterns we explore in this paper, application of the White reality test suggests that simple data mining is unlikely to be responsible for the results.

6 Sub-Sample Stability

Campbell (1990) observes that until 1952 short term Treasury rates were fixed by the Treasury and did not respond to market pressures. After 1951, auctions were held for bills, resulting in rates that arguably better reflect competitive pricing (although open market operations conducted by the Treasury still heavily influence short-end rates, as we describe above in Sections 1 and 3.1). In contrast, until 1971 the Treasury offered *notes and bonds* strictly in fixed-price sales (see Garbade (2007)). In 1971 the Treasury began experimenting with a variety of auction methods and slowly introduced note and bond auctions. The use of auctions in Treasury market offerings was standardized by 1982, and little has disturbed the competitive process of Treasury price-setting since that time, at least through to the end of our sample period, 2007. (See Appendix C for much more

[44]In the U.S. data we employ, the onset/recovery pattern is the not the model most strongly correlated with returns; a model with the onset/recovery variable lagged five months is the most strongly correlated. Given this, we modify the White procedure to determine the distribution of the second-to-maximum test statistic.

[45]This is the significance of the correlation of the twelve onset/recovery values with the twelve monthly mean returns over our entire sample, relative to similar correlations achievable with the wide range of alternative patterns outlined above.

institutional detail and historical information about the Treasury security market.) Altogether this suggests that the influence of seasonally varying risk aversion on Treasury prices through investor behavior should be a less prominent feature of the Treasury market before 1971, and a more stable feature since 1971, and especially since 1982.[46] To explore the impact of these institutional changes in the Treasury market on our findings, and to ensure our results are not driven by features of the data observed prior to the Treasury's effort to stabilize government offerings of notes and bonds, we performed two additional sets of analysis. First, we performed rolling-window estimation of Model 1. Second, we conducted sub-period analysis for 1952–1970 (pre-auction period), 1971–1981 (transition period), 1982–1994 (first half of modern auction period), and 1995–2007 (second half of modern auction period).

Consider first the rolling-window estimation. We estimated the onset/recovery coefficient based on Model 1 for each of the 5-, 7-, 10-, and 20-year Treasury security excess return series. We estimated the model as a system of equations using GMM as described for Equation (1). Instead of running this regression using the full sample of data, we re-estimated the model each month using a rolling window of 300 months of data. (For the regression associated with month t, we estimated the model using data for month t and the 299 months prior to t.) In order to provide the largest set of rolling windows estimates possible for this window width, we exploited the full range of data available to us, starting in January 1942 and extending to December 2012. We provide plots of the time series of onset-recovery coefficient estimates emerging from this rolling-window estimation in Figures 5 and 6.[47]

Figures 5 and 6 incorporate vertical bars to highlight changes to the process used by the Treasury to market its securities. Events associated with the vertical bars are as follows. Regular auctions for Treasury notes and bonds began in November 1970 with an 18-month note. In October 1972, the Treasury introduced auctions for a 2-year note, followed by auctions

[46]The influence of Operation Twist, run by the Federal Reserve Board since 2011, may overwhelm the influence of investor risk aversion on Treasury returns in recent data.

[47]To calculate the standard errors we follow Newey and West (1994) and use the Bartlett kernel and an automatic bandwidth parameter (autocovariance lags) equal to the integer value of $4 \cdot (T/100)^{2/9}$. The instruments used to form the GMM moments include the constant, the onset/recovery variable, the contemporaneous 30-day T-bill rate, and a lag of the value-weighted CRSP return.

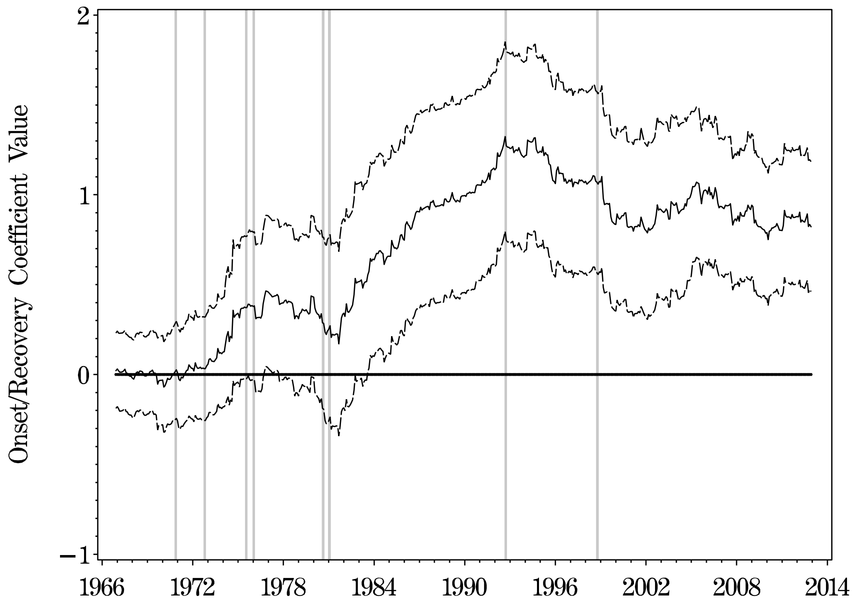


Figure 5: Rolling-window coefficient estimates restricted to be the same across Treasury series.

Description: We plot onset/recovery coefficient estimates, $\mu_{\hat{OR}}$, obtained from rolling-window estimations of Model 1 based on 300 months of data, with the first window starting in January 1942 and ending in December 1965. Each rolling-window coefficient estimate is restricted to be identical across maturities. The solid line represents $\mu_{\hat{OR}}$, and the dotted lines represent a 90 percent confidence interval. The coefficient estimates are based on estimating Equation (1) as a system of equations with GMM, but with rolling windows. Vertical bars highlight significant changes to the processes used by the Treasury to market its various securities, as detailed in Section 6.

