Journal of Financial Econometrics, 2023, Vol. 21, No. 5, 1728–1758
https://doi.org/10.1093/jjfinec/nbac021
Advance Access Publication Date: 24 June 2022
Article

OXFORD

Anatomy of a Sovereign Debt Crisis: Machine Learning, Real-Time Macro Fundamentals, and CDS Spreads*

Pierluigi Balduzzi[1], Roberto Savona [2], and Lucia Alessi[3]

[1]Boston College, Newton, MA, USA, [2]Department of Economics and Management, University of Brescia, Brescia, Italy and [3]European Commission—Joint Research Centre, Ispra, Italy

Address correspondence to Pierluigi Balduzzi, Carroll School of Management, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, or e-mail: balduzzp@bc.edu.

Received July 31, 2020; revised June 1, 2022; editorial decision June 7, 2022; accepted June 7, 2022

Abstract

We employ a Least Absolute Shrinkage and Selection Operator (LASSO)-based extension of the Fama–MacBeth procedure to characterize the time-varying dependence of sovereign Credit Default Swap (CDS) spreads on macro indicators during the samples 2009–2013 and 2013–2020. While CDS spreads are mainly reflective of fundamentals, this relationship varies substantially over time, leading to price variation that appears unrelated to fundamentals. The estimated LASSO coefficients are used to endogenously identify macro-sensitivity "regimes" of variation, consistently with a multiple-equilibrium view of the sovereign debt markets.

Key words: CDS spreads, LASSO, macroeconomic fundamentals

JEL classification: G12

This article investigates the possibly time-varying relationship between prices—quoted spreads on sovereign Credit Default Swap (CDS) contracts—and underlying economic fundamentals—country-specific macroeconomic indicators—using a machine-learning extension of the well-known Fama–Macbeth (Fama and MacBeth 1973) cross-sectional regression approach. Two are the innovations relative to the Fama–Macbeth methodology. First, we let the data identify the set of covariates relevant for pricing. We do so by

* Part of this work was done while Lucia Alessi was affiliated with the European Central Bank. The views in this article are those of the authors and do not necessarily reflect those of the European Commission. We thank Silvia Ardagna, Patrick Augustin, Jean-Paul Renne, Marco Sandri, and seminar participants at the Joint Research Centre of the European Commission and the Single Resolution Board, as well as participants at the 2018 Annual Conference of the Community of Practice in Financial Research, for useful comments.

implementing a series of cross-sectional Least Absolute Shrinkage and Selection Operator (LASSO) regressions, one for each daily or monthly time-series observation. Second, rather than modeling a cross-section of returns, we effectively model a cross-section of *prices*. Returns on financial assets are notoriously noisy and they reflect *innovations* in fundamentals. Thus, modeling returns is intrinsically challenging and requires taking a stand on market expectations. Prices, on the other hand, simply reflect the most current market information, without the need to disentangle the news component. Moreover, by focusing on prices, we can potentially identify relevant covariates whose effect evolves only smoothly over time and that could be missed by the cross-sectional analysis of returns, which requires pronounced time-series variability in such effects.

Our main application is the study of the "anatomy" of the eurozone sovereign debt crisis. Specifically, we carry out a detailed analysis of the relationship between the pricing of sovereign risk and a comprehensive set of real-time macroeconomic fundamentals. To our knowledge, we are the first to perform this type of exercise. While machine-learning techniques have been employed to explain equity returns, both in the time-series and the cross-section (e.g., see Chinco, Clark-Joseph, and Ye 2019; Feng, Giglio, and Xiu 2019), there are no applications to the cross-section of CDS spreads. The LASSO is especially useful in this context, as the size of the cross-section of sovereign CDSs is small relative to the number of potentially relevant macroeconomic indicators, making the standard OLS estimator unfeasible.

Our procedure highlights the role of macro fundamentals in explaining the relative level of CDS spreads during the sovereign debt crisis. Importantly, the explanatory power of the different macro variables changes over time, both across and within "regimes." Indeed, we show that what matters for the pricing of sovereign CDS spreads is *both* the *level* of a country's economic fundamentals and the *relevance* that markets attribute to different fundamentals. We show that the relevance of fundamentals is extremely low at the outset of the crisis: in this phase, markets "panic," with some countries being penalized for their mere belonging to the GIPSI (Greece, Ireland, Portugal, Spain, and Italy) group of countries identified as vulnerable. This explains why an abrupt, substantial repricing of risks may take place against unchanged, or only marginally deteriorated, economic conditions. On the contrary, at the height of the crisis, attention to economic fundamentals becomes extreme. Past the peak, we get back to a virtual disconnect between market developments and macro fundamentals, eventually returning to low CDS spreads after Brexit and during the Covid-19 pandemic. Our results highlight the role of the ECB's unconventional monetary policy in reducing spreads across the board in a time of crisis, by providing the proverbial "tide that lifts all boats" (e.g., De Grauwe and Ji 2013). However, individual countries can ultimately improve their funding costs only by intervening in their macro fundamentals.

It is worth noting that the existing work on the relation between CDS spreads and macro fundamentals has used *revised* macroeconomic data. The very few papers employing real-time data have focused on macro-news announcements (e.g., Beetsma et al. 2013; Kim, Salem, and Wu 2015), relating changes in sovereign CDS spreads to the "distance" between released and expected quantities. However, this approach suffers from two main limitations. First, given that macro announcements are not synchronized across countries, it is not possible to implement a pure cross-sectional analysis of the responses of CDS spreads to the news. Hence, if there is time variation in these responses, the time variation needs to be modeled explicitly. Second, modeling changes in CDS spreads are tantamount to modeling returns and, as argued at the beginning of this section, this requires data on market expectations. However, the data on consensus forecasts needed to construct the "surprise" components of the announcements are available for only a few macroeconomic variables and countries. Hence, any feasible empirical analysis can only be limited in scope.

In our analysis, on the other hand, we can use data for *all* trading days and a *rich* cross-section of macro indicators; see Section 1 for a full description of our real-time data set used in the analysis. This approach allows us to accommodate time-variation in the relation between spreads and fundamentals, without having to model such relation explicitly. This is important, as we document that the variation in macro-sensitivities—the coefficients relating the pricing of default risk to macro fundamentals—is substantial, with the set of relevant variables itself varying drastically from one "regime" to another. Moreover, since we model the level of CDS spreads, rather than their variation, we can include in the analysis a broad set of indicators, for which consensus forecasts are not available and surprises cannot be easily computed. This is also important because, at times, we find that less-known macro indicators play an important role in explaining the cross-section of CDS spreads. Specifically, we use the European Central Bank (ECB) e-archives to construct a unique real-time, daily-frequency data set on 19 macroeconomic fundamentals, for 11 eurozone countries over the period from May 11, 2009 to April 25, 2013. We then relate the cross-section of sovereign CDS spreads of different maturity to the macro fundamentals, employing data for all the trading days in the sample.

In the following, we summarize our main contributions. Our *first* contribution is to characterize and interpret the cross-section of sovereign CDS spreads based on our real-time macroeconomic data set. Analytically, we implement a Fama–MacBeth procedure in which, for each day, the sovereign CDS spreads for the 11 eurozone countries at 3, 5, 7, and 10 years are regressed on the 19 country-specific macro fundamentals, controlling for the level and volatility of an indicator of banking valuations, and controlling for being part of the GIPSI group of countries. These covariates are employed individually, to capture "level" effects, and then they are interacted with the maturities of the CDS contracts, to capture "slope" effects.

Given the large initial set of covariates, we implement a LASSO approach to the regression analysis (Tibshirani 1996). In doing this, we reduce the dimensionality of the space of covariates, making the estimation procedure feasible. Moreover, the day-by-day variable selection procedure eliminates those covariates that do not contribute to explaining the cross-section of CDSs. The daily LASSO cross-sectional coefficients are then stacked together, producing the time series of macro-sensitivities whose values tell us which variables are important at any given time.[1] Our empirical strategy allows for the parameters of the relationship between CDS spreads and macro fundamentals to vary over time, resulting in an exceptionally good cross-sectional fit, with an average cross-sectional R^2 of 98.5%. For comparison, Aizenman, Hutchison, and Jinjarak (2013) employ a panel-data estimator

1 The use of LASSO to identify the "salient" macroeconomic information that market participants respond to is supported by recent literature that focuses on the role of rational inattention and information rigidities; see, for example, Sims (2003), Ball, Mankiw, and Reis (2005), and Gabaix (2014). This new paradigm relaxes the assumption of common information or full information processing based on costs associated with acquiring and processing information.

with time-fixed effects, but *constant* coefficients, explaining only between 46% and 52% of the time-series variability of CDS spreads.

Hence, a flexible approach allowing for time-variation in the dependence of CDS spreads on macro fundamentals is important. Indeed, the pronounced time variation in the macro sensitivities that we document reconciles the high degree of co-movement in sovereign CDS spreads with the relatively low correlations between domestic macro fundamentals.[2]

Our *second* contribution is to use the daily time series of the LASSO coefficients to identify homogeneous clusters of observations. Our approach differs from existing papers which instead identify non-crisis vs. crisis regimes based on the behavior of CDS/bond spreads alone (e.g., Blommestein, Eijffinger, and Qian 2016; Delatte, Fouquau, and Portes 2017).[3] Starting from April 2010, the time series of the macro-sensitivities lead us to identify three main crisis regimes. Notably, the GIPSI dummy turns out to be the key variable at the onset of the crisis, while at the peak of the crisis the cross-sections of sovereign CDS were mostly driven by GDP growth and employment. Interestingly, it is during the highest risk regime that CDS spreads reflect macro fundamentals the most. In other words, it is precisely when the overall perception of sovereign default risk is greatest that country-specific macro information impacts CDS spreads the most, consistent with the notion that these are the times when the returns from information production are also the highest.

Our analysis and results are related to Bernoth and Erdogan (2013), who employ a time-varying coefficients regression model to explain the level of sovereign CDS spreads. However, the pronounced volatility in the macro-sensitivities that we document is at odds with their conclusions that the relation between CDS spreads and fundamentals is "changing gradually over time, rather than having a discrete break-point between regimes." Afonso et al. (2018) employ a similar approach and extend the analysis to the sample period following Draghi's "whatever it takes" speech. They find that the introduction of a program of open market transactions on the part of the ECB significantly reduced the sensitivity of CDS spreads to macro fundamentals. Our econometric approach uncovers a different pattern of variation, with distinct "regimes" and substantial variation in the estimated macro-sensitivities, both within and across regimes.

Our empirical results are supportive of the implications of theoretical models that account for feedback from the cost of borrowing to the probability of default, leading to a multiplicity of equilibria (see Calvo 1988; Kehoe and Cole 2000; Corsetti and Dedola 2016). Some authors have argued that a significant part of the surge in the euro area sovereign yield spreads over the period 2010–2012 was disconnected from underlying fiscal fundamentals (De Grauwe and Ji 2012; Favero 2013; De Santis 2014; Di Cesare et al. 2015; Dewachter et al. 2015), others have argued that sovereign bond prices and CDS spreads exhibited excessive sensitivity to macroeconomic indicators (Aizenman, Hutchison, and Jinjarak 2013; Bernoth and Erdogan 2013), and, finally, other authors have tried to

2 Indeed, Longstaff et al. (2011) show that sovereign CDS spreads can be strongly correlated across markets, even if local equity returns are not, while Augustin (2018) finds that the importance of local equity returns in explaining changes in CDS spreads decreases with the slope of the term structure of CDS spreads.

3 As a practical matter, it turns out that the main driver of the identification of the regimes is the cross-sectional average level of CDS spreads.

disentangle fundamental-based spreads from trading momentum behavior, which contributed massively to the observed higher spreads during the crisis (Chiarella et al. 2015).[4]

We also extend our analysis to the May 2013–December 2020 period, where we see much lower cross-sectional dispersion and volatility in CDS spreads overall, relative to the earlier sample.[5] The main patterns are a temporary increase in macro-sensitivities due to the outcome of the U.K. EU membership referendum (Brexit) and in anticipation of the U.S. Presidential election. Macro-sensitivities are otherwise moderate and stable, even during the Covid-19 pandemic.

