

Article

The Effect of Macroeconomic Announcements on U.S. Treasury Markets: An *Autometric* General-to-Specific Analysis of the Greenspan Era

James J. Forest ^{1,2} ¹ School of Business Administration, SUNY New Paltz, New Paltz, NY 12561, USA; forestj@newpaltz.edu² Center for International Securities and Derivatives Markets, Isenberg School of Management, University of Massachusetts—Amherst, Amherst, MA 01003, USA

Abstract

This research studies the impact of macroeconomic announcement surprises on daily U.S. Treasury excess returns during the heart of Alan Greenspan's tenure as Federal Reserve Chair, addressing the possible limitations of standard static regression (SSR) models, which may suffer from omitted variable bias, parameter instability, and poor mis-specification diagnostics. To complement the SSR framework, an automated general-to-specific (Gets) modeling approach, enhanced with modern indicator saturation methods for robustness, is applied to improve empirical model discovery and mitigate potential biases. By progressively reducing an initially broad set of candidate variables, the Gets methodology steers the model toward congruence, dispenses unstable parameters, and seeks to limit information loss while seeking model congruence and precision. The findings, herein, suggest that U.S. Treasury market responses to macroeconomic news shocks exhibited stability for a core set of announcements that reliably influenced excess returns. In contrast to computationally costless standard static models, the automated Gets-based approach enhances parameter precision and provides a more adaptive structure for identifying relevant predictors. These results demonstrate the potential value of incorporating interpretable automated model selection techniques alongside traditional SSR and Markov switching approaches to improve empirical insights into macroeconomic announcement effects on financial markets.



Academic Editor: Julien Chevallier

Received: 7 March 2025

Revised: 26 May 2025

Accepted: 4 June 2025

Published: 21 June 2025

Citation: Forest, J. J. (2025). The Effect of Macroeconomic Announcements on U.S. Treasury Markets: An *Autometric* General-to-Specific Analysis of the Greenspan Era. *Econometrics*, 13(3), 24. <https://doi.org/10.3390/econometrics13030024>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: U.S. Treasury Market; general-to-specific modeling; macroeconomic announcements; encompassing; bias correction; model selection; autometrics; market efficiency

JEL Classification: C22; C58; C52; E44; E4; G14; G12; G10

1. Introduction

The relationship between macroeconomic announcements and U.S. Treasury returns has long been a focal point of empirical research, offering critical insights into financial market dynamics and the transmission of monetary policy.

This paper extends the literature on macroeconomic announcement effects by leveraging recent methodological advancements in financial econometrics—specifically, the automated general-to-specific (Gets) modeling framework implemented via Autometrics.¹ The Gets methodology, rooted in the London School of Economics (LSE/Oxford) econometric approach, provides a systematic and robust method for model selection that

complements conventional static regression models, which often suffer from omitted variable bias, overfitting, and model and parameter instability. Unlike traditional empirical approaches, which rely on pre-specified models that may ignore key structural shifts, Gets employs an adaptive process that selects relevant variables while steering toward model congruence. This is particularly desirable in high-frequency financial markets, where news-driven regime shifts and time-varying risk premia complicate model estimation.^{2,3}

The contribution of this study is twofold. First, it demonstrates the empirical benefits of Gets for modeling financial time series, specifically in identifying those announcements offering stable market responses to macroeconomic surprises. Using a comprehensive dataset of macroeconomic announcements and U.S. Treasury excess returns during the heart of the Greenspan era, this analysis reveals which indicators were consistently the most relevant to market participants compared to others, where their effects evolved over time. Second, this study suggests the broader applicability of Gets modeling in financial econometrics, advocating for its use as a rigorous alternative to traditional static models in contexts where parameter precision, dynamic variable selection, and structural robustness are critical.

The Greenspan era presents a particularly compelling setting, not just because of evolving market structures and macroeconomic shifts, but also due to the Federal Reserve's deliberate opacity in communication. Unlike later Fed regimes that embraced formal forward guidance, Greenspan's leadership was characterized by a discretionary, often obfuscatory, approach, requiring market participants to parse public statements, tone, and macroeconomic data for signals about future monetary policy (Blinder et al., 2008; Gürkaynak et al., 2007; Swanson, 2006).^{4,5}

As a result, market sensitivity to macroeconomic news was heightened, making this period an ideal testing ground for models that capture dynamic responses to economic information. Under Alan Greenspan's leadership, the Federal Reserve maintained a discretionary, data-driven approach, with market participants parsing public statements and economic indicators for clues about future monetary policy. The resulting uncertainty heightened Treasury market sensitivity to macroeconomic news, making this period an ideal testing ground for models that can efficiently capture dynamic responses to economic information.

By integrating methodological innovation with empirical application, this study advances both econometric practice and financial market analysis. The Greenspan era serves not only as a rich historical case for evaluating announcement effects but also as a proving ground for econometric techniques capable of addressing the complexities of modern financial data.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature on macroeconomic announcement effects in financial markets. Section 3 details the econometric methodology, emphasizing the application of Gets and indicator saturation techniques. Section 4 describes the data. Section 5 presents the empirical results, including model selection diagnostics, robustness checks, and implications for financial market efficiency. Section 6 concludes with a summary of key findings and directions for future research.

2. Review of the Literature

Numerous studies have examined the effects of macroeconomic announcements on financial markets, focusing on how such news affects interest rates, equity markets, and exchange rates. This section reviews the relevant literature, highlighting the significant findings and identifying gaps that this study aims to fill.

First, the literature related to the Treasury market is presented. Second, equity market macro announcement studies are surveyed. Finally, Gets modeling literature in economics and finance are summarized. To the best of the author's knowledge, this is the first study of macroeconomic announcements in U.S. Treasury markets to employ automated Gets modeling.

2.1. Treasury Markets

Early research, such as (T. Urich & Wachtel, 1984), investigated the effects of money supply and inflation on interest rates, finding that unanticipated results had an immediate impact on short-term rates. (Ederington & Lee, 1993) used intraday data to show that macroeconomic announcements are responsible for most of the observed time-of-day and day-of-week volatility in the Treasury bond, Eurodollar, and Deutsche mark futures markets.

(Jones et al., 1998) examined the effect of employment and producer price index data on daily Treasury bond prices, concluding that announcement day volatility does not persist beyond the announcement day. (Li & Engle, 1998) explored the effects of macroeconomic announcements on the volatility of U.S. Treasury futures, while (Fleming & Remolona, 1999) examine the effect of public information on price formation and liquidity in the Treasury market.

(Bollerslev et al., 2000) find that the employment report, PPI, employment cost index (ECI), retail sales, and National Association of Purchasing Managers' survey have the greatest effect on the volatility of Treasury futures. Similarly, (Balduzzi et al., 2001) studied surprises in 17 public news releases of economic data. Christie-David et al. (2002) demonstrated that unanticipated macroeconomic news significantly affects interest rates of futures markets, highlighting the need to incorporate such controls in empirical models. See also: (Balduzzi & Moneta, 2017).

Recent studies, such as Amin and Tédongap (2023), have controlled for auction demand while modeling the TIPS auction cycle, but did not account for macroeconomic announcements. Both (Smales, 2021) and (Forest & Mackey, 2023) incorporated a wide range of macroeconomic announcements surprises, but did not apply automated model discovery methods.

(Gigante et al., 2024) formulated EGARCH models for Treasury market returns to identify regime-specific effects of news arrival on asymmetric volatility, but they relied on dummy variables for announcement days rather than the information contained in announcement surprises. The findings suggest that allowing for asymmetric responses to macroeconomic variables may be important to the return-generating process.

2.2. Equity Markets

One of the earliest high-frequency studies on announcement effects was conducted by (Jain, 1988), who found that money supply and CPI significantly affected stock prices, with adjustments completed within an hour. (Andersen et al., 2007) explored global financial markets' responses to U.S. macroeconomic data releases, noting that markets react differently to the same news depending on the state of the U.S. economy. During economic expansion, bad news positively impacted equity markets, while during recessions, the impact was negative. This runs contrary to what is expected in U.S. Treasury markets, where market participants generally expect a positive reaction to bad economic news as it suppresses the inflation risk premium.

(Connolly & Stivers, 2005) examined the effect of macroeconomic announcements on stock turnover and volatility clustering using a sample of daily data for 29 firms over a 15-year period. They found that volatility clustering tends to be stronger during periods of greater uncertainty, as measured by the dispersion of beliefs regarding economic announce-

ments. Increasingly, asymmetries in expectations and responses have become of interest to financial economists from both theoretical and empirical perspectives. For example, see (C. J. Campbell et al., 1999; Aktas et al., 2004; Bessembinder et al., 1996; Brockman et al., 2009).

2.3. Gets in Finance and Elsewhere

The fundamentals of the LSE/Oxford econometric approach, Gets, are articulated in Hendry (1995). The methods were automated by (Hoover & Perez, 1999), leading to the development of the PcGets software, version 9, employed in (Hendry & Krolzig, 1999, 2001; Campos et al., 2003). The algorithm has since evolved and was reformulated in the OxMetrics platform as Autometrics, as discussed in detail in (Doornik, 2009). Hendry and Doornik (2014) provided an extensive text on automated model discovery. Hendry (2024) summarized the history of the LSE/Oxford approach and the evolution of the methodology.

Although the Gets approach was initially employed in macroeconomics, it has since been adopted across a range of fields. Recently, these methods have been employed in environmental studies, such as (Pretis et al., 2016), who used Gets to detect volcanic eruptions. Likewise, (Ericsson et al., 2022) studied structural climate change using Autometrics.

Also recently, a growing number of applications have been developed in the field of finance. For instance, (Engle et al., 2012) investigated how news arrival propagates both volatility and volatility clustering. (Choi, 2013) employed Gets to study the dynamics of the market risk for value and growth stocks. (Billio et al., 2017) employed Gets to relate financial integration to international portfolio diversification. More recently, (Eijffinger & Pieterse-Bloem, 2023) applied the Gets methodology to study Eurozone government bond spreads.⁶ For example, Gómez-Puig et al. (2023) studied announcement effects under ECB regimes, but their analysis lacked a robust selection methodology. Our work extends this literature by incorporating automated model discovery, which is especially valuable when policy signals are unclear.

Collectively, this literature review suggests that the growing acceptance of automated model discovery methods offers the potential to revisit existing studies in a robust manner, which may lead to deeper empirical insights, particularly with respect to high-frequency analysis of announcement effects.⁷

3. Econometric Methodology

This study employs two distinct econometric frameworks to analyze the relationship between macroeconomic announcements and U.S. Treasury returns during the Greenspan era: (1) a traditional standard static regression (SSR) model, and (2) the general-to-specific (Gets) modeling framework implemented via Autometrics. The term static regression follows (Hendry et al., 1984) (Handbook of Econometrics, vol. II, Chapter 18), where a model with only contemporaneous regressors—no lags—is called static despite being estimated on time-series data. Both approaches are presented here, emphasizing their respective methodological strengths and limitations.

3.1. Equation (1)—Standard Static Regression Model (SSR)

The first model employed is a traditional SSR framework, regressing daily excess returns on contemporaneous standardized announcement surprises. Most studies of macro announcements in fixed-income markets have adopted some variation of this framework. It serves as a benchmark for comparison with the Gets model. The general specification for SSR is as follows:

$$R_t = \mu + \sum_{i=1}^{26} \beta_i x_{i,t} + e_t \quad (1)$$

where

R_t = is the daily excess return of the US Treasury security at time t

$x_{i,t}$ = standardized surprise in macro indicator i at time t

μ = is the intercept term

β_i = are the coefficients on the explanatory variables

$e_t \sim \text{NIID}(0, \sigma^2)$ gaussian i.i.d error term

It will be shown that the simplicity of this model comes at the cost of robustness to omitted variable bias, structural breaks, and overfitting. The SSR approach relies heavily on prior economic theory and researcher judgment for variable selection, potentially exposing it to mis-specification errors.

Although SSR models often fail specification tests and often struggle with omitted variable bias (OVB), they remain computationally efficient and widely used in empirical research. Their simplicity makes them a valuable benchmark, particularly in settings where near-costless estimation is a priority. However, automated Gets models appear to be an attractive complement as they tend to overfit less thanks to popular penalty-based shrinkage selection models, such as Lasso. See (Tibshirani, 1996, 2011; Desboulets, 2018; Epprecht et al., 2019; Becker et al., 2021; Muhammadullah et al., 2022).

3.2. Equation (2)—The General-to-Specific (Gets) Framework

The Gets modeling framework is introduced as an alternative to the SSR. The approach begins with specifying a general unrestricted model (GUM), which includes all potentially relevant variables suggested in theory. The equation takes the form of typical distributed lag models based on the dynamic model typology of dynamic models in (Hendry et al., 1984; Koyck, 1954; Almon, 1965; Dhrymes, 1971; Forest & Turner, 2013).⁸

The Gets algorithm systematically reduces the GUM by eliminating statistically insignificant variables across multiple search paths, seeking a terminal model that remains parsimonious while preserving explanatory power. Diagnostic checks for congruence, such as residual normality, autocorrelation, and heteroscedasticity, guide this reduction process. The final model is chosen from the set of terminal models by means of information criteria (SIC). A visual summary of the Gets reduction algorithm is available in Supplementary Materials, excerpted from the author's dissertation, to aid readers unfamiliar with the Gets reduction logic.

Equation (2) shows the form of the GUM used in this study. We can see that the model nests the model in Equation (1) but also allows a first-order autoregressive term (to allow for momentum or mean-reverting effects), lagged announcement surprises (which would suggest a violation of the Efficient Market Hypothesis), choices between positive and negative surprise series (to identify possible asymmetric responses), and indicator saturates (to control for outliers and level shifts).

It will be shown that, despite the extremely high dimensionality of the initial GUM, the algorithm retains a sensible model that is consistent with financial theory.

$$R_t = \alpha + \phi R_{t-1} + \sum_{i=1}^{26} \beta_i x_{i,t} + \sum_{i=1}^{26} \delta_i x_{i,t-1} + \sum_{i=1}^{26} \varphi_i x_{i,t}^+ + \sum_{i=1}^{26} \eta_i x_{i,t}^- + \sum_z \theta_i z_t + e_t \quad (2)$$

where

R_t = daily excess return of bond at time t

α = a constant (abnormal return)

ϕ = a first-order autoregressive term (persistence)

$x_{i,t}$ = standardized surprise

$x_{i,t}^+$ = positive standardized surprise

$x_{i,t}^-$ = negative standardized surprise

z_t = saturation dummy variables (IIS + SIS)

$e_t \sim NIID(0, \sigma^2)$ gaussian i.i.d error term

Standardized surprise = $\frac{(Actual - Forecast)}{\sigma_{Actual - Forecast}}$

The Autometrics functions in PcGive were used to automate the Gets procedure.⁹ This research places great emphasis on attaining a congruent model, while imposing a strict selection criteria. Model reductions were conducted with a target size of 0.01 to limit the false retention rate to one per 100 parameters.¹⁰ Mis-specification tests were also targeted at a p -value of 0.01. The tests employed are as follows:

1. AR 1-2 test is a Lagrange multiplier test for r th-order autocorrelation, based on (Godfrey, 1978);
2. ARCH test is based on (Engle, 1982);
3. Normality Test—Chi² form of (Doornik & Hansen, 2008);
4. Hetero test for heteroscedastic errors based on (White, 1980);
5. Regression Specification Error Test (RESET) based on (Ramsey, 1969).