Interpretation: The rolling-window coefficient estimates are generally significant from the 1980s onward, consistent with the Treasury having standardized the use of auctions and competitive pricing in their marketing of securities, with auction introductions and refinements indicated by the vertical shaded bars.

for a 4-year note in July 1975. The auction for the 5-year maturity began in January 1976, the 10-year maturity in August 1980, and the 7 and 20-year maturities in January 1981. Garbade (2007) reports the view that by 1982 market participants had finally concluded that the Treasury had wholeheartedly adopted a strategy of regular and predictable issuance. In September 1992, the Treasury started conducting single price auctions for

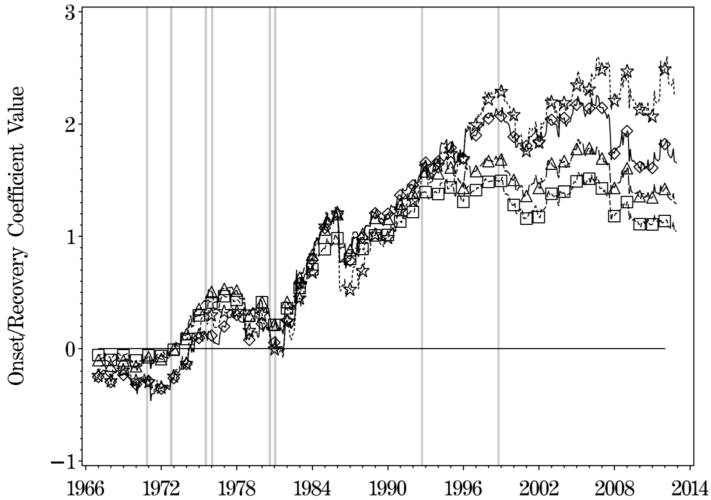


Figure 6: Rolling-window coefficient estimates for each Treasury series.

Description: We plot onset/recovery coefficient estimates, $\mu_{\hat{OR}}$, obtained from rolling-window estimations of Model 1 based on 300 months of data, with the first window starting in January 1942 and ending in December 1965. Each set of rolling-window estimates represents the unrestricted coefficient estimates for each of the 5-year (△), 7-year (□), 10-year (⋆), and 20-year (⋄) maturities. The coefficient estimates are based on estimating Equation (1) as a system of equations with GMM, but with rolling windows. Vertical bars highlight significant changes to the processes used by the Treasury to market its various securities, as detailed in Section 6.

Interpretation: The rolling-window coefficient estimates are generally significant from the 1980s onward, consistent with the Treasury having standardized the use of auctions and competitive pricing in their marketing of securities, with auction introductions and refinements indicated by the vertical shaded bars.

the 2- and 5-year notes. In November 1998, the Treasury adopted the single-price method for all auctions.[48]

Figure 5 shows the rolling-window onset/recovery coefficient estimate $\mu_{\hat{OR}}$ restricted to be identical across the 5-, 7-, 10-, and 20-year maturities, with the dotted lines representing the 90 percent confidence bound around the restricted coefficient estimates. Figure 6 shows the onset-recovery coefficient estimates for each of the individual Treasury security maturities,

[48] An additional point worth mentioning is Federal Reserve Chairman Volcker's announcement on October 6, 1979, that the discount rate was rising by 2 percentage points, leading to a crash in bond and equity markets.

unrestricted. Triangles, squares, stars, and diamonds are associated with the 5-, 7-, 10-, and 20-year maturities, respectively.

Consistent with the maturation of the Treasury market, the average onset/ recovery coefficient estimate (shown in Figure 5) is indistinguishable from zero prior to the 1980s though it appears to be steadily rising in magnitude from the early 1970s. As post-1982 data begins entering the rolling-window sample, the onset/recovery coefficient estimate becomes significantly greater than zero, and it remains that way until the present day. Coefficient estimates for each of the individual maturities, shown in Figure 6, show the same pattern.

Consider now our second set of sub-period analysis, based on fixed-window estimation for the periods 1952–1970, 1971–1981, 1982–1994, and 1995–2007. We provide the full set of results in Appendix E; a summary of the findings is as follows. The latter two sub-samples, 1982–1994 and 1995–2007, show very similar onset/recovery coefficients; they equal about 1 and are statistically significant in both sub-samples. The onset/recovery coefficient is of similar magnitude in the 1971–1981 sub-period, a little below 1, but it is not statistically significant. Data over the 1952–1970 sub-sample is not well captured by the seasonally varying risk aversion model, with the onset/recovery variable showing little or no statistical significance, and taking on a negative value, albeit of small magnitude (roughly −0.02). Overall, this evidence is consistent with a break in the process driving Treasury prices during the 1970s. Knowing that the Treasury switched to a competitive auction process during the 1970s, and that the non-competitive nature of Treasury issuance prior to 1971 was a matter of great concern to economists as early as the late 1950s, both for its impact on market-clearing and on attrition of Treasury buyers (see Garbade (2004, endnote 15) and related text), we view this as a cautionary note in interpreting regression results that include the pre-1971 period. To allay concerns that Model 1 outperforms the alternative models on the full sample period simply as a function of changes in the Treasury price-setting process, in Appendix E we replicate our full analysis for all twelve models using only the post-1970 data, and in untabulated results we replicate the full analysis for all twelve models using only the post-1981 data. In both cases, our findings are qualitatively identical to those we report here: the coefficient estimate on the onset/recovery variable is positive and statistically significant, and it captures the seasonal variation in Treasury returns, whereas the alternative models do not.

7 Discussion and Conclusion

We identify a striking seasonal pattern in the U.S. Treasury market in which average returns are statistically and economically significantly varying through the seasons. The pattern is present both by measure of jointly testing across series for monthly seasonality and by measure of testing for conditional seasonality (correlation with the timing of the onset of and recovery from seasonal depression in individuals). Monthly returns are approximately 80 basis points higher in October than in April, which is anomalously large by any measure. Relative to patterns documented by Kamstra *et al.* (2003) in *equity* returns, the conditional correlation between onset/recovery and Treasury returns is oppositely signed, despite the unconditional positive correlation that is empirically observed for equity and Treasury returns, and in contrast to the theoretical implications of standard asset pricing models.