The remainder of this article is organized as follows. Section 1 describes our real-time, macro indicator data set. Section 2 derives the empirical specification used in the cross-sectional regression analysis and discusses the methodology. Section 3 presents the results of the analysis for the 2009–2013 sample. Section 4 discusses the methodology employed for the identification of the different regimes and the related evidence. Section 5 presents the results of the analysis for the 2013–2020 sample. Section 6 explores extensions and robustness checks of our analysis. Section 7 concludes.

1 Macroeconomic Data in Real-Time

The dependent variable is the premium or *spread* on dollar-denominated, ISDA 2003, sovereign CDSs at the 3-, 5-, 7-, and 10-year maturity, over the May 11, 2009–April 25, 2013 period, for a set of 11 eurozone countries.[6] CDSs are financial contracts aimed at protecting the buyer of a bond from the default risk of the issuer. As pointed out in Longstaff et al. (2011), an important advantage of using sovereign CDS data (rather than sovereign bond data) is that the sovereign CDS market is typically more liquid than the corresponding sovereign bond market, resulting in more accurate estimates of perceived default risk. Our analysis covers the following eurozone countries: Austria, Belgium, Cyprus, Germany, Finland, France, Ireland, Italy, the Netherlands, Spain, and Portugal. As in other studies using sovereign CDS data (e.g., De Santis 2019), Greece is not covered as reliable data on Greek CDSs is not available after September 2011.[7]

We aim at explaining the dynamics in sovereign CDS prices based on a novel real-time data set comprising country-specific macroeconomic fundamentals and macro-financial indicators. Several studies have tried to explain sovereign debt crises by linking some measures of sovereign stress to macro fundamentals. While earlier papers in this literature have used standard, low-frequency data sets (e.g., Manasse and Roubini 2009), only recently higher-frequency data have been used for this purpose (Beber, Brandt, and Luisi 2014).

As to the advantage of using a real-time dataset, there is a growing literature showing the importance of considering non-revised macroeconomic data for forecasting not only

4 See also the role of fundamentals' contagion (Beirne and Fratzscher 2013) and feedback loops between sovereign and domestic bank risks (Bolton and Jeanne 2011; Acharya et al. 2018).

5 For this period we rely on publicly available data at the *monthly frequency*.

6 Based on conversations with market participants, we use data on dollar-denominated contracts, as they are more liquid than the euro-denominated counterparts, and we avoid using the 1- and 2-year contracts, as they are illiquid.

7 The Greece CDS was "triggered" in March of 2012 and then the market effectively disappeared until 2014, when negotiations restarted, but with generally very low notional principal outstanding.

macro variables themselves (e.g., Giannone, Reichlin, and Small 2008, and, more recently, Beber, Brandt, and Luisi 2014) but also financial variables (see Ghysels, Horan, and Moench 2018). Real-time data sets are also needed to construct credible early warning models for financial crises (e.g., Alessi and Detken 2011; Alessi and Detken 2014), as these models are intended to be used by policymakers based on the information set that is available to them at each point in time.

For these reasons, real-time data sets are becoming increasingly popular also beyond the macro-econometric literature. However, to our knowledge, no existing real-time data sets cover many countries and macroeconomic variables at high frequency. For the euro area, in particular, the most well-known real-time database is the one maintained by the Euro Area Business Cycle Network (see Giannone et al. 2012), which also started publishing real-time data for some individual European countries. This data set, however, reports macroeconomic indicators at a monthly frequency, and hence it is not suitable to pin down the high-frequency impact of data releases on the financial markets.

Against this background, we construct a novel real-time data set at a daily frequency, covering 11 eurozone countries and including the 19 macroeconomic and macro financial indicators (see Table A.1 of the Appendix for details). Specifically, our real-time data set covers the following indicators:

- Labor market: unemployment and employment rates.
- Prices and costs: inflation rate, industrial producer prices (% change), hourly labor cost (% change).
- Money, credit, and debt: growth of M3, loans to the private sector, loans to the government, total credit to the private sector, and total credit to government, as well as public sector deficit over GDP.
- Output: real GDP, private consumption, government expenditures, investment, exports, imports, and industrial production (all rates of growth) and changes in inventories over nominal GDP.

The panel of time series is balanced, with all the above variables being available for each of the considered countries over the whole time span (see Tables A.2 and A.3 of the Appendix for summary statistics). In contrast, the data set used, for example, by Beber, Brandt, and Luisi (2014) includes a comparable number of indicators for Germany, but many less for the other countries.

2 Sovereign Risk Pricing

2.1. Empirical Specification

We start from a general framework for the pricing of CDSs in an arbitrage-free setting, which motivates the specification used in the empirical analysis. For given country n, we express the *upfront* spread on τ-maturity sovereign CDS with spread $s_{n\tau t}$ as

$$s_{n\tau t} = \delta_t(\tau)^\top x_{nt}, \tag{1}$$

where x_{nt} denotes a $(K+2)\times 1$ vector of K country-specific macro variables plus the GIPSI dummy and a constant. We also assume:

$$\delta_t(\tau) = \delta_{1t} + \delta_{2t}\tau. \tag{2}$$

In this way, we provide a simple model of the term structure of sovereign CDS spreads, where the coefficients δ_{1t} and δ_{2t} capture a capture a "level" and a "slope" effect, respectively.

The empirical specification used in our analysis results directly from the expression derived for the τ-maturity CDS spread. Based on Equations (1) and (2), we have

$$s_{nmt} = \delta_{1t}^{\top} x_{nt} + \delta_{2t}^{\top}(x_{nt}\tau_m) + \epsilon_{nmt},\ t = 1, \ldots, T, \tag{3}$$

where, in the daily analysis, $T = 1034$. Note that the dimension of each cross-section at time t is given by $N \times M$, where N is the number of countries and M is the number of CDS maturities. Therefore, since in our sample we have 11 countries and 4 maturities, we then have a total of $11 \times 4 = 44$ cross-sectional observations at every time t. This relatively small cross-section complicates the estimation process as the $K+2$ (21 macro indicators + 1 GIPSI dummy + constant =) 23 country-specific variables in x_{nt} entering in Equation (3) both alone and then interacted with τ, for a total of 46 covariates.

We address the issue above by implementing the LASSO penalty regression (Tibshirani 1996). In this way, we characterize the cross-sections of sovereign CDS spreads while selecting a small number of covariates that explain most of the variation in the response. Agnostically, we assume that although all of the available macro variables are in principle equally important, we expect that in any cross-section only a small number matters and that over time the sub-set of relevant covariates can change based on changing investors' views.

2.2. Econometric Methodology

Analytically, the LASSO algorithm estimates regression parameters by imposing a constraint on the sum of absolute values of the coefficients (excluding the constant). In our context, cross-sectional regressions (Equation (3)) are estimated by solving the following problem:

$$\min_{\delta_{1t},\delta_{2t}} \sum_{n=1}^{N}\sum_{m=1}^{M}\{s_{nmt} - [\delta_{1t}^{\top}x_{nt} + \delta_{2t}^{\top}(x_{nt}\times\tau_m)]\}^2,\ t = 1, \ldots, T, \tag{4}$$

subject to $\sum_{k=1}^{K+1}|\delta_{1kt}| + \sum_{k=1}^{K+2}|\delta_{2kt}| \leq c$, where c is the tuning parameter which "shrinks" coefficient estimates and in some cases forces them to equal zero.[8] Smaller values of the tuning parameter c in Equation (4) restrict the dimension of the parameter space by forcing more coefficients to zero, while larger values tend to include more covariates up until convergence to the OLS solution.

Since c controls the complexity of the model, a key issue is how to select the best value for this parameter. As pointed out in Chinco, Clark-Joseph, and Ye (2019), there is no a priori theoretically optimal value for c. Therefore, we rely on the standard cross-validation procedure, through which the data set (the cross-section of sovereign CDS contracts) is split into two sub-sets, using one sub-set (the *training set*) to estimate the model and then judge the goodness of the prediction based on the remainder of the data (the *test set*) (further details on the estimation are provided in Section A of the Online Appendix).

Note that our approach imposes no penalty nor restriction on the differences in covariate selection and in the values of the associated coefficients, between any two observations.

8 The intercept term does not enter the penalty function and is estimated separately, by minimizing the objective for a given value of the other coefficient estimates.

We view this as an *advantage* of our approach, in that it allows us to be agnostic regarding the homogeneity of different groups of observations and lets us separate the estimation of the model (Equation (3)) from the identification of different regimes of variation.

3 Results over the Period 2009–2013

3.1. Time-Varying Sensitivities to the Macro Factors

The daily time series of the most important cross-sectional LASSO coefficients are displayed in Figure 1 (Figures IA.1–IA.4 of the Online Appendix report the daily time series for all other coefficients). Since the covariates are standardized by their cross-sectional standard deviation before running the procedure, coefficients are scale-independent, which helps to assess the importance of each macro-fundamental based on the absolute value of the coefficient itself.

The significance of the LASSO estimates, computed each day through a resampling-approximation permutation test (see the next section for details), is confirmed for *all* coefficients at 5% level, and for most coefficients at the 1% level. Columns "0.05 level" and "0.01 level" in Tables 1 and 2 report the number of times in which the coefficient estimate is outside the $1-99\%$ and $5-95\%$ bounds, respectively, expressed as the ratio over the total number of observations in which the variable was selected by the LASSO algorithm. Also, note that in Figure 1 the estimated cross-sectional coefficients are most of the times outside the 1% confidence bands (gray lines). A similar approach is implemented in Figure 2 to evaluate the joint significance of all non-zero LASSO coefficients, by comparing the R^2s of the model selected by the LASSO to the distribution of R^2s under the null. As in the case of the individual point estimates, these indicators of cross-sectional fit are strongly significant.

At first glance, the dynamics of the stacked coefficients exhibit significant time variation, with a prominent role played by the average term-structure level effect (δ_{01t}). The macro-sensitivities exhibit an increased variability in 2010, while they are essentially dormant in 2009, to explode later in 2011–2012, and calm down in 2013.

As noted in the introduction, our findings are generally consistent with Bernoth and Erdogan (2013), who show that the impact of fiscal policy variables and general investors' risk aversion on sovereign yield spreads in Europe was not constant over time. However, our results are in contrast with the authors' view that it is plausible to think of the time-varying sensitivities as changing gradually over time. Our results indicate, instead, that the time-varying macro-sensitivities exhibit substantial jumpiness in their dynamics, with sudden and discrete changes in "regime." The column "Zeros" of Tables 1 and 2 is informative as to the number of times, expressed as a ratio over the total cross-sections, in which the variable did not contribute to explaining the cross-section of CDS spreads and was discarded by the LASSO algorithm. This number should be read carefully, as it is informative only about the "importance persistence" of the variable, regardless of how much the specific explanatory power of that variable was—this analysis is performed in the next section. The bank risk indicator was the most selected level-effect variable, being discarded in less than one-third of the estimations, while among the slope-effect variables, GIPSI exhibits a ratio of 0.688, the lowest value among all slope-effect coefficient estimates.