Additionally, the algorithm will perform two additional tests before the optimal model is chosen from the set of terminal models based on Schwarz information criteria (SIC). These are as follows:

6. Chow Predictive Failure Test based on (Chow, 1960);
7. Encompassing Tests between competing terminal models as in (Sargan, 1959; Govaerts et al., 1994).

This approach is particularly interesting when the precision of the parameters is of a primary concern and when the sample is large. However, in practice, it is found that congruence is difficult to achieve with only macroeconomic announcement surprises as with candidate variables in fixed-income asset pricing regressions.

The stochastic nature of bond market returns, which is subject to random jumps and level shifts, requires a robust estimation method to achieve a reasonable level of congruence. The methods employed in this study are described next.

3.3. Indicator Saturation Techniques

Indicator saturation methods address outliers and structural breaks that are pervasive in financial time series. (Ericsson, 2017) provides a clear, non-technical discussion. He states, “IIS is a generic test for an unknown number of breaks, occurring at unknown times anywhere in the sample, with unknown duration, magnitude, and functional form”. The method involves repeated split-sample estimation of the model with dummy variables and retention of those that are significant.

SIS is conducted similarly, with steps creating $S_{it} = 1$ for $t \geq I$ and 0 otherwise. Step saturates capture the effects of level shifts in the data. The use of both IIS and SIS has been termed *supersaturation*, which is the approach employed in this study. See (Ericsson, 2017, Section 4). See also, (Castle et al., 2013, 2015; Ericsson et al., 2022). The use of indicator saturation methods in fixed-income markets is discussed in (Forest, 2018; Forest et al., 2024a, 2024b).

3.4. Comparing Models

The two approaches, SSR and Gets, are compared based on their ability to produce plausible models, achieve congruence, handle omitted variable bias, estimate stable invariant parameters, and adjust to outliers and structural shifts. While SSR relies on predefined variables and is prone to mis-specification, it is simple to apply and is computationally near-costless.

Gets demonstrates greater flexibility and robustness, particularly when analyzing high-frequency datasets subject to structural changes. However, it is computationally expensive when using supersaturation and customized tests. This study used settings that required several hours per Gets regression model, although it was found that the “quick” blocking setting in Autometrics could reduce this to just several minutes.¹¹

As an additional robustness check for potential latent nonlinear dynamics, we also estimate two-state Markov-switching models with intercept and variance regimes.¹² These models, presented in Section 5, allow us to examine whether the Gets-selected regressors provide incremental information in a nonlinear context. This comparison serves to highlight the potential complementarity between robust selection methods like Gets and flexible regime-based specifications.

4. Data

This section describes the data used to analyze the relationship between macroeconomic announcements and U.S. Treasury excess returns, covering the period from 2 January 1990 to 10 September 2001. The data are drawn from two key sources:

8. MMS Macroeconomic Announcement Dataset: This dataset includes detailed macroeconomic announcement data, capturing release timings, expectations, and as-reported actual values for key economic indicators.¹³
9. CRSP U.S. Treasuries Database: Provides daily U.S. Treasury returns and facilitates the calculation of excess returns as the difference between observed returns and the risk-free rate. This study uses daily returns data for U.S. Treasury securities.

4.1. Macroeconomic Announcement Data

The macroeconomic announcement data covers 26 important data releases that are known to influence financial markets, including:

- Employment reports (e.g., NonFarm Payrolls);
- Inflation measures (e.g., CPI, PPI);
- GDP growth;
- Consumer confidence and housing indices.

Announcements are paired with market expectations from the MMS dataset to calculate the surprise component of each release. These surprises are defined as the difference between the actual announcement value and the consensus forecast, standardized by the standard deviation of the forecast errors. This standardization ensures comparability across announcement types. Figure 1 provides graphs of the macroeconomic standardized surprise series.

Figure 1 shows the distribution of standardized surprises across the 26 macroeconomic indicators. The staggered calendar of releases leads to time-varying clustering of surprise events, which contributes to persistent return variation across trading days.

Table 1 provides descriptive data for standardized surprises in the MMS macroeconomic announcement data. This table offers insights into the accuracy of market expectations for macroeconomic variables. Overall, the data fails to reveal significant abnormalities that would suggest that the market’s reaction to any one indicator is due to the systematic inability of economists to forecast indicators.

However, the range of observations for standardized surprises offers a few interesting observations. The Index of Leading Economic Indicators and Core Producer Price Index (excluding food and energy) showed the largest negative standardized surprises at -5.22 and -4.43 , respectively. With respect to positive standardized surprises, capacity utilization is the largest at 4.19 .

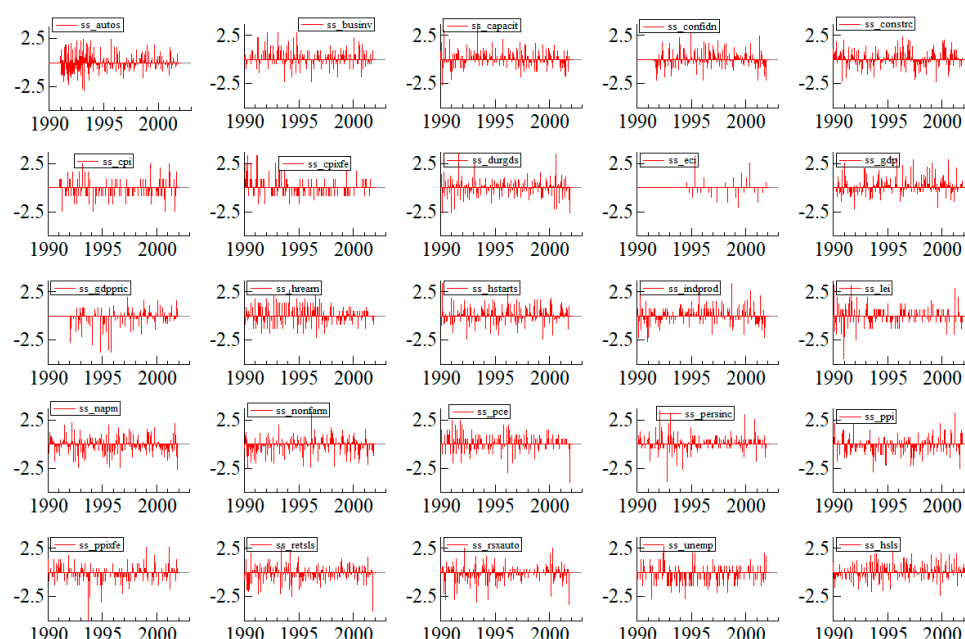


Figure 1. Standardized surprises of macroeconomic announcements. Note: The y-axis represents the standard deviation of the surprise in the individual macroeconomic indicator.

Table 1. Properties of the consensus forecasts.

Indicators	Abbreviation	# Obs.	Avg. Abs. SS	Range		Distribution	
				Min. SS	Max. SS	Skewness	Kurtosis
Auto Sales	AUTOS	199	0.78	−2.92	2.47	0.17	3.05
Business Inventories ^{ab}	BUSINV	142	0.76	−2.30	2.76	0.17	3.31
Capacity Utilization ^{ab}	CAPACIT	141	0.78	−2.69	4.19	0.37	4.32
Consumer Confidence ^{ab}	CONFIDN	130	0.76	−2.16	2.71	0.12	2.93
Construction Spending ^{ab}	CONSTRC	142	0.78	−2.29	2.33	−0.09	2.61
Consumer Price Index ^{r,ab}	CPI	130	0.74	−2.48	2.48	0.24	3.35
Core CPI (ex: food and energy) ^{ab}	CPIXFE	141	0.66	−1.68	3.36	1.07	4.46
Durable Goods Orders ^{nr,ab}	DURGDS	141	0.76	−2.60	3.46	0.13	4.51
Employment Cost Index	ECI	30	0.83	−2.03	3.04	0.50	3.41
Gross Domestic Product	GDP	138	0.77	−2.19	3.10	0.59	3.48
GDP Price Deflator	GDPPRICE	117	0.63	−3.79	1.89	−1.33	5.95
Goods and Services ^{r,nb}	GDSSERV	142	0.79	−2.40	3.79	0.29	3.78
Hourly Earnings ^{ab}	HREARN	142	0.82	−2.22	2.66	0.02	2.39
Home Sales ^{nb}	HSLS	141	0.81	−2.42	2.19	−0.04	2.69
Housing Starts ^{r,ab}	HSTARTS	142	0.80	−2.42	3.41	0.17	3.05
Industrial Production ^{nr,ab}	INDPROD	142	0.77	−2.62	3.37	0.06	3.69
Index of Lead. Econ. Ind. ^{nr,ab}	LEI	142	0.72	−4.43	3.16	−0.22	5.68
Nat. Assoc. of Purch. Mgrs. ^{ab}	NAPM	142	0.80	−2.65	2.25	−0.05	2.81
Nonfarm Payrolls ^{r,nb}	NONFARM	143	0.77	−2.53	3.30	0.01	3.43
Personal Cons. Expenditures ^{ab}	PCE	140	0.77	−3.96	2.48	−0.65	4.75
Personal Income ^{r,nb}	PERSINC	141	0.69	−3.90	3.47	0.06	5.93
Producer Price Index ^{nr,ab}	PPI	143	0.77	−2.88	3.24	0.20	3.60
Core PPI (ex: food and energy) ^{ab}	PPIXFE	143	0.71	−5.22	2.61	−0.89	7.38
Retail Sales ^{nr,ab}	RETSLS	142	0.78	−4.02	2.68	−0.35	4.07
Retail Sales (ex: Auto Sales) ^{ab}	RSXAUTO	142	0.72	−3.36	2.52	−0.52	4.40
Unemployment Rate ^{r,nb}	UNEMP	143	0.76	−2.72	2.72	0.16	3.04

Note: SS indicates a standardized surprise based on the standard deviation of the forecast surprise. Descriptives of the as-reported economic indicators announced from 1/1990 to 11/2001. The superscript ^r indicates rational expectations, ^{nr} indicates nonrationality, ^{ab} indicates anchoring bias, and ^{nb} indicates no anchoring bias. These findings are based on (Aggarwal et al., 1995; Schirm, 2003; S. D. Campbell & Sharpe, 2009; Hess & Orbe, 2013; Forest & Doronkina, 2025).

In the last two columns, we see evidence of non-normality in the distributions of standardized surprises. Notably, we see that there is notable negative skewness for the GDP Deflator and positive skewness for Core CPI. Likewise, most economic indicators demonstrate excess kurtosis in standardized surprises during the sample period. Again, these appear in the inflation indicators, especially for Core PPI with a kurtosis of 7.38. This is reflective of the period, which was marked by notable inflation concerns, despite it proving to be well controlled. It is also not surprising, given the purposeful obfuscation of the critical driving factors to by policy Fed Chair Greenspan's public testimonies.

The MMS data have been rigorously tested for violations of the rational expectations hypothesis (REH) in (Aggarwal et al., 1995; Schirm, 2003). These include stationarity of forecasts and actuals, cointegration tests for nonstationary cases, tests of forecast bias, and more recent tests of anchoring bias by (S. D. Campbell & Sharpe, 2009; Hess & Orbe, 2013).

Specific deviations from rationality have been detected in Durable Goods Orders, Industrial Production and the Producer Price Index (PPI). Anchoring bias, although potentially rational in the case of an asymmetric loss function, has been found to be widespread across most macro announcements. Table 1 footnotes show available test results, while Appendix A.2 provides greater detail. It should be noted that irrationality of forecasts and various biases complicate the task of interpreting macroeconomic announcements.¹⁴

4.2. Treasury Return Data

Daily excess returns at various maturities of U.S. Treasury securities were obtained from the CRSP database. In financial studies, excess returns are computed by subtracting the risk-free rate (6-month T-bill yield) from the observed return of the respective Treasury security. Focusing on excess returns isolates the component of Treasury performance attributable to macroeconomic factors rather than baseline risk-free rates and aligns our specification with traditional asset pricing models in finance. This study employs both 30-year bonds and 10-year notes in this study and considers the auction cycle.

The most recently issued security at a given maturity is considered the "on-the-run" issue (hereafter OTR), while the first "off-the-run" (FTR) issue refers to the second-most recently issued security. Bond market participants typically show a marked preference for more liquid OTR issues. This phenomenon, called the *bond/old bond spread*, was explored in detail by (Krishnamurthy, 2002).¹⁵ It has also been well documented as a factor in the famous failure of the hedge fund Long Term Capital Management. See (Krishnamurthy, 2002, p. 464).

The sample ends on 10 September 2001, the last full trading day before the 9/11 terrorist attacks. Appendix A.1 displays the evolution of Treasury yields and the bond–bill term spread over the sample period. This cutoff ensures that the analysis reflects a consistent monetary policy regime under Chairman Greenspan, prior to the onset of post-crisis liquidity interventions and wartime policy adjustments. Although Greenspan remained Fed Chair until 2006, the announcement dynamics in the immediate post-9/11 period reflected a fundamentally different regime.

The quality and consistency of the datasets provide a sturdy foundation for evaluating the econometric models outlined in Section 3 and supports the empirical findings discussed next in Section 5.

5. Empirical Results

This section presents the empirical results of the analysis, focusing on the application of the standard static regression (SSR) models and their comparison to the general-to-specific (Gets) modeling results. The discussion is structured as follows: First, the selected models are described (Section 5.1), comparing the SSR and Gets approaches, and focusing

on key insights from the variables selected. Next, tests for mis-specification are evaluated (Section 5.2). The stability of the selected models is then examined (Section 5.3). A dynamic analysis of market efficiency is discussed (Section 5.4). The relative information loss between models is further examined through encompassing tests (Section 5.4), followed by an analysis of model selection bias and its corrections (Section 5.5).

This systematic evaluation demonstrates the complementary benefits of the Gets framework in achieving parsimonious, congruent, and encompassing models, particularly in the high-frequency, high-dimensional setting of U.S. Treasury returns.