The seasonal patterns in Treasury returns are largely unaffected when we control for a range of contemporaneous proxies for macroeconomic cycles and risk factors. These controls include both shocks and predictable movements in the macroeconomy (exploiting real-time vintage data, the most recent measures of macroeconomic data, and seasonally unadjusted data), suggesting that the seasonality we demonstrate is not related in any obvious way to time-varying risk or macroeconomic cyclicality. Employing turnover and stock market volatility measures suggested by Connolly *et al.* (2005) does not account for the seasonality. (In Appendix K we present evidence suggesting that Treasury market volume is also unlikely to account for the seasonal pattern observed in Treasury returns.) Investor sentiment as described by Baker and Wurgler (2006) and Baker and Wurgler (2007) could in principle lead to the sort of seasonality in Treasuries that we find, but we find the Baker-Wurgler sentiment index does not explain the seasonal patterns we observe here, nor does the Michigan consumer sentiment measure. The Fama-French and momentum factors also do not account for the seasonal pattern. Finally, accounting for various regularities including the Treasury auction schedule, the FOMC announcement cycle, and the supply of Treasury debt does not explain the large seasonal cycles we demonstrate. Based on seasonality tests discussed in Section 4, we find none of the models are able to explain a meaningful portion of the seasonal variation in returns except for the simplest model, Model 1, which explains more than 60 percent of the swing in returns from October to April.

Robustness checks confirm that the statistical and economic significance of the seasonal pattern is not an artifact of estimation technique and that the effects are apparent across the term structure. Further, whether we use raw returns or excess returns, the seasonal pattern in returns is evident, as is the ability of the onset/recovery variable to explain the seasonality in an economically meaningful way. Additionally, there are no seasonal artifacts induced in Treasury returns by maturity mismatches, which occur when there is no security available with maturity identical to the target maturity. (Many auctions have issues that mature mid-month, and consequently with monthly returns we cannot match targets any closer than within two weeks. In Appendix C we split our sample into cases that match by more or less than one month and find the properties of the two subsamples are very similar.)

Overall, the observed seasonal pattern in Treasury returns is consistent with seasonally varying investor risk aversion impacting financial markets through the depression that arises with seasonally lower daylight in fall and winter. Certainly our tests of the hypothesis, that seasonal depression onset/recovery correlates with an annual cycle in Treasury returns, result in a clear rejection of the null of no seasonality. Use of the White (2000) reality test demonstrates that the correlation of return seasonality with the clinical incidence of seasonal depression is unlikely to be the result of data snooping.

Building on prior research that firmly links seasonally varying investor mood with seasonally varying investor risk aversion, these findings contribute to a growing literature that finds financial markets are influenced by seasonally varying investor risk aversion in a manner that is statistically significant and economically meaningful. For example, Kamstra *et al.* (2013) find the flow of funds between risky and safe categories of mutual funds varies seasonally, with funds flowing from risky to safe categories in the fall and reversing in the spring. They study both net flows and net exchanges between funds within a mutual fund family. (We describe those results more fully in Appendix L.) Further, Kamstra *et al.* (2014) provide theoretical foundations for the seasonal patterns in Treasury and equity returns in an asset pricing model with seasonally varying risk aversion and seasonally varying elasticity of intertemporal substitution (EIS). Calibrating to consumption, they find that in order to match the observed seasonal patterns in equity and Treasury market returns, agents must have inelastic consumption (i.e., they must become more eager to consume in the present) during the fall/winter seasons when they are also more risk averse.[49]

Note that in this paper, and in related papers of ours, we focus our analysis on returns to equity and Treasury securities, taking those asset classes as representative of safe and risky ends of the risk spectrum. It is plausible that investors may seasonally shift their holdings not only between stocks and bonds but also between other investment alternatives. For example, some investors who experience seasonal changes in risk aversion may move money into or out of cash, annuities, insurance policies, collectibles, gold, real estate, or investments in human capital. We focus on equities and Treasuries due to the size and liquidity of these markets, and the ease of availability of reliable data. Future research may explore the implications of seasonally varying risk aversion for other asset classes.

One might reasonably wonder whether an investor could implement a trading strategy based on seasonality in the equity and/or Treasury markets to improve performance. An investor who exhibits seasonally varying risk aversion would realize lower returns on average if s/he followed his/her instincts and reallocated investments across equities and Treasury securities by season. An investor who was interested in attempting to “exploit” the seasonal patterns in returns would not be able to earn an “arbitrage” return by using a long/short strategy, since the expected equity premium is positive year round (albeit smaller on average in the spring/summer than in the fall/winter). A speculator could use leverage to try to increase returns in the fall/winter. S/he could also seasonally shift his/her stock holdings across the hemispheres to take advantage of seasonal differences in equity index returns that Kamstra *et al.* (2003) show exist across the hemispheres. We do not know of a comparable seasonal allocation strategy on the fixed income side. It is difficult to find a foreign Treasury security considered by investors to be as safe as U.S. Treasuries, so any cross-hemisphere strategy would be fraught with relatively more ambiguity and/or risk.

Finally, we must emphasize that the seasonal pattern we find in Treasury returns does not necessarily imply seasonal variation in risk itself. If

[49]The intuition is as follows. During the fall/winter season, the inelasticity of consumption puts downward pressure on both equity and risk-free returns while the increased risk aversion lowers equity but raises risk-free returns, and the impact of risk aversion overwhelms the impact of EIS on bond returns. Together the incorporation of seasonally varying risk aversion and EIS allows Kamstra *et al.* (2014) to match both the direction and the magnitude of seasonality in equity and risk-free returns.

a seasonal influence moves relatively predictably through the year in a pattern that corresponds to the fluctuations in the clinical onset of and recovery from seasonal depression, it is unlikely that smooth variations in *risk* through the course of the year are responsible. Certainly the macroeconomic variables and asset pricing factors we control for are the most plausible sources of time-varying risk. In spite of accounting for all of these effects, we still find remarkably strong, economically and statistically significant evidence of a seasonal effect in Treasury returns.

Appendices

A Data Sources

Here we provide the source of each data series used in the paper. Unless indicated otherwise, data sourced from the Board of Governors of the Federal Reserve System were collected from the Federal Reserve Bank of St. Louis, Economic Data (FRED): http://research.stlouisfed.org/fred2 Much of the data sourced from the U.S. Department of Labor: Bureau of Labor Statistics were also downloaded from FRED.

A.1 Treasury Index Return Series

20-year, 10-year, 7-year, 5-year Treasury Returns CRSP US Treasury and Inflation — Monthly. Sample period: 01/1952–12/2007.

A.2 Model 1: Onset/Recovery

Onset/Recovery ($\hat{\mathbf{OR}}_t$) Constructed with onset/recovery data from Lam (1998). Data available from http://www.markkamstra.com.

