Ignorance and Indifference: Decision-Making in the Lab and in the Market. Supplement*

Sergiy Verstyuk†

This version: 8 October 2024

Abstract

These supplemental materials contain additional plots of empirical and model-implied certainty equivalents as well as posterior mean and probability weighting functions. These materials also offer potential refinements pertaining to the posterior mean function. They present statistical analysis of detected regime changes. They also include additional bibliographical and technical details on the concept of shrinkage. Proofs for the lemmas and propositions stated in the main text are collected here too. A description of data sources is also given.

*Acknowledgements: see the main text.
†Contact address: verstyuk@cmsa.fas.harvard.edu.

Contents

A Decisions from description and from experience: Figures **1**

B Decisions from description and under ambiguity: Figures **4**

C Posterior mean function: Refinement **6**

D Analysis of detected regime changes **8**

E Shrinkage **19**

F Proofs **20**

F.1 Proof of Proposition 1 20
F.2 Proof of Proposition 2 20
F.3 Proof of Proposition 3 20
F.4 Proof of Lemma 1 20
F.5 Proof of Lemma 2 20
F.6 Proof of Proposition 4 21

G Data **22**

A Decisions from description and from experience: Figures

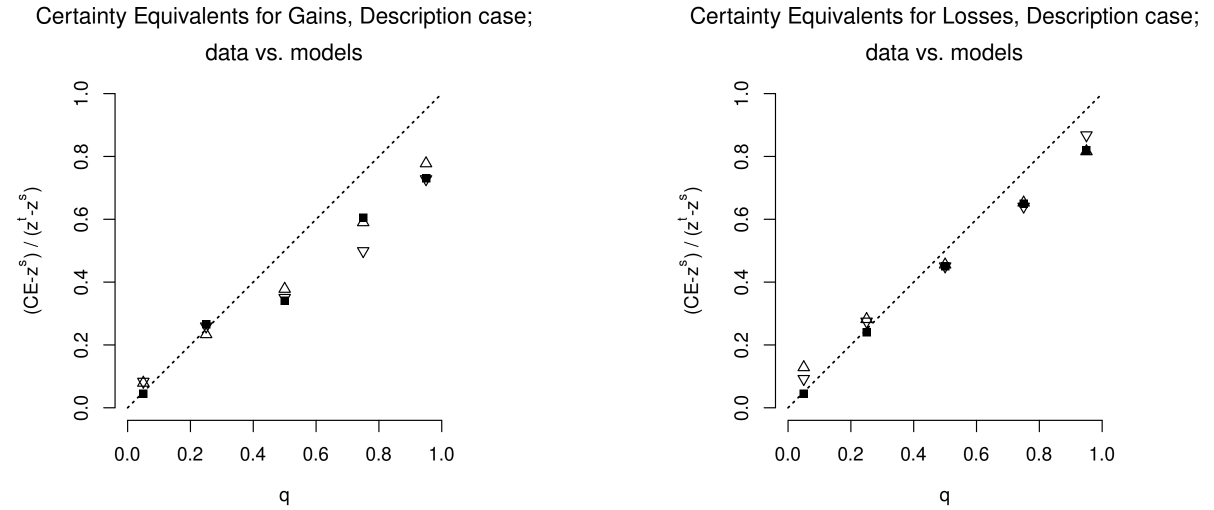

Figure 1: Ratios of CEs to lottery payoffs, standardized, median values (vertical axis) vs. declared true tail probabilities (horizontal axis), where values of CEs are either measured empirically (solid squares), implied by the CPT model (triangles pointing down) or implied by the encompassing revealed-prior model (triangles pointing up). Lotteries with gains and with losses, information obtained from description.

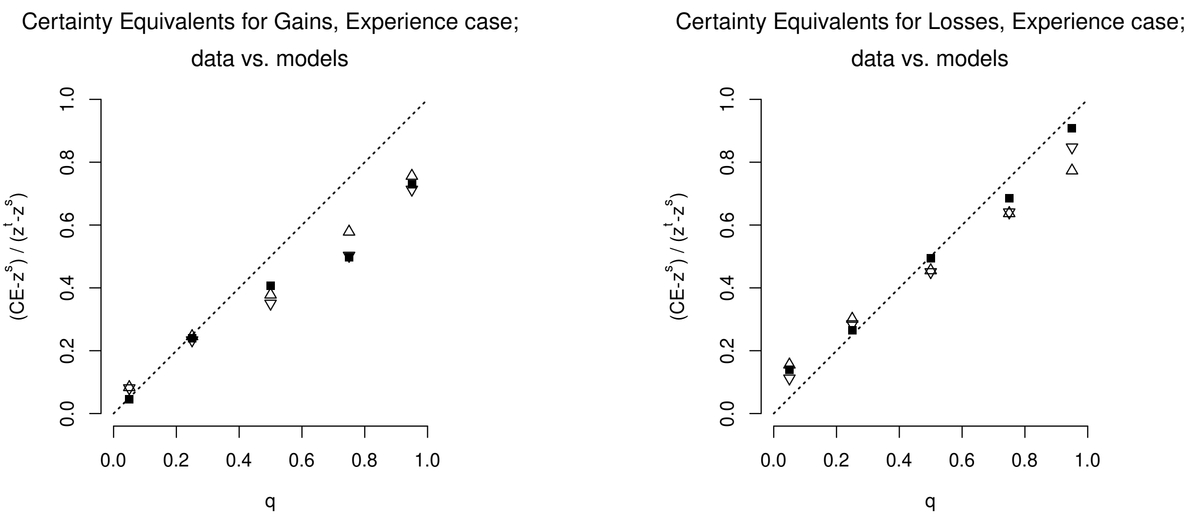

Figure 2: Ratios of CEs to lottery payoffs, standardized, median values (vertical axis) vs. declared true tail probabilities (horizontal axis), where values of CEs are either measured empirically (solid squares), implied by the CPT model (triangles pointing down) or implied by the encompassing revealed-prior model (triangles pointing up). Lotteries with gains and with losses, information obtained from experience.

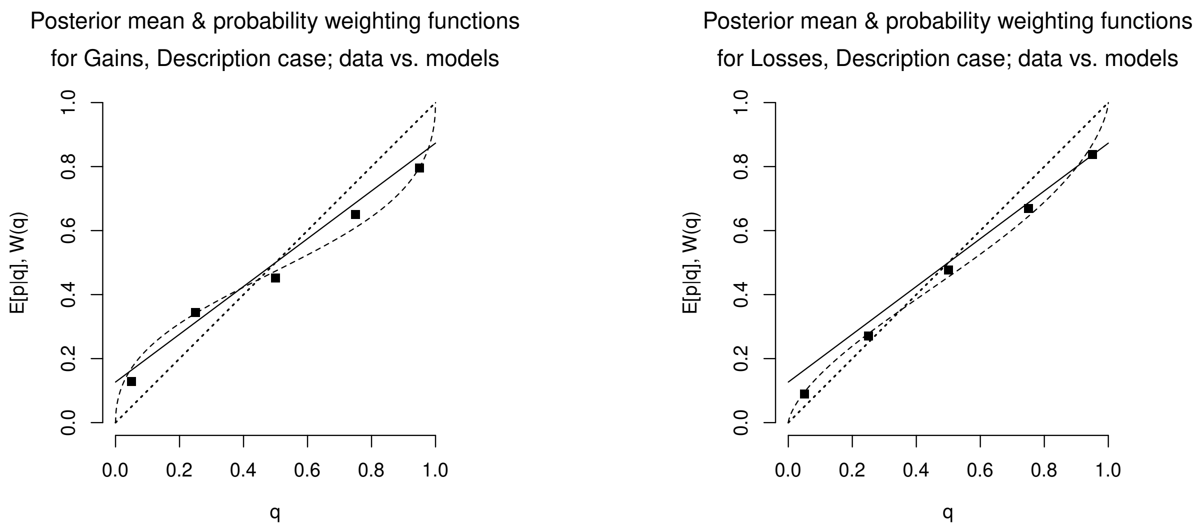

Figure 3: Posterior mean and probability weighting functions (vertical axis) vs. declared true tail probabilities (horizontal axis), with posterior mean function produced by the encompassing revealed-prior model from Table 3 (solid) and with probability weighting functions produced by the CPT model from Table 3 (dashes) as well as by agnostic data fitting (solid squares). Lotteries with gains and with losses, information obtained from description.