5.1. Selected Models

The Gets models in Table 2 offer a significant improvement over the SSR models by removing unstable or unimportant variables while retaining key predictors across different securities and maturities. Importantly, the Gets approach consistently selects similar sets of variables for 30-year bonds and 10-year notes for both on-the-run (OTR) and first-off-the-run (FTR) securities. This consistency shows the robustness of the methodology in identifying the most economically relevant variables. Aligning with the Efficient Market Hypothesis (EMH), all lagged dependent variables are reduced out of the Gets models. This finding highlights the rapid incorporation of macroeconomic surprises into Treasury returns and affirms the efficiency of these markets.

Additionally, the Gets models do not yield overly complex specifications. Aside from the inclusion of numerous indicator saturation dummies to address structural breaks and outliers, the retained regressors are limited, focusing on stable and economically meaningful predictors.

The retained regressors include variables that were found to be both rational and irrational. For example, Durable Goods Orders were found to be irrational during this same period but were retained. Notably, Nonfarm Payrolls (found to be rational) were also highly significant and retained in all cases. These were focal points of Chairman Greenspan's policy during this era.

One particularly salient result is the role of the Employment Cost Index (ECI), which is the only macroeconomic announcement showing clear evidence of asymmetric effects on Treasury returns. This asymmetry, observed across both bonds and notes, suggests that ECI held a unique importance in shaping time-varying risk premia when it was first introduced during the Greenspan era. This aligns with broader inflationary concerns at the time, particularly those stemming from labor market pressures, which were a hallmark of the period.

In comparing the relative size and significance of regressors across models, Gets models consistently show fewer but more impactful predictors. The inclusion of indicator saturation dummies dramatically improves model fit, with substantial increases in adjusted R-squared. Approximately 90 saturates are retained in each model, striking a balance between flexibility and parsimony.

Given the dataset's size—over 2900 observations—this level of saturation does not lead to overfitting but rather enhances the robustness of the results. These findings emphasize the utility of Gets in refining model accuracy while preserving economic interpretability.

Table 2. 30-Year bond estimation results. This table estimates Equation (1), the SSR model, and Equation (2), the Gets model with saturations, for 30-year OTR and FTR U.S. Treasury securities.

Indicators	Abbreviation	30-Year On-the-Run		30-Year 1st Off-the-Run	
		SSR Coefficient	GETS Coefficient	SSR Coefficient	GETS Coefficient
Constant	C	0.105 *		0.116 **	
Auto Sales	SS_AUTOS	−0.266		−0.253	
Business Inventories	SS_BUSINV	0.051		0.077	
Capacity Utilization	SS_CAPACIT	−0.474		−0.498	
Consumer Confidence	SS_CONFIDN	−0.848 **	−0.986 **	−0.816 **	−0.900 **
Construction Spending	SS_CONSTRC	−0.058		−0.041	
Consumer Price Index	SS_CPI	−0.247		−0.253	
Core CPI (ex: food and energy)	SS_CPIXFE	−0.664 **	−0.738 **	−0.657 **	−0.700 **
Durable Goods Orders	SS_DURGDS	−0.680 **	−0.529 **	−0.675 **	−0.617 **
Employment Cost Index	SS_ECI	−1.174 **		−1.156 **	−2.557 **
Gross Domestic Product	SS_GDP	−0.122		−0.102	
GDP Price Deflator	SS_GDPPRIC	−0.393		−0.344	
Goods and Services	SS_GDSSERV	0.030		0.050	
Hourly Earnings	SS_HREARN	−0.886 **	−0.662 **	−0.864 **	−0.780 **
Home Sales	SS_HSLS	−0.715 **	−0.752 **	−0.737 **	−0.856 **
Housing Starts	SS_HSTARTS	−0.057		−0.030	
Industrial Production	SS_INDPROD	−0.021		0.087	
Index of Lead. Econ. Ind.	SS_LEI	−0.139		−0.160	
Nat. Assoc. of Purch. Mgrs.	SS_NAPM	−1.138 **	−1.355 **	−1.085 **	−1.287 **
Nonfarm Payrolls	SS_NONFARM	−1.438 **	−1.243 **	−1.467 **	−1.063 **
Personal Cons. Expenditures	SS_PCE	−0.038		−0.057	
Personal Income	SS_PERSINC	−0.318		−0.320	
Producer Price Index	SS_PPI	0.229		0.228	
Core PPI (ex: food and energy)	SS_PPIXFE	−0.582 **		−0.622 **	
Retail Sales	SS_RETSLs	−0.697 *		−0.707 **	
Retail Sales (ex: Auto Sales)	SS_RSXAUTO	0.085		0.090	
Unemployment Rate	SS_UNEMP	0.244		0.262	
Negative ECI	NEG_SS_ECI		−2.726 **		
Positive ECI	POS_SS_ECI				2.311 **
sigma		2.26	1.98	2.22	1.94
log-likelihood		−6523.74	−6090.81	−6468.16	−6031.68
#. of observations		2928	2927	2927	2927
RSS		14,769.32	11,000.149	14,235.678	10,564.589
#. of parameters		27	113	28	115
Adj. R ²		0.064	0.309	0.065	0.313

Note: SS indicates standardized surprise. **, * indicate significance at 0.01 and 0.05, respectively. Bias correction based on code courtesy of Hendry, Doornik, and Castle.

5.2. Diagnostic Tests

The diagnostic tests demonstrate the Gets framework's ability to address critical model specification issues and produce models that approach congruence. While the specific tests are outlined in Section 3.2, this section focuses on the implications of Gets vs. SSR models.

The Gets models consistently passed key diagnostic tests, including assessments for autocorrelation, heteroskedasticity, functional form, and normality. In contrast, the SSR models frequently failed multiple tests, particularly in handling normality, heteroskedasticity, and (as we will see in Section 5.3) parameter constancy. Mis-specification tests for the 10-year notes were particularly concerning results across four out of five tests.

The ability of indicator saturation techniques contributes to the refined congruence results. By addressing structural breaks and outliers, saturated models may achieve a

higher level of robustness that is crucial for reliable inference in finance. The results suggest that the residuals of the Gets models are well behaved, meeting the assumptions required for valid statistical inference.

Table 3 provides a summary of the diagnostic results for both static regression (SSR) and Gets models, detailing the improvement in model congruence achieved by the latter. It should be noted that the automated Gets procedure with supersaturation proved successful at achieving a primary goal of the London School of Economics Approach (hereafter the LSE/Oxford Approach)—i.e., to minimize mis-specification to at most a single violation.¹⁶

Table 3. Mis-specification test results.

Panel A. 30-Year Bond.	30-Year OTR Bonds			30-Year FTR Bonds		
	SSR		GETS	SSR		GETS
Congruent	No		Yes	No		Yes
AR 1-2 test	0.5138		0.7579	0.8790		0.9120
ARCH 1-1 test	0.0000	**	0.4069	0.0000	**	0.3816
Normality test	0.0000	**	0.3021	0.0000	**	0.9165
Hetero test	0.0058	**	0.1181	0.0004	**	0.0428
RESET23 test	0.2951		0.2919	0.1642		0.3003
Panel B. 10-Year Note.	10-Year OTR Notes			10-Year FTR Notes		
	SSR		GETS	SSR		GETS
Congruent	No		Yes	No		Yes
AR 1-2 test	0.0093	**	0.1084	0.0125	*	0.1096
ARCH 1-1 test	0.0000	**	0.8413	0.0000	**	0.4914
Normality test	0.0000	**	0.7605	0.0000	**	0.8567
Hetero test	0.0002	**	0.0101	0.0034	**	0.0213
RESET23 test	0.0657		0.2033	0.0911		0.1548

Note: ** and * indicate significance at 0.01 and 0.05, respectively. All tests are F-tests except for the Hetero test, which is the Chi² test of (Doornik & Hansen, 2008). Shaded indicates FTR securities.

While some heteroscedasticity remains, heteroscedasticity-consistent standard errors can be computed. Proponents of the LSE/Oxford Approach have long advocated employing such devices *only in cases* of a single mis-specification test violation, as multiple violations seen in the SSR regressions are seen as a larger problem and are suggestive of omitted variable bias (OVB).

Figure 2 provides a complementary visual comparison of model performance for 30-year OTR bonds. The top panel displays actual versus fitted values, with the static regression (SSR) results on the left and the saturated Gets results on the right. The Gets model exhibits a closer alignment between actual and fitted values, suggesting better model performance and the ability to capture the relevant features of the underlying local data-generating process.

The bottom panels display QQ plots, describing the normality of the residuals. The SSR model deviates significantly from normality, particularly in the tails, as shown in the left QQ plot. This result would be concerning to a practitioner relying on accurate assessments of large negative returns—or downside risk.

In contrast, the Gets model (right QQ plot) aligns more closely with the theoretical normal distribution. These visual comparisons further imply enhancements achieved by employing automated Gets. The troubling left-tail behavior of the SSR models will be revisited in Section 6.

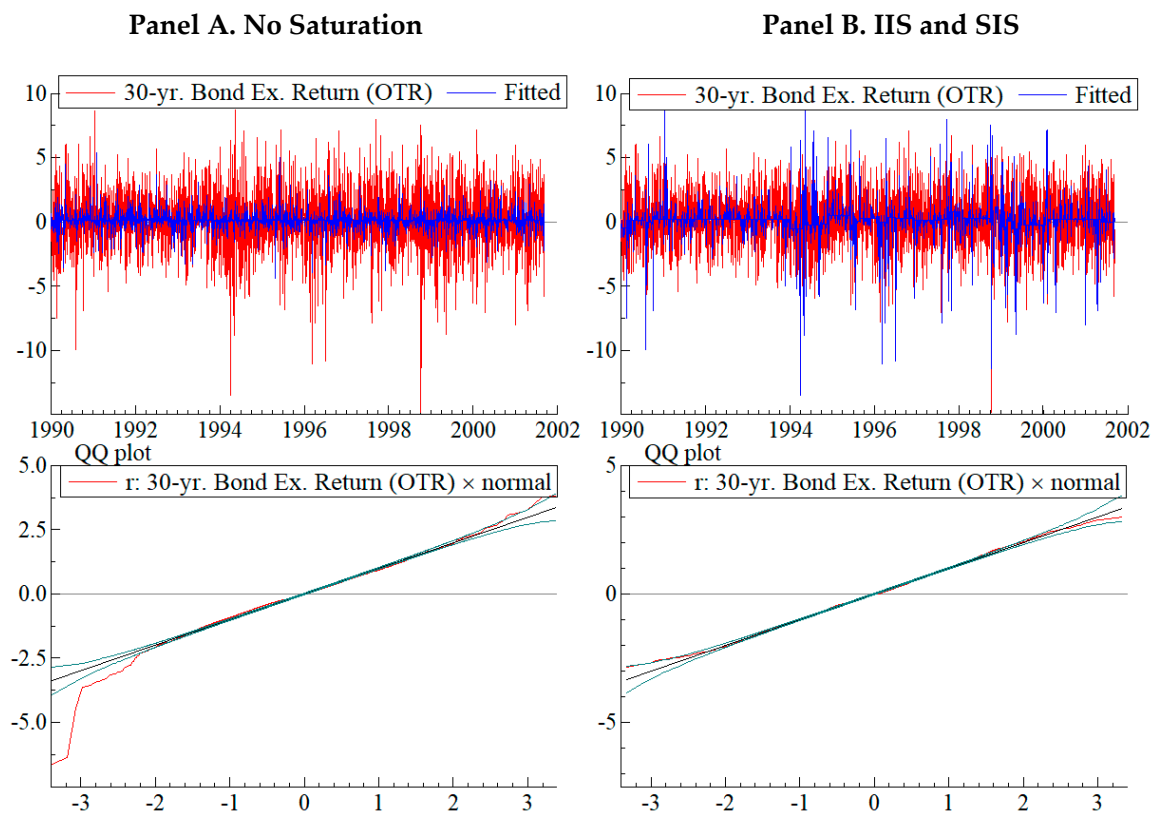


Figure 2. Graphical analysis—SSR vs. saturated Gets (30-year OTR).

5.3. Model and Parameter Stability

The stability of model parameters is also crucial for the reliability of econometric models, particularly in financial applications where structural breaks and market volatility are common. Figure 3 illustrates the behavior of the 30-year OTR bond model coefficients in the context of expanding window-recursive estimates in the SSR regressions. At each point, the parameter estimate is provided based on the data sample since the beginning of the sample.

The set of graphs highlight the sign-switching behavior of certain regressors in SSR regressions, where the coefficients alternate between positive and negative values. This instability is particularly evident for four macroeconomic surprises, with frequent sign changes undermining the reliability of the estimated parameters.

While these were widely followed economic benchmarks, the market did not respond consistently across the sample. The parameters on these variables evolved during the overall sample, possibly due to the opacity of the Fed's operating procedures and the Fed Chair's public comments, but perhaps also simply as a function of the state of the economy or in the existence of collinearity.

The GDP deflator's importance to market participants was likely affected by the introduction of ECI in the early half of the sample. Importantly, collinearity should be considered for these variables. Correlation analysis shows that, while most macroeconomic announcements are orthogonal, there are several key pairs of contemporaneously released announcements that have high correlations.¹⁷

For example, there are three cases where correlations are both significant and large. Personal Consumption Expenditures (PCEs) and Industrial Production (IP) had a correlation coefficient of 71%. Gets removes both variables.

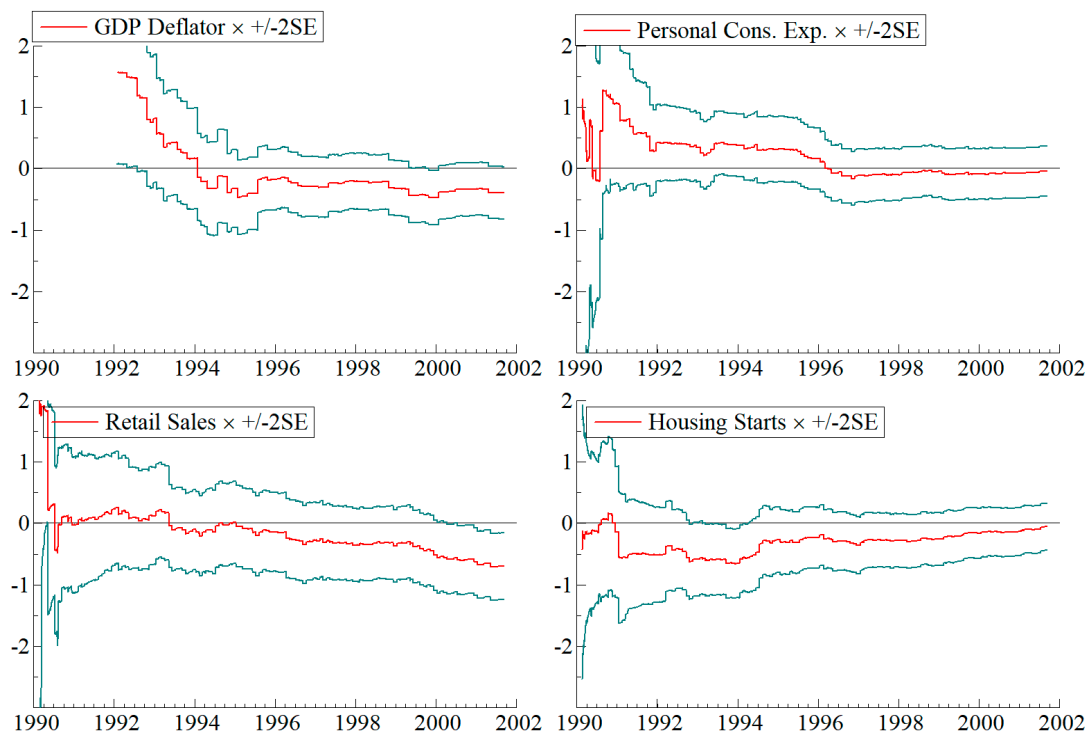


Figure 3. Recursive parameters—Sign-switching SSR parameters.