A few findings are worthy of note. *First*, the coefficient for GDP growth has the highest average value (in absolute terms), both for level and slope effects. Since all regressors are

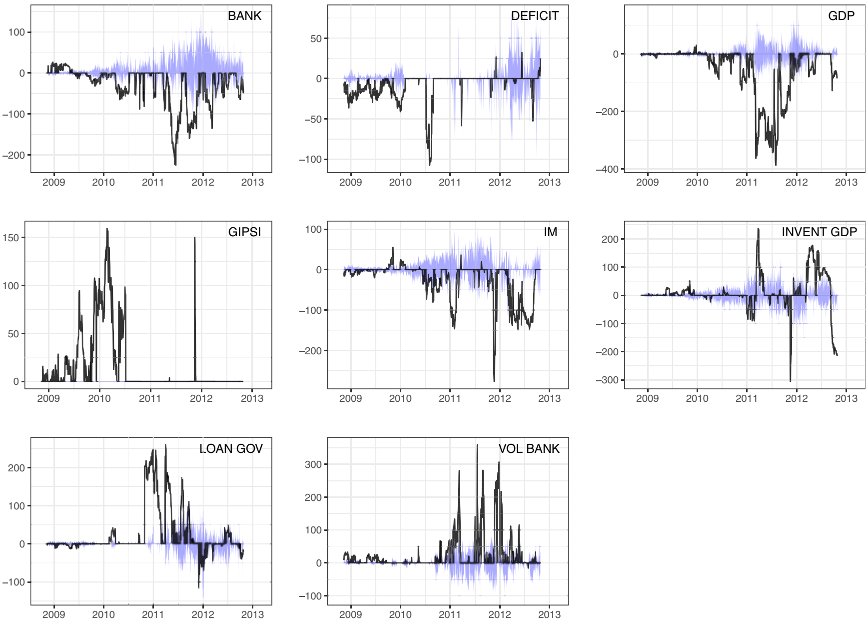

Figure 1 Most important coefficients: 2009–2013. The figure shows the LASSO coefficients of the cross-sectional model (Equation (3)) over the May 11, 2009–April 25, 2013 period for the most important covariates, as identified by a cross-sectional covariance decomposition exercise by regime (see Online Appendix). We report the time patterns of the coefficient estimates (black line), as well as the 1st and the 99th percentiles of empirical distribution computed each day (pink line) with 500 random samples.

cross-sectionally standardized and de-meaned, the intercept represents the average CDS spread across all countries and maturities, whereas the level and slope coefficients represent the effect, in basis points, of a one (cross-sectional) standard deviation increase in the corresponding covariates. GDP growth impacts negatively (positively), on average, the level (slope) of the term structure of CDS spreads. Assuming, in the first approximation, a constant expected loss given default, this evidence is consistent with the notion that an increase in GDP growth reduces the (risk-adjusted) marginal probability of default, but that this effect is stronger at short horizons. *Second*, the term structure slope effect is flattening-oriented for most of the variables, except for GIPSI, inflation, credit to the private sector, credit to government, and exports. For these macro-variables, all showing positive level effect-type coefficients, the longer the CDS maturity, the higher the impact on the CDS spread. *Third*, Min and Max denote high values for all coefficients, thereby reflecting substantial spikes and, in turn, "jumpy" dynamics of macro-sensitivities.

While the analysis above provides us with a full description of the time-series dynamics of the macro-sensitivities, the number of covariates, and their corresponding time-varying coefficients, complicate the understanding of underlying economic developments. To deal with this dimensionality problem, in Section 4, we introduce a simple statistical procedure to detect homogeneous groups of observations for the cross-sectional regression coefficients which, in turn, identify regimes in the macro-sensitivity behavior. They allow us to come

Table 1 Summary statistics, level-effect LASSO coefficients

	Zeros	Min	Max	Mean	Std Dev	0.01 level	0.05 level
Panel A: alpha, bank and GIPSI							
alpha	0.00	41.80	499.37	226.08	140.34	–	–
gipsi	0.69	0.00	158.97	15.38	32.20	1.00	1.00
bank	0.31	−222.89	27.94	−28.10	45.65	0.80	1.00
vol_bank	0.59	−13.89	357.49	23.67	54.55	0.80	1.00
Panel B: employment							
unempl	0.63	−85.22	80.38	0.34	13.23	0.48	1.00
empl	0.60	−345.59	26.91	−30.74	60.04	0.80	1.00
Panel C: prices and costs							
infl	0.62	−53.92	174.68	9.77	30.13	0.58	1.00
ind_price	0.49	−89.57	300.77	16.07	57.63	0.46	1.00
labor	0.60	−282.82	62.14	−5.67	25.87	0.44	1.00
Panel D: money, debt, and credit							
m3	0.73	−167.29	71.26	−5.74	25.65	0.64	1.00
loan_priv	0.67	−135.11	56.17	−3.04	17.81	0.84	1.00
loan_gov	0.59	−110.80	253.17	23.12	59.68	0.72	1.00
cr_priv	0.74	−43.67	93.70	2.40	8.49	0.72	1.00
cr_gov	0.52	−209.26	98.28	2.52	37.38	0.74	1.00
deficit	0.60	−106.57	30.94	−8.03	16.44	0.88	1.00
Panel E: output							
gdp	0.46	−380.47	30.77	−52.73	89.24	0.87	1.00
cons	0.59	−201.39	95.77	1.45	26.13	0.69	1.00
gov_cons	0.65	−234.76	248.03	3.98	47.83	0.70	1.00
inv	0.64	−205.10	68.45	−14.21	31.48	0.80	1.00
invent_gdp	0.49	−289.82	233.72	7.86	58.36	0.64	1.00
ex	0.61	−68.96	224.38	13.52	38.95	0.60	1.00
im	0.45	−274.77	55.39	−24.51	45.36	0.71	1.00
ind_prod	0.54	−189.70	41.46	−15.06	29.65	0.63	1.00

Notes: This table reports summary statistics for the level-effect cross-sectional LASSO coefficient estimates. All regressors are cross-sectionally standardized and de-meaned: the intercept (alpha) represents the average CDS spread across all countries and maturities, whereas the coefficients represent the effect, in basis points, of a one (cross-sectional) standard deviation increase in the corresponding covariates. The column "Zeros" reports the number of times, expressed as ratio over the total number of cross-sections, in which the variable was discarded by the LASSO algorithm. Under "0.01 level" and "0.05 level," we report the frequency of cases in which the coefficient estimate is outside the 0.01–0.99 and 0.05–0.95 bounds, respectively, of the empirical distribution of the 500 LASSO coefficients.

up with a parsimonious characterization of the changing nature of sovereign risk in Europe during the period 2009–2013.

3.2. A Placebo Test

To verify the ability of the LASSO approach to avoid spurious relations among the variables of interest and to assess the statistical significance of our results, we propose a "placebo test" based on Monte Carlo random sampling from all possible permutations of the observed data. The general idea is to replicate the LASSO regressions on synthetic samples

Table 2 Slope-effect LASSO coefficients

	Zeros	Min	Max	Mean	Std Dev	Slope	0.01 level	0.05 level
Panel A: bank, GIPSI, and tau								
gipsi	0.60	−63.42	44.45	1.13	7.94	Steep	0.55	1.00
bank	0.93	0.00	4.65	0.15	0.66	Flat	0.65	1.00
vol_bank	0.76	−52.83	17.09	−2.64	9.56	Flat	0.48	1.00
tau	0.79	−40.95	15.47	0.24	3.54	Steep	0.51	1.00
Panel B: employment								
unempl	0.95	−30.69	16.92	−0.21	1.86	Flat	0.16	1.00
empl	0.62	−0.86	67.61	2.85	6.80	Flat	0.35	1.00
Panel C: prices and costs								
infl	0.78	−45.44	13.08	0.97	4.19	Steep	0.38	1.00
ind_price	0.82	−54.11	7.68	−2.17	7.56	Flat	0.40	1.00
labor	0.81	−18.25	24.84	1.02	3.36	Flat	0.06	1.00
Panel D: money, debt, and credit								
m3	0.71	−1.63	90.52	3.22	11.08	Flat	0.28	1.00
loan_priv	0.90	−10.50	57.80	0.32	3.12	Flat	0.64	1.00
loan_gov	0.67	−97.50	3.09	−7.29	16.82	Flat	0.73	1.00
cr_priv	0.83	−22.59	14.70	0.16	2.65	Steep	0.26	1.00
cr_gov	0.69	−10.46	47.31	3.67	9.65	Steep	0.46	1.00
deficit	0.86	−0.74	70.76	2.55	9.68	Flat	0.88	1.00
Panel E: output								
gdp	0.79	−1.25	86.08	7.56	19.46	Flat	0.54	1.00
cons	0.85	−66.66	24.35	−0.96	10.11	Flat	0.38	1.00
gov_cons	0.81	−41.55	23.75	−0.29	4.95	Flat	0.18	1.00
inv	0.86	−1.97	30.05	1.34	4.94	Flat	0.46	1.00
invent_gdp	0.63	−41.74	36.98	−1.43	8.88	Flat	0.32	1.00
ex	0.71	−27.49	29.65	1.65	7.46	Steep	0.28	1.00
im	0.63	−6.61	61.31	2.26	8.84	Flat	0.46	1.00
ind_prod	0.69	−6.83	78.09	2.04	7.31	Flat	0.17	1.00

Notes: This table reports summary statistics for the slope-effect cross-sectional LASSO coefficient estimates. All regressors are cross-sectionally standardized and de-meaned: the coefficients represent the effect, in basis points, of a one (cross-sectional) standard deviation increase in the corresponding covariates. The column "Zeros" reports the number of times, expressed as ratio over the total number of cross-sections, in which the variable was discarded by the LASSO algorithm. Under "0.01 level" and "0.05 level," we report the frequency of cases in which the coefficient estimate is outside the 0.01–0.99 and 0.05–0.95 bounds, respectively, of the empirical distribution of the 500 LASSO coefficients.

where the outcome (the CDS spread for that specific day/country/maturity) is replaced by a pseudo-outcome that is known not to be affected by the treatment (a randomly selected CDS spread for any day/country/maturity). The method is based on Lindgren et al. (1996) and involves a repetitive reordering of all entries in the response variable—the CDS spread—while maintaining covariates fixed—the macro fundamentals.[9] Note that we do

9 In a recent application of LASSO to stock returns, Freyberger, Neuhierl, and Weber (2020) implement a similar simulation exercise to verify the ability of Adaptive Group LASSO to select the correct model in small samples, for the prediction of individual stock returns.

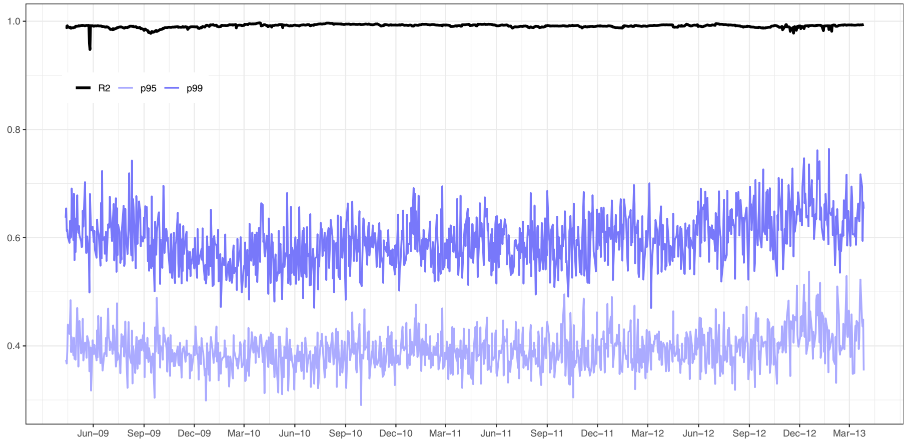

Figure 2 Cross-sectional R^2s: 2009–2013. The figure shows the daily cross-sectional R^2s of the model (Equation (3)) over the May 11, 2009–April 25, 2013 period. We report the time patterns of the point estimate (black line), as well as the 99th (P99) and the 95th (P95) percentiles of the simulated distribution of the cross-sectional R^2s, for each day (gray line) with 500 random samples.

not have the ambition to "make inference on the coefficients," at this point. We shall implement a more formal approach to inference in Section 6.3, in the context of Adaptive LASSO.