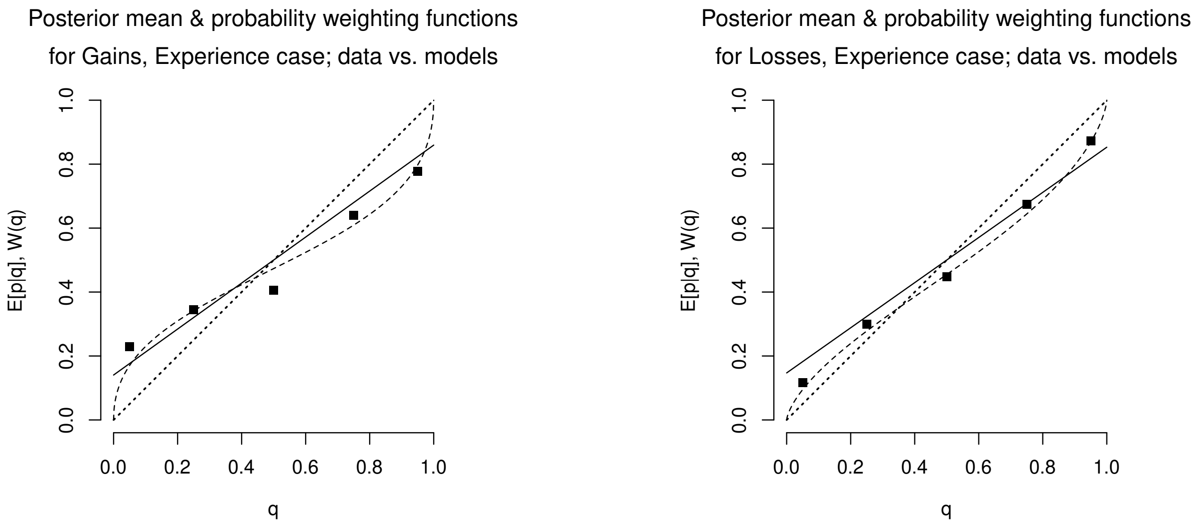

Figure 4: Posterior mean and probability weighting functions (vertical axis) vs. declared true tail probabilities (horizontal axis), with posterior mean function produced by the encompassing revealed-prior model from Table 3 (solid) and with probability weighting functions produced by the CPT model from Table 3 (dashes) as well as by agnostic data fitting (solid squares). Lotteries with gains and with losses, information obtained from experience.

B Decisions from description and under ambiguity: Figures

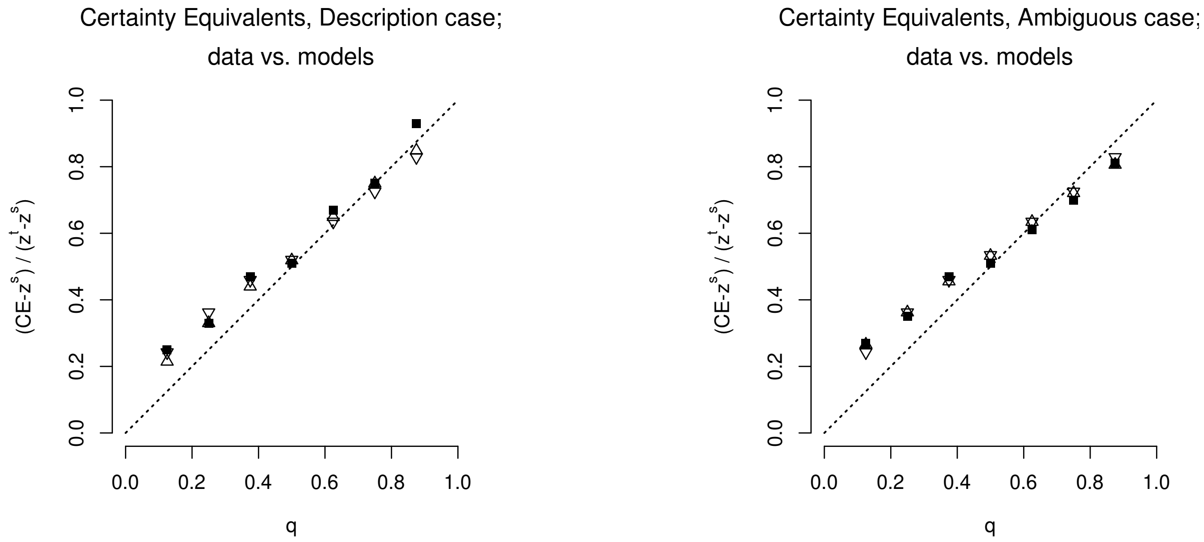

Figure 5: Ratios of CEs to lottery payoffs, standardized, median values (vertical axis) vs. declared true tail probabilities (horizontal axis), where values of CEs are either measured empirically (solid squares), implied by the CPT model (triangles pointing down) or implied by the encompassing revealed-prior model (triangles pointing up). Lotteries with gains, information obtained from description and under ambiguity.

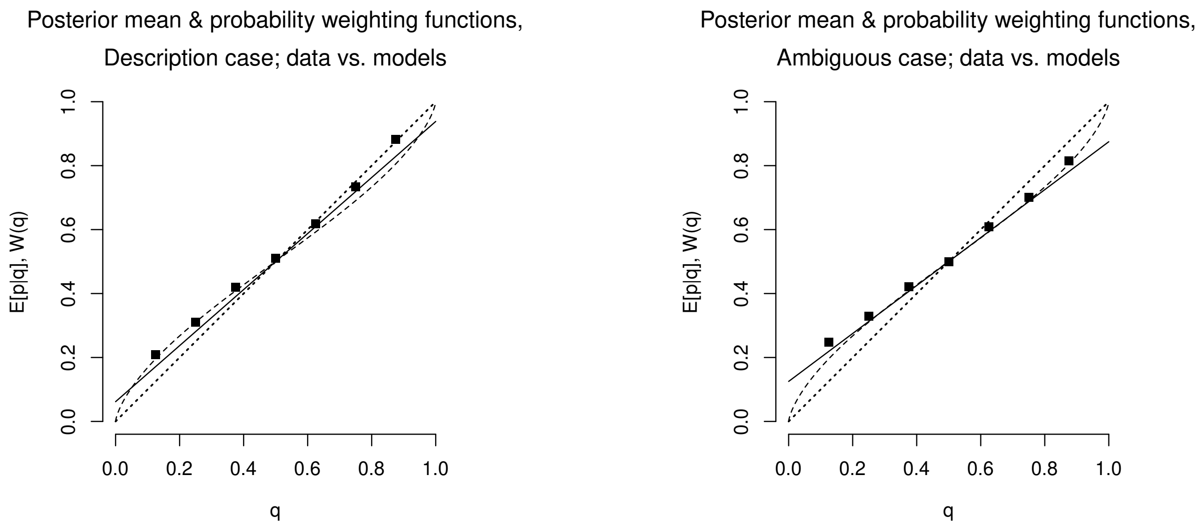

Figure 6: Posterior mean and probability weighting functions (vertical axis) vs. declared true tail probabilities (horizontal axis), with posterior mean function produced by the encompassing revealed-prior model from Table 4 (solid) and with probability weighting functions produced by the CPT model from Table 4 (dashes) as well as by agnostic data fitting (solid squares). Lotteries with gains, information obtained from description and under ambiguity.

C Posterior mean function: Refinement

In the presented derivations (leading to equation 10) and the resulting plot (Figure 4), (i) the probability distortions exhibit a linear pattern; (ii) they have a fixed, or inflection, point at $p = 1/2$; and (iii) distorted probabilities in the corners of $p = 0$ and $p = 1$ do not converge continuously, or "paste smoothly", with the original probabilities (so a degenerate, non-stochastic lottery becomes stochastic).

These features may be contradicting the common view on the relevant stylized facts (as given in, say, Kahneman and Tversky, 1979, 1992; or in Camerer and Ho, 1994): an inverse-S-shaped pattern; a fixed point around $p = 1/3$; and the "smooth pasting" at 0 and 1 (e.g., see Figure 1). More recent estimates, however, that use less restrictive non-parametric methods and focus on representative samples of participants are in fact more supportive of the linear pattern for distortions as well as of the inflection at the midpoint (for some evidence, see Fehr-Duda and Epper, 2012). Moreover, degenerate lotteries are rarely tested in the experiments; hence, the smooth pasting result may be an artifact of the parametric estimation and fitting procedure and, strictly speaking, warrants a confirmation in a separate, narrowly-focused study.

In section §4.1 of the main text we show that the above restrictions stemming from our derivations do not stand on the way of good empirical fit. Nevertheless, we also present some further refinements of the general approach that would rationalize potentially desirable stylized empirical facts.

First, consider the matching of declared and posterior (i.e., of original and distorted) probabilities at the corners $p = 0$ and $p = 1$. This would be ensured for an agent who uses, instead of the Jeffreys prior $\mathcal{B}(1/2, 1/2)$, a mixture prior that combines the Jeffreys prior with the so-called Haldane prior $\mathcal{B}(0, 0)$.[1] Instead of a parameterization invariance, Haldane's approach is motivated by the fact that it effectively forces the posterior mean of the distribution of the parameter in question to coincide with its maximum-likelihood point estimate. Still, it does imply a uniform distribution for a particular parameterization: specifically, in terms of the logarithm of the odds ratio, $\ln(p/(1-p))$. Intuitively, it can be interpreted as representing 0 initial pseudo-observations (hence, the probability distribution corresponding to Haldane prior does not integrate to 1 and is "improper").[2] Standard Bayesian updating in this case produces the following posterior:

$$\pi^M(p|q) = \sum_{m=1}^{M} \pi(m|q) \times \frac{p^{\alpha_m + nq - 1}(1-p)^{\beta_m + n(1-q) - 1}}{\mathrm{B}(\alpha_m + nq, \beta_m + n(1-q))}, \tag{1}$$

[1]The distribution that corresponds to the Haldane prior is essentially a symmetric mixture of two Dirac delta functions on each end of the $[0, 1]$ interval.