With respect to the recursive parameter estimates, PCE alternated between a positive coefficient, which was nearly significant at a loose level, and became negative and nearly zero roughly halfway through the sample. Likewise, retail sales were highly insignificant and sometimes positive in the early part of the 90s but became substantially negative and significant at the end of the decade.

This announcement, however, is released at the same time as a competing measurement of core retail sales, which excludes auto sales.¹⁸ The correlation between these two variables is 67% and, again, Gets reduces both measures out.¹⁹ This is likely due to the Chow stability test that is part of the Autometrics reduction. Therefore, it appears that there may be unmodelled state dependence for this variable.²⁰

Similarly, CPI and PPI are also announced with their core components and show correlations with these subcomponents, which remove volatile food and energy prices. The correlations of these measures are 55% and 49%, respectively. Here, however, the Gets models consistently favor the core rates, which are retained.

These results are complemented by Appendix A.6, which presents model and parameter stability tests of (Hansen, 1992) for both SSR and the Gets model while excluding saturates. While the SSR regressions are shown to contain several unstable parameters in each regression, such as Housing Starts and Core PPI, the parameters retained under Gets estimation are all tested as invariant, even when the saturates are excluded.

By contrast, we could look at Appendix A.5, which provides graphs of selected common coefficients in both the SSR and Gets regressions.²¹ Results show that common coefficients are quite stable in either setting. Thus, it appears that Gets improves on SSR in this setting by vetting out troublesome unstable parameters but retaining those that are both significant and invariant.

This ability to avoid overfitting seems particularly advantageous in financial settings. It also tends to dispense with one or both contemporaneous covariates, suggesting the more dominant of the competing measures is retained.²² This appears consistent with the discussion of collinearity in (Granger & Hendry, 2005).²³

5.4. Efficient Markets—Momentum and Mean Reversion

Figure 4 examines the dynamic behavior of a lagged dependent variable coefficient in the 30-year OTR bond model using expanding window recursive graphs with parameter estimates plotted alongside ± 2 standard error bands. Both Equation (1), the SSR, and Equation (2), supersaturated Gets, were re-estimated with an autoregressive term.²⁴

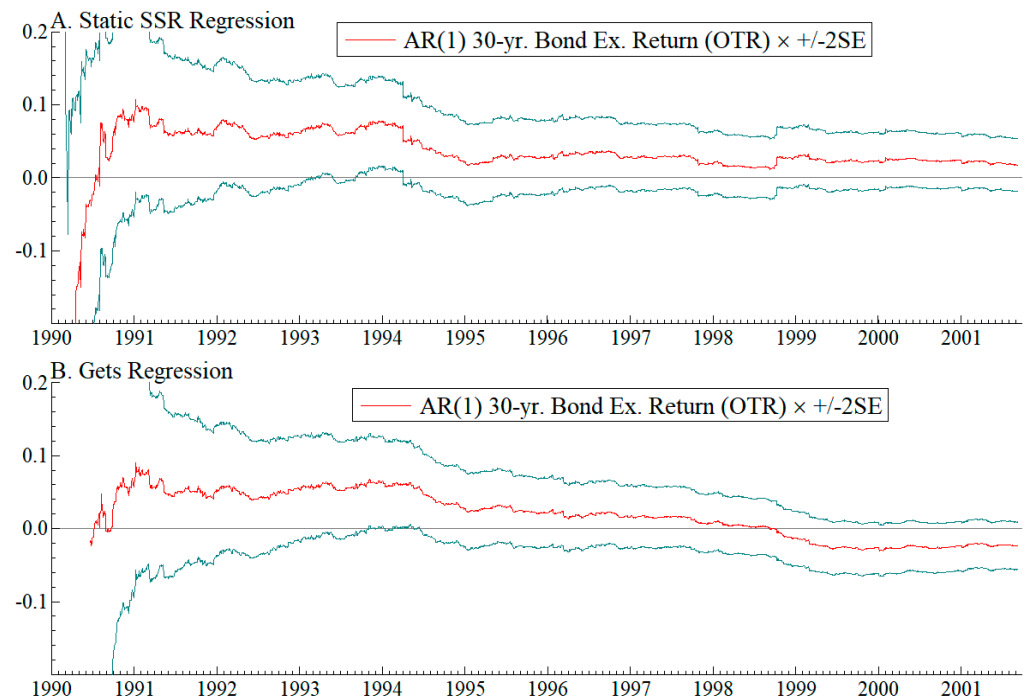


Figure 4. Graphical analysis—Dynamic analysis of market efficiency.

The AR(1) parameter is crucial in asset pricing models for evaluating market efficiency: a positive AR(1) indicates momentum, while a negative AR(1) suggests mean-reverting behavior. Neither should be significant under weak-form efficiency, as past returns should not be a reliable predictor of future returns. The significance of the coefficient is determined by whether both error bands are positive or negative up to a given point. In other words, whether the interval between them contains zero.

The SSR regression suggests near-significant to significant momentum in the early sample, with the AR(1) parameter being often positive but typically insignificant, as indicated by the error bands. Some significance or near-significance is observed in 1993–1994, corresponding to heightened market activity during the “Great Bond Massacre” bear market of 1994, when long-term yields rose from roughly 6 percent to more than 8 percent. While the AR(1) parameter lacks statistical significance across most of the samples, there is no suggestion of mean reversion in the SSR model.

The Gets regression, however, shows a fascinating sign change for this parameter. Later in the sample, the AR(1) coefficient switches from positive to negative, suggesting a shift from momentum to near-significant mean reversion in the Treasury market, starting roughly at the point of the famous (LTCM) Crisis. This analysis fails to reject market efficiency but also suggests a fascinating disagreement between the SSR and the automated, supersaturated Gets model. The latter’s ability to capture dynamic changes in market efficiency that may occur during times of financial crisis and may deserve more attention.²⁵

To illustrate how Gets modeling can complement nonlinear approaches in financial econometrics, Figure 5 presents an example of dynamic intercept behavior. The top panel shows the evolution of the abnormal return (alpha) from the Gets model, based on Au-

tometrics reduction. The middle and bottom panels display two-state Markov-switching models—with and without the Gets-selected regressors—applied to the same excess return series. While the Markov models capture regime-dependent variation in mean and volatility, the Gets framework offers a clearer insight into the structure of alpha under congruence. The comparison highlights how information selected through Gets may refine or inform switching specifications, underscoring the value of combining robust model selection with flexible state-dependent frameworks.

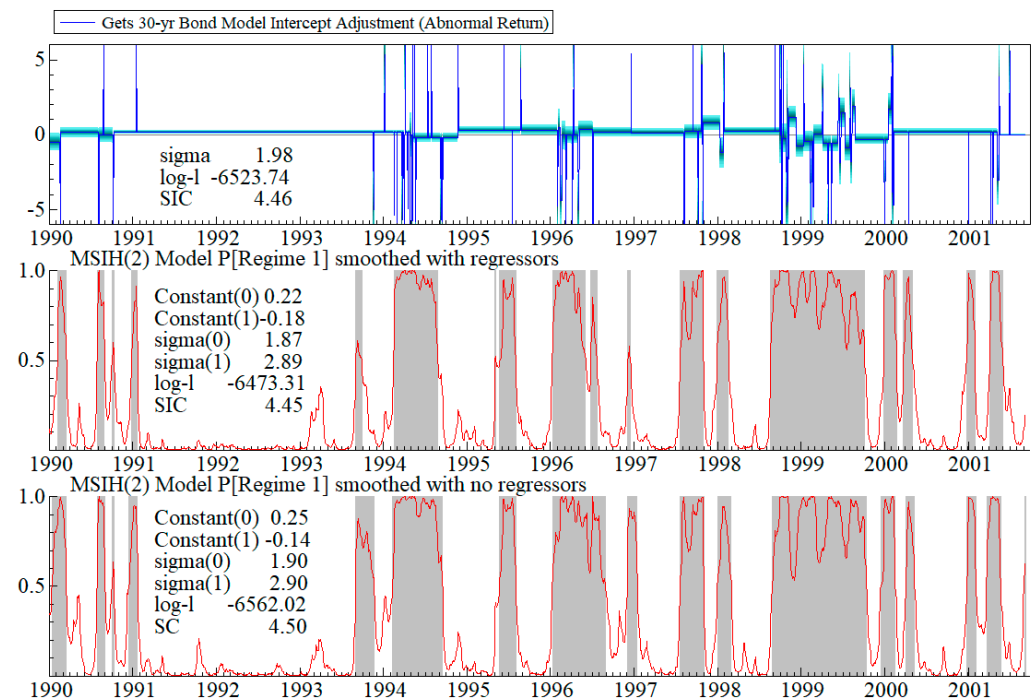


Figure 5. Graphical analysis—Dynamic analysis of abnormal return and variance regimes.

5.5. Corrections for Model Selection Bias

A point of contention for critics of Gets model discovery methods is that parameter estimates are not perfectly unbiased. This selection bias is verifiable. Notwithstanding, advocates of the Gets approach argue that estimates are *approximately* unbiased, as the selection bias is well understood and easily corrected using a routine bias adjustment procedure. See (Hendry & Doornik, 2014, Chapter 10). The correction process is described in great detail in (Hendry & Krolzig, 2005).

Because of sampling, some relevant variables will likely have $t^2 < C_\alpha^2$ in a particular sample (where C is the critical value at a given alpha). Conditional estimates will be biased away from the origin as variables are based on the condition $t^2 < C_\alpha^2$. By chance, approximately $\alpha(N - n)$ irrelevant variables will be retained due to adventitiously significant $t^2 < C_\alpha^2$.

However, as shown in (Hendry & Krolzig, 2005), bias correction will achieve *approximate* unbiasedness of the relevant variables while driving the coefficients on the irrelevant variables to zero. The two-step bias correction procedure can be applied to parameter estimates and requires only the estimated parameters, t-statistics, sample size, and significance level from the Gets estimation.²⁶

When compared to the effects of the approximate omitted variable bias, which is likely to plague an empirical model that does not adequately represent the local data-generating process, this study suggests that the small and manageable bias of the Gets estimates poses a more-acceptable risk level.

Table 4 provides a sense of the degree of the Gets estimate selection bias versus estimated OVB in the SSR. This analysis draws heavily from (Hendry & Doornik, 2014). The results are favorable to the Gets model, while the SSR model suffers in terms of bias relative to the uncorrected Gets model coefficients (columns D. vs. E.).

Table 4. Analysis of approximate bias in common coefficients—30-year bond and 10-year note. Columns A and B present parameter estimates from the SSR and Gets models, respectively. Column C shows the two-pass, bias-corrected coefficients. Columns D and E show the approximate bias as a percentage of the bias-corrected coefficients.

Panel A. 30-Year Bond.										
	30-Year OTR					30-Year FTR				
	A. SSR	B. Gets	C. Gets Bias Corr.	D. Gets Bias %	E. SSR OV Bias %	A. SSR	B. Gets	C. Gets Bias Corr.	D. Gets Bias %	E. SSR OV Bias %
SS_CONFIDN	−0.85	−0.99	−0.98	1.0%	−13.3%	−0.82	−0.90	−0.90	0.0%	−8.9%
SS_CPIXFE	−0.66	−0.74	−0.72	2.8%	−8.3%	−0.66	−0.70	−0.67	4.5%	−1.5%
SS_DURGDS	−0.68	−0.53	−0.39	35.9%	74.4%	−0.68	−0.62	−0.56	10.7%	21.4%
SS_HREARN	−0.89	−0.66	−0.62	6.5%	43.5%	−0.86	−0.78	−0.77	1.3%	11.7%
SS_HSLs	−0.72	−0.75	−0.73	2.7%	−1.4%	−0.74	−0.86	−0.85	1.2%	−12.9%
SS_NAPM	−1.14	−1.36	−1.36	0.0%	−16.2%	−1.09	−1.29	−1.29	0.0%	−15.5%
SS_NONFARM	−1.44	−1.24	−1.24	0.0%	16.1%	−1.47	−1.06	−1.06	0.0%	38.7%
Panel B. 10-Year Note.										
	10-Year OTR					10-Year FTR				
	A. SSR	B. Gets	C. Gets Bias Corr.	D. Gets Bias %	E. SSR OV Bias %	A. SSR	B. Gets	C. Gets Bias Corr.	D. Gets Bias %	E. SSR OV Bias %
SS_CONFIDN	−0.65	−0.63	−0.63	0.0%	3.2%	−0.63	−0.69	−0.69	0.0%	−8.7%
SS_CPIXFE	−0.43	−0.57	−0.57	0.0%	−24.6%	−0.39	−0.58	−0.58	0.0%	−32.8%
SS_DURGDS	−0.42	−0.39	−0.36	8.3%	16.7%	−0.41	−0.40	−0.42	−4.8%	−2.4%
SS_HREARN	−0.58	−0.63	−0.63	0.0%	−7.9%	−0.56	−0.41	−0.38	7.9%	47.4%
SS_HSLs	−0.52	−0.49	−0.47	4.3%	10.6%	−0.55	−0.61	−0.61	0.0%	−9.8%
SS_NAPM	−0.82	−0.88	−0.88	0.0%	−6.8%	−0.77	−0.77	−0.77	0.0%	0.0%
SS_NONFARM	−1.07	−0.77	−0.77	0.0%	39.0%	−1.03	−0.80	−0.80	0.0%	28.8%

Note: ss = standardized surprise. Bias correction based on code, courtesy of Hendry, Doornik, and Castle. p -value = 0.01. OV bias % = (SSR coefficient − Gets 2-step bias-corrected coefficient)/Gets 2-step bias-corrected coefficient. Gets bias % = (unadjusted Gets coefficient − Gets 2-step bias-corrected coefficient)/Gets 2-step bias-corrected coefficient. Shading indicates the FTR securities.