Essentially, our procedure is a reality check to see how far our model estimates are "from being a coincidence correlation" (with the language of Lindgren et al. 1996), thereby removing any doubt that the model, and specifically the non-zero LASSO coefficients estimates, might be the result of pure chance. In principle, the robustness check of a regression model through permutation test consists of reshuffling the dependent variable proving no or very limited association with the covariates. Computationally, this is carried out by re-estimating the model in each permutation, then recording diagnostics for each coefficient estimate and goodness of fit (e.g., R^2). In this way, we obtain distributions for each diagnostic, that we then compare with those calculated from the original data. Specifically, for each day, we keep fixed the covariates, while we randomly select 500 samples from all possible permutations of the NM CDS spreads (i.e., $NM!$ corresponding to $2.66E + 54$ samples). We then implement our empirical analysis on the 500 re-shuffled datasets, and we generate the distribution of the LASSO cross-sectional regression coefficients.

In the vast majority of cases, if a covariate is selected by the LASSO, the corresponding estimate falls *outside* of the range between the 1st and 99th percentiles of the empirical distribution of the LASSO coefficients under the null reported in Figure 1, thus making it very unlikely that the variable has been selected by pure chance. Conversely, when a covariate is not selected, the value of zero is always inside the 1st–99th percentile range, indicating that the LASSO is correctly setting that variable to zero.

In Figure 2, on the other hand, we report the sample R^2-s together with the 1st and 99th percentiles of the empirical distribution of the post-LASSO R^2-s under the null of no association between the outcome and the covariates. In this case, in all instances, the sample value of the R^2-s is largely outside the percentile bounds, indicating that it is highly unlikely that the high in-sample fit is due to chance.

3.3. Ruling Out "Spurious" Variable Selection

Another concern with the LASSO is that the algorithm may arbitrarily select one or more covariates within a cluster of correlated covariates (labor market; price and costs; money, credit and debt; output, banking). Hence, going from one cross-sectional regression to another, we could see different covariates being selected, while the other covariates are set to zero. To address this concern, we constructed an indicator variable for each covariate, with the value of 1 when the LASSO selects the covariate (non-zero coefficient), and 0 when it is discarded (zero coefficients). Next, for each cluster we compute the pairwise correlations between indicator variables, using Pearson's ϕ coefficient, which is a common measure of association between two binary variables with values ranging from -1 (negative association) and $+1$ (positive association).[10]

The rationale behind the test described above is that a high negative association is suggestive of the LASSO "mining" the data, by choosing one covariate rather than another just to maximize the fit and not because of a causal relationship with the dependent variable. Our results (available upon request) denote a low level of association, with values ranging from -0.3046 and 0.3459, for level-effect coefficients, and from -0.1773 and 0.3427, for slope-effect coefficients.[11] This is a further indication of the robustness of the LASSO variable-selection procedure in our context.

4 Macro-sensitivity Regimes

Existing studies of sovereign CDS regimes look at CDS/bond spreads to identify regime changes and incorporate structural changes in the econometric relationships between the spreads and the macroeconomic covariates (e.g., Blommestein, Eijffinger, and Qian 2016; Delatte, Fouquau, and Portes 2017). Other authors (e.g., Afonso et al. 2018) identify regimes based on the ECB policy intervention decisions and then explore how sovereign risk sensitivity changed once these measures took place.

10 Pearson's ϕ coefficient is commonly used to measure the degree of association between two binary variables (e.g., see Agresti 2007). For each pair of coefficients, we computed a 2×2 contingency table reporting the numbers in which both coefficients are 0 and 1 on the diagonal, and off-diagonal the numbers in which the coefficients are 0 for one variable and 1 for the other (and vice versa). Computationally, the ϕ coefficient is computed as follows. Denote by T_{11}, T_{10}, T_{01}, T_{00}, the non-negative counts of numbers of observations in each 2×2 contingency table cell that sum to T, the total number of observations. The coefficient is:

$$\phi = \frac{T_{11} - T_{00}}{(T_{11} + T_{10})(T_{01} + T_{00})(T_{10} + T_{00})(T_{11} + T_{01})}. \tag{7}$$

The two dichotomous time-varying LASSO coefficients are considered positively associated if most of the data fall on the diagonal cells (Pearson's ϕ coefficient is positive), whereas the two binary variables are considered negatively associated if most of the data fall off the diagonal (Pearson's ϕ coefficient is negative).

11 In categorical data analysis, it is common practice to evaluate the corresponding Cramer's V, that in the case of a 2×2 contingency table is equivalent to the squared Pearson's ϕ. A Cramer's V in the range of $[0, 0.3]$ is considered as weak; in the range of $[0.3, 0.7]$ as a medium; and greater than 0.7 as strong (maximum is 1).

Our approach differs, as we use the data on macro-sensitivities to identify regimes conceived as homogeneous groups of observations over the entire observed period. Macro-sensitivity regime identification is based on Kaufman and Rousseeuw (1990)'s clustering algorithm: Partitioning Around Medoids (PAM). This algorithm identifies clusters that, in our context, denote homogeneous time dynamics of the macro-sensitivities, and as such identify specific "regimes." These regimes have distinctive features that we explore by focusing on the changing structure of the sovereign risk sources due to both a shift in macroeconomic fundamentals and changes in risk pricing.

Computationally, the procedure needs to pre-specify the number of clusters before running the algorithm. In principle, we may view this number as given and related to some a priori theoretical reasoning or empirical evidence. Alternatively, the number could be data-driven, based on some of the existing criteria proposed in the literature (see Kaufman and Rousseeuw 1990). We choose the number of regimes by running a specific *F*-test-based clustering method which looks at the percentage of the explained variance as a function of the number of clusters. More specifically, the clusters are chosen by comparing the percentage of variance explained by the clusters against their number: the appropriate number of clusters corresponds to the point in which the marginal gain, expressed in terms of additional explained variance, drops, thereby signaling no significant information added by the last cluster. The criterion is commonly used in the literature and is based on the between-group variance, consistent with the concept of "distance" used to identify the homogeneous cluster of observations. Having the objective to pre-specify the number of crisis regimes, we run the test over the April 1, 2010–April 25, 2013 period, thereby arbitrarily establishing the May 11, 2009–March 31, 2010 sub-period as the "pre-crisis" regime. This is consistent with the empirical evidence we discussed in the introduction, as the surge of CDS/bond spreads of GIPSI countries occurs in April 2010, when Greece activates the 45 billion Euros EU-IMF bailout and S&P downgrades Greek debt to junk status. The criterion points to the existence of three clusters, or regimes, over the crisis period.

We display the regime timeline in Figure 3, where we see the three crisis regimes (regimes 1–3), as well as the pre-crisis regime (regime 0), where the different shaded areas identify the different regimes. In the same figure, we also display the time-varying intercept, which corresponds, by construction, to the cross-sectional average of sovereign CDS spreads.

The regimes identified by the PAM procedure are broadly consistent with the chronology of the eurozone crisis. The pre-crisis regime is characterized by moderate, but increasing CDS spreads, until March 2010. At that point, we have the Greece-driven transition to the first crisis regime, which continues until the third quarter of 2011. We then have fluctuations between the first and second regime until November 2011, when we enter the third and most risky regime, in terms of the cross-sectional average of CDS spreads. Starting with September 2012, the average CDS spread comes down as we first transition to the second regime, and finally, we revert to the first regime, starting in March 2013, when the average CDS spread is around 220 basis points.

To better interpret the pre-crisis and crisis regimes, Tables 3 and 4 report summary statistics of the LASSO coefficients conditional on regimes, and Online Appendix Table IA.4 shows the value of the cross-sectional CDS spread variance explained by each variable (see Section B of the Online Appendix). The values of explained variance are also normalized by the highest value as $\frac{v_{kt}}{v_{kth}} \times 100$, where v_k is the value of the cross-sectional CDS variance

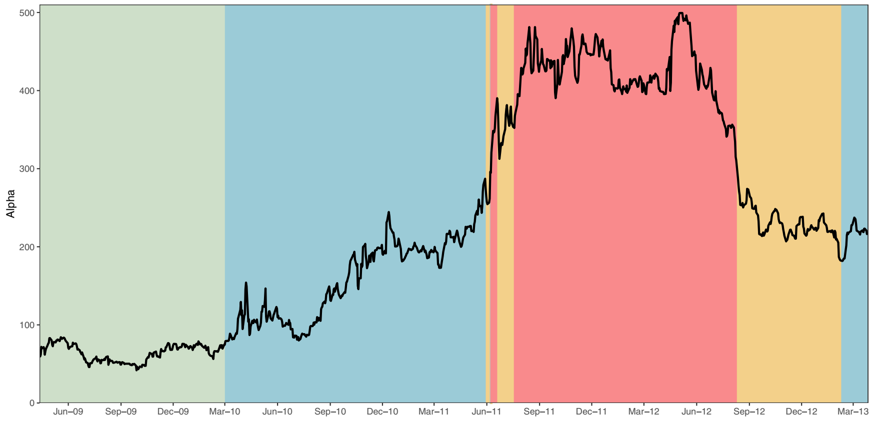

Figure 3 Alphas and regimes: 2009–2013. The figure displays the daily intercepts of the model (Equation (3)) over the May 11, 2009–April 25, 2013 period. Pre-crisis and crisis regimes are colored as green (regime 0), light blue (regime 1), orange (regime 2), and red (Regime 3).

explained by variable k and v_{kth} is highest value of the explained variance out of all $K+1$ covariates. We set 50% as the cut-off value for this normalized measure to identify the most informative variables for each regime, thereby highlighting only those variables showing an explanatory power not less than 50%, compared to the most informative variable.

The following patterns emerge:

- The pre-crisis regime, from May 2009 to March 2010, displays moderate while increasing, macro-sensitivities, with the deficit and GIPSI variables being the main determinants of CDS spreads, followed by the banking risk measure and the banking risk volatility indicator.
- The first crisis regime covers April 2010 to July 2011 period—which includes the 40 billion Euros Greek bailout and the ensuing widening of the spreads of peripheral eurozone countries—and the mid-March to April 2013 period, when Cyprus secured a 10 billion Euros bank bailout from the European Union and the IMF. During this regime, GIPSI and Loans-to-Government were the main drivers of the cross-section of sovereign CDS spreads, which averaged around 167 bps.
- The second crisis regime comprises the June to August 2011 and September 2012 to mid-March 2013, periods, namely the periods before the peak of the eurozone crisis and following the third turning point (Draghi's July 2012 speech and the announcement of the OMT program in September 2012), during which we also have the agreement on the part of European leaders for the European Stability Mechanism to directly recapitalize banks, rather than having to act through national governments (October 19, 2012). This is an intermediate regime, as the eurozone transitions from regimes 1–3 and then also back from 3 to 1. The average value of sovereign CDS spreads was around 248 bps. During this regime, imports and changes in inventories over GDP were the most influential variables, because of their effect on GDP growth through the balance of payment pressures and increased macroeconomic volatility. Our coefficient estimates for changes in inventories over GDP, on average positive during the regime, are consistent with recent evidence (European Commission 2015) documenting how firms viewing their