[2]For the sake of completeness, the maximum entropy interpretation here implies a geometric mean value restriction of 0 both for the probability of success as well as for the probability of failure (another instance of flexible choices).

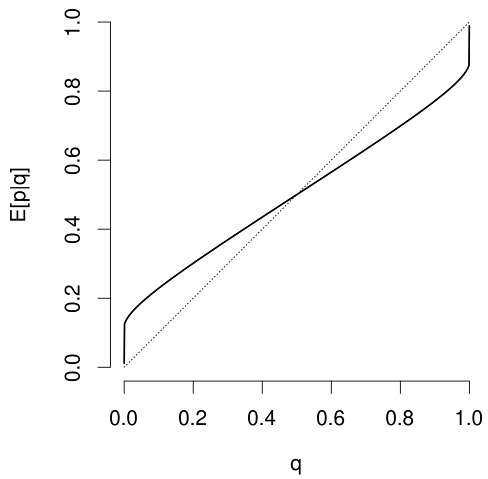

Figure 7: Posterior mean (vertical) vs. declared (horizontal) probabilities, refinement (parameterization used is $\mathrm{E}[p|q]$ utilizing equation (1) with $\alpha_m, \beta_m \in \{0.50, 0.40, 0.31, 0.23, 0.16, 0.10, 0.06, 0.03, 0.01, 0.00\}$ and $\pi_m \in \{0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.91\}$ as well as $n = 1$).

with

$$\pi(m|q) := \frac{\pi(m)\mathrm{B}(\alpha_m + nq, \beta_m + n(1-q))/\mathrm{B}(\alpha_m, \beta_m)}{\sum_k \pi(k)\mathrm{B}(\alpha_k + nq, \beta_k + n(1-q))/\mathrm{B}(\alpha_k, \beta_k)},$$

where $\pi(m)$ is the initial prior weight of the mixture component m. For a sufficiently high weight $\pi(m)$ assigned to the Haldane prior, and with $M = 2$, the corner probabilities are matched, whereas the interior ones are unaffected, because a “dogmatic” Haldane prior has 0 posterior weight in those regions (hence, Figure 4 is unchanged, except for its two end points).

Still, in this case the corner probabilities do not paste smoothly, that is, there are discontinuous jumps on each end.

Second, consider the smooth pasting of the declared and posterior probabilities. This would be achieved by expanding the set of the mixture prior's components with the prior distributions that are located (in the hyperparameter space) between the Jeffreys prior and the Haldane prior. More concretely, one may use the same Bayesian updating formula (1) with $\{\alpha_m, \beta_m\}$ spread between $\{0, 0\}$ and $\{1/2, 1/2\}$, as well as with the initial prior weights $\pi(m)$ equally allocated between mixture components (except for the disproportionately high weight on the Haldane component in the corner). Then, the corner probabilities are smoothly pasted along with an inverse-S-shaped pattern of probability distortions. Specific mixture weights are merely an empirical question (as a proof of concept, see Figure 7).

D Analysis of detected regime changes

This analysis uses the outputs of three changepoint detection procedures summarized in Table 5. We are interested in asking (i) which variables may be the drivers behind regime changes $\mathbf{1}_{\star,t}$ as well as (ii) which variables may be driven by regime changes (in the weak sense of statistical dependence rather than causal relationships). To this end, we consider such economic variables as the real GDP growth y_t, the magnitude of the cyclical component of the real GDP ("output gap") $y_{c,t}$, the unemployment rate u_t, the NBER recessions indicator $\mathbf{1}_{\searrow,t}$, the CPI inflation rate i_t, the federal funds rate $r_{ff,t}$, the market return r_t, the Volatility Index (VIX) ς_t.

Note that we added the following variables to the list of otherwise self-explanatory economic time series. Variable $\mathbf{1}_{\star,t}$ is a binary indicator that equals 1 if there is a regime change at period t; a regime change is defined as a changepoint produced by a given changepoint detection algorithm. Variable $\mathbf{1}_{\searrow,t+i}$ is a binary indicator that equals 1 in the event of U.S. recession at period $t+i$; formally, it is defined as $\mathbf{1}_{\searrow,t+i} := \left(\mathbf{1}_{\searrow}(t+i)\right)_t$ for any $i \in \mathbb{Z}$, where $\mathbf{1}_A(x)$ is an indicator function.

First, we look at the pairwise point-biserial correlation coefficients between the corresponding binary regime change variables and various economic variables. Table 1 presents the results for a parametric frequentist detection method: it reveals that regime changes are associated with lags and leads of market returns and volatility variables, particularly at daily frequencies. Table 2 presents the results for a parametric Bayesian detection method: it documents the association of regime changes with lags and leads of daily and/or monthly unemployment rate, inflation rate, federal funds rate, market returns and volatility variables, as well as the leads of (i.e., future) NBER recession indicator. Table 3 presents the results for a non-parametric detection method: it shows that regime changes are associated with a wide set of variables, especially lags and leads of daily and/or monthly federal funds rate, market returns and volatility, as well as lags of (i.e., past) monthly inflation rate and leads of monthly/quarterly real GDP growth changes, output gap changes, NBER recession indicator.

Second, we look at the results on multivariate logistic regressions of regime changes on the above economic variables (focusing on regression specifications with regressor timing adjusted in light of pairwise correlations' results). Table 5 presents the results for a parametric frequentist detection method: among the market variables, which suggest the determinants behind the given method's detection of a regime change, both market returns and return volatilities have some effect on the regime change variable, but the corresponding goodness-of-fit statistics are not strong; among the macroeconomic and policy variables, which may be indirectly related to detected regime changes, real GDP growth and unemployment rate variables are somewhat connected to regime changes, although the goodness-of-fit measures are fairly weak. Table 7 presents the results for a parametric Bayesian method: among the market variables, both market returns and return volatilities are at play, though not showing strong fits; among the macroeconomic and policy variables, the unemployment rate, NBER recessions indicator and inflation rate are connected to regime changes, although fits are weak. Table 9 presents the results for a non-parametric method: among the market variables, both market returns and return volatilities have an effect, and the fits are relatively strong; among the macroeconomic and policy variables, the output gap, NBER recessions and inflation rate are connected to regime changes, the fits are again relatively strong.

Table 1: Correlates of regime changes (parametric frequentist detection)

Freq	Regime Change	$\lvert\Delta y_t\rvert$	$\lvert\Delta\Delta y_t\rvert$	$\lvert y_{c,t}\rvert$	$\lvert\Delta y_{c,t}\rvert$	u_t	$\lvert\Delta u_t\rvert$	$\mathbf{1}_{\searrow,t}$	$\lvert i_t\rvert$	$\lvert\Delta i_t\rvert$	$r_{ff,t}$	$\lvert\Delta r_{ff,t}\rvert$	$\lvert r_t\rvert$	$\lvert\Delta r_t\rvert$	ς_t	$\lvert\Delta\varsigma_t\rvert$
D	$\mathbf{1}_{\star,t+3}$										-0.011	-0.006	0.040***	0.015**	0.001	0.015
	$\mathbf{1}_{\star,t+2}$										-0.011	-0.003	0.014*	0.033***	0.000	-0.004
	$\mathbf{1}_{\star,t+1}$										-0.012	-0.008	0.016**	0.017**	0.007	0.025**
	$\mathbf{1}_{\star,t}$										-0.012	-0.005	0.038***	0.031***	0.005	0.005
	$\mathbf{1}_{\star,t-1}$										-0.011	-0.007	0.056***	0.072***	0.020*	0.057***
	$\mathbf{1}_{\star,t-2}$										-0.012	-0.007	0.020***	0.029**	0.029**	0.046***
	$\mathbf{1}_{\star,t-3}$										-0.012	-0.003	0.046***	0.038***	0.030***	0.058***
M	$\mathbf{1}_{\star,t+3}$					-0.024	0.060*	0.028	-0.038	0.003	-0.056	-0.051	0.080**	0.016	0.036	0.023
	$\mathbf{1}_{\star,t+2}$					-0.023	0.081**	0.013	-0.034	-0.022	-0.057	-0.040	0.041	0.041	0.046	-0.029
	$\mathbf{1}_{\star,t+1}$					-0.018	-0.039	0.013	-0.052	-0.013	-0.060*	-0.010	0.006	0.041	0.039	-0.022
	$\mathbf{1}_{\star,t}$					-0.017	-0.035	0.028	-0.035	0.006	-0.064*	-0.025	0.120***	0.118***	0.047	0.141***
	$\mathbf{1}_{\star,t-1}$					-0.005	0.038	0.030	-0.027	0.005	-0.063*	-0.047	0.077**	0.023	0.153***	0.329***
	$\mathbf{1}_{\star,t-2}$					0.001	-0.012	0.015	-0.001	-0.003	-0.065*	-0.033	0.026	-0.005	0.145***	0.031
	$\mathbf{1}_{\star,t-3}$					-0.003	0.027	0.000	0.001	-0.002	-0.067*	-0.041	0.013	0.012	0.069	0.031
Q	$\mathbf{1}_{\star,t+2}$	0.075	-0.021	-0.051	0.054	-0.029	0.010	0.025	-0.028	-0.009	-0.093	-0.063	0.020	-0.044	0.037	0.065
	$\mathbf{1}_{\star,t+1}$	0.011	-0.003	-0.036	-0.033	-0.028	0.039	0.026	-0.079	0.054	-0.102	-0.092	0.015	0.070	0.066	0.099
	$\mathbf{1}_{\star,t}$	0.024	-0.095	-0.001	-0.026	-0.016	0.066	0.049	-0.090	-0.002	-0.118*	-0.053	0.183***	0.256***	0.152*	0.251***
	$\mathbf{1}_{\star,t-1}$	0.118**	0.066	0.059	0.122**	0.010	0.038	-0.002	-0.009	0.031	-0.128**	-0.058	0.147**	0.047	0.134	0.189**
	$\mathbf{1}_{\star,t-2}$	0.050	0.075	0.082	-0.004	0.005	0.126**	-0.059	-0.028	0.031	-0.125**	-0.021	-0.019	-0.064	0.069	0.044