Gets bias is low in absolute terms, often producing estimates with no *approximate* bias at all, if we consider the corrected Gets as *approximately* unbiased. When model selection bias exists in the uncorrected coefficients, the size of the estimated bias tends to be less than half that of the suggested omitted variable bias in SSR regressions.

5.6. Encompassing Tests

A cornerstone of the Gets approach to econometric modeling is the idea of encompassing and advocating a progressive research strategy. See (Mizon & Richard, 1986; Chong & Hendry, 1986; Govaerts et al., 1994; Bontemps & Mizon, 2003, 2008; Doornik, 2008; Ericsson, 2008). The concept of encompassing is that a model should be able to explain the results of a competing model. This concept is discussed in Chapter 14 of (Hendry, 1995).

When two or more explanations compete in describing a phenomenon, one or more must be incorrect. This is because models are simply reductions in the data-generating process, and therefore, they are reduced re-combinations of the data. If a model, M_1 , purports to explain the data, then it should be able to explain re-combinations of the data that rival models seek to explain.

This study treats the SSR model as M_1 and the Gets model as M_2 and performs formal encompassing tests on whether M_1 encompasses M_2 and, conversely, whether M_2 encompasses M_1 . In terms of notation, the epsilon symbol, ϵ , is used to indicate

encompassing—i.e., this study tests $M_1 \in M_2$ and $M_2 \in M_1$, respectively (Hendry, 1995, p. 502).

This research suggests that estimates produced by SSR models may be improved, and that automated Gets models offer a statistically admissible addition, especially when applying bias correction. It can also be seen that the 10-year OTR Gets model can explain the results of the typical SSR model. Furthermore, it shows that the SSR model fails to encompass all the final Gets models.

Gets modeling employs encompassing tests in the model reduction process; specifically, Autometrics tests whether the reduced model encompasses the nesting general unrestricted model (GUM). However, encompassing testing of non-nested models is also a valuable tool. In the SSR and Gets models, M1 and M2 are mutually non-nested.

The output shown in Table 5 suggest that the results favor the Gets models. The test of Sargan (1959) checks whether the restricted reduced form of a structural model encompasses the unrestricted reduced form, including exogenous regressors from the rival models. The second test is a Joint Model F-test, which checks whether each model parsimoniously encompasses the linear nesting model. Both the Sargan Test and the Joint Model F-test suggest that only the 30-year OTR M2 fails to encompass M1 at 1% significance. Both the 30- and 10-year FTR models reject at 5% significance, though the test statistics and sigmas of the models improve with the Gets models and approach that of the Joint Model. Overall, the Gets models suggest a reduction in information loss relative to that of the SSR.

Table 5. Encompassing Tests. The (Sargan, 1959) restricted/unrestricted reduced-form χ^2 test checks whether the reduced form of a structural model encompasses the unrestricted reduced form, including exogenous regressors from rival models. The Joint Model F-test checks whether each model parsimoniously encompasses the linear nesting model. Model 1 is the static SSR regression and Model 2 is the final supersaturated Gets model.

Test	Model 1 vs. Model 2	Model 2 vs. Model 1
30-Year OTR Bonds		
Sargan	$\chi^2(106) = 771.62 [0.0000]^{**}$	$\chi^2(20) = 41.113 [0.0036]^{**}$
Joint Model	$F(106,2794) = 9.556 [0.0000]^{**}$	$F(20,2794) = 2.0713 [0.0035]^{**}$
sigma(M1) = 2.25673	sigma(M2) = 1.97714	sigma(Joint) = 1.96965
30-Year FTR Bonds		
Sargan	$\chi^2(107) = 774.99 [0.0000]^{**}$	$\chi^2(19) = 35.361 [0.0126]^*$
Joint Model	$F(107,2793) = 9.5197 [0.0000]^{**}$	$F(19,2793) = 1.8721 [0.0123]^*$
Sigma(M1) = 2.21563	sigma(M2) = 1.93829	sigma(Joint) = 1.9326
10-Year OTR Notes		
Sargan	$\chi^2(152) = 1029 [0.0000]^{**}$	$\chi^2(20) = 27.436 [0.1234]$
Joint Model	$F(152,2748) = 9.9429 [0.0000]^{**}$	$F(20,2748) = 1.3755 [0.1227]$
Sigma(M1) = 1.47236	sigma(M2) = 1.21655	sigma(Joint) = 1.2149
10-Year FTR Notes		
Sargan	$\chi^2(154) = 1038.1 [0.0000]^{**}$	$\chi^2(19) = 32.957 [0.0243]^*$
Joint Model	$F(154,2746) = 9.9414 [0.0000]^{**}$	$F(19,2746) = 1.7435 [0.0239]^*$
Sigma(M1) = 1.41527	sigma(M2) = 1.16836	sigma(Joint) = 1.16539

Note: * Significant at 5%, ** significant at 1%.

6. Remarks

The empirical findings demonstrate several relative strengths of the Gets methodology in modeling U.S. Treasury responses to macroeconomic announcements. These may be computationally expensive but financially beneficial. First, greater congruence is achieved as the Gets models consistently pass key specification tests, whereas traditional static

regressions suffer from heteroskedasticity, autocorrelation, and non-normality. This might allow the modeler to have greater faith in actionable parameter estimates.

Second, stability is enhanced, as Gets eliminates unstable parameters, leading to models with greater parameter constancy over time. This information would be particularly useful to risk managers who seek to immunize financial positions from announcement risk. Third, the Gets models exhibit stronger encompassing properties, suggesting minimal information loss relative to static models, which often retain spurious relationships while omitting key determinants. Additionally, bias reduction is suggested, and the selection bias in uncorrected Gets is often zero or smaller than the approximate OVB seen in the conventional approach.

A particularly noteworthy result is that the Gets models excluded all lagged macroeconomic variables and removed the autoregressive (AR1) term, reinforcing the validity of the Efficient Markets Hypothesis (EMH). This outcome suggests that Treasury markets rapidly incorporate new macroeconomic information, leaving little room for predictable autocorrelation in excess return. The ability of Gets to generate parsimonious yet encompassing models, while preserving theoretical consistency with EMH, implies its value as a robust econometric tool for financial market analysis.²⁷

The scenario analysis in Appendix A.7 quantifies the estimated impact of positive three-sigma macroeconomic surprises in key indicators on U.S. Treasury securities with maturities greater than 20 years. The analysis is based on a modern average trade of just under five million U.S. dollars on bond maturities greater than 20 years—i.e., 30-year bonds auctioned in the prior decade. Four scenarios consider different combinations of downside shocks (positive surprises) to key economic indicators—Nonfarm Payrolls, Hourly Earnings, and the Employment Cost Index (ECI)—and compare the estimated post-trade transaction value under the Gets and SSR models. The results are summarized as follows:

- Nonfarm Payrolls Shock (Scenario 1): The Gets model estimates a smaller negative price impact (−0.95 vs. −1.17 for SSR), leading to a \$10,599 lower estimated trading loss for a typical transaction.
- Hourly Earnings Shock (Scenario 2): Again, the SSR model exhibits excessive sensitivity (−0.75 vs. −0.20 in Gets), resulting in a \$26,800 larger estimated loss than the Gets model.
- Simultaneous Shocks to both Nonfarm Payrolls and Hourly Earnings (Scenario 3): The compounded impact is significantly smaller under Gets (−1.26 vs. −1.90 in SSR), translating to \$31,173 less trading loss.
- Employment Cost Index (ECI) Shock (Scenario 4): Interestingly, the Gets model estimates a larger impact (−2.36 vs. −0.93 in SSR), suggesting that ECI shocks drive a more pronounced repricing under Gets. This results in an estimated \$69,073 greater trading loss if Gets is more accurate, implying that traditional SSR models may have underestimated the market's reaction to labor cost shocks.

These results offer three key takeaways. First, Gets models demonstrated lower sensitivity to three-sigma shocks in most cases, mitigating the potential overreaction present in static models. Again, this finding may draw the interest of financial risk managers. Second, the Gets model identifies ECI as a meaningful driver of downside risk in this sample, whereas SSR appears to misestimate its impact. It appears that the Gets models perform better in the left tail of the return distribution, which was shown to be concerning for the SSR model and is the predominant concern of market participants with long (ownership) positions in security (as opposed to those who are in short positions).

Finally, in terms of economic measurement, this scenario analysis quantified the expected model risk cost associated with the potentially less robust SSR models, relative

to that of bias-corrected Gets. While the dollar amounts are a small percentage of the average trade, we should also consider that they are economically meaningful and, perhaps more importantly, that large players (like hedge funds) often use large transactions and substantial amounts of leverage. This would serve to magnify risk. Further research might focus on quantifying model risk in more exotic trading strategies under different leverage scenarios.

7. Conclusions

This study examines the impact of macroeconomic announcements on U.S. Treasury returns during the Greenspan era, a period characterized by a profound market preoccupation with economic data and Alan Greenspan's well-documented ability to shape expectations through his communication style. It has since been revealed that Greenspan's calculated ambiguity in his communications was, by design, a measure to mitigate market volatility that occurred during the prior decade, during Chair Paul Volcker's short-lived experiment with explicit money supply targeting.

By applying an automated Gets approach, with dual forms of indicator saturation, this research tested the viability of systematic model discovery in a most-challenging environment—where economic indicators were under intense scrutiny and where even subtle shifts in tone or emphasis could drive asset price movements.

The empirical results demonstrate that automated Gets provides a structured mechanism for identifying both stable and unstable parameters over the data sample. While certain relationships remained persistent, others appear to have evolved in response to changing macroeconomic conditions and Federal Reserve communication strategies.

Although Gets does not guarantee congruence, it steers towards it by systematically refining models through mis-specification testing. Compared to alternative machine learning approaches such as LASSO, Gets retains a stronger emphasis on *interpretability*, aligning with the LSE/Oxford tradition of econometric modeling. Unlike purely algorithmic selection techniques, it integrates economic reasoning into the model discovery process, offering transparency in variable selection and hypothesis testing.

While the main analysis is based on linear regression models, we also explore nonlinear robustness using Markov-switching specifications. The results suggest that Gets-selected regressors can meaningfully inform regime-switching behavior, particularly by clarifying abnormal return dynamics. This underscores the potential for combining Gets with nonlinear models to enrich empirical asset pricing frameworks.

Phillips (2005) envisioned a role for automation in econometrics, noting the potential in finance, where the volume and complexity of data necessitate efficient model discovery procedures and benefit from more precise parameter estimation. The findings herein align with this perspective, illustrating how automated Gets may complement traditional econometric methodologies like common static regression by providing a disciplined framework for specification search. This is particularly relevant in financial markets, such as the U.S. Treasury market, where the interaction between policy signals and asset prices demands adaptive modeling techniques.

While no single methodology is universally optimal, the Gets approach provides a robust alternative that balances statistical rigor with economic intuition. As financial markets continue to evolve in response to central bank policies and macroeconomic conditions, tools that facilitate systematic and interpretable model selection will likely remain attractive tools for applied empirical research.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/econometrics13030024/s1>.

Funding: This research was funded by the Center for International Securities and Derivatives Markets at the University of Massachusetts, Amherst and the School of Business Administration at SUNY New Paltz.

Data Availability Statement: The MMS data were downloaded from Harvard Dataverse at <https://library.harvard.edu/services-tools/harvard-dataverse> (accessed on 16 July 2018), US Treasury return data are from CRSP and are proprietary. <https://www.crsp.org/research/crsp-us-treasury-database/> (accessed on 16 July 2018).

Acknowledgments: I would like to thank Hossein Kazemi, Ben Branch, and Sanjay Nawalkha of the University of Massachusetts—Amherst for helpful comments and suggestions. Additionally, I would like to thank Mahnaz Mahdavi of Smith College and Yi Zheng from SUNY New Paltz for helpful comments, and seminar participants at the University of New Hampshire for additional suggestions. I thank Svetlana Doronkina for research assistance, manuscript preparation, and formatting. Any remaining oversights or errors are completely the responsibility of the author.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Appendix A.1. US Treasury Yields and Term Spread

The following table presents a graph of U.S. Treasury interest rates for four maturities, as well as an interest rate spread between bonds and bills, over the sample range.

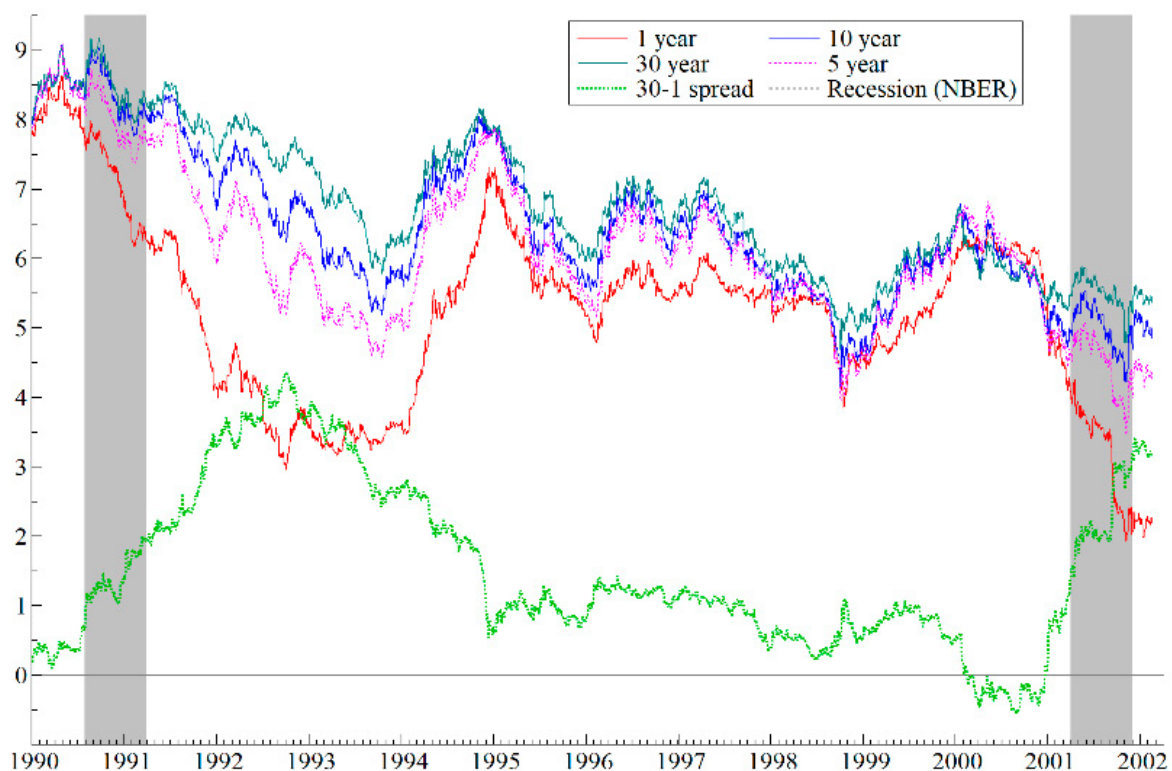


Figure A1. Source: St. Louis Fed FRED Database <https://fred.stlouisfed.org/> (accessed on 2 January 2024).