A.3 Model 2: Treasury Debt Supply Factors

Debt-to-GDP Ratio (DebtToGDP) Obtained from the Federal Reserve Bank of St. Louis, series IDs GFDEBTN and GDP, quarterly data linearly interpolated to the monthly frequency. Sample period: 01/1970–11/2007.

FOMC Board of Governors of the Federal Reserve System, http://www.federalreserve.gov/monetarypolicy/fomccalendars.htm Sample period: 01/1970–11/2007.

A.4 Models 3/5: Chen, Roll, and Ross Macroeconomic Risk Variables

Industrial Production Growth (IP) We obtained an index of industrial production and capacity utilization, series ID INDPRO, seasonally adjusted, percent change, from the Board of Governors of the Federal Reserve System. Sample period: 01/1952–12/2007.

Default Spread (Default) The Aaa and Baa bond yield data, used in constructing the Default variable, were obtained from the Board of Governors of the Federal Reserve System. The series we used are Moody's Seasoned Aaa Corporate Bond Yield and Moody's Seasoned Baa Corporate Bond Yield. Sample period: 01/1952–12/2007.

Term Spread (Term) The data we used to construct the Term variable are the 20-year Treasury bond and 30-day Treasury bill return series. Both series are from CRSP. We computed the monthly spread as the difference between the 20-year and 30-day values for each month, lagged. Sample period: 01/1952–12/2007.

Expected Inflation (Inf) and Surprise Inflation (InfSurp) We obtained the consumer price index for all urban consumers, all items, seasonally adjusted, series ID CPIAUCSL, percent change, from the U.S. Department of Labor: Bureau of Labor Statistics. Using an ARMA(1,1) model, we formed predicted and surprise inflation variables. Sample period: 01/1952–12/2007.

A.5 Models 4/5: Seasonally Unadjusted Macro Variables

GDP Growth ($\mathbf{GDP_{SU}}$) We obtained the quarterly GDP growth rate data from the Bureau of Economic Analysis and linearly interpolated to the monthly frequency. Sample period: 01/1952–12/2006.

Inflation based on PPI ($\mathbf{PPI_{SU}}$) We calculated the monthly percent change in the producer price index using PPI data, for all commodities, obtained from the U.S. Department of Labor: Bureau of Labor Statistics. Sample period: 01/1952–12/2006.

Industrial Production Growth ($\mathbf{IP_{SU}}$) We calculated the monthly growth rate in the industrial production total index using data obtained from Global Insight. Sample period: 01/1952–12/2006.

Unemployment Growth ($\mathbf{UEG_{SU}}$) We calculated the monthly unemployment growth rate based on data obtained from Global Insight (series LZHUR, 16 years of age and older). Sample period: 01/1952–12/2006.

Inflation based on CPI ($\mathbf{CPI_{SU}}$) We calculated the monthly percent change in CPI based on CPI for all urban consumers (series CPIAUCNS), obtained from the U.S. Department of Labor: Bureau of Labor Statistics. Sample period: 01/1952–12/2006.

A.6 Model 6: Real-Time Macro Variables

Unemployment Rate Surprise, Contraction (USurpC) and Expansion (USurpE) We obtained seasonally unadjusted unemployment rates for individuals 16 years of age and older from the Bureau of Labor Statistics, and we obtained real-time unemployment rates from the Philadelphia Federal Reserve Bank. We used these series to construct the expected change in the unemployment rate and the surprise in the change in the unemployment rate, as described in Appendix I. Sample period: 12/1965–12/2003.

Industrial Production: Expected (IP) and Surprise (IPSurp) We obtain an index of industrial production from the Board of Governors of the Federal Reserve System, and real-time data come from the Philadelphia Federal Reserve Bank. As we detail in Appendix I, we used these series to construct the expected growth in industrial production and the surprise in the industrial production growth rate. Sample period: 12/1965–12/2003.

Change in Default Spread (ΔDefault) The Aaa and Baa bond yield data, used in constructing the Default variable, were obtained from the Board of Governors of the Federal Reserve System. The series we used are Moody's Seasoned Aaa Corporate Bond Yield and Moody's Seasoned Baa Corporate Bond Yield. Sample period: 12/1965–12/2003.

Term Spread (ΔTerm) The data we used to construct the Term variable are the 20-year Treasury bond and 30-day Treasury bill return series. Both series are from CRSP. We computed the monthly spread as the difference between the 20-year and 30-day values for each month, then we computed the monthly change in the spread by taking the difference from one month to the next. Sample period: 12/1965–12/2003.

Probability of Contraction (ProbC) We obtained the Stock and Watson (1989) experimental coincident recession index from the National

Bureau of Economic Research. This series is a real-time indicator, making use of real-time information only in determining whether the economy is expanding or contracting at a given point in time. Sample period: 12/1965–12/2003.

Inflation Surprise (InfSurp) and Predicted (Inf) We obtained two CPI-based inflation rate series from the Philadelphia Federal Reserve Bank. The first is real-time inflation, announced quarterly, available in real-time format only from mid-1994. The second is most recently revised inflation, available from 1965. As we explain in Appendix I, we used these series to construct the inflation surprise variable two different ways. Sample period: 12/1965–12/2003.

A.7 Models 7/8: Factors Related to Cross-Market Hedging

Turnover (Turnover) Using the CRSP monthly stock file, we calculated the monthly total volume and total shares outstanding of all stocks, formed the ratio of volume to shares outstanding, then calculated the deviation of this ratio from the (rolling) one-year average of this ratio. Sample period: 08/1960–12/2007.

Conditional Volatility (CondVar) This is the fitted (conditional) value from a GARCH(1,1) model estimated on monthly S&P 500 returns. Sample period: 01/1952–12/2007.

Treasury Liquidity (Liquidity) We formed the proxy for Treasury market liquidity using proportional bid-ask spread data on short-term Treasury securities, maturity less than or equal to 1 year. We follow Goyenko *et al.* (2011) and adjusted this measure by removing a time trend and the square of the time trend. We obtained the raw monthly data from the CRSP Treasury Quotes file. Sample period: 08/1960–12/2007.

Treasury Volatility (TreasuryVol) We calculated the volatility of daily yields for 5-year, 7-year, 10-year, and 20-year constant maturity Treasury securities acquired from the Board of Governors of the Federal Reserve System through the St. Louis Federal Reserve, series IDs DGS5, DGS7, DGS10, and DGS20. Sample periods for the 5-year, 7-year, 10-year, and 20-year series, respectively: 04/1962–12/2007, 10/1969–12/2007, 04/1962–12/2007, and 01/1994–12/2007.