Notes: Three vertical panels feature data with different frequencies: daily (D), monthly (M) and quarterly (Q), as indicated in the first column from the left. The rows present point-biserial correlation coefficients between a regime change variable (produced by parametric frequentist detection method described in text) at the time period stated in the second column from the left and one of the economic variables listed in the remaining columns on the right, namely the absolute real GDP growth (i.e., temporal difference in the log of), the absolute magnitude of the cyclical component of the log of real GDP, the unemployment rate (as a fraction), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the federal funds rate (as a fraction), the absolute log market return, the Volatility Index VIX value (divided by 100), with each economic variable (except the US recessions variable) being additionally accompanied by the absolute value of its temporal difference. Superscripts *, **, *** indicate p-values in the two-sided paired samples correlation test below 0.10, 0.05, 0.01 levels, respectively. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the unemployment rate, in which case the data sample is 1948:M1:D1–2019:M12:D31; the federal funds rate, in which case it is 1954:M7:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 2: Correlates of regime changes (parametric Bayesian detection)

Freq	Regime Change	$\lvert\Delta y_t\rvert$	$\lvert\Delta\Delta y_t\rvert$	$\lvert y_{c,t}\rvert$	$\lvert\Delta y_{c,t}\rvert$	u_t	$\lvert\Delta u_t\rvert$	$\mathbf{1}_{\searrow,t}$	$\lvert i_t\rvert$	$\lvert\Delta i_t\rvert$	$r_{ff,t}$	$\lvert\Delta r_{ff,t}\rvert$	$\lvert r_t\rvert$	$\lvert\Delta r_t\rvert$	ς_t	$\lvert\Delta\varsigma_t\rvert$
D	$\mathbf{1}_{\star,t+3}$										-0.035***	-0.005	0.136***	0.147***	0.228***	0.112***
	$\mathbf{1}_{\star,t+2}$										-0.034***	-0.014*	0.144***	0.150***	0.242***	0.128***
	$\mathbf{1}_{\star,t+1}$										-0.033***	-0.019**	0.167***	0.155***	0.246***	0.150***
	$\mathbf{1}_{\star,t}$										-0.035***	-0.020***	0.489***	0.462***	0.261***	0.354***
	$\mathbf{1}_{\star,t-1}$										-0.035***	-0.018**	0.127***	0.286***	0.251***	0.115***
	$\mathbf{1}_{\star,t-2}$										-0.035***	-0.018**	0.141***	0.152***	0.246***	0.119***
	$\mathbf{1}_{\star,t-3}$										-0.034***	-0.023***	0.134***	0.146***	0.239***	0.112***
M	$\mathbf{1}_{\star,t+3}$					0.131***	-0.021	0.032	0.125***	0.084**	0.035	0.061*	0.095***	0.029	0.178***	0.080
	$\mathbf{1}_{\star,t+2}$					0.125***	-0.046	0.039	0.106***	0.042	0.041	0.023	0.123***	0.078**	0.209***	0.078
	$\mathbf{1}_{\star,t+1}$					0.126***	-0.045	0.061*	0.095***	0.071**	0.047	0.026	0.116***	0.093***	0.241***	0.064
	$\mathbf{1}_{\star,t}$					0.128***	-0.034	0.089***	0.116***	0.065*	0.049	0.068*	0.108***	0.175***	0.306***	0.123**
	$\mathbf{1}_{\star,t-1}$					0.131***	0.030	0.089***	0.132***	0.074**	0.049	0.089**	0.168***	0.173***	0.295***	0.180***
	$\mathbf{1}_{\star,t-2}$					0.132***	-0.016	0.111***	0.115***	0.096***	0.042	0.071**	0.126***	0.115***	0.264***	0.099*
	$\mathbf{1}_{\star,t-3}$					0.145***	0.008	0.118***	0.086**	0.105***	0.031	0.037	0.044	0.074**	0.231***	0.100*
Q	$\mathbf{1}_{\star,t+2}$	-0.006	0.016	-0.015	0.029	0.043	0.008	-0.045	0.116**	0.115*	0.030	-0.026	0.091	0.066	0.148	0.039
	$\mathbf{1}_{\star,t+1}$	0.010	0.000	-0.013	-0.046	0.032	-0.003	-0.046	0.061	0.089	0.028	-0.013	0.081	0.073	0.142	0.089
	$\mathbf{1}_{\star,t}$	-0.016	-0.024	-0.015	-0.041	0.016	-0.053	0.034	0.097*	0.123**	0.023	0.034	-0.019	0.098*	0.182**	0.079
	$\mathbf{1}_{\star,t-1}$	-0.111*	-0.003	-0.046	-0.073	0.022	-0.060	0.073	0.063	0.088	0.006	0.072	0.083	0.018	0.101	-0.027
	$\mathbf{1}_{\star,t-2}$	-0.032	0.015	-0.001	0.085	0.030	0.095	0.074	0.048	0.050	0.004	0.034	-0.059	0.069	0.142	0.056

Notes: Three vertical panels feature data with different frequencies: daily (D), monthly (M) and quarterly (Q), as indicated in the first column from the left. The rows present point-biserial correlation coefficients between a regime change variable (produced by parametric Bayesian detection method described in text) at the time period stated in the second column from the left and one of the economic variables listed in the remaining columns on the right, namely the absolute real GDP growth (i.e., temporal difference in the log of), the absolute magnitude of the cyclical component of the log of real GDP, the unemployment rate (as a fraction), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the federal funds rate (as a fraction), the absolute log market return, the Volatility Index VIX value (divided by 100), with each economic variable (except the US recessions variable) being additionally accompanied by the absolute value of its temporal difference. Superscripts *, **, *** indicate p-values in the two-sided paired samples correlation test below 0.10, 0.05, 0.01 levels, respectively. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the unemployment rate, in which case the data sample is 1948:M1:D1–2019:M12:D31; the federal funds rate, in which case it is 1954:M7:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 3: Correlates of regime changes (non-parametric detection)

Freq	Regime Change	$\|\Delta y_t\|$	$\|\Delta\Delta y_t\|$	$\|y_{c,t}\|$	$\|\Delta y_{c,t}\|$	u_t	$\|\Delta u_t\|$	$\mathbf{1}_{\searrow,t}$	$\|i_t\|$	$\|\Delta i_t\|$	$r_{ff,t}$	$\|\Delta r_{ff,t}\|$	$\|r_t\|$	$\|\Delta r_t\|$	ς_t	$\|\Delta\varsigma_t\|$
D	$\mathbf{1}_{\star,t+3}$										0.017**	0.021***	0.019***	0.006	0.056***	0.053***
	$\mathbf{1}_{\star,t+2}$										0.019**	0.027***	0.020***	0.008	0.063***	0.053***
	$\mathbf{1}_{\star,t+1}$										0.022***	0.023***	0.035***	0.009	0.065***	0.062***
	$\mathbf{1}_{\star,t}$										0.022***	0.008	0.052***	-0.002	0.073***	0.057***
	$\mathbf{1}_{\star,t-1}$										0.019**	0.022***	0.051***	0.109***	0.067***	0.083***
	$\mathbf{1}_{\star,t-2}$										0.018**	0.019**	0.048***	0.053***	0.060***	0.041***
	$\mathbf{1}_{\star,t-3}$										0.020***	0.031***	0.022***	0.023***	0.066***	0.037***
M	$\mathbf{1}_{\star,t+3}$					-0.041	-0.011	0.017	0.050	-0.028	0.089**	0.025	0.019	0.018	0.012	-0.052
	$\mathbf{1}_{\star,t+2}$					-0.043	-0.033	0.036	0.073**	0.009	0.087**	0.001	0.028	0.012	0.093*	0.176***
	$\mathbf{1}_{\star,t+1}$					-0.038	-0.020	0.074**	0.072**	-0.004	0.087**	0.058	0.041	-0.003	0.180***	0.134**
	$\mathbf{1}_{\star,t}$					-0.033	-0.002	0.055	0.124***	0.062*	0.093***	0.123***	0.197***	0.203***	0.265***	0.254***
	$\mathbf{1}_{\star,t-1}$					-0.029	0.025	0.083**	0.074**	0.063*	0.092***	0.071**	0.155***	0.217***	0.266***	0.251***
	$\mathbf{1}_{\star,t-2}$					-0.019	0.014	0.102***	0.047	-0.008	0.077**	0.094***	0.083**	0.077**	0.206***	0.086
	$\mathbf{1}_{\star,t-3}$					-0.007	0.056*	0.131***	0.026	-0.015	0.062*	0.078**	0.089***	0.057*	0.161***	0.004
Q	$\mathbf{1}_{\star,t+2}$	0.066	0.133**	0.016	0.114*	-0.022	0.021	-0.004	0.127**	0.007	0.100	-0.027	0.007	-0.021	0.108	0.150
	$\mathbf{1}_{\star,t+1}$	-0.016	0.109*	0.017	-0.063	-0.026	-0.082	0.039	0.119**	-0.066	0.112*	-0.024	0.012	-0.010	0.084	0.072
	$\mathbf{1}_{\star,t}$	-0.028	0.056	0.061	0.031	-0.023	-0.011	0.096	0.147**	-0.015	0.119*	0.135**	0.123**	0.275***	0.395***	0.359***
	$\mathbf{1}_{\star,t-1}$	0.086	0.095	0.086	0.187***	0.020	0.100*	0.158***	0.120**	0.029	0.092	0.072	0.219***	0.280***	0.173*	0.109
	$\mathbf{1}_{\star,t-2}$	0.037	0.206***	0.114*	0.097	0.054	0.101*	0.055	0.118**	0.012	0.093	0.054	0.037	0.054	0.113	0.045