Table 3 Level-effect LASSO coefficients within regimes

	Regime 0		Regime 1		Regime 2		Regime 3	
	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat	Mean	*t*-stat
Panel A: alpha, bank and GIPSI								
alpha	64.23	10.45	166.74	8.97	247.01	4.32	423.97	21.67
gipsi	12.42	1.66	34.54	2.70	0.00	–	1.67	0.80
bank	5.24	3.98	−19.42	2.17	−13.10	1.765	−74.90	2.80
vol_bank	8.87	3.98	4.67	2.17	22.71	1.76	60.76	2.80
Panel B: employment								
unempl	−4.70	−2.41	−1.67	−0.85	9.85	2.45	1.84	0.68
empl	0.10−	−0.28	−18.06	−1.54	−5.45	−2.16	−86.08	−3.50
Panel C: prices and costs								
infl	−0.29	−2.13	9.10	1.94	−7.34	−0.72	28.28	3.06
ind_price	1.47	2.59	−8.24	−2.12	4.90	1.25	65.39	2.39
labor	1.99	2.53	−0.13	−0.05	−5.64	−1.46	−19.12	−2.75
Panel D: money, debt, and credit								
m3	2.50	0.83	−13.16	−1.84	−1.66	−1.58	−5.26	−0.45
loan_priv	4.45	2.96	−14.35	−2.86	6.42	1.03	0.12	0.14
loan_gov	−1.83	−2.90	27.42	0.92	30.91	1.27	33.87	1.64
cr_priv	6.83	5.89	0.35	0.75	1.77	0.70	1.73	1.13
cr_gov	3.84	3.50	22.48	4.46	−45.74	−1.146	2.24	0.93
deficit	−14.26	−8.95	−12.21	−1.51	−2.06	−1.09	−0.82	−1.72
Panel E: output								
gdp	−0.39	−0.51	−23.97	−2.87	−14.57	−1.51	−153.63	−3.61
cons	−5.63	−1.80	14.13	1.47	−15.72	−1.32	0.43	0.31
gov_cons	−1.41	−1.89	−3.82	−0.86	−11.10	−0.71	26.68	0.82
inv	−3.41	−1.30	−6.70	−0.76	−38.65	−2.45	−19.37	−1.86
invent_gdp	3.95	1.42	−14.67	−0.42	72.74	2.31	4.45	0.27
ex	3.61	3.63	−1.26	−1.82	9.53	2.47	42.84	2.11
im	−3.47	−2.99	−9.17	−1.70	−107.66	−17.59	−15.98	−1.71
ind_prod	−0.75	−1.92	−18.46	−4.10	7.04	1.38	−34.60	−4.23

Notes: This table presents summary statistics for the daily cross-sectional level-effect LASSO coefficient estimates conditional on pre-crisis (Regime 0) and crisis regimes (Regimes 1–3). The pre-crisis regime is arbitrarily set from May 11, 2009 to March 31, 2010, whereas crisis regimes are identified based on the PAM clustering algorithm (Kaufman and Rousseeuw 1990) executed on the time-varying parameters from Equation (3). For each regime, the table reports the arithmetic average (Mean) and the corresponding t-stat computed with Newey–West robust standard errors (non-parametric kernel).

inventory stocks as "too large" are expected to react by cutting production in the following months, thereby exacerbating economic downturn during crisis periods.

- The third crisis regime takes place between July 2011 (when we have a rebound between regimes 2 and 3) and August 2012. This regime corresponds to the highest risk phase when average spreads reached 400 bps and the cross-section of spreads was mostly explained by GDP growth and employment. The regime includes the Italian government crisis (November 2011) and the release of the results of the second round of pan-European stress tests (eight European banks failed the stress tests, while 16 were in a "danger zone").

Table 4 Slope-effect LASSO coefficients within regimes

	Regime 0			Regime 1			Regime 2			Regime 3		
	Mean	*t*-stat	Slope	Mean	*t*-stat	Slope	Mean	*t*-stat	Slope	Mean	*t*-stat	Slope
Panel A: Bank, GIPSI, and tau												
gipsi	6.04	4.04	Steep	0.39	1.37	Steep	4.72	1.52	Steep	−3.94	−1.84	Flat
bank	0.50	1.73	Steep	0.12	0.66	Flat	0.00	–	–	0.00	–	
vol_bank	0.39	3.16	Steep	0.27	0.61	Steep	0.18	0.89	Steep	−10.41	−2.96	Flat
tau	0.33	3.17	Steep	0.75	1.28	Steep	1.65	2.11	Steep	−1.25	−1.57	Flat
Panel B: Employment												
unempl	0.00	–	–	−0.22	−2.36	Steep	0.11	0.978	Steep	−0.54	−1.02	Flat
empl	1.99	2.66	Flat	1.35	2.52	Flat	2.13	2.59	Flat	5.86	2.45	Flat
Panel C: Prices and costs												
infl	0.01	1.01	Flat	3.56	3.89	Steep	0.00	–	–	−1.02	−0.60	Flat
ind_price	−0.07	−1.09	Flat	0.27	2.08	Flat	−0.83	−1.21	Flat	−7.76	−3.18	Flat
labor	0.04	2.13	Steep	0.81	0.87	Flat	2.67	3.23	Flat	1.18	1.53	Flat
Panel D: Money, debt, and credit												
m3	0.39	4.25	Steep	4.02	1.81	Flat	−0.02	−1.45	Steep	6.32	1.16	Flat
loan_priv	0.23	1.40	Steep	0.07	1.05	Flat	−0.23	−1.41	Flat	0.99	1.23	Steep
loan_gov	−0.58	−1.44	Steep	−7.96	−2.13	Flat	−9.92	−1.40	Flat	−10.52	−1.80	Flat
cr_priv	0.15	0.87	Steep	0.98	2.99	Steep	−1.84	−1.77	Flat	0.20	1.91	Steep
cr_gov	0.01	1.35	Steep	0.19	0.15	Steep	15.69	6.90	Flat	4.58	1.37	Steep
deficit	−0.01	−2.04	Steep	0.40	1.53	Flat	3.21	0.81	Flat	7.05	2.57	Flat
Panel E: Output												
gdp	−0.05	−0.99	Steep	0.17	1.14	Flat	0.08	1.33	Flat	27.42	2.30	Flat
cons	−0.03	−1.64	Steep	−0.24	−1.25	Flat	4.06	2.24	Flat	−5.40	−0.86	Flat
gov_cons	0.23	2.29	Flat	−0.64	−2.39	Steep	2.86	1.08	Flat	−1.98	−1.59	Flat
inv	−0.09	−1.63	Steep	2.46	0.85	Flat	0.34	1.58	Flat	1.63	2.056	Flat
invent_gdp	0.62	1.81	Steep	2.34	1.093	Flat	−12.55	−3.36	Flat	−1.87	−1.95	Flat
ex	−0.36	−1.23	Flat	1.13	0.76	Flat	13.85	5.87	Steep	−2.70	−2.64	Flat
im	−2.296	−4.348	Steep	−0.18	−1.34	Steep	2.07	2.67	Flat	9.26	3.25	Flat
ind_prod	−0.82	−3.29	Steep	0.48	1.14	Flat	2.33	1.42	Steep	6.24	2.80	Flat

Notes: This table presents summary statistics for the daily cross-sectional slope-effect LASSO coefficient estimates conditional on pre-crisis (Regime 0) and crisis regimes (Regimes 1–3). Pre-crisis regime is arbitrarily set from May 11, 2009 to March 31, 2010, whereas crisis regimes are identified based on the PAM clustering algorithm (Kaufman and Rousseeuw 1990) executed on the time-varying parameters from Equation (3). For each regime, the table reports the arithmetic average (Mean) and the corresponding *t*-stat computed with Newey–West robust standard errors (non-parametric kernel). Column Slope denotes the flattening (flat) or steepeing (step) oriented term structure effect for each variable.

Interestingly, the key variables selected based on their contribution to the explained cross-sectional variance of CDS spreads are also those covariates with the higher LASSO coefficient averages in absolute terms (see Table IA.4 in the Online Appendix for details).

Our characterization of non-crisis and crisis regimes offers new insights into the economic mechanisms underlying the eurozone sovereign debt crisis. As pointed out in De Grauwe and Ji (2013), one view of the crisis is that the surging spreads from 2010 to mid-2012 were the result of deteriorating fundamentals and the market was just a messenger of bad news. A second view is that beginning in 2010, the spreads were driven away from country fundamentals. The first view would explain why austerity-based measures should be the right measure of policy intervention. The second view implies that in times of market panic, central banks should act as liquidity providers.

Our findings accommodate both views. We show CDS spreads were disconnected from fundamentals, but only in regime 1, when being in the GIPSI group of countries was the key driver of the surge of CDS spreads. In regime 3, on the other hand, at the peak of the crisis, markets restored a fundamental-based connection with GDP growth, even when the ECB intervened to provide essentially unlimited support to the government bond markets. Then, we move toward regime 2, finally returning to regime 1, where GIPSI is again the main risk factor, but at a lower average level of spreads and lower GIPSI sensitivity. In this context, regime 2 appears to be a transition regime, in which we do have a connection between spreads and risk signals from imports and inventory dynamics.

5 Results over the 2013–2020 Sample

This section extends the analysis of the drivers of the cross-section of CDS spreads to the May 2013 to December 2020 period. This later sample is of potential interest, given the occurrence of systematic shocks such as Brexit and the Covid-19 pandemic.

As discussed in Section 1, our real-time daily macroeconomic data set was initially constructed for the sample ending April 30, 2013, when we had access to the ECB data libraries. Since then, we lost access and we could not extend the data set. Hence, we had to perform the analysis over the period from May 2013 to December 2020 using publicly available data for which we had no information about the exact first release date. For this reason, we used monthly end-of-month sovereign CDS spreads and lagged country-specific fundamentals. Data were collected by merging different databases (Datastream, ECB Statistical Data Warehouse, Facset Standardized Economics, and National Statistical Databases), thereby maintaining almost the same set of covariates used over the period 2009–2013, except for Hourly labor cost (labor), Loans to the private sector (loan-priv), Loans to government sector (loan-gov), and Consumption growth (cons).

Figure 4 shows the time series of the cross-sectional LASSO coefficients for the variables that proved to be most important during the first sub-period, while Tables 5 and 6 report summary statistics computed over the second sub-period for all the cross-sectional coefficient estimates.

To assess the significance of the LASSO estimates, for both single and all non-zero coefficients jointly, we followed the same approach used in Section 3. In this regard, note first in Tables 5 and 6 that the significance at 5% level is confirmed for all coefficients every month, except for Unemployment rate (unempl) and Banking risk proxy (bank), which were significant (i.e., the coefficient estimates were outside the $5 - 95\%$ bound) for the 31% and around the 98% of the total number of observations, respectively. At 1% level, the significance drops substantially (on average, 65% and 49% for level-type and slope-type coefficients, respectively); see, also, Figure 4 in which the estimated cross-sectional coefficients are almost half of the times outside the 1% confidence bands (pink areas) (Figures IA.5–IA.8 in the Online Appendix report graphs for the time-series of all other LASSO coefficients).

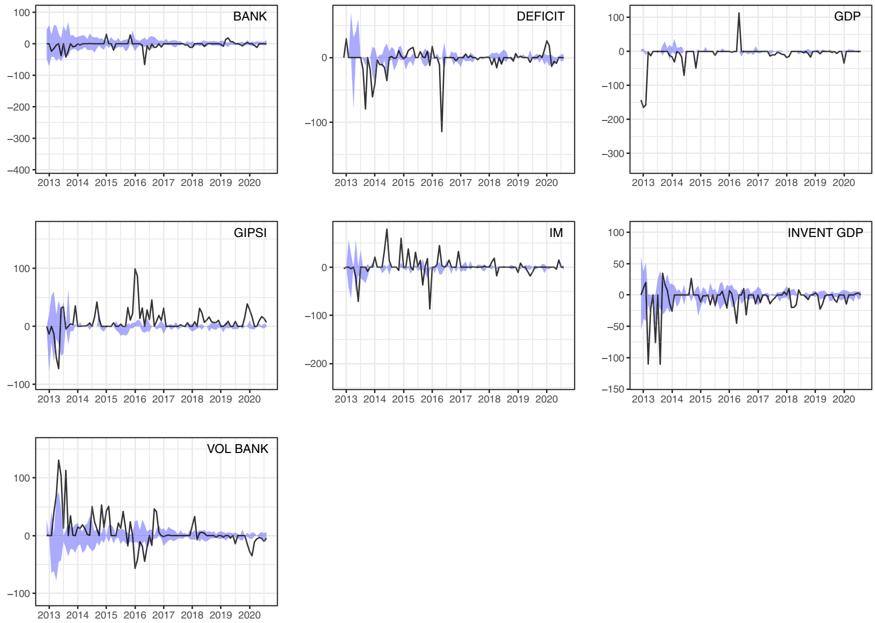

Figure 4 Most important coefficients: 2013–2020. The figure shows the LASSO coefficients of the cross-sectional model (Equation (3)) over the May 2013–December 2020 period that over the period May 11, 2009–April 25, 2013 have proven to be the most important, as identified by the covariance decomposition exercise by the regime. We report the time patterns of the coefficient estimates (black line), as well as the 1st and the 99th percentiles of empirical distribution computed each day (pink line) with 500 random samples.