A.8 Models 9/10: Sentiment

Baker-Wurgler Sentiment (BWSentiment) We obtained these data from Jeff Wurgler's Web site, http://pages.stern.nyu.edu/ jwurgler/ Sample period: 03/1966–12/2005.

Michigan Consumer Sentiment (MSentiment) We obtained the Michigan consumer sentiment data from the Board of Governors of the Federal Reserve System through the St. Louis Federal Reserve, series IDs UMCSENT (mostly quarterly 11/1952 to 11/1977) and UMCSENT1 (monthly 01/1978 to 01/2008). We linearly interpolate the levels of the 11/1952 to 11/1977 index to monthly frequency, we splice the interpolated 1952–1977 monthly series with the 1978–2007 monthly series, and then we calculate the monthly change. MSentiment_t is defined as the lag of the change in the monthly series. Sample period: 02/1953–12/2007.

A.9 Model 11: Fama and French Model

Size (SMB) We obtained these data from Ken French's Web site: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ Sample period: 01/1952–12/2007.

Book-to-Market (HML) We obtained these data from Ken French's Web site. Sample period: 01/1952–12/2007.

Momentum (MOM) We obtained these data from Ken French's Web site. Sample period: 01/1952–12/2007.

Default Spread (Default) We obtained these data from Ken French's Web site. Sample period: 01/1952–12/2007.

Term Spread (Term) This is the 20-year Treasury monthly return in excess of the 30-day T-bill return, obtained from CRSP. Sample period: 01/1952–12/2007.

Orthogonalized Market Return (MKT) We produced this series as described in footnote 35. Sample period: 01/1952–12/2007.

Term Spread (Term90) As defined by Harvey (1989), this is the 90-day T-bill monthly return in excess of the 30-day T-bill return, both obtained from CRSP. Sample period: 01/1952–12/2007.

A.10 Model 12: Conditional CAPM

Excess Dividend Yield (XDP) This series was constructed by subtracting the CRSP equal-weighted index returns without dividends from the CRSP equal-weighted index returns with dividends. Sample period: 01/1952–12/2007.

Excess Market Return (MKT) We obtained these data from Ken French's Web site. Sample period: 01/1952–12/2007.

Default Spread (Default) The Aaa and Baa bond yield data, used in constructing the Default variable, are obtained from the Board of Governors of the Federal Reserve System. The series we use are Moody's Seasoned Aaa Corporate Bond Yield and Moody's Seasoned Baa Corporate Bond Yield. Sample period: 01/1952–12/2007.

Term Spread (Term90) This is the 90-day T-bill monthly return in excess of the 30-day T-bill return, both obtained from CRSP. Sample period: 01/1952–12/2007.

A.11 Other Data

Value-Weighted U.S. Market Return We obtained this series from CRSP. It is the NYSE/Amex/NASDAQ value-weighted return, including distributions. Sample period: 01/1952–12/2007.

30-day Treasury-Bill Return CRSP US Treasury and Inflation-Monthly. Sample period: 01/1952–12/2007.

B Summary Statistics on Series Used in Models 2–12

Tables B1 through B8 contain summary statistics on the variables used in Models 2-12, with format analogous to that used for the Treasury and equity return series shown in Table 1. Appendix A reports sources for all of the series, and Appendix D contains a broader set of summary statistics for each of the series, as well as bootstrap standard errors for each of the seasonality tests.

Series	Mean	Std	Seasonality test: Asymptotic p-values			
			Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Debt-to-GDP Ratio	0.50	0.13	1.000	0.831	0.875	0.904
FOMC	0.75	0.43	**<0.001**	**0.002**	**<0.001**	**0.001**

Table B1: Summary statistics — Model 2 series: Treasury debt supply factors.
Description: See Table 1 for details. The sample period is 01/1970–11/2007 ($N = 455$) for the Debt-to-GDP series and 01/1970–12/2007 ($N = 456$) for the FOMC series.

Series	Mean	Std	Seasonality test: Asymptotic p-values			
			Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Industrial Production Growth (IP)	0.268	0.93	0.962	0.692	0.762	0.698
Expected Inflation (Inf)	0.310	0.20	1.000	0.753	0.960	0.637
Surprise Inflation (InfSurp)	0.000	0.24	0.692	0.546	0.268	0.601
Default Spread (Default)	0.934	0.41	0.819	0.805	0.804	0.811
Term Spread (Term)	0.135	2.64	0.300	0.479	0.457	0.131

Table B2: Summary statistics — Models 3/5 series: CRR macro factors.
Description: See Table 1 for details. The sample period for these series is 01/1952–12/2007 ($N = 672$).

Series	Mean	Std	Seasonality test: Asymptotic p-values			
			Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Inflation based on CPI (CPI_{SU})	0.309	0.34	**<0.001**	**<0.001**	0.453	0.219
GDP Growth (GDP_{SU})	0.018	0.03	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Industrial Production Growth (IP_{SU})	0.003	0.02	**<0.001**	**<0.001**	0.221	**0.001**
PPI Inflation (PPI_{SU})	0.261	0.72	**<0.001**	**0.002**	0.661	0.301
Unemployment Growth (UEG_{SU})	0.005	0.10	**<0.001**	**<0.001**	**0.092**	**<0.001**

Table B3: Summary statistics — Models 4/5 series: Seasonally unadjusted macro variables.
Description: See Table 1 for details. The sample period for these series is 01/1952–12/2006 ($N = 660$).

Series	Mean	Std	Seasonality test: Asymptotic p-values Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Unemployment Surprise, Contraction (USurpC)	0.013	0.08	0.645	0.205	0.730	0.941
Unemployment Surprise, Expansion (USurpE)	−0.020	0.17	0.337	0.537	0.948	0.833
Probability of Contraction (ProbC)	0.158	0.28	0.849	0.726	0.701	0.579
Industrial Production Surprise (IPSurp)	−0.107	0.72	**0.028**	0.988	**0.047**	**0.084**
Expected Growth in Industrial Production (IP)	0.255	0.33	0.168	0.317	0.143	0.252
Expected Change in Unemployment (EUEG)	1.210	10.28	0.475	0.590	0.327	0.485
Change in Default Spread (ΔDefault)	−0.138	11.20	**0.004**	**<0.001**	0.344	0.157
Term Spread (ΔTerm)	0.170	3.00	0.307	0.402	0.213	0.120
Inflation Surprise (InfSurp)	0.001	0.21	0.352	0.124	0.392	0.928
Predicted Inflation (Inf)	0.386	0.23	1.000	0.913	0.815	0.847

Table B4: Summary statistics—Model 6 series: Real-time macro variables.
Description: See Table 1 for details. The sample period for these series is 12/1965–12/2003 ($N = 457$).