Notes: Three vertical panels feature data with different frequencies: daily (D), monthly (M) and quarterly (Q), as indicated in the first column from the left. The rows present point-biserial correlation coefficients between a regime change variable (produced by non-parametric detection method described in text) at the time period stated in the second column from the left and one of the economic variables listed in the remaining columns on the right, namely the absolute real GDP growth (i.e., temporal difference in the log of), the absolute magnitude of the cyclical component of the log of real GDP, the unemployment rate (as a fraction), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the federal funds rate (as a fraction), the absolute log market return, the Volatility Index VIX value (divided by 100), with each economic variable (except the US recessions variable) being additionally accompanied by the absolute value of its temporal difference. Superscripts *, **, *** indicate p-values in the two-sided paired samples correlation test below 0.10, 0.05, 0.01 levels, respectively. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the unemployment rate, in which case the data sample is 1948:M1:D1–2019:M12:D31; the federal funds rate, in which case it is 1954:M7:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 4: Logistic regression with correlates of regime changes (parametric frequentist detection) — contemporaneous regressors

	Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	constant	-6.26	-3.49	-2.72	-2.44	-2.47	-2.59	-1.35	-1.12	-1.16	-1.49	-1.25	-5.93	-2.94	-1.72	-2.95	-1.74
		(0.295)	(0.406)	(0.48)	(0.651)	(0.649)	(0.265)	(0.436)	(0.396)	(0.332)	(0.324)	(0.292)	(0.22)	(0.24)	(0.26)	(0.16)	(0.17)
D	$\|\Delta r_t\|$	39.27															
		(10.48)															
	$\|\Delta \varsigma_t\|$	-22.22															
		(19.16)															
M	$\|\Delta r_t\|$		5.11														
			(5.72)														
	$\|\Delta \varsigma_t\|$		10.76														
			(5.48)														
Q	$\|\Delta r_t\|$			6.09													
				(3.16)													
	$\|\Delta \varsigma_t\|$			10.68													
				(6.10)													
M	$\|u_t\|$				-2.68	-1.97											
					(10.51)	(10.48)											
	$\mathbf{1}_{\searrow,t}$				0.77		0.76										
					(0.466)		(0.466)										
	$\|i_t\|$				-135.87	-118.48	-135.93										
					(79.96)	(80.42)	(79.88)										
	$\|\Delta r_{ff,t}\|$				-36.04	-12.60	-37.63										
					(62.98)	(55.7)	(62.56)										
Q	$\|\Delta y_t\|$							19.17	-15.66								
								(39.68)	(28.89)								
	$\|\Delta y_t^c\|$							-95.63		-27.26							
								(54.58)		(39.09)							
	$\|\Delta u_t\|$							145.12			102.64						
								(82.82)			(67.54)						
	$\mathbf{1}_{\searrow,t}$							1.11				0.97					
								(0.608)				(0.525)					
	$\|i_t\|$							-61.12	-55.83	-54.56	-50.75	-66.43					
								(33.95)	(32.91)	(33.14)	(32.33)	(33.02)					
	$\|\Delta r_{ff,t}\|$							-18.31	-2.44	3.48	-21.03	-18.17					
								(36.6)	(32.17)	(33.5)	(35.73)	(34.55)					
	Pseudo-R^2	0.007	0.004	0.085	0.008	0.003	0.008	0.046	0.013	0.015	0.016	0.030	0.000	0.000	0.000	0.000	0.000
	Accuracy	0.997	0.947	0.849	0.950	0.950	0.950	0.843	0.851	0.851	0.851	0.851	0.997	0.950	0.849	0.950	0.851
	Observations	7552	359	119	785	785	785	261	261	261	261	261	7552	359	119	785	261

Notes: The explained variable is a regime change variable $\mathbf{1}_{\star,t}$ (produced by parametric frequentist detection method described in text). The top rows present coefficient estimates for the following explanatory variables: firstly, market variables (separately for daily, D, monthly, M, and quarterly, Q, frequency) including the absolute log market return, the absolute Volatility Index VIX growth (i.e., temporal difference in the log of VIX over 100); secondly, macroeconomic and policy variables (separately for M and Q frequency) including the absolute real GDP growth (i.e., difference in the log of), the absolute change of the cyclical component of the log of real GDP (i.e., its temporal difference), the unemployment rate (i.e., taken as a fraction and then used in levels or in absolute values of temporal differences), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the absolute change of the federal funds rate (i.e., taken as a fraction and then temporally differenced); in both cases, the timing of explanatory variables coincides with that of the explained variable. The bottom three rows provide: goodness-of-fit measures including Pseudo-R^2 (using Efron's definition) and the fraction of accurate predictions (measured as #Correct/#Total), as well as the number of observations. The columns contain different data frequency and regression specifications, with the last five columns providing results for relevant null models. The estimation is conducted by maximum likelihood. Standard errors (in parenthesis) are obtained from the Fisher information matrix. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the federal funds rate, in which case the data sample is 1954:M7:D1–2019:M12:D31; the unemployment rate, in which case it is 1948:M1:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 5: Logistic regression with correlates of regime changes (parametric frequentist detection) — asynchronous regressors

	Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	constant	-6.54	-4.16	-2.72	-3.66	-3.92	-2.95	-2.12	-1.95	-1.76	-1.73	-1.47	-5.98	-3.00	-1.72	-2.98	-1.76
		(0.293)	(0.502)	(0.48)	(1.76)	(1.641)	(0.538)	(0.417)	(0.401)	(0.325)	(0.305)	(0.286)	(0.23)	(0.25)	(0.26)	(0.17)	(0.18)
D	$\|\Delta r_{t+1}\|$	23.14															
		(11.34)															
	$\|\Delta \varsigma_{t+1}\|$	15.76															
		(9.88)															
M	$\|\Delta r_{t}\|$		5.22														
			(5.83)														
	$\|\Delta \varsigma_{t+1}\|$		26.87														
			(6.8)														
Q	$\|\Delta r_{t}\|$			6.09													
				(3.16)													
	$\|\Delta \varsigma_{t}\|$			10.68													
				(6.10)													
M	$\|u_{t+3}\|$				4.72	4.84											
					(10.26)	(10.26)											
	$\mathbf{1}_{\searrow,t+3}$				0.39		0.39										
					(0.526)		(0.525)										
	$\|i_{t}\|$				-122.82	-111.55	-120.81										
					(83.59)	(82.55)	(83.56)										
	$\|\Delta r_{ff,t+2}\|$				-44.53	-34.50	-41.87										
					(70.79)	(67.05)	(70.76)										
Q	$\|\Delta y_{t+1}\|$							20.88	45.54								
								(37.05)	(25.56)								
	$\|\Delta y^{c}_{t+1}\|$							21.35		56.75							
								(47.62)		(29.53)							
	$\|\Delta u_{t+2}\|$							165.01			145.86						
								(85.32)			(62.7)						
	$\mathbf{1}_{\searrow,t+1}$							-0.70				0.30					
								(0.821)				(0.557)					
	$\|i_{t-2}\|$							-26.94	-22.83	-31.02	-38.62	-30.85					
								(33.07)	(32.39)	(32.08)	(32.1)	(33.21)					
	$\|\Delta r_{ff,t}\|$							-21.07	-15.77	-20.30	-17.88	-13.17					
								(38.04)	(36.28)	(36.52)	(37.56)	(36.1)					
	Pseudo-R^2	-0.005	0.116	0.085	0.005	0.004	0.004	0.047	0.024	0.033	0.035	0.010	0.000	0.000	0.000	0.000	0.000
	Accuracy	0.997	0.958	0.849	0.952	0.952	0.952	0.857	0.853	0.853	0.853	0.853	0.997	0.953	0.849	0.952	0.853
	Observations	7552	359	119	784	784	784	259	259	259	259	259	7552	359	119	784	259