Regarding the joint significance of all non-zero LASSO coefficients, in Figure 6 we compare the R^2s of the model selected by the LASSO to the distribution of R^2s under the null. As in the first sub-period, we confirm a strongly significant cross-sectional fit also for the period 2013–2020. As a whole, the dynamics of the LASSO coefficients over the second sub-period exhibit less time variability relative to the first. In 2013, 2014, and 2016 macro-sensitivities denote a relatively increased variability, while they go back to being dormant from 2017 onward as in 2009 when the LASSO coefficients are very low.

To better understand the behavior of macro-sensitivities over the period 2013–2020, we classified the monthly LASSO coefficients within the same regimes 0–3, identified and discussed in Section 4. Specifically, for each month we computed the Euclidean distances between the monthly LASSO coefficients and the corresponding mean values within regimes (Tables 3 and 4), next assigning the regime whose distance was the lowest. Note that this approach is consistent with the PAM algorithm used before to identify the macro-sensitivity regimes, being based on a distance-minimization criterion, also allowing a regime-based comparative analysis between the first and the second sub-period.

Figure 5 shows the regime timeline along the same shaded areas used in Figure 3 also displaying the time-varying intercept. Our LASSO coefficients, which ended the first sub-period in regime 1, continued in that regime in the latter part of 2013, but with two jumps

Table 5 Summary statistics, level-effect LASSO coefficients 2013–2020

	Zeros	Min	Max	Mean	Std Dev	0.01 level	0.05 level
Panel A: alpha, bank, and GIPSI							
alpha	0.00	−12.96	368.78	83.23	73.56	–	–
gipsi	0.36	−81.13	98.87	8.36	20.65	0.81	1.00
bank	0.47	−66.02	29.95	−3.21	11.69	0.38	0.98
vol_bank	0.34	−56.69	126.27	8.14	28.31	0.65	1.00
Panel B: employment							
unempl	0.10	−19.62	9.20	−0.39	4.17	0.08	0.31
empl	0.48	−292.60	23.93	−13.48	48.95	0.68	1.00
Panel C: prices and costs							
infl	0.41	−128.42	120.08	−4.75	31.12	0.83	1.00
ind_price	0.41	−63.90	43.85	−0.98	13.80	0.41	1.00
labor	Na	na	na	na	na	na	na
Panel D: money, debt, and credit							
m3	0.36	−120.53	64.94	−6.57	18.34	0.53	1.00
loan_priv	Na	na	na	na	na	na	na
loan_gov	Na	na	na	na	na	na	na
cr_priv	0.44	−104.37	27.08	−10.25	17.89	0.91	1.00
cr_gov	0.53	−78.50	28.61	−0.94	12.32	0.64	1.00
deficit	0.44	−114.48	29.74	−3.19	18.39	0.65	1.00
Panel E: output							
gdp	0.61	−164.45	113.42	−6.44	27.91	0.88	1.00
cons	Na	na	na	na	na	na	na
gov_cons	0.44	−89.53	59.22	−6.86	18.72	0.67	1.00
inv	0.62	−51.47	105.50	−1.50	17.25	0.76	1.00
invent_gdp	0.45	−110.04	34.67	−5.89	20.57	0.52	1.00
ex	0.46	−62.52	162.64	6.90	31.28	0.80	1.00
im	0.59	−86.58	79.66	1.77	18.86	0.79	1.00
ind_prod	0.48	−162.67	24.24	−4.93	27.67	0.78	1.00

Notes: This table reports summary statistics for the level-effect cross-sectional LASSO coefficient estimates over the 2013–2020 period. All regressors are cross-sectionally standardized and de-meaned: the intercept (alpha) represents the average CDS spread across all countries and maturities, whereas the coefficients represent the effect, in basis points, of a one (cross-sectional) standard deviation increase in the corresponding covariates. The column "Zeros" reports the number of times, expressed as ratio over the total number of cross-sections, in which the variable was discarded by the LASSO algorithm. Under "0.01 level" and "0.05 level," we report the frequency of cases in which the coefficient estimate is outside the 0.01–0.99 and 0.05–0.95 bounds, respectively, of the empirical distribution of the 500 LASSO coefficients.

into regime 3 in June and September, possibly to be attributed to uncertainty coming from the United States—"markets began to contemplate the possibility that the Federal Reserve might taper-off its bond purchase programme in the near future," wrote the ECB Monthly Bulletin 09/2013, pg. 43. In 2014, the average CDS spread falls toward 100 basis points, as it was in 2009, and indeed we transition to the pre-crisis regime 0. However, during April-May and October 2014 we see transitions to regime 1. This is most likely due to a cross-country spillover of bank risk around the European Central Bank's release of the Comprehensive Assessment (CA) results on October 26, 2014. As documented in

Table 6 Summary statistics, slope-effect LASSO coefficients 2013–2020

	Zeros	Min	Max	Mean	Std Dev	Slope	0.01 level	0.05 level
Panel A: bank, GIPSI, and tau								
gipsi	0.03	0.00	31.45	13.78	7.72	Steep	0.72	1.00
bank	0.70	−12.80	12.84	0.51	3.06	Flat	0.30	1.00
vol_bank	0.46	−9.95	28.19	5.15	7.94	Steep	0.57	1.00
tau	0.36	0.00	28.00	5.78	7.28	Steep	0.70	1.00
Panel B: employment								
unempl	0.79	−18.92	8.43	0.44	2.66	Flat	0.63	1.00
empl	0.66	−29.53	19.32	0.16	5.77	Flat	0.65	1.00
Panel C: prices and costs								
infl	0.61	−19.02	27.69	0.69	5.81	Flat	0.31	1.00
ind_price	0.38	−17.46	11.51	0.25	4.37	Flat	0.24	1.00
labor	Na	na	na	na	na	na	na	
Panel D: money, debt, and credit								
m3	0.45	−16.41	10.58	−0.60	4.56	Steep	0.33	1.00
loan_priv	Na	na	na	na	na	na	na	
loan_gov	Na	na	na	na	na	na	na	
cr_priv	0.60	−17.79	15.04	−0.59	4.54	Steep	0.51	1.00
cr_gov	0.55	−22.91	16.37	1.07	4.35	Flat	0.34	1.00
deficit	0.45	−27.15	8.84	−2.95	5.22	Steep	0.37	1.00
Panel E: output								
gdp	0.69	−10.36	40.56	0.15	6.80	Flat	0.82	1.00
cons	Na	na	na	na	na	na	na	
gov_cons	0.51	−13.78	31.79	0.81	6.79	Flat	0.31	1.00
inv	0.67	−10.46	21.40	0.12	3.80	Flat	0.63	1.00
invent_gdp	0.50	−15.59	24.44	0.25	5.37	Flat	0.27	1.00
ex	0.61	−31.90	10.54	−0.71	5.26	Flat	0.57	1.00
im	0.71	−24.54	22.75	0.34	4.84	Steep	0.74	1.00
ind_prod	0.45	−18.96	44.23	2.58	9.06	Flat	0.41	1.00

Notes: This table reports summary statistics for the slope-effect cross-sectional LASSO coefficient estimates over the 2013–2020 period. All regressors are cross-sectionally standardized and de-meaned: the coefficients represent the effect, in basis points, of a one (cross-sectional) standard deviation increase in the corresponding covariates. The column "Zeros" reports the number of times, expressed as ratio over the total number of cross-sections, in which the variable was discarded by the LASSO algorithm. Under "0.01 level" and "0.05 level," we report the frequency of cases in which the coefficient estimate is outside the 0.01–0.99 and 0.05–0.95 bounds, respectively, of the empirical distribution of the 500 LASSO coefficients.

Breckenfelder and Schwaab (2018), bank risk uncovered by the CA in stressed countries was not reflected in the prices of their sovereign bonds, but spilled over to countries with non-stressed banking sectors, leading to an increase in sovereign CDS spreads across the eurozone. The LASSO coefficients remain in the pre-crisis regime 0 until June 2016, when many eurozone countries experienced a marked surge in credit risk spreads after the Brexit results. We jump first into regime 1, then move into regime 2 in September 2016, with the intercept at around 300 basis points, caused by the results of the U.S. Presidential vote election, because of concerns that inflation might increase during Donald Trump's presidency.

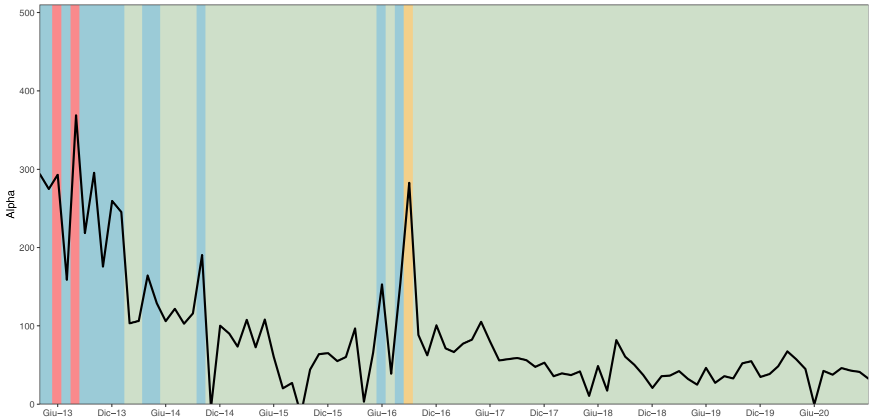

Figure 5 Alphas and regimes: 2013–2020. The figure displays the daily intercepts of the model (Equation (3)) over the May 2013–December 2020 period. Pre-crisis and crisis regimes are colored as green (regime 0), light blue (regime 1), orange (regime 2), and red (Regime 3).

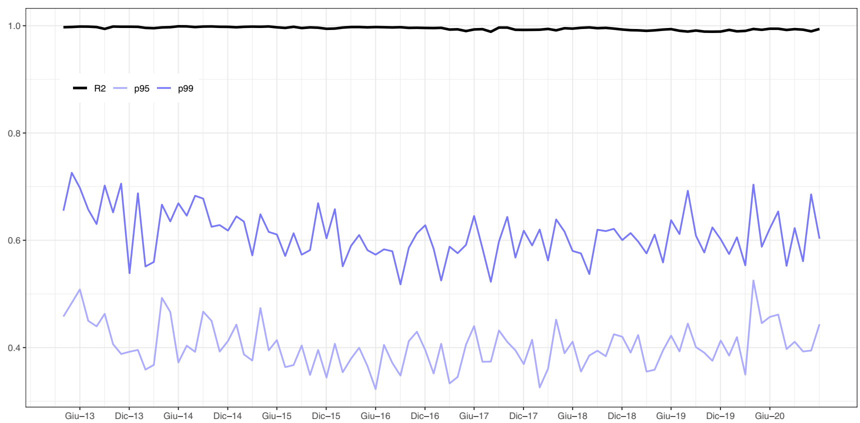

Figure 6 Cross-sectional R^2s: 2013–2020. The figure shows the daily cross-sectional R^2s of the model (Equation (3)) over the May 2013–December 2020 period. We report the time patterns of the point estimate (black line), as well as the 99th (P99) and the 95th (P95) percentiles of the simulated distribution of the cross-sectional R^2s, for each day (gray line) with 500 random samples.