Series	Mean	Std	Seasonality test: Asymptotic p-values Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Turnover (Turnover)	0.029	0.13	**<0.001**	0.350	0.132	**<0.001**
Conditional Volatility (CondVar)	19.60	11.82	0.963	0.924	0.590	0.847
Treasury Liquidity (Liquidity)	0.00	0.02	0.850	0.852	0.929	0.877
5-Year Forecasted Realized Volatility ($\text{TreasuryVol}_{5\text{Year}}$)	1.34	0.71	0.918	0.662	0.541	0.935
7-Year Forecasted Realized Volatility ($\text{TreasuryVol}_{7\text{Year}}$)	2.01	0.92	0.774	0.623	0.433	0.944
10-Year Forecasted Realized Volatility ($\text{TreasuryVol}_{10\text{Year}}$)	2.39	1.35	0.975	0.794	0.677	0.948
20-Year Forecasted Realized Volatility ($\text{TreasuryVol}_{20\text{Year}}$)	4.47	0.65	0.913	0.458	**0.096**	0.634

Table B5: Summary statistics — Models 7/8 series: Factors related to cross-market hedging.

Description: See Table 1 for details. The sample period is 08/1960–12/2007 ($N = 569$) for Turnover, CondVar, and Liquidity; 04/1962–12/2007 ($N = 549$) for $\text{TreasuryVol}_{5\text{year}}$ and $\text{TreasuryVol}_{10\text{year}}$; 10/1969–12/2007 ($N = 459$) for $\text{TreasuryVol}_{7\text{year}}$; and 01/1994–12/2007 ($N = 168$) for $\text{TreasuryVol}_{20\text{year}}$.

Series	Mean	Std	Seasonality test: Asymptotic p-values Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Baker-Wurgler Sentiment (BWSentiment)	0.00	0.42	0.342	0.605	0.225	0.702
Michigan Consumer Sentiment (MSentiment)	−0.02	3.08	**0.004**	**0.032**	0.773	0.340

Table B6: Summary statistics—Models 9/10 series: sentiment.

Description: See Table 1 for details. The sample period is 03/1966–12/2005 ($N = 478$) for BWSentiment and 02/1953–12/2007 ($N = 659$) for MSentiment.

			Seasonality test: Asymptotic p-values			
Series	Mean	Std	Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Size (SMB)	0.168	2.97	**<0.001**	**<0.001**	0.769	0.290
Book-to-Market (HML)	0.394	2.71	**0.005**	**<0.001**	0.206	**0.079**
Momentum (MOM)	0.853	3.69	**<0.001**	**0.009**	0.197	0.965
Default Spread (Default)	0.934	0.41	0.819	0.805	0.804	0.811
Term Spread (Term)	0.135	2.64	0.300	0.479	0.457	0.131
Orthogonalized Market Return (MKT)	0.005	3.81	**0.005**	0.693	**0.005**	0.337

Table B7: Summary statistics — Model 11 series: Fama and French model.
Description: See Table 1 for details. The sample period is 01/1952–12/2007 ($N = 672$).

			Seasonality test: Asymptotic p-values			
Series	Mean	Std	Nonspec. Monthly	Fall vs. Winter	Sep. vs. Mar.	Oct. vs. Apr.
Excess Dividend Yield (XDP)	0.261	0.11	**<0.001**	0.390	0.493	0.215
Lagged Excess Market Return (XMKT)	0.577	4.21	**0.012**	0.668	0.319	0.106
Term Spread (Term90)	0.042	0.09	**0.016**	0.680	0.819	0.716

Table B8: Summary statistics — Model 12 series: Conditional CAPM.
Description: See Table 1 for details. The sample period is 01/1952–12/2007 ($N = 672$).

Appendices C – L

Appendices C through L are available on the journal's Web site and at http://www.markkamstra.com

References

Andersen, T. G. and L. Benzoni. 2009. "Realized Volatility". In: *Handbook of Financial Time Series*. Ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiβ, and T. Mikosch. Berlin Heidelberg: Springer Verlag. 555–575.

Andersen, T. G. and L. Benzoni. 2010. "Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models". *Journal of Finance*. 65(2): 603–653.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2003. "Modeling and Forecasting Realized Volatility". *Econometrica*. 71(2): 529–626.

Andrews, D. W. K. and B. Lu. 2001. "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models". *Journal of Econometrics*. 101(1): 123–164.

Ang, A. and M. Piazzesi. 2003. "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables". *Journal of Monetary Economics*. 50(4): 745–787.

Athanassakos, G. 2008. "Seasonal Patterns in Canadian Financial Markets and the Impact of Professional Portfolio Rebalancing: Evidence of Profitable Opportunities". *Journal of Financial and Economic Practice*. 9(1): 73–96.

Baker, M. and J. Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns". *Journal of Finance*. 61(4): 1645–1680.

Baker, M. and J. Wurgler. 2007. "Investor Sentiment in the Stock Market". *Journal of Economic Perspectives*. 21(2): 129–151.

Bassi, A., R. Colacito, and P. Fulghieri. 2013. "'O Sole Mio: An Experimental Analysis of Weather and Risk Attitudes in Financial Decisions". *Review of Financial Studies*. 26(7): 1824–1852.

Bekaert, G. and C. R. Harvey. 1995. "Time-Varying World Market Integration". *Journal of Finance*. 50(2): 403–444.

Boudoukh, J. 1993. "An Equilibrium Model of Nominal Bond Prices with Inflation-Output Correlation and Stochastic Volatility". *Journal of Money, Credit, and Banking*. 25(3): 636–665.

Boyd, J. H., J. Hu, and R. Jagannathan. 2005. "The Stock Market's Reaction to Unemployment News: Why Bad News is Usually Good for Stocks". *Journal of Finance*. 60(2): 649–672.