Notes: The explained variable is a regime change variable $\mathbf{1}_{\star,t}$ (produced by parametric frequentist detection method described in text). The top rows present coefficient estimates for the following explanatory variables: firstly, market variables (separately for daily, D, monthly, M, and quarterly, Q, frequency) including the absolute log market return, the absolute Volatility Index VIX growth (i.e., temporal difference in the log of VIX over 100); secondly, macroeconomic and policy variables (separately for M and Q frequency) including the absolute real GDP growth (i.e., difference in the log of), the absolute change of the cyclical component of the log of real GDP (i.e., its temporal difference), the unemployment rate (i.e., taken as a fraction and then used in levels or in absolute values of temporal differences), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the absolute change of the federal funds rate (i.e., taken as a fraction and then temporally differenced); in both cases, the timing of explanatory variables is adjusted following the results in Table 1. The bottom three rows provide: goodness-of-fit measures including Pseudo-R^2 (using Efron's definition) and the fraction of accurate predictions (measured as #Correct/#Total), as well as the number of observations. The columns contain different data frequency and regression specifications, with the last five columns providing results for relevant null models. The estimation is conducted by maximum likelihood. Standard errors (in parenthesis) are obtained from the Fisher information matrix. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the federal funds rate, in which case the data sample is 1954:M7:D1–2019:M12:D31; the unemployment rate, in which case it is 1948:M1:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 6: Logistic regression with correlates of regime changes (parametric Bayesian detection) — contemporaneous regressors

	Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	constant	-3.41	0.41	2.21	-0.50	-0.52	0.50	2.39	2.45	2.13	2.08	2.24	-1.60	1.23	3.13	0.92	2.66
		(0.070)	(0.228)	(0.765)	(0.329)	(0.327)	(0.13)	(0.622)	(0.569)	(0.475)	(0.487)	(0.418)	(0.03)	(0.13)	(0.46)	(0.08)	(0.25)
D	$\lvert\Delta r_t\rvert$	105.88															
		(4.00)															
	$\lvert\Delta \varsigma_t\rvert$	36.54															
		(3.42)															
M	$\lvert\Delta r_t\rvert$		14.57														
			(4.64)														
	$\lvert\Delta \varsigma_t\rvert$		13.56														
			(7.42)														
Q	$\lvert\Delta r_t\rvert$			6.24													
				(8.91)													
	$\lvert\Delta \varsigma_t\rvert$			24.09													
				(29.70)													
M	$\lvert u_t\rvert$				17.91	18.33											
					(5.44)	(5.41)											
	$\mathbf{1}_{\searrow,t}$				0.98		1.01										
					(0.332)		(0.332)										
	$\lvert i_t\rvert$				101.27	117.28	103.71										
					(37.72)	(36.67)	(37.5)										
	$\lvert\Delta r_{ff,t}\rvert$				-6.38	11.24	2.19										
					(25.05)	(25.87)	(25.01)										
Q	$\lvert\Delta y_t\rvert$							-30.67	-22.76								
								(54.36)	(38.03)								
	$\lvert\Delta y_t^c\rvert$							27.19		21.30							
								(83.03)		(56.03)							
	$\lvert\Delta u_t\rvert$							20.38			69.62						
								(136.74)			(125.95)						
	$\mathbf{1}_{\searrow,t}$							15.62				15.86					
								(1104.7)				(1105.8)					
	$\lvert i_t\rvert$							42.95	51.83	53.28	56.90	42.82					
								(53.27)	(47.74)	(48.54)	(49.40)	(51.52)					
	$\lvert\Delta r_{ff,t}\rvert$							-18.03	4.90	-3.06	-6.25	-13.71					
								(49.53)	(48.13)	(48.68)	(48.73)	(46.73)					
	Pseudo-R^2	0.271	0.050	0.011	0.043	0.031	0.030	0.017	0.009	0.007	0.006	0.015	0.000	0.000	0.000	0.000	0.000
	Accuracy	0.865	0.774	0.958	0.715	0.715	0.715	0.935	0.935	0.935	0.935	0.935	0.832	0.774	0.958	0.715	0.935
	Observations	7552	359	119	785	785	785	261	261	261	261	261	7552	359	119	785	261

Notes: The explained variable is a regime change variable $\mathbf{1}_{\star,t}$ (produced by parametric Bayesian detection method described in text). The top rows present coefficient estimates for the following explanatory variables: firstly, market variables (separately for daily, D, monthly, M, and quarterly, Q, frequency) including the absolute log market return, the absolute Volatility Index VIX growth (i.e., temporal difference in the log of VIX over 100); secondly, macroeconomic and policy variables (separately for M and Q frequency) including the absolute real GDP growth (i.e., difference in the log of), the absolute change of the cyclical component of the log of real GDP (i.e., its temporal difference), the unemployment rate (i.e., taken as a fraction and then used in levels or in absolute values of temporal differences), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the absolute change of the federal funds rate (i.e., taken as a fraction and then temporally differenced); in both cases, the timing of explanatory variables coincides with that of the explained variable. The bottom three rows provide: goodness-of-fit measures including Pseudo-R^2 (using Efron's definition) and the fraction of accurate predictions (measured as #Correct/#Total), as well as the number of observations. The columns contain different data frequency and regression specifications, with the last five columns providing results for relevant null models. The estimation is conducted by maximum likelihood. Standard errors (in parenthesis) are obtained from the Fisher information matrix. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the federal funds rate, in which case the data sample is 1954:M7:D1–2019:M12:D31; the unemployment rate, in which case it is 1948:M1:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 7: Logistic regression with correlates of regime changes (parametric Bayesian detection) — asynchronous regressors

	Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	constant	-2.32	0.15	2.21	-0.67	-0.68	0.52	1.83	2.75	2.25	1.13	2.09	-1.60	1.23	3.13	0.92	2.66
		(0.051)	(0.24)	(0.765)	(0.33)	(0.328)	(0.13)	(0.665)	(0.581)	(0.468)	(0.542)	(0.41)	(0.03)	(0.13)	(0.46)	(0.08)	(0.25)
D	$\lvert\Delta r_{t+1}\rvert$	68.33															
		(3.35)															
	$\lvert\Delta\varsigma_{t+1}\rvert$	-13.89															
		(3.15)															
M	$\lvert\Delta r_{t}\rvert$		14.69														
			(4.70)														
	$\lvert\Delta\varsigma_{t+1}\rvert$		30.06														
			(9.36)														
Q	$\lvert\Delta r_{t}\rvert$			6.24													
				(8.91)													
	$\lvert\Delta\varsigma_{t}\rvert$			24.09													
				(29.70)													
M	$\lvert u_{t+3}\rvert$				21.64	21.72											
					(5.59)	(5.56)											
	$\mathbf{1}_{\searrow,t+3}$				1.20		1.21										
					(0.356)		(0.355)										
	$\lvert i_{t}\rvert$				82.22	109.02	91.30										
					(39.17)	(37.65)	(38.69)										
	$\lvert\Delta r_{ff,t+2}\rvert$				-7.92	6.91	2.95										
					(25.96)	(27.01)	(26.04)										
Q	$\lvert\Delta y_{t+1}\rvert$							-61.14	-62.53								
								(53.87)	(34.48)								
	$\lvert\Delta y^{c}_{t+1}\rvert$							-25.92		-34.87							
								(71.98)		(43.47)							
	$\lvert\Delta u_{t+2}\rvert$							613.30			621.22						
								(278.57)			(268.27)						
	$\mathbf{1}_{\searrow,t+1}$							14.40				15.68					
								(1599.4)				(1115.2)					
	$\lvert i_{t-2}\rvert$							104.34	84.31	88.38	97.42	70.45					
								(65.70)	(55.16)	(53.95)	(62.65)	(56.10)					
	$\lvert\Delta r_{ff,t}\rvert$							-27.26	-17.74	-14.14	-32.15	-22.28					
								(48.25)	(48.74)	(48.77)	(47.48)	(47.56)					
	Pseudo-R^2	0.097	0.075	0.011	0.050	0.035	0.033	0.075	0.033	0.015	0.036	0.019	0.000	0.000	0.000	0.000	0.000
	Accuracy	0.837	0.774	0.958	0.714	0.714	0.714	0.934	0.934	0.934	0.934	0.934	0.832	0.774	0.958	0.714	0.934
	Observations	7552	359	119	784	784	784	259	259	259	259	259	7552	359	119	784	259