Afterward, macro-sensitivities return to regime 0 and remain there throughout the remaining years 2017–2020. The intercept estimate never exceeds 100 basis points, even during the period from mid-2018 to November 2018, when the Italian government budget negotiations caused tensions on their sovereign CDSs, while spreads for most EU sovereigns remained at relatively low levels. Regime 0 persists in 2020, when the outbreak of the Covid-19 pandemic generated worries of a severe global economic recession, exacerbating public debt problem concerns. Consistent with recent empirical evidence (Corradin, Grimm, and Schwaab 2020), which documented how the EU's fiscal policy announcements lowered yields more uniformly, our estimates indicate that the pandemic had virtually no effect on the market-implied sovereign risk as measured by macro-sensitivities.

Finally, note that the analysis at the monthly frequency differs from the analysis at the daily frequency only in that we are using values for the macro variables that may have been revised, whereas the analysis with daily data uses the original values that were disclosed to the public and that we assume to matter for CDS valuation. As a result, there may be measurement errors in our monthly macro variables that, in principle, can bias our estimates. However, with more than one covariate, a priori the direction of the bias is unknown even in a standard OLS setting; see, for example, Abel (2018).

6 Extensions and Robustness Checks

In this section, we discuss several extensions and robustness checks of our analysis (Tables IA.1–IA.8 and Figure IA.9 in the Online Appendix present detailed results for all the exercises.).

6.1 LASSO on Panel Data

To better understand the role of the time variation in the covariate selection and the corresponding macro-sensitivities, we implement a panel-data estimation approach, implemented separately for the four maturities, imposing *constant* macro-sensitivities, while the variable selection is still performed using LASSO.[12] This alternative approach results in a large number of covariates being selected, but, despite rich specification, the fit of the panel-data estimator is much lower than that of our cross-sectional approach, at about 67% for all the maturities considered. These results highlight the importance of allowing for the macro-sensitivities to be time-varying, as in our Fama–Macbeth approach.

6.2 OLS and Elastic Net

We compare the daily cross-sectional LASSO with daily pooled OLS-style regressions, estimated separately month-by-month—hence, where coefficients are allowed to vary every month—and the Elastic Net, which is the natural alternative to LASSO as it combines a penalty from LASSO with that of the Ridge regression (Hastie, Tibshirani, and Friedman 2010).[13] As pointed out in Zou and Hastie (2005), although the LASSO usually performs well in variable selection problems, it may run into issues when the number of predictors is larger than the number of observations, and when there is a group of variables whose pairwise correlations are high. In the first case, the LASSO tends to select at most all the predictor variables before it saturates (Zou and Hastie 2005). In the second case, the algorithm tends to select only one variable at random from the group. In these scenarios, the Elastic Net seems to be more efficient than the LASSO while maintaining a similar sparsity of representation.

12 In the estimation, we control for time- and country-fixed effects. We also implemented a panel-data LASSO approach which combines the four maturities and where the covariates are interacted with the maturity of the contract, as in our main analysis. In this case, we control for maturity by time, country by time, and contract-fixed effects. The R^2 of the model is 68%.

13 We run monthly pooled OLS-style regressions because the number of covariates exceeds the size of the cross-section of contracts.

Both the OLS and the Elastic Net approaches lead to results remarkably close to LASSO for both level-type and slope-type coefficients. In particular, the correlation between the monthly averages of the daily coefficients from the OLS and the LASSO is around 0.82, although OLS presents more volatile estimates.

6.3 Adaptive LASSO

Adaptive LASSO (Zou 2006) has been advocated as an alternative to standard LASSO because of its oracle property; that is, its ability to identify the correct model with probability one as the sample size increases. Zou (2006) derives the asymptotic normality of the Adaptive LASSO estimates.[14]

Formally, Adaptive LASSO modifies the constraint on the regression coefficients as follows:

$$\sum_{k=1}^{K+1} \frac{1}{|\hat{\delta}_{1kt}|^{\lambda}} |\delta_{1kt}| + \sum_{k=1}^{K+2} \frac{1}{|\hat{\delta}_{2kt}|^{\lambda}} |\delta_{2kt}| + \leq c, \tag{5}$$

where $\hat{\delta}_{kt}$ are initial estimates typically obtained using either the OLS or the Ridge estimator. As λ increases, the Adaptive LASSO increases its "shrinkage" of the estimates toward zero. For $\lambda = 0$, on the other hand, we are back to the standard LASSO set-up. Common choices for λ are 0.5, 1, and 2.

We report results for $\lambda = 0.5$, as the other two parameter choices lead to a very unsatisfactory fit (low R^2-s). We find that all coefficient estimates are positively correlated with those of the standard LASSO results, but more strongly so for the level-effect coefficients. However, the means of the estimates are generally higher, in absolute value, for standard LASSO. Moreover, the volatility of the estimates is always higher for the standard LASSO. These differences are because Adaptive LASSO places a higher penalty on large (in absolute value) coefficient estimates, as discussed above.

We also perform post-selection inference based on the asymptotic results of Lee et al. (2016) and Taylor and Tibshirani (2018), where we condition on the choices of λ and c.[15] We find that the significance of the individual coefficient estimates is lower than what we reported for the standard LASSO estimates, which was based on our simulation-based approach to inference. This discrepancy should be attributed, at least in part, to the lower absolute magnitude of the Adaptive LASSO estimates.

6.4 Group LASSO

Another alternative to the standard LASSO is Group LASSO (Yuan and Lin 2006), which is an extension of LASSO performing variable selection on predefined *groups* of variables in linear regression models.[16] In our setting, Group LASSO amounts to

14 For the asymptotic properties of sparse regression model see also Belloni, Chernozhukov, and Hansen (2014); Zhang and Zhang (2014); and Belloni, Chernozhukov, and Kato (2015).

15 Specifically, we employ the function fixedLASSOInf in the R Package selectiveInference; see Tibshirani et al. (2019).

16 Note that Fused LASSO, introduced in Tibshirani et al. (2005), would also be a candidate estimator in our setting. However, Fused LASSO adds to the standard constraint on the sum of the absolute values of the coefficients a second constraint on the sum of the absolute values of *changes* in the

$$\min_{\delta_0,\delta_{1t},\dots,\delta_{2t}} \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\{s_{nmt} - \delta_0 - [\delta_{1t}^{\top}x_{nt} + \delta_{2t}^{\top}(x_{nt}\times\tau_m)]\}^2, \tag{6}$$

subject to $\sum_{t=1}^{T}|\delta_{1kt}| \leq c_1$, $\sum_{t=1}^{T}|\delta_{2kt}| \leq c_2$, $k = 1,\dots,K+2$, where a variable is included or excluded for the entire sample. Hence, in this setting, there is a link between macro-sensitivity estimates at different points in time that is absent from our purely cross-sectional approach.[17]

Group LASSO leads to numerical issues when implemented on our main daily data set, as we have $(K+2)\times 2\times T = 46\times 1{,}034 = 47{,}564$ covariates—K macro indicators plus the GIPSI dummy and the constant, affecting both the level and slope of the term structure of CDS spreads, times the number of daily observations. However, Group LASSO can be implemented on the second sample of 92 monthly observations, from 2013 to 2020. In this case, the initial number of covariates is $(17+2)\times 2\times 92 = 3{,}496$ that we sort into 46 groups. The results of the analysis are overall consistent with our analysis based on the standard version of LASSO, although Group LASSO is less flexible as it selects covariates for the entire sample.

7 Conclusions

We adapt the LASSO-regression machine-learning algorithm to the classical Fama–MacBeth approach to cross-sectional inference. We apply our methodology to the time-varying relationship between CDS spreads on sovereign bonds and the macroeconomic fundamentals of a panel of eurozone countries during the May 11, 2009–December 31, 2020 sample.

Our main findings are as follows. First, we document pronounced time-variation in the sensitivity of CDS spreads to the country-specific macro indicators. Second, we identify three distinct risk regimes based on the general level of CDS premia, the sensitivity of CDS premia to different macro indicators, and the GIPSI connotation. It is during the regime corresponding to the most intense phase of the eurozone crisis that CDS spreads reflected macro fundamentals the most, whereas before the crisis and at its onset it was only the GIPSI connotation to matter. We document another period of high macro sensitivities at the time of Brexit and more recently at the time of the U.S. presidential election. The Covid-19 pandemic, on the other hand, had virtually no effect on overall sovereign default spreads.

Supplemental Data

Supplemental data are available at https://www.datahostingsite.com

Appendix: Details of Data Construction

Our real-time data set is mostly based on European Central Bank e-archives. These e-archives contain historical records of the information supplied to the public by the ECB. In

coefficients. Hence, fused LASSO induces the coefficient estimates to be smooth over time by *construction*, when instead we could have "jumps" in their temporal dynamics.

17 Note that in the case where there is only one variable in each group, Group LASSO reduces to our standard cross-sectional LASSO.

constructing the data set, we have considered the various lags with which new data are released by the ECB, compared to the moment they are released by national statistical institutes and national central banks. The latter date is the release date that matters, as it corresponds to when new information reaches the markets for the first time. For this purpose, official release dates have been retrieved or double-checked using information from Bloomberg, Money Market Services (MMS), as well as information from national central banks and statistics offices.

The structure of the data set differs from standard, lower-frequency, real-time data sets, as it does not exhibit "vintages." This is due to the different frequencies at which the variables of interest are released (up to quarterly) and of the data set itself (daily). The data set is structured as a standard panel of mixed-frequency time series. The difference relative to a standard data set is that, at each date, each of the macro variables listed above takes the *latest released value*, instead of the value for the reference period, which is not known in real-time. For example, current eurozone GDP growth will only be available 30 days after the end of this quarter ($T+30$), in the form of a preliminary flash estimate, which will be revised 15 days later ($T+45$), while the second GDP release will be published 60 days after the end of this quarter ($T+60$). Eurostat improved the timeliness of the eurozone GDP

Table A.1 Macroeconomic variables in real time

Variable	Description	Cluster	Frequency
unempl	Unemployment rate	Labor market	Monthly
empl	Employment rate, total	Labor market	Quarterly
infl	Inflation rate	Price and costs	Monthly
ind_price	Industrial Producer Prices (% change)	Price and costs	Monthly
Labor	Hourly labor cost (price index) (% change)	Price and costs	Quarterly
m3	M3 (variation)	Money, credit, and debt	Monthly
loan_priv	Loans to private sector (variation)	Money, credit, and debt	Monthly
loan_gov	Loans to government (variation)	Money, credit, and debt	Monthly
cr_priv	Credit to private sector (variation)	Money, credit, and debt	Monthly
cr_gov	Credit to government (variation)	Money, credit, and debt	Monthly
deficit	Public sector deficit over GDP	Money, credit, and debt	Quarterly
gdp	Real GDP growth	Output	Quarterly
cons	Consumption growth	Output	Quarterly
gov_cons	Government consumption growth	Output	Quarterly
inv	Investment growth	Output	Quarterly
invent_gdp	Changes in inventories over nominal GDP	Output	Quarterly
ex	Exports growth	Output	Quarterly
im	Imports growth	Output	Quarterly
ind_prod	Industrial production growth (price index)	Output	Monthly
bank	Banking risk proxy	Banking	Daily
vol_bank	20 days rolling windows realized volatility of bank	Banking	Daily

Notes: This table reports the list of macroeconomic variables used in our analysis. Variables are grouped by type, as indicated in the "Cluster" column: (i) labor market; (ii) price and costs; (iii) money, credit and debt; (iv) output; and (v) banking.