Appendix A.2. Rational Expectations of Macro Announcements Literature

This table aggregates the findings of four key studies on rational expectations of macroeconomic announcements.

	A. Rational Expectations		B. Anchoring Bias	
	Aggarwal et al. (1995)	Schirm (2003)	S. D. Campbell and Sharpe (2009)	Hess and Orbe (2013)
Auto Sales				
Business Inventories				X
Capacity Utilization				X
Consumer Confidence			X	X
Construction Spending				X
CPI	✓	✓	Mixed case	X
Core CPI			Mixed case	X
Durable Goods Orders	X	X	X	X
Employment Cost Index				
Gross Domestic Product				
GDP Price Deflator				
Trade Balance	✓	✓		✓
Hourly Earnings				X
Home Sales				✓
Housing Starts	✓	✓		X
Industrial Production	X	X	X	X
Index of Lead. Econ. Ind.	X			X
ISM Manufacturing (NAPM)			✓	X
Nonfarm Payrolls		✓	✓	✓
Personal Cons. Expenditures				X
Personal Income	✓	Mixed case		✓
PPI	X	X		X
Core PPI (ex: food and energy)				X
Retail Sales	X	✓	X	X
Retail Sales (ex: Auto Sales)			X	X
Unemployment Rate	✓	✓		✓
Sample Start	Varies	05/1990	Varies	Varies
Sample End	1993	12/2000	03/2006	2009
Method	DF and ADF, Engle-Yoo	ADF, Engle-Yoo	AR(5), anchoring bias and Wald tests	ARIMA models, anchoring bias tests
Survey	MMS	MMS and TF	MMS	MMS
Key:	✓ = RE X = no RE	✓ = RE X = no RE	✓ = no anchoring bias X = anchoring bias	✓ = no anchoring bias X = anchoring bias

Note: These findings are based on (Aggarwal et al., 1995; Schirm, 2003; S. D. Campbell & Sharpe, 2009; Hess & Orbe, 2013; Forest & Doronkina, 2025).

Appendix A.3. 10-Year Note Estimation Results

This table estimates Equation (1), the SSR model, and Equation (2), the Gets model with saturations, for ten-year OTR and FTR U.S. Treasury securities.

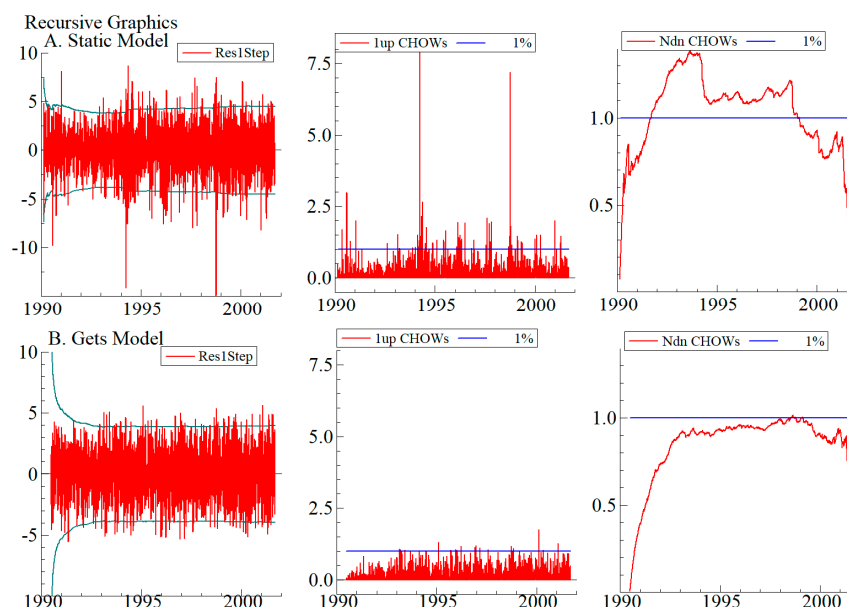
Indicators	Abbreviation	10-Year On-the-Run		10-Year 1st Off-the-Run	
		SSR Coefficient	GETS Coefficient	SSR Coefficient	GETS Coefficient
Constant	C	0.097 **		0.107 **	
Auto Sales	SS_AUTOS	−0.144		−0.150	
Business Inventories	SS_BUSINV	0.013		0.021	
Capacity Utilization	SS_CAPACIT	−0.383 *		−0.391 *	
Consumer Confidence	SS_CONFIDN	−0.651 **	−0.629 **	−0.628 **	−0.689 **
Construction Spending	SS_CONSTRC	−0.039		−0.032	
Consumer Price Index	SS_CPI	−0.135		−0.153	
Core CPI (ex: food and energy)	SS_CPIXFE	−0.428 **	−0.567 **	−0.394 **	−0.581 **
Durable Goods Orders	SS_DURGDS	−0.421 **	−0.393 **	−0.410 **	−0.420 **
Employment Cost Index	SS_ECI	−0.779 **		−0.730 **	−1.150 **
Gross Domestic Product	SS_GDP	−0.139		−0.094	
GDP Price Deflator	SS_GDPPRIC	−0.197		−0.215	

Indicators	Abbreviation	10-Year On-the-Run		10-Year 1st Off-the-Run	
		SSR Coefficient	GETS Coefficient	SSR Coefficient	GETS Coefficient
Goods and Services	SS_GDSSERV	0.033		0.040	
Hourly Earnings	SS_HREARN	−0.578 **	−0.630 **	−0.557 **	−0.405 **
Home Sales	SS_HSLS	−0.524 **	−0.487 **	−0.549 **	−0.608 **
Housing Starts	SS_HSTARTS	−0.045		−0.052	
Industrial Production	SS_INDPROD	0.058		0.082	
Index of Lead. Econ. Ind.	SS_LEI	−0.070		−0.071	
Nat. Assoc. of Purch. Mgrs.	SS_NAPM	−0.824 **	−0.883 **	−0.767 **	−0.774 **
Nonfarm Payrolls	SS_NONFARM	−1.070 **	−0.769 **	−1.030 **	−0.797 **
Personal Cons. Expenditures	SS_PCE	−0.058		−0.052	
Personal Income	SS_PERSINC	−0.175		−0.189	
Producer Price Index	SS_PPI	0.277		0.274 *	
Core PPI (ex: food and energy)	SS_PPIXFE	−0.342 *		−0.295 *	
Retail Sales	SS_RETSLs	−0.494 **		−0.502 **	
Retail Sales (ex: Auto Sales)	SS_RSXAUTO	0.024		0.030	
Unemployment Rate	SS_UNEMP	0.272 *		0.271 *	
Negative ECI	NEG_SS_ECI		−1.648 **		
sigma		1.472	1.217	1.415	1.168
# of observations		2928	2927	2928	2927
RSS		6286.739	4096.613	5808.688	3774.402
log-likelihood		−5273.33	−4645.24	−5157.54	−4525.35
#. of parameters		27	159	27	162
Adj. R ²		0.075	0.403	0.076	0.406

Note: SS indicates standardized surprise. **, * indicate significance at 0.01 and 0.05, respectively. Bias correction based on code courtesy of Hendry, Doornik, and Castle.

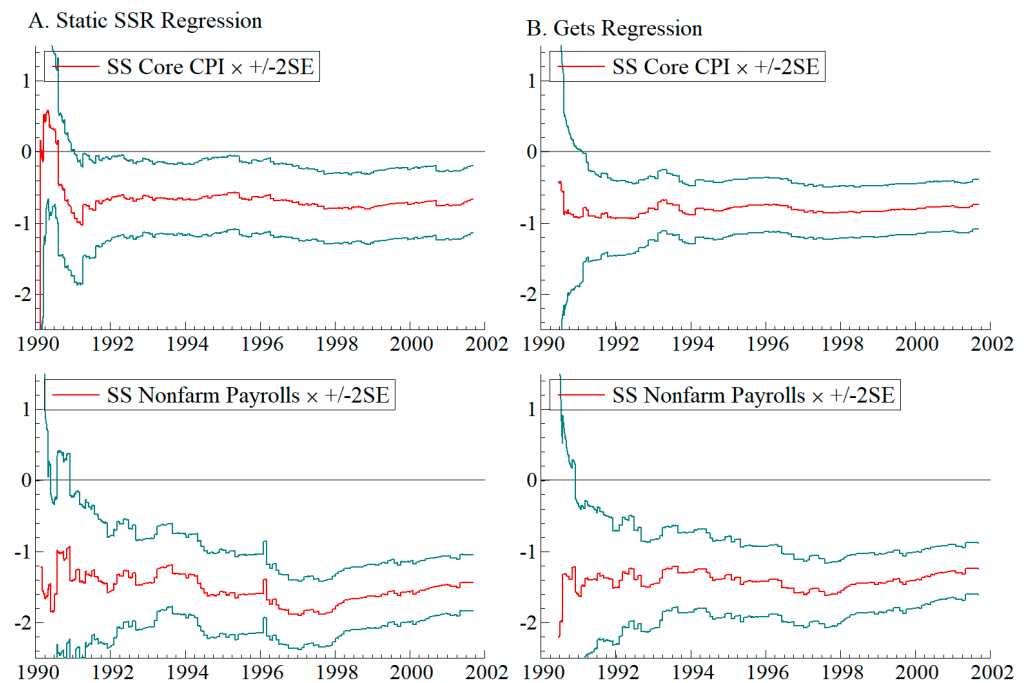
Appendix A.4. Recursive Stability Graphics

One-Step Residuals: The tighter spread in the Gets model reflects reduced misspecification and better handling of structural breaks. **1-Up Chow Tests:** This test assesses stability by including one additional observation at a time. For the SSR model, numerous spikes exceed the 1% critical value, highlighting instability during several key periods. However, the Gets model shows much lower and consistent test statistics, rarely exceeding the threshold. **N-Down Chow Tests:** Evaluates stability by progressively removing observations. The SSR model exhibits frequent instability, particularly in the mid-1990s, while the Gets model demonstrates more consistent parameter estimates throughout the sample period.



Appendix A.5. Recursive Parameters—Common Coefficients

The graphs below provide recursive parameter estimates of selected common coefficients in the SSR (panel A on left) and Gets (panel B on right). The graphics confirm stability of Core CPI and Nonfarm Payrolls in both estimation methodologies.



Appendix A.6. Parameter Stability

Parameter instability based on Hansen (1992). Because the test is not appropriate for models containing dummy variables, panel B is estimated from the selected Gets models with indicator saturation dummies removed.

Panel A. SSR Models.				
	OTR 30-Year	FTR 30-Year	OTR 10-Year	FTR 10-Year
Hansen Instability Tests				
Variance	1.779 **	2.184 **	0.770 *	0.822 **
Joint	6.890 **	7.748 **	6.254 *	6.241 *
Individual Instability Tests				
Constant	0.037	0.035	0.043	0.043
SS_AUTOS	0.114	0.108	0.175	0.143
SS_BUSINV	0.103	0.086	0.185	0.138
SS_CAPACIT	0.460	0.381	0.470	0.485 *
SS_CONFIDN	0.093	0.090	0.101	0.099
SS_CONSTRC	0.175	0.218	0.247	0.277
SS_CPI	0.054	0.054	0.054	0.054
SS_CPIXFE	0.063	0.059	0.079	0.082
SS_DURGDS	0.139	0.163	0.102	0.184
SS_ECI	0.074	0.080	0.111	0.110
SS_GDP	0.047	0.055	0.052	0.056
SS_GDPPRIC	0.125	0.115	0.237	0.215
SS_GDSSERV	0.148	0.157	0.158	0.169
SS_HREARN	0.127	0.126	0.071	0.090
SS_HSLs	0.066	0.064	0.057	0.058

Panel A. SSR Models.				
	OTR 30-Year	FTR 30-Year	OTR 10-Year	FTR 10-Year
SS_HSTARTS	0.575 *	0.572 *	0.749 *	0.671 *
SS_INDPROD	0.329	0.175	0.311	0.292
SS_LEI	0.107	0.114	0.104	0.108
SS_NAPM	0.249	0.303	0.138	0.157
SS_NONFARM	0.157	0.146	0.138	0.134
SS_PCE	0.438	0.525 *	0.508 *	0.514 *
SS_PERSINC	0.152	0.172	0.108	0.099
SS_PPI	0.175	0.165	0.226	0.189
SS_PPIXFE	0.488 *	0.400	0.582 *	0.534 *
SS_RETSLs	0.455	0.383	0.548 *	0.479 *
SS_RSXAUTO	0.052	0.046	0.061	0.063
SS_UNEMP	0.082	0.104	0.080	0.076
Panel B. Gets Models with Indicator Saturates Removed.				
	OTR 30-Year	FTR 30-Year	OTR 10-Year	FTR 10-Year
Hansen Instability Tests				
Variance	1.862 **	2.254 **	0.862 **	0.920 **
Joint	3.150 **	3.771 **	1.900	2.069
Individual Instability Tests				
SS_CONFIDN	0.109	0.107	0.118	0.113
SS_CPIXFE	0.069	0.074	0.093	0.103
SS_DURGDS	0.176	0.205	0.121	0.191
SS_ECI		0.040		0.113
SS_HREARN	0.140	0.143	0.086	0.108
SS_HSLs	0.067	0.065	0.061	0.061
SS_NAPM	0.237	0.293	0.140	0.158
SS_NONFARM	0.164	0.153	0.145	0.140
NEG_SS_ECI	0.112		0.124	
POS_SS_ECI		0.077		

Note: Larger values indicate parameter non-constancy (significance marked by * for 5% and ** for 1%).

Appendix A.7. Scenario Analysis—Selected Downside Announcement Surprises (U.S. Treasury OTR > 20 Years)

The table below quantifies the estimated difference in post-trade value of an average U.S. Treasury transaction. Scenarios measure positive 3-sigma shocks in one or more economic announcements, listed in the Shock column. Model differences between measures Gets and SSR. Scenario one represents shocks to Nonfarm Payrolls. Scenario two is for shocks to Hourly Earnings. Scenario three is for contemporaneous shocks of the same size to both Nonfarm Payrolls and Hourly Earnings. Scenario four is for shocks to ECI.