Brandt, M. W. and K. Q. Wang. 2003. "Time-Varying Risk Aversion and Unexpected Inflation". *Journal of Monetary Economics*. 50(7): 1457–1498.

Campbell, J. Y. 1990. "Measuring the Persistence of Expected Returns". *American Economic Review*. 80(2): 43–47.

Campbell, J. Y. and R. J. Shiller. 1991. "Yield Spreads and Interest Rate Movements: A Bird's Eye View". *Review of Economic Studies*. 58(3): 495–514.

Carton, S., R. Jouvent, C. Bungenera, and D. Widlöcher. 1992. "Sensation Seeking and Depressive Mood". *Personality and Individual Differences*. 13(7): 843–849.

Carton, S., P. Morand, C. Bungenera, and R. Jouvent. 1995. "Sensation-Seeking and Emotional Disturbances in Depression: Relationships and Evolution". *Journal of Affective Disorders*. 13(3): 219–225.

Chang, E. C. and R. D. Huang. 1990. "Time-Varying Return and Risk in the Corporate Bond Market". *Journal of Financial and Quantitative Analysis*. 25(3): 323–340.

Chen, N., R. Roll, and S. A. Ross. 1986. "Economic Forces and the Stock Market". *Journal of Business*. 59(3): 383–403.

Cochrane, J. H. 2005. *Asset Pricing*. 2nd. Princeton, NJ: Princeton University Press.

Cochrane, J. H. and M. Piazzesi. 2005. "Bond Risk Premia". *American Economic Review*. 95(1): 138–160.

Connolly, R., C. Stivers, and L. Sun. 2005. "Stock Market Uncertainty and the Stock-Bond Return Relation". *Journal of Financial and Quantitative Analysis*. 40(1): 161–194.

De Bondt, W. F. M. and M. M. Bange. 1992. "Inflation Forecast Errors and Time Variation in Term Premia". *Journal of Financial and Quantitative Analysis*. 27(4): 479–496.

Dupont, D. and B. Sack. 1999. "The Treasury Security Market: Overview and Recent Developments". *Federal Reserve Bulletin*. 85: 785–806.

Engle, R. F. 1982. "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation". *Econometrica*. 50: 987–1008.

Fama, E. F. and R. R. Bliss. 1987. "The Information in Long-Maturity Forward Rates". *American Economic Review*. 77(4): 680–692.

Fama, E. F. and K. R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds". *Journal of Financial Economics*. 33: 3–56.

Ferson, W. E. and S. R. Foerster. 1994. "Finite Sample Properties of Generalized Methods of Moments in Tests of Conditional Asset Pricing Models". *Journal of Financial Economics*. 36(1): 29–55.

Ferson, W. E. and C. R. Harvey. 1992. "Seasonality and Consumption-Based Asset Pricing". *Journal of Finance*. 47(2): 511–552.

Fleming, M. J. and K. D. Garbade. 2007. "Deal Behavior in the Specials Market for US Treasury Securities". *Journal of Financial Intermediation*. 16(2): 204–228.

Fleming, M. J., K. D. Garbade, and F. Keane. 2005. "Anomalous Bidding in Short-Term Treasury Bill Auctions". *Journal of Financial Research*. 28(2): 165–176.

Fridson, M. S. 2000. "Semiannual Seasonality in High-Yield Bond Returns". *Journal of Portfolio Management*. 26(4): 102–111.

Garbade, K. D. 2004. "The Institutionalization of Treasury Note and Bond Auctions, 1970–75". *FRBNY Economic Policy Review*. 10(1): 29–45.

Garbade, K. D. 2007. "The Emergence of 'Regular and Predictable' as a Treasury Debt Management Strategy". *FRBNY Economic Policy Review*. 13(1): 53–71.

Garrett, I., M. J. Kamstra, and L. A. Kramer. 2005. "Winter Blues and Time Variation in the Price of Risk". *Journal of Empirical Finance*. 12(2): 291–316.

Gibson, W. E. 1970. "Interest Rates and Monetary Policy". *Journal of Political Economy*. 78(3): 431–455.

Godfrey, L. G. 1978a. "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables". *Econometrica*. 46(6): 1293–1301.

Godfrey, L. G. 1978b. "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables". *Econometrica*. 46(6): 1303–1310.

Gonçalves, S. and H. White. 2002. "The Bootstrap of the Mean for Dependent Heterogeneous Arrays". *Econometric Theory*. 18: 1367–1384.

Gonçalves, S. and H. White. 2005. "Bootstrap Standard Error Estimates for Linear Regressions". *Journal of the American Statistical Association*. 100: 970–979.

Goyenko, R., A. Subrahmanyam, and A. Ukhov. 2011. "The Term Structure of Bond Market Liquidity and its Implications for Expected Bond Returns". *Journal of Financial and Quantitative Analysis*. 46(1): 111–139.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moments Estimators". *Econometrica*. 50(4): 1029–1054.

Harmatz, M. G., A. D. Well, C. E. Overtree, K. Y. Kawamura, M. Rosal, and I. S. Ockene. 2000. "Seasonal Variation of Depression and Other Moods: A Longitudinal Approach". *Journal of Biological Rhythms*. 15(4): 344–350.

Harvey, C. R. 1989. "Time-Varying Conditional Covariances in Tests of Asset Pricing Models". *Journal of Financial Economics*. 24(2): 289–317.

Hirshleifer, D. and T. Shumway. 2003. "Good Day Sunshine: Stock Returns and the Weather". *Journal of Finance*. 58(3): 1009–1032.

Holland, A. S. and M. Toma. 1991. "The Role of the Federal Reserve as "Lender of Last Resort" and the Seasonal Fluctuation of Interest Rates". *Journal of Money, Credit, and Banking*. 23(4): 659–676.

Ilmanen, A. 1995. "Time-Varying Expected Returns in International Bond Markets". *Journal of Finance*. 50(2): 481–506.

Jegadeesh, N. and S. Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency". *Journal of Finance*. 48(1): 65–91.

Jin, L. and A. Scherbina. 2011. "Inheriting Losers". *Review of Financial Studies*. 24(3): 786–820.