Notes: The explained variable is a regime change variable $\mathbf{1}_{\star,t}$ (produced by parametric Bayesian detection method described in text). The top rows present coefficient estimates for the following explanatory variables: firstly, market variables (separately for daily, D, monthly, M, and quarterly, Q, frequency) including the absolute log market return, the absolute Volatility Index VIX growth (i.e., temporal difference in the log of VIX over 100); secondly, macroeconomic and policy variables (separately for M and Q frequency) including the absolute real GDP growth (i.e., difference in the log of), the absolute change of the cyclical component of the log of real GDP (i.e., its temporal difference), the unemployment rate (i.e., taken as a fraction and then used in levels or in absolute values of temporal differences), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the absolute change of the federal funds rate (i.e., taken as a fraction and then temporally differenced); in both cases, the timing of explanatory variables is adjusted following the results in Table 2. The bottom three rows provide: goodness-of-fit measures including Pseudo-R^2 (using Efron's definition) and the fraction of accurate predictions (measured as #Correct/#Total), as well as the number of observations. The columns contain different data frequency and regression specifications, with the last five columns providing results for relevant null models. The estimation is conducted by maximum likelihood. Standard errors (in parenthesis) are obtained from the Fisher information matrix. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the federal funds rate, in which case the data sample is 1954:M7:D1–2019:M12:D31; the unemployment rate, in which case it is 1948:M1:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 8: Logistic regression with correlates of regime changes (non-parametric detection) — contemporaneous regressors

	Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	constant	-5.50	-3.32	-2.77	-1.35	-1.35	-2.22	-0.98	-1.05	-1.20	-1.14	-1.20	-5.34	-2.43	-1.38	-1.74	-0.71
		(0.223)	(0.353)	(0.48)	(0.417)	(0.416)	(0.163)	(0.302)	(0.293)	(0.243)	(0.245)	(0.222)	(0.17)	(0.19)	(0.23)	(0.1)	(0.13)
D	$\|\Delta r_t\|$	-19.98															
		(15.30)															
	$\|\Delta \varsigma_t\|$	30.10															
		(6.70)															
M	$\|\Delta r_t\|$		7.30														
			(4.55)														
	$\|\Delta \varsigma_t\|$		18.01														
			(5.79)														
Q	$\|\Delta r_t\|$			6.82													
				(3.00)													
	$\|\Delta \varsigma_t\|$			20.55													
				(7.40)													
M	$\|u_t\|$				-15.78	-15.66											
					(7.10)	(7.09)											
	$\mathbf{1}_{\searrow,t}$				0.10		0.07										
					(0.299)		(0.298)										
	$\|i_t\|$				108.25	111.20	105.62										
					(36.42)	(35.36)	(36.15)										
	$\|\Delta r_{ff,t}\|$				42.41	43.65	34.38										
					(21.45)	(21.14)	(20.51)										
Q	$\|\Delta y_t\|$							-20.96	-17.05								
								(27.64)	(21.31)								
	$\|\Delta y_t^c\|$							23.13		-0.46							
								(37.33)		(25.79)							
	$\|\Delta u_t\|$							-59.47			-33.61						
								(66.92)			(56.22)						
	$\mathbf{1}_{\searrow,t}$							0.33				0.33					
								(0.46)				(0.407)					
	$\|i_t\|$							33.05	40.12	40.41	39.76	36.36					
								(20.62)	(19.65)	(19.70)	(19.63)	(20.26)					
	$\|\Delta r_{ff,t}\|$							22.23	22.81	21.67	25.92	18.82					
								(22.19)	(20.75)	(21.71)	(22.26)	(20.81)					
	Pseudo-R^2	-0.008	0.063	0.182	0.036	0.036	0.029	0.044	0.039	0.036	0.037	0.038	0.000	0.000	0.000	0.000	0.000
	Accuracy	0.995	0.919	0.824	0.851	0.851	0.851	0.682	0.670	0.670	0.674	0.690	0.995	0.919	0.798	0.851	0.670
	Observations	7552	359	119	785	785	785	261	261	261	261	261	7552	359	119	785	261

Notes: The explained variable is a regime change variable $\mathbf{1}_{\star,t}$ (produced by non-parametric detection method described in text). The top rows present coefficient estimates for the following explanatory variables: firstly, market variables (separately for daily, D, monthly, M, and quarterly, Q, frequency) including the absolute log market return, the absolute Volatility Index VIX growth (i.e., temporal difference in the log of VIX over 100); secondly, macroeconomic and policy variables (separately for M and Q frequency) including the absolute real GDP growth (i.e., difference in the log of), the absolute change of the cyclical component of the log of real GDP (i.e., its temporal difference), the unemployment rate (i.e., taken as a fraction and then used in levels or in absolute values of temporal differences), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the absolute change of the federal funds rate (i.e., taken as a fraction and then temporally differenced); in both cases, the timing of explanatory variables coincides with that of the explained variable. The bottom three rows provide: goodness-of-fit measures including Pseudo-R^2 (using Efron's definition) and the fraction of accurate predictions (measured as #Correct/#Total), as well as the number of observations. The columns contain different data frequency and regression specifications, with the last five columns providing results for relevant null models. The estimation is conducted by maximum likelihood. Standard errors (in parenthesis) are obtained from the Fisher information matrix. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the federal funds rate, in which case the data sample is 1954:M7:D1–2019:M12:D31; the unemployment rate, in which case it is 1948:M1:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

Table 9: Logistic regression with correlates of regime changes (non-parametric detection) — asynchronous regressors

	Regressors	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	constant	-6.06	-3.39	-2.77	-1.59	-1.62	-2.20	-1.36	-1.37	-1.49	-1.17	-1.02	-5.34	-2.39	-1.38	-1.74	-0.72
		(0.222)	(0.364)	(0.48)	(0.407)	(0.405)	(0.163)	(0.317)	(0.305)	(0.261)	(0.236)	(0.217)	(0.17)	(0.19)	(0.23)	(0.1)	(0.13)
D	$\lvert\Delta r_{t+1}\rvert$	37.10															
		(7.31)															
	$\lvert\Delta\varsigma_{t+1}\rvert$	7.73															
		(7.84)															
M	$\lvert\Delta r_t\rvert$		8.82														
			(4.30)														
	$\lvert\Delta\varsigma_{t+1}\rvert$		18.82														
			(5.80)														
Q	$\lvert\Delta r_t\rvert$			6.82													
				(3.00)													
	$\lvert\Delta\varsigma_t\rvert$			20.55													
				(7.40)													
M	$\lvert u_{t+3}\rvert$				-11.09	-10.47											
					(6.93)	(6.88)											
	$\mathbf{1}_{\searrow,t+3}$				0.74		0.72										
					(0.275)		(0.274)										
	$\lvert i_t\rvert$				94.06	123.80	88.27										
					(38.66)	(36.86)	(38.32)										
	$\lvert\Delta r_{ff,t+2}\rvert$				12.84	19.43	7.79										
					(21.40)	(20.89)	(20.96)										
Q	$\lvert\Delta y_{t+1}\rvert$							-8.80	34.82								
								(27.45)	(20.67)								
	$\lvert\Delta y^c_{t+1}\rvert$							96.72		95.83							
								(35.81)		(26.37)							
	$\lvert\Delta u_{t+2}\rvert$							-31.01			88.39						
								(70.56)			(52.74)						
	$\mathbf{1}_{\searrow,t+1}$							0.60				0.93					
								(0.561)				(0.403)					
	$\lvert i_{t-2}\rvert$							1.42	17.47	9.14	8.33	0.50					
								(23.07)	(21.44)	(21.89)	(21.58)	(22.40)					
	$\lvert\Delta r_{ff,t}\rvert$							22.98	29.23	21.42	30.30	32.01					
								(24.05)	(23.15)	(24.02)	(22.42)	(22.70)					
	Pseudo-R^2	0.001	0.056	0.182	0.035	0.026	0.033	0.082	0.033	0.075	0.032	0.043	0.000	0.000	0.000	0.000	0.000
	Accuracy	0.995	0.914	0.824	0.852	0.852	0.853	0.699	0.691	0.699	0.676	0.676	0.995	0.916	0.798	0.851	0.672
	Observations	7552	359	119	784	784	784	259	259	259	259	259	7552	359	119	784	259

Notes: The explained variable is a regime change variable $\mathbf{1}_{\star,t}$ (produced by non-parametric detection method described in text). The top rows present coefficient estimates for the following explanatory variables: firstly, market variables (separately for daily, D, monthly, M, and quarterly, Q, frequency) including the absolute log market return, the absolute Volatility Index VIX growth (i.e., temporal difference in the log of VIX over 100); secondly, macroeconomic and policy variables (separately for M and Q frequency) including the absolute real GDP growth (i.e., difference in the log of), the absolute change of the cyclical component of the log of real GDP (i.e., its temporal difference), the unemployment rate (i.e., taken as a fraction and then used in levels or in absolute values of temporal differences), the binary variable capturing US recessions, the absolute CPI inflation rate (i.e., difference in the log of), the absolute change of the federal funds rate (i.e., taken as a fraction and then temporally differenced); in both cases, the timing of explanatory variables is adjusted following the results in Table 3. The bottom three rows provide: goodness-of-fit measures including Pseudo-R^2 (using Efron's definition) and the fraction of accurate predictions (measured as #Correct/#Total), as well as the number of observations. The columns contain different data frequency and regression specifications, with the last five columns providing results for relevant null models. The estimation is conducted by maximum likelihood. Standard errors (in parenthesis) are obtained from the Fisher information matrix. The data sample is U.S. 1947:M4:D1–2019:M12:D31 (with the exception of the federal funds rate, in which case the data sample is 1954:M7:D1–2019:M12:D31; the unemployment rate, in which case it is 1948:M1:D1–2019:M12:D31; and VIX, in which case it is 1990:M1:D2–2019:M12:D31), at a daily, monthly and quarterly frequency, starting from the highest frequency available and aggregating appropriately when necessary (see Appendix §G for a description of data sources).