Scenario	Model	Const	Beta 1	Beta 2	Shock	SS	YhatDiff	Value	Loss
1	Gets	0.000	−0.32		Nonfarm Payrolls	3	−0.95	\$ 4,817,689	\$ 46,256
	SSR	0.013	−0.39		Nonfarm Payrolls	3	−1.17	\$ 4,807,091	\$ 56,855
	Difference					3	0.22	\$ 10,599	
2	Gets	0.000	−0.07		Hourly Earnings	3	−0.20	\$ 4,854,461	\$ 9485
	SSR	−0.035	−0.24		Hourly Earnings	3	−0.75	\$ 4,827,661	\$ 36,285
	Difference					3	0.55	\$ 26,800	

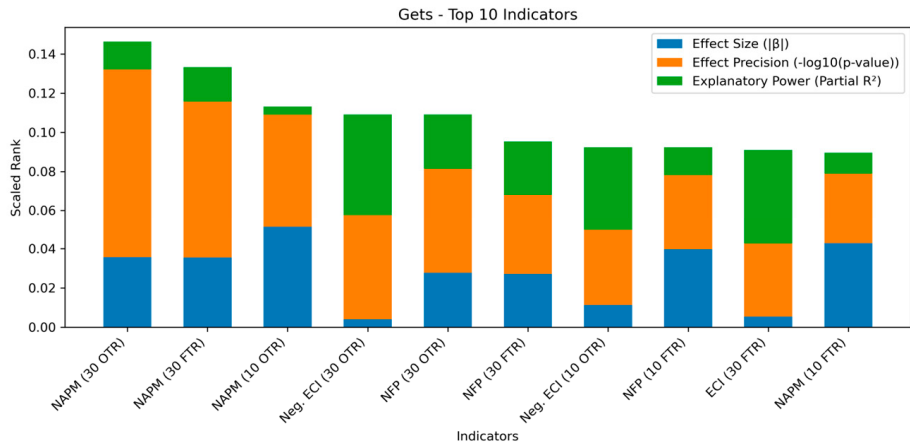
Scenario	Model	Const	Beta 1	Beta 2	Shock	SS	YhatDiff	Value		Loss
3	Gets	0.000	−0.32	−0.10	Nonfarm Payrolls and Hourly Earnings	3	−1.26	\$	4,802,806	\$ 61,140
	SSR	0.013	−0.39	−0.24		3	−1.90	\$	4,771,633	\$ 92,313
	Difference					3	0.64	\$	31,173	
4	Gets	0.000	−0.79		ECI	3	−2.36	\$	4,749,400	\$ 114,546
	SSR	0.013	−0.32		ECI	3	−0.93	\$	4,818,473	\$ 45,473
	Difference					3	−1.42	\$	(69,073)	

Note: Calculations are based on the modern trade size of hypothetical dealer to customer OTC transaction: \$4,863,945.58. Average dealer to customer transaction size estimated from TRACE volumes—2 July 2024. Diff stands for the difference between parameter estimates. Source: <https://www.finra.org/finra-data> (accessed on 1 May 2025). Shaded region highlights calculated differences in Y hat and value.

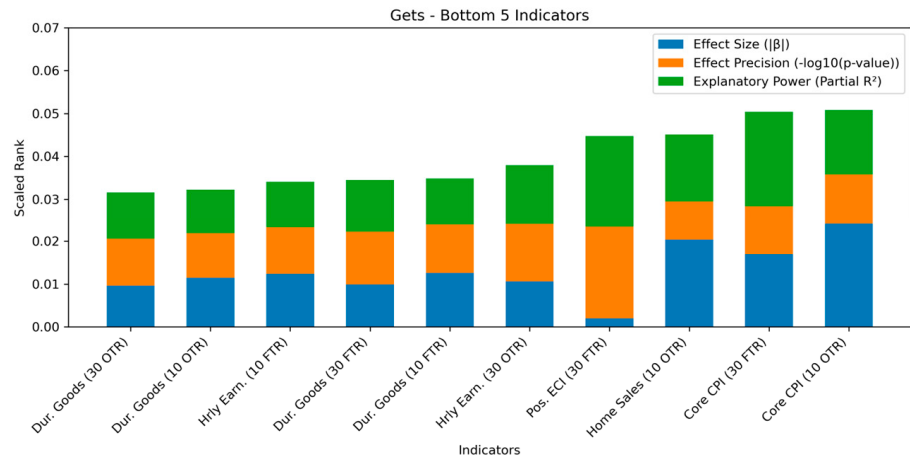
Appendix A.8. Effect Size, Precision, and Explanatory Power Analysis

The following graphics summarize macroeconomic indicaor quality based on an equal weighted average of effect size (absolute beta), effect precision (−log10(p-value)), and explanatory power (partial R²). The rankings are created by aggregating results from the regression analyses.

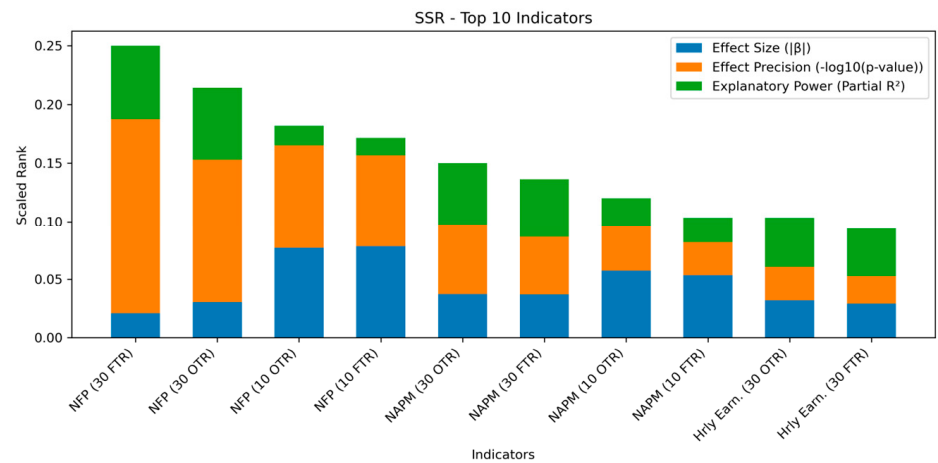
A.



B.



C.



Notes

- 1 The Autometrics algorithm is discussed in detail in (Doornik, 2009) and the implementation is demonstrated in (Doornik & Hendry, 2022).
- 2 Indicator saturation is a statistical technique that enables models to account for structural breaks and outliers systematically, enhancing robustness and model accuracy.
- 3 Autometrics, part of the PcGive software suite, automates the model selection (Hendry & Doornik, 2014).
- 4 Details of Greenspan's career can be found in (Sicilia & Cruikshank, 2000) and in (Greenspan, 2007). Note, a graphic depicting the path of US Treasury rates and spreads appears in Appendix A.1.
- 5 For examples of Greenspan's pre-FRB scholarly works, see: (Greenspan et al., 1958; Greenspan, 1964, 1971, 1978, 1980; Hymans et al., 1973).
- 6 Other financial applications of Gets include: (Sucarrat & Escribano, 2012; Bekaert et al., 2012; Bekaert & Hoerova, 2014; Stillwagon, 2016, 2017; Frydman & Stillwagon, 2018; Frydman et al., 2020; Bekaert & Mehl, 2019; Bonnier, 2022; Gómez-Puig et al., 2023; Forest et al., 2024a; Marçal, 2024).
- 7 Here, 'high-frequency' refers to the Treasury return data, which are daily. While macro announcements are generally monthly or quarterly, their staggered timing across days produces continuous flows of information that can influence bond returns throughout each month.
- 8 For a more detailed description of the multiple path tree search used in the modern Autometrics package, see (Doornik, 2009).
- 9 See www.doornik.com (accessed on 1 May 2025) for additional details on the software. Given the large amount of data and candidate regressors, the computation can take several hours for each model.
- 10 The model selection literature often uses the term gauge to describe the false retention probability.
- 11 Computations were performed on an AMD Ryzen 5 laptop with 8 GB of RAM with 4 cores. Additional memory and processing power would likely improve computation time.
- 12 The Markov-switching models used here are standard two-state models with regime-dependent means and variances, estimated using PcGive. These models are intended to test whether regime-sensitive behavior is amplified or mitigated.
- 13 Notable studies using MMS Survey Data include: (T. Ulrich & Wachtel, 1982, 1984; T. J. Ulrich, 1982; Jain, 1988; Aggarwal et al., 1995; Li & Engle, 1998; Almeida et al., 1998; Balduzzi et al., 2001; Andersen et al., 2003, 2007; Ramchander et al., 2005; Kilian & Vega, 2012).
- 14 Stationarity of both the dependent variables (Treasury excess returns) and the standardized macroeconomic surprise regressors was confirmed using ADF-Fisher unit root tests. The null hypothesis of a unit root was rejected at the 1% level for all series.
- 15 See also: (Pasquariello & Vega, 2007) regarding the "on-the-run liquidity phenomenon".
- 16 Recent work has rebranded the moniker to include Oxford, as the automation of model discovery methods were pioneered at Oxford University by Prof. Sir David F. Hendry and his coauthors.
- 17 Alternative Appendix A.1, available upon request.
- 18 Auto sales are subject to structural changes, particularly during labor strikes, and excluding them from retail sales may provide a better estimate of underlying consumer demand.
- 19 It is possible, but not tested here, that such key contemporaneous pairs may be best combined with an interaction term.

- This is consistent with the concerns of (Smales, 2021), who provided supplementary regression for robustness for the same reason. Within the LSE/Oxford modelling framework, it is advised to reformulate the GUM to account for state dependencies and/or interactions of interest to explore this phenomena deeper.
- Additional recursive stability diagnostics are given in Appendix A.5.
- Visual inspection of terminal models suggests strong agreement amongst competing terminal models with respect to significant macro variables. Disagreement between competing models was observed to be concentrated in the adjacent dates of IIS and SIS alternatives.
- See answer 5.
- Autometrics allows the user to force retention of unrestricted fixed variables that are theoretically meaningful for evaluation. Therefore, we re-estimated Equation (2) with the first order lagged dependent variable fixed.
- Although a rich literature on bond market reversals and momentum exists, I am not aware of any examples where Gets and saturation methods are used. See the following: (Khang & King, 2004; Zaremba & Kambouris, 2018; Li & Galvani, 2021; Zhang et al., 2021).
- It is notable that the shrinkage of parameters under Gets is done post estimation, in contrast with penalty-based methods, such as Lasso.
- It is also notable that the market did not appear to be affected by the anchoring bias suggested in the literature for several of the retained regressors. This implies that market participants adeptly adjust to the predictable bias of those economists participating in the MMS survey.

References

- Aggarwal, R., Mohanty, S., & Song, F. (1995). Are survey forecasts of macroeconomic variables rational? *The Journal of Business*, 68(1), 99–119. [\[CrossRef\]](#)
- Aktas, N., de Bodt, E., & Levasseur, M. (2004). Heterogeneity effects from market interventions. *The European Journal of Finance*, 10(5), 412–436. [\[CrossRef\]](#)
- Almeida, A., Goodhart, C., & Payne, R. (1998). The effects of macroeconomic news on high frequency exchange rate behavior. *The Journal of Financial and Quantitative Analysis*, 33(3), 383–408. [\[CrossRef\]](#)
- Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1), 178–196. [\[CrossRef\]](#)
- Amin, S., & Tédongap, R. (2023). The changing landscape of treasury auctions. *Journal of Banking & Finance*, 148, 106714. [\[CrossRef\]](#)
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2003). Micro effects of macro announcements: Real-time price discovery in Foreign exchange. *American Economic Review*, 93(1), 38–62. [\[CrossRef\]](#)
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, 73(2), 251–277. [\[CrossRef\]](#)
- Balduzzi, P., Elton, E. J., & Green, T. C. (2001). Economic news and bond prices: Evidence from the U.S. treasury market. *The Journal of Financial and Quantitative Analysis*, 36(4), 523–543. [\[CrossRef\]](#)
- Balduzzi, P., & Moneta, F. (2017). Economic risk premia in the fixed-income markets: The intraday evidence. *Journal of Financial and Quantitative Analysis*, 52(5), 1927–1950. [\[CrossRef\]](#)
- Becker, W., Paruolo, P., & Saltelli, A. (2021). Variable selection in regression models using global sensitivity analysis. *Journal of Time Series Econometrics*, 13(2), 187–233. [\[CrossRef\]](#)
- Bekaert, G., Hodrick, R. J., & Zhang, X. (2012). Aggregate idiosyncratic volatility. *Journal of Financial and Quantitative Analysis*, 47(6), 1155–1185. [\[CrossRef\]](#)
- Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2), 181–192. [\[CrossRef\]](#)
- Bekaert, G., & Mehli, A. (2019). On the global financial market integration “swoosh” and the trilemma. *Journal of International Money and Finance*, 94, 227–245. [\[CrossRef\]](#)
- Bessembinder, H., Chan, K., & Seguin, P. J. (1996). An empirical examination of information, differences of opinion, and trading activity. *Journal of Financial Economics*, 40(1), 105–134. [\[CrossRef\]](#)
- Billio, M., Donadelli, M., Paradiso, A., & Riedel, M. (2017). Which market integration measure? *Journal of Banking & Finance*, 76, 150–174. [\[CrossRef\]](#)
- Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., & Jansen, D.-J. (2008). Central Bank communication and monetary policy: A survey of theory and evidence. *Journal of Economic Literature*, 46(4), 910–945. [\[CrossRef\]](#)
- Bollerslev, T., Cai, J., & Song, F. M. (2000). Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US Treasury bond market. *Journal of Empirical Finance*, 7(1), 37–55. [\[CrossRef\]](#)
- Bonnier, J.-B. (2022). Forecasting crude oil volatility with exogenous predictors: As good as it Gets? *Energy Economics*, 111, 106059. [\[CrossRef\]](#)