Jordan, S. D. and B. D. Jordan. 1991. "Seasonality in Daily Bond Returns". *Journal of Financial and Quantitative Analysis*. 26(2): 269–285.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2001. "Losing Sleep at the Market: The Daylight Saving Anomaly". *American Economic Review*. 90(4): 1005–1011.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2003. "Winter Blues: A SAD Stock Market Cycle". *American Economic Review*. 93(1): 324–343.

Kamstra, M. J., L. A. Kramer, and M. D. Levi. 2012. "A Careful Re-Examination of Seasonality in International Stock Markets: Comment on Sentiment and Stock Returns". *Journal of Banking and Finance*. 36(4): 934–956.

Kamstra, M. J., L. A. Kramer, M. D. Levi, and T. Wang. 2014. "Seasonally Varying Preferences: Theoretical Foundations for an Empirical Regularity". *Review of Asset Pricing Studies*. 4(1): 39–77.

Kamstra, M. J., L. A. Kramer, M. D. Levi, and R. Wermers. 2013. "Seasonal Asset Allocation: Evidence from Mutual Fund Flows". Manuscript, University of Toronto. Available at SSRN: URL: http://ssrn.com/abstract=1907904.

Kaplanski, G. and H. Levy. 2008. "Seasonal Affective Disorder (SAD) and Perceived Market Risk". Manuscript, University of Jerusalem.

Kramer, L. A. and J. M. Weber. 2012. “This is Your Portfolio on Winter: Seasonal Affective Disorder and Risk Aversion in Financial Decision-Making”. *Social Psychological and Personality Science*. 3(2): 193–199.

Krishnamurthy, A. 2002. “The Bond/Old-Bond Spread”. *Journal of Financial Economics*. 66: 463–506.

Krishnamurthy, A. and A. Vissing-Jorgensen. 2012. “The Aggregate Demand for Treasury Debt”. *Journal of Political Economy*. 120(2): 233–267.

Lam, R. W. 1998. “Seasonal Affective Disorder: Diagnosis and Management”. *Primary Care Psychiatry*. 4: 63–74.

Lauricella, T., M. Phillips, and D. Enrich. 2011. “Stocks Dive Again on Europe, Economy Fears”. *Wall Street Journal*. URL: http://online.wsj.com/ (accessed on 08/11/2011).

Levi, M. D. 1973. “Errors in the Variables Bias in the Presence of Correctly Measured Variables”. *Econometrica*. 41(5): 985–986.

MacKinnon, J. G. and H. White. 1985. “Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties”. *Journal of Econometrics*. 29(3): 305–325.

Merton, R. C. 1980. “On Estimating the Expected Return on the Market”. *Journal of Financial Economics*. 8: 323–361.

Molin, J., E. Mellerup, T. Bolwig, T. Scheike, and H. Dam. 1996. “The Influence of Climate on Development of Winter Depression”. *Journal of Affective Disorders*. 37(2–3): 151–155.

Murfin, J. and M. Petersen. 2014. “Loans on Sale: Credit Market Seasonality, Borrower Need, and Lender Rent Seeking”. NBER Working Paper No. w20310. Available at SSRN: URL: http://ssrn.com/abstract=2471203.

Newey, W. K. and K. D. West. 1987. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”. *Econometrica*. 55(3): 703–708.

Newey, W. K. and K. D. West. 1994. “Automatic Lag Selection in Covariance Matrix Estimation”. *Review of Economic Studies*. 61: 631–653.

Palm, F. C., S. Smeekes, and J.-P. Urbain. 2011. “Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests”. *Journal of Econometrics*. 163(1): 85–104.

Pietromonaco, P. R. and K. S. Rook. 1987. “Decision Style in Depression: The Contribution of Perceived Risks Versus Benefits”. *Journal of Personality and Social Psychology*. 52(2): 399–408.

Politis, D. N. and J. P. Romano. 1994. “The Stationary Bootstrap”. *Journal of the American Statistical Association*. 89(428): 1303–1313.

Qiu, L. and I. Welch. 2006. "Investor Sentiment Measures". NBER Working Paper No. W10794. Available at SSRN: URL: http://ssrn.com/abstract=595193.

Rilstone, P. and M. Veall. 1996. "Using Bootstrapped Confidence Intervals for Improved Inferences with Seemingly Unrelated Regression Equations". *Econometric Theory*. 12(3): 569–580.

Schneeweis, T. and J. R. Woolridge. 1979. "Capital Market Seasonality: The Case of Bond Returns". *Journal of Financial and Quantitative Analysis*. 14(5): 939–958.

Shefrin, H. 2008. *A Behavioral Approach to Asset Pricing*. 2nd. New York, NY: Elsevier Academic Press.

Smoski, M. J., T. R. Lynch, M. Z. Rosenthal, J. S. Cheavens, A. L. Chapman, and R. R. Krishnan. 2008. "Decision-Making and Risk Aversion among Depressive Adults". *Journal of Behavior Therapy*. 39: 567–576.

Staiger, D. and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments". *Econometrica*. 65(3): 557–586.

Statman, M., K. L. Fisher, and D. Anginer. 2008. "Affect in a Behavioral Asset Pricing Model". *Financial Analysts Journal*. 64(2): 20–29.

Stock, J. H. and M. W. Watson. 1989. "New Indexes of Coincident and Leading Economic Indicators". In: *NBER Macroeconomics Annual*. Ed. by O. J. Blanchard and S. Fischer. Cambridge, MA: MIT Press.

Stock, J. H. and J. H. Wright. 2000. "GMM with Weak Identification". *Econometrica*. 68(5): 1055–1096.

Sullivan, R., A. Timmermann, and H. White. 2001. "Dangers of Data Mining: The Case of Calendar Effects in Stock Returns". *Journal of Econometrics*. 105(1): 249–286.

White, H. 2000. "A Reality Check for Data Snooping". *Econometrica*. 68(5): 1097–1126.

Wilson, J. W. and C. P. Jones. 1990. "Is There a January Effect in Corporate Bond and Paper Returns?" *Financial Review*. 25(1): 55–71.

Young, M. A., P. M. Meaden, L. F. Fogg, E. A. Cherin, and C. I. Eastman. 1997. "Which Environmental Variables Are Related to the Onset of Seasonal Affective Disorder?" *Journal of Abnormal Psychology*. 106(4): 554–562.

Zeng, M. 2011. "Safe Haven? Investors Still Run to Treasurys". *Wall Street Journal*. URL: http://online.wsj.com/ (accessed on 08/09/2011).