E Shrinkage

In the multidimensional setup, variance-covariance matrix shrinkage implies regularizing its eigenvalues by squeezing them together (reducing the largest and amplifying the smallest ones), which improves matrix conditioning and helps its inversion and hence usage in, say, portfolio optimization.

(There is an established result in random matrix theory that eigenvalues of sample valiance-covariance matrixes are overdispersed: the largest sample eigenvalue asymptotically overestimates the largest population eigenvalue, and the smallest sample eigenvalue underestimates its population counterpart. This result is based on Marchenko-Pastur and Wigner's semicircle distribution laws, for reference see Stein (1975 or 1986) and Johnstone (2001), also see Ledoit and Wolf (2013).)

In practice a regularized variance-covariance matrix is usually obtained as a convex combination between the sample variance-covariance matrix and some "well-behaved" counterpart (e.g., an identity matrix). Importantly, a matrix condition number (defined as the ratio of the largest and the smallest eigenvalues) quantifies the asymmetry, the relative spread along its principal axes and is, roughly speaking, inversely related to its distance from singularity (see, e.g., Horn and Johnson, 1985). So, improving the variance-covariance matrix conditioning by reducing this number in some sense amplifies the corresponding random variable's dispersion.

This resonates strongly with our result on a Student's t mixture posterior, suggesting that in this case random matrix theory-motivated eigenvalue regularization and Bayesian indifference priors are mutually consistent approaches leading in the same direction, which gives another possible reason behind the popularity and good performance of the variance-covariance matrix shrinkage in finance.

References for Shrinkage

[1] Horn, Roger A. and Charles R. Johnson. (1985) *Matrix Analysis*, Cambridge University Press.

[2] Johnstone, Ian M. (2001) "On the Distribution of the Largest Eigenvalue in Principal Components Analysis." *Annals of Statistics*, 29(2): 295–327.

[3] Ledoit, Olivier and Michael Wolf. (2013) "Spectrum Estimation: A Unified Framework for Covariance Matrix Estimation and PCA in Large Dimensions", *SSRN Working Paper*, No. 2198287.

[4] Stein, Charles M. (1975) "Estimation of a Covariance Matrix", Rietz Lecture, 39th Annual Meeting of the Institute of Mathematical Statistics. Atlanta, GA.

[5] Stein, Charles M. (1986) "Lectures on the Theory of Estimation of Many Parameters." *Journal of Soviet Mathematics*, 34(1): 1373–1403.

F Proofs

F.1 Proof of Proposition 1

Proof. Integrating p with respect to measure (9) produces equation (10). From the resulting expression it follows that the mean will be larger than q as long as $q < \alpha/(\alpha+\beta)$, and it will be smaller otherwise. $\square$

F.2 Proof of Proposition 2

Proof. Using likelihood $L(p_2|p_1)$ from equation (11) and a Beta prior updated with the communicated information $\mathcal{B}(\alpha+n_1 q, \beta+n_1(1-q))$ yields the first equation (using instead the full likelihood $L(\iota|p_2)L(p_2|p_1)$ also works). Using $L(\iota|p_2)$ from (11) and a prior updated with the processed information $\mathcal{B}(\alpha + n_2\mathrm{E}[p_1|q], \beta + n_2(1 - \mathrm{E}[p_1|q]))$ yields the second equation. The correspondence between values of $n_\cdot$ is established by finding the value that equalizes the solutions for $\mathrm{E}[p_2|\mathrm{E}[p_1|q]]$ and for $\mathrm{E}[p_2|q]$. $\square$

F.3 Proof of Proposition 3

Proof. Take the initial Beta prior and update it as elaborated in the text (if necessary, sequentially), which gives $\mathcal{B}(\alpha+\sum_s n_s q_s, \beta+\sum_s n_s(1-q_s))$; combine the earlier expression with the likelihood (7) or (11). Normalize the resulting joint probability distribution to obtain a proper conditional distribution $\pi(p|\boldsymbol{q})$ or $\pi(p_2|\boldsymbol{q})$. Integrate p or p_2 with respect to the conditional distribution obtained earlier. $\square$

F.4 Proof of Lemma 1

Proof. Equations (16)–(19) are definitions, and they can be verified to be consistent with the outlined procedure. The posterior distribution $\pi(\mu, \sigma^2|\ell_\tau, \lambda_\tau, \alpha_\tau, \beta_\tau)$ is computed from the likelihood (13) as well as priors (14) and (15), utilizing the Bayes rule and four equations above. The marginal posterior distribution (20) can be disentangled from the joint distribution exploiting the conjugacy, then the marginal distribution (21) can be identified after integrating out σ^2 from the joint distribution. $\square$

F.5 Proof of Lemma 2

Proof. Multiplying the Gaussian likelihood (13) for period $(\tau+1)$ by the $\mathcal{IG}(\alpha_\tau, \beta_\tau)$ times $\mathcal{N}(\ell_\tau, \sigma^2/\lambda_\tau)$ prior updated with equations (16–19) from Lemma 1 and integrating out μ and σ^2 gives the result. $\square$

F.6 Proof of Proposition 4

Proof. The derivation of each of the mixture's Student's t components has been provided earlier in the text, see Lemma 2 and the arguments leading to this Proposition. In case of $p < 1$, tails of the mixture are heavier than the Gaussian ones because this is true for each of its components; the case of $p = 1$ is trivial. $\square$

G Data

Consumption-based asset pricing calculations

Real per capita consumption growth series is constructed using quarterly data on nominal seasonally adjusted at annual rates Personal Consumption Expenditures for Nondurable Goods and Services, corresponding seasonally adjusted Price Indexes, as well as Population size from U.S. Bureau for Economic Analysis's NIPA tables. Time period covered: 1947:Q2—2019:Q4.

Real risk-free and market returns are constructed using nominal Risk-free and Market returns from Fama-French online data library, converted from monthly into quarterly frequency, as well as the Personal Consumption Expenditures for Nondurable Goods and Services Deflator applied to nominal Consumption series as described above. Time period covered: 1947:Q2—2019:Q4.

Real dividend growth series is constructed using nominal Dividends for S&P500 Composite at quarterly frequency from Robert Shiller's online data collection as well as the Personal Consumption Expenditures for Nondurable Goods and Services Deflator applied to nominal Consumption series as described above. Time period covered: 1947:Q2—2019:Q4.

Changepoint detection calculations

Real GDP growth series is constructed using quarterly data on Real seasonally adjusted at annual rate Gross Domestic Product from U.S. Bureau for Economic Analysis's NIPA tables. Time period covered: 1947:Q2—2019:Q4.

The cyclical component of the log of real GDP is constructed from quarterly Real Gross Domestic Product data described above using the Hodrick-Prescott filter with the smoothing parameter $\lambda = 1600$. Time period covered: 1947:Q2—2019:Q4.

The unemployment rate series is monthly seasonally adjusted data from U.S. Bureau of Labor Statistics's Current Population Survey (Household Survey), converted from monthly into quarterly frequency when necessary. Time periods covered: 1948:M1—2019:M12 and 1948:Q1—2019:Q4.

The U.S. recessions binary variable is monthly and quarterly data on NBER based Recession Indicators for the United States from the Peak through the Period preceding the Trough constructed by the Federal Reserve Bank of St. Louis as an interpretation of US Business Cycle Expansions and Contractions data provided by the National Bureau of Economic Research. Time periods covered: 1947:M4—2019:M12 and 1947:Q2—2019:Q4.

The CPI inflation rate is constructed from monthly seasonally adjusted data on "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average" from the

U.S. Bureau of Labor Statistics, converted from monthly into quarterly frequency when necessary. Time periods covered: 1947:M4—2019:M12 and 1947:Q2—2019:Q4.

The federal funds rate is the Federal Funds Effective Rate at daily and monthly frequencies from the Board of Governors of the Federal Reserve System, converted from monthly into quarterly frequency when necessary. Time periods covered: 1954:M7:D1–2019:M12:D31, 1954:M7–2019:M12 and 1954:Q2–2019:Q4.

Market returns are nominal market returns at daily and monthly frequencies from Fama-French online data library, converted from monthly into quarterly frequency when necessary. Time periods covered: 1947:M4:D1–2019:M12:D31, 1947:M4—2019:M12 and 1947:Q2—2019:Q4.

The Volatility Index (VIX) change series is constructed using daily data on the CBOE Volatility Index from Chicago Board Options Exchange, converted from daily into monthly and quarterly frequencies when necessary. Time periods covered: 1990:M1:D2–2019:M12:D31, 1990:M1—2019:M12 and 1990:Q1—2019:Q4.