- Bontemps, C., & Mizon, G. E. (2003). Congruence and encompassing. In *Econometrics and the philosophy of economics: Theory-data confrontations in economics*. Princeton University Press. [CrossRef]
- Bontemps, C., & Mizon, G. E. (2008). Encompassing: Concepts and implementation*. *Oxford Bulletin of Economics and Statistics*, 70(s1), 721–750. [CrossRef]
- Brockman, P., Chung, D. Y., & Pérignon, C. (2009). Commonality in liquidity: A global perspective. *Journal of Financial and Quantitative Analysis*, 44(4), 851. [CrossRef]
- Campbell, C. J., Kazemi, H. B., & Nanisetty, P. (1999). Time-Varying risk and return in the bond market: A test of a new equilibrium pricing model. *The Review of Financial Studies*, 12(3), 631–642. [CrossRef]
- Campbell, S. D., & Sharpe, S. A. (2009). Anchoring bias in consensus forecasts and its effect on market prices. *Journal of Financial and Quantitative Analysis*, 44(2), 369–390. [CrossRef]
- Campos, J., Hendry, D. F., & Krolzig, H.-M. (2003). Consistent model selection by an automatic Gets approach. *Oxford Bulletin of Economics and Statistics*, 65, 803–819. [CrossRef]
- Castle, J. L., Doornik, J., Hendry, D., & Pretis, F. (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, 3(2), 240–264. [CrossRef]
- Castle, J. L., Doornik, J. A., & Hendry, D. F. (2013). Evaluating automatic model selection. *Journal of Time Series Econometrics*, 3(3), 1941–1928. [CrossRef]
- Choi, J. (2013). What drives the value premium?: The role of asset risk and leverage. *Review of Financial Studies*, 26(11), 2845–2875. [CrossRef]
- Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, 53(4), 671–690. [CrossRef]
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28, 591–605. [CrossRef]
- Christie-David, R., Chaudhry, M., & Khan, W. (2002). News releases, market integration, and market leadership. *Journal of Financial Research*, 25, 223–245. [CrossRef]
- Connolly, R., & Stivers, C. (2005). Macroeconomic News, stock turnover, and volatility clustering in daily stock returns. *Journal of Financial Research*, XXVIII(2), 235–259. [CrossRef]
- Desboulets, L. D. D. (2018). A review on variable selection in regression analysis. *Econometrics*, 6(4), 45. [CrossRef]
- Dhrymes, P. J. (1971). *Distributed lags: Problems of estimation and formulation*. Holden-Day.
- Doornik, J. A. (2008). Encompassing and automatic model selection*. *Oxford Bulletin of Economics and Statistics*, 70(s1), 915–925. [CrossRef]
- Doornik, J. A. (2009). *Autometrics*. Oxford University Press. [CrossRef]
- Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality*. *Oxford Bulletin of Economics and Statistics*, 70(s1), 927–939. [CrossRef]
- Doornik, J. A., & Hendry, D. F. (2022). *PcGive 16*. Timberlake Consultants.
- Ederington, L., & Lee, J. H. (1993). How markets process information: News releases and volatility. *Journal of Finance*, 48(4), 1161–1191. [CrossRef]
- Eijffinger, S. C. W., & Pieterse-Bloem, M. (2023). Eurozone government bond spreads: A tale of different ECB policy regimes. *Journal of International Money and Finance*, 139, 102965. [CrossRef]
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007. [CrossRef]
- Engle, R. F., Hansen, M. K., & Lunde, A. (2012). *And now, the rest of the News: Volatility and firm specific news arrival*. CREATES Research Papers. Department of Economics and Business Economics.
- Epprecht, C., Guégan, D., Veiga, Á., & Correa da Rosa, J. (2019). Variable selection and forecasting via automated methods for linear models: Lasso/Adalasso and autometrics. *Communications in Statistics—Simulation and Computation*, 50(1), 103–122. [CrossRef]
- Ericsson, N. R. (2008). The fragility of sensitivity analysis: An encompassing perspective. *Oxford Bulletin of Economics and Statistics*, 70(s1), 895–914. [CrossRef]
- Ericsson, N. R. (2017). How biased are U.S. government forecasts of the federal debt? *International Journal of Forecasting*, 33(2), 543–559. [CrossRef]
- Ericsson, N. R., Dore, M. H. I., & Butt, H. (2022). Detecting and quantifying structural breaks in climate. *Econometrics*, 10(4), 33. [CrossRef]
- Fleming, M., & Remolona, E. (1999). Price formation and liquidity in the U.S. treasury market: The response to public information. *Journal of Finance*, 54(5), 1901–1915. [CrossRef]
- Forest, J. J. (2018). *Essays in financial economics: Announcement effects in fixed income markets* [Doctoral dissertation, University of Massachusetts-Amherst]. Available online: https://scholarworks.umass.edu/dissertations_2/1339 (accessed on 1 May 2025).
- Forest, J. J., Branch, B. S., & Berry, B. T. (2024a). Trading activity in the corporate bond market: A sad tale of macro-announcements and behavioral seasonality? *Risks*, 12(5), 80. [CrossRef]

- Forest, J. J., Chang, D. M., Doronkina, S., & Panahova, A. (2024b). *Midas meets Fsher & Friedman: Assessing QTM as a short-term forecasting device*. SUNY New Paltz. [\[CrossRef\]](#)
- Forest, J. J., & Doronkina, S. (2025). *Rational expectations in housing markets: The case of survey forecasts*. Working Paper. SUNY New Paltz. [\[CrossRef\]](#)
- Forest, J. J., & Mackey, S. (2023). The effect of treasury auction announcements on interest rates: 1990–1999. *Academy of Economics and Finance Journal*, 14, 1–10. [\[CrossRef\]](#)
- Forest, J. J., & Turner, P. (2013). Alternative estimators of cointegrating parameters in models with Non-stationary data: An application to US export demand. *Applied Economics*, 45(5), 629–636. [\[CrossRef\]](#)
- Frydman, R., Mangee, N., & Stillwagon, J. (2020). How market sentiment drives forecasts of stock returns. *Journal of Behavioral Finance*, 22(4), 351–367. [\[CrossRef\]](#)
- Frydman, R., & Stillwagon, J. R. (2018). Fundamental factors and extrapolation in stock-market expectations: The central role of structural change. *Journal of Economic Behavior & Organization*, 148, 189–198. [\[CrossRef\]](#)
- Gigante, G., Guarniero, P., & Pasini, S. (2024). Markovian analysis of U.S. Treasury volatility: Asymmetric responses to macroeconomic announcements. *Economics Letters*, 239, 111723. [\[CrossRef\]](#)
- Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, 46(6), 1303–1310. [\[CrossRef\]](#)
- Govaerts, B., Hendry, D. F., & Richard, J.-F. (1994). Encompassing in stationary linear dynamic models. *Journal of Econometrics*, 63(1), 245–270. [\[CrossRef\]](#)
- Gómez-Puig, M., Pieterse-Bloem, M., & Sosvilla-Rivero, S. (2023). Dynamic connectedness between credit and liquidity risks in euro area sovereign debt markets. *Journal of Multinational Financial Management*, 68, 100800. [\[CrossRef\]](#)
- Granger, C. W. J., & Hendry, D. F. (2005). A dialogue concerning a new instrument for econometric modeling. *Econometric Theory*, 21(01), 278–297. [\[CrossRef\]](#)
- Greenspan, A. (1964). Liquidity as a determinant of industrial prices and interest rates. *The Journal of Finance*, 19(2), 159–169. [\[CrossRef\]](#)
- Greenspan, A. (1971). A model of capital expenditures and internal rates of return for the U.S. Economy. *Business Economics*, 6(3), 44–49.
- Greenspan, A. (1978). Inflation and economic activity. *Business Economics*, 13(3), 11–14.
- Greenspan, A. (1980). The great malaise. *Challenge*, 23(1), 37–40. [\[CrossRef\]](#)
- Greenspan, A. (2007). *The age of turbulence*. Penguin Books.
- Greenspan, A., Simpson, P. B., & Cutler, A. T. (1958). Discussion. *The American Economic Review*, 48(2), 171–177.
- Gürkaynak, R. S., Sack, B. P., & Swanson, E. T. (2007). Market-Based measures of monetary policy expectations. *Journal of Business & Economic Statistics*, 25(2), 201–212. [\[CrossRef\]](#)
- Hansen, B. E. (1992). Tests for parameter instability in regressions with 1(1) processes. *Journal of Business & Economic Statistics*, 10(3), 321–335. [\[CrossRef\]](#)
- Hendry, D. F. (1995). *Dynamic econometrics*. Oxford University Press. [\[CrossRef\]](#)
- Hendry, D. F. (2024). A Brief History of General-to-specific Modelling*. *Oxford Bulletin of Economics and Statistics*, 86, 1–20. [\[CrossRef\]](#)
- Hendry, D. F., & Doornik, J. A. (2014). *Empirical model discovery and theory evaluation*. MIT Press. [\[CrossRef\]](#)
- Hendry, D. F., & Krolzig, H. M. (1999). Improving on ‘Data mining reconsidered’ by K.D. Hoover and S.J. Perez. *The Econometrics Journal*, 2(2), 202–219. [\[CrossRef\]](#)
- Hendry, D. F., & Krolzig, H.-M. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25, 831–866. [\[CrossRef\]](#)
- Hendry, D. F., & Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *The Economic Journal*, 115(502), C32–C61. [\[CrossRef\]](#)
- Hendry, D. F., Pagan, A. R., & Sargan, J. D. (1984). Chapter 18 dynamic specification. In Z. Griliches, & D. Michael (Eds.), *Handbook of econometrics* (pp. 1023–1100). Intriligator. [\[CrossRef\]](#)
- Hess, D., & Orbe, S. (2013). Irrationality or efficiency of macroeconomic survey forecasts? Implications from the anchoring bias test*. *Review of Finance*, 17(6), 2097–2131. [\[CrossRef\]](#)
- Hoover, K. D., & Perez, S. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, 2(2), 167–191. [\[CrossRef\]](#)
- Hymans, S. H., Greenspan, A., Shiskin, J., & Early, J. (1973). On the use of leading indicators to predict cyclical turning points. *Brookings Papers on Economic Activity*, 1973(2), 339–384. [\[CrossRef\]](#)
- Jain, P. C. (1988). Response of hourly stock prices and trading volume to economic News. *Journal of Business*, 61(2), 219–231. [\[CrossRef\]](#)
- Jones, C. M., Lamont, O., & Lumsdaine, R. L. (1998). Macroeconomic news and bond market volatility. *Journal of Financial Economics*, 47(3), 315–337. [\[CrossRef\]](#)
- Khang, K., & King, T. (2004). Return reversals in the bond market: Evidence and causes. *Journal of Banking and Finance*, 28, 569–593. [\[CrossRef\]](#)
- Kilian, L., & Vega, C. (2012). *Replication data for: Do energy prices respond to U.S. macroeconomic News? A test of the hypothesis of predetermined energy prices*. Harvard Dataverse. [\[CrossRef\]](#)

- Koyck, L. M. (1954). *Distributed lags and investment analysis*. North-Holland. [CrossRef]
- Krishnamurthy, A. (2002). The bond/old-bond spread. *Journal of Financial Economics*, 66(2–3), 463–506. [CrossRef]
- Li, L., & Engle, R. F. (1998). *Macroeconomic announcements and volatility of treasury futures*. UCSD Economics Discussion Papers. University of California San Diego. Available online: <https://ssrn.com/abstract=145828> (accessed on 1 May 2025).
- Li, L., & Galvani, V. (2021). Informed trading and momentum in the corporate bond market*. *Review of Finance*, 25(6), 1773–1816. [CrossRef]
- Marçal, E. F. (2024). Testing rational expectations in a cointegrated VAR with structural change. *International Review of Financial Analysis*, 95, 103435. [CrossRef]
- Mizon, G. E., & Richard, J.-F. (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica*, 54(3), 657–678. [CrossRef]
- Muhammadullah, S., Urooj, A., Khan, F., Alshahrani, M. N., Alqawba, M., Al-Marzouki, S., & Zhu, P. (2022). Comparison of weighted lag adaptive Lasso with autometrics for covariate selection and forecasting using time-series data. *Complexity*, 2022, 2649205. [CrossRef]
- Pasquariello, P., & Vega, C. (2007). Informed and strategic order flow in the bond markets. *Review of Financial Studies*, 20(6), 1975–2019. [CrossRef]
- Phillips, P. C. B. (2005). Automated discovery in econometrics. *Econometric Theory*, 21(1), 3–20. [CrossRef]
- Pretis, F., Schneider, L., Smerdon, J. E., & Hendry, D. F. (2016). Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation. *Journal of Economic Surveys*, 30(3), 403–429. [CrossRef]
- Ramchander, S., Simpson, M. W., & Chaudhry, M. K. (2005). The influence of macroeconomic news on term and quality spreads. *The Quarterly Review of Economics and Finance*, 45(1), 84–102. [CrossRef]
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 31(2), 350–371. [CrossRef]
- Sargan, J. D. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 21(1), 91–105. [CrossRef]
- Schirm, D. C. (2003). A comparative analysis of the rationality of consensus forecasts of U.S. Economic indicators. *The Journal of Business*, 76(4), 547–561. [CrossRef]
- Sicilia, D. B., & Cruikshank, J. L. (2000). *The greenspan effect*. McGraw-Hill.
- Smales, L. A. (2021). Macroeconomic news and treasury futures return volatility: Do treasury auctions matter? *Global Finance Journal*, 48, 100537. [CrossRef]
- Stillwagon, J. R. (2016). Non-linear exchange rate relationships: An automated model selection approach with indicator saturation. *The North American Journal of Economics and Finance*, 37, 84–109. [CrossRef]
- Stillwagon, J. R. (2017). TIPS and the VIX: Spillovers from financial panic to breakeven inflation in an automated, nonlinear modeling framework. *Oxford Bulletin of Economics and Statistics*, 80(2), 218–235. [CrossRef]
- Sucarrat, G., & Escribano, A. (2012). Automated model selection in finance: General-to-specific modelling of the mean and volatility specifications. *Oxford Bulletin of Economics and Statistics*, 74(5), 716–735. [CrossRef]
- Swanson, E. T. (2006). Have increases in federal reserve transparency improved private sector interest rate forecasts? *Journal of Money, Credit, and Banking*, 38(3), 791–819. [CrossRef]
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. [CrossRef]
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: A retrospective. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73(3), 273–282. [CrossRef]
- Urlich, T., & Wachtel, P. (1984). The effects of inflation and money supply announcements on interest rates. *Journal of Finance*, 39(4), 1177–1188. [CrossRef]
- Urlich, T., & Wachtel, P. (2012). Market response to the weekly money supply announcements in the 1970s. *The Journal of Finance*, 36(5), 1063–1072. [CrossRef]
- Urlich, T. J. (1982). The information content of weekly money supply announcements. *Journal of Monetary Economics*, 10(1), 73–88. [CrossRef]
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817–838. [CrossRef]
- Zaremba, A., & Kambouris, G. (2018). The sources of momentum in international government bond returns. *Applied Economics*, 51, 848–857. [CrossRef]
- Zhang, W., Wang, P., & Li, Y. (2021). Bond intraday momentum. *Journal of Behavioral and Experimental Finance*, 31, 100515. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.