Table A.2 Descriptive statistics, non-GIPSI

Variable	AT		BE		CY		DE		FI		FR		NL		All non-GIPSI	
	avg	std	avg	std	avg	std	Avg	std	avg	std	avg	std	Avg	std	avg	std
infl	2.14	1.22	2.10	1.48	2.39	1.45	1.58	0.91	2.46	0.90	1.62	0.95	1.89	1.06	2.02	1.14
gdp	0.66	2.49	0.24	2.14	−0.61	1.78	0.57	3.54	−0.22	4.57	0.32	1.65	−0.48	2.43	0.07	2.66
unempl	4.42	0.43	7.73	0.54	8.31	2.58	6.42	0.87	8.10	0.48	9.95	0.41	4.51	0.74	7.06	0.86
ind price	0.83	5.15	0.59	9.20	0.27	10.14	−1.62	8.62	3.18	5.99	0.23	7.19	−0.96	11.52	0.36	8.26
empl	0.86	1.50	0.75	0.91	0.82−	1.76	0.79	0.76	−0.09	2.21	1.02	2.45	−0.16	1.09	0.34	1.53
cons	2.69	4.66	2.99	6.47	2.81	6.82	2.76	3.03	2.70	4.61	2.26	3.60	2.10	5.95	2.62	5.02
gov cons	3.76	4.30	5.02	7.03	7.65	13.11	3.10	1.91	5.62	7.77	4.69	4.83	7.12	7.11	5.28	6.58
inv	1.67	9.60	−0.71	5.20	−5.35	9.82	−1.73	6.54	−0.04	11.44	0.25	6.16	−1.06	10.86	−1.00	8.52
invent gdp	0.26	2.94	−0.25	2.24	0.68	8.79	0.10	2.75	0.15	2.88	−0.14	1.91	−0.09	1.28	0.10	3.26
export	2.45	12.58	1.24	12.27	0.72	14.03	4.76	13.76	−0.25	16.47	2.79	9.75	4.19	11.86	2.27	12.96
import	1.18	9.87	0.69	9.17	−3.98	9.13	2.77	8.45	1.02	11.12	−0.13	7.29	2.08	7.51	0.52	8.94
ind prod	−1.35	13.20	0.13	17.89	−8.41	6.61	0.30	12.65	−3.70	13.43	−1.71	7.42	−1.05	7.84	−2.25	11.29
m3	2.48	3.05	2.44	2.31	7.15	8.29	3.53	2.78	2.88	2.67	2.08	3.63	4.72	2.66	3.61	3.63
loan priv	2.51	1.93	−3.05	3.29	8.59	5.04	0.95	1.52	5.63	2.49	2.76	2.79	1.68	3.45	2.73	2.93
loan gov	2.76	3.14	2.23	13.33	−3.77	4.18	1.50	8.87	13.49	3.60	2.92	7.39	5.13	11.33	3.47	7.40
cr priv	2.82	1.49	−0.83	1.96	8.24	7.06	−0.53	2.34	5.33	2.20	1.95	2.75	1.40	2.24	2.62	2.86
cr gov	9.12	7.39	−1.62	4.10	12.68	48.88	5.79	8.67	18.19	12.56	2.00	11.74	7.93	8.69	7.73	14.57
labor	−1.18	6.38	−1.58	9.98	−6.89	15.01	−1.27	5.50	−4.35	12.26	−3.47	10.92	−5.03	12.38	−3.40	10.35
deficit	−3.02	1.00	−4.04	1.06	−4.88	2.14	−1.96	1.47	−0.93	2.06	−6.08	1.23	−3.89	2.03	−3.54	1.57
bank	75.51	20.65	84.73	29.70	77.47	32.19	72.73	16.31	96.21	14.33	77.78	20.24	107.33	14.71	84.54	21.16
vol bank	0.02	0.01	0.03	0.01	0.03	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01

Notes: This table reports summary statistics (in percentage points) for the macro indicators for the non-GIPSI countries, for the May 11, 2009–April 25, 2013 sample. "avg" denotes the average and "std" denotes the standard deviation. AT is Austria, BE is Belgium, CY is Cyprus, DE is Germany, FI is Finland, and FR is France.

Table A.3 Descriptive statistics, (G)IPSI

Variable	ES		IE		IT		PT		(G)IPSI	
	Avg	std	avg	std	avg	std	avg	std	avg	std
infl	1.92	1.35	0.01	1.82	2.22	1.11	1.75	1.77	1.47	1.51
gdp	−1.08	1.65	−2.05	3.39	−1.38	2.45	−1.44	1.95	−1.49	2.36
unempl	21.60	2.59	13.85	0.99	8.88	1.25	12.38	2.52	14.18	1.84
ind price	−0.31	8.36	0.43	5.47	−1.12	8.28	0.45	8.00	−0.14	7.53
empl	−2.43	3.93	−3.93	3.42	−0.67	0.73	−2.35	1.21	−2.34	2.32
cons	2.92	8.81	2.32	9.68	2.71	6.92	1.80	6.10	2.44	7.88
gov cons	5.03	7.05	6.14	15.17	3.53	7.25	3.78	10.21	4.62	9.92
inv	−3.57	14.77	−5.01	25.97	−1.30	9.21	−3.16	5.64	−3.26	13.90
invent gdp	0.27	4.40	0.03	1.38	−0.20	1.46	−0.30	1.32	−0.05	2.14
ex	4.77	11.97	4.02	4.41	1.60	13.63	7.21	9.94	4.40	9.99
im	−3.07	9.88	−0.49	6.17	1.25	11.96	−0.61	8.10	−0.73	9.03
ind prod	−4.76	8.56	−5.72	13.06	−2.42	7.01	−2.89	5.08	−3.95	8.43
m3	−1.28	3.53	−5.06	9.51	2.01	3.71	−2.05	4.27	−1.60	5.26
loan priv	−1.88	3.01	−10.86	5.05	3.30	3.45	−0.70	4.08	−2.54	3.90
loan gov	22.06	8.39	81.94	193.86	3.64	2.05	22.98	42.50	32.65	61.70
cr priv	−0.13	5.68	−7.09	4.31	3.87	5.60	1.28	7.10	−0.52	5.67
cr gov	19.96	13.32	7.15	22.35	11.96	5.42	35.07	26.45	18.53	16.89
labor	−5.48	12.39	−8.15	12.84	−1.63	11.40	−4.83	9.89	−5.02	11.63
deficit	−9.11	1.92	−16.55	7.13	−4.26	0.67	−6.84	2.46	−9.19	3.05
bank	82.85	17.83	43.11	43.49	77.71	18.66	78.62	22.60	70.57	25.64
vol bank	0.02	0.01	0.04	0.02	0.02	0.01	0.02	0.01	0.03	0.01

Notes: This table reports summary statistics (in percentage points) for the macro indicators (see Table A.1), GIPSI countries, for the May 11, 2009–April 25, 2013 sample. "avg" denotes the average and "std" denotes the standard deviation. ES is Spain, IE is Ireland, IT is Italy, and PT is Portugal. Values are in percentage form.

flash estimate including a preliminary release at $T+30$ days in 2016. Some countries nowadays publish preliminary GDP flash estimates, while some only publish GDP figures 60–70 days after the end of the reference quarter. Focusing on the countries covered in this study, no flash estimate is provided for Ireland. Moreover, data for GDP components may be released together with a flash estimate or only with the second GDP release, depending on the country. Even monetary and credit aggregates, which are released in a timelier manner compared to macroeconomic statistics, are published in the month following the reference month.

Given publication lags, market participants never really know the current state of the economy. They base their decisions on a continuous flow of information, where data on various macroeconomic and macro-financial indicators are released with different timeliness and revised afterward. The real-time data set that we develop reflects the information set available to market participants at each point in time, based on which they form their expectations. In this respect, our data set is similar to those used in "news" studies, such as Balduzzi, Elton, and Green (2001), Ehrmann and Fratzscher (2005) and, more recently, Beber, Brandt, and Luisi (2014).

Finally, we complement our macroeconomic and macro-financial real-time data set with a market-based indicator, namely a proxy for country-specific banking risk. As documented in the existing literature, the doom loop between sovereign and bank credit risk was indeed the hallmark of the 2009–2012 sovereign debt crisis in the eurozone (Brunnermeier et al. 2016). To obtain a market-based proxy for banking risk daily, we used the country-specific banking equity index. More specifically, since sovereign CDSs and banking equity indices are strongly correlated, we orthogonalize the daily banking equity returns by regressing them on the contemporaneous daily change in sovereign CDS premia and saving the residuals. The cumulative sums of residuals were used to construct our measure of banking risk, which, by construction, is uncorrelated with the variations in CDS premia, thereby reflecting the health status of the banking systems as measured by the market. While the banking risk indicator, computed as a daily cumulative sum of residuals, reflects the "level" of health of the banking systems, we also computed the 20-day rolling window realized volatility of the residuals, which gives us a measure of the uncertainty associated with banks' health.

References

Abel, A. 2018. "Classical Measurement Error with Several Regressors." Working paper. Wharton School of the University of Pennsylvania.

Acharya, V., T. Eisert, C. Eufinger, and C. Hirsch. 2018. Real Effects of the Sovereign Debt Crisis in Europe: Evidence from Syndicated Loans. *Review of Financial Studies* 34: 2855–2896.

Afonso, A., M. Arghyrou, M. D. Gadea, and A. Kontonikas. 2018. "Whatever It Takes" to Resolve the European Sovereign Debt Crisis? Bond Pricing Regime Switches and Monetary Policy Effects. *Journal of International Money and Finance* 86: 1–30.

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. 2nd edn. Hoboken, NJ: John Wiley and Sons.

Aizenman, J., M. Hutchison, and Y. Jinjarak. 2013. What is the Risk of European Sovereign Debt Defaults? Fiscal Space, CDS Spreads and Market Pricing of Risk. *Journal of International Money and Finance* 34: 37–59.

Alessi, L., and C. Detken. 2011. Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity. *European Journal of Political Economy* 27: 520–533.

Alessi, L., and C. Detken. 2014. On Policymakers' Loss Functions and the Evaluation of Early Warning Systems. *Comment. Economics Letters* 124: 338–340.

Augustin, P. 2018. The Term Structure of CDS Spreads and Sovereign Credit Risk. *Journal of Monetary Economics* 96: 53–76.

Balduzzi, P., E. Elton, and C. Green. 2001. Economic News and Bond Prices: Evidence from the U.S. Treasury Market. *The Journal of Financial and Quantitative Analysis* 36: 523–543.

Ball, L., G. Mankiw, and R. Reis. 2005. Monetary Policy for Inattentive Economies. *Journal of Monetary Economics* 52: 703–725.

Beber, A. M. Brandt, and M. Luisi. 2014. "Eurozone Sovereign Yield Spreads and Diverging Economic Fundamentals." Working paper. CEPR.

Beetsma, R., M. Giuliodori, F. de Jong, and D. Widijanto. 2013. Spread the News: The Impact of News on the European Sovereign Bond Markets during the Crisis. *Journal of International Money and Finance* 34: 83–101.

Beirne, J., and M. Fratzscher. 2013. The Pricing of Sovereign Risk and Contagion during the European Sovereign Debt Crisis. *Journal of International Money and Finance* 34: 60–82.

Belloni, A., V. Chernozhukov, and C. Hansen. 2014. Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies* 81: 608–650.

Belloni, A., V. Chernozhukov, and K. Kato. 2015. Uniform Post Selection Inference for Least Absolute Deviation Regression and Other z-Estimation Problems. *Biometrika* 102: 77–94.

Bernoth, K., and B. Erdogan. 2013. Sovereign Bond Yield Spreads: A Time-Varying Coefficient Approach. *Journal of International Money and Finance* 34: 639–656.

Blommestein, H., S. Eijffinger, and Z. Qian. 2016. Regime-Dependent Determinants of Euro Area Sovereign CDS Spreads. *Journal of Financial Stability* 22: 10–21.

Bolton, P., and O. Jeanne. 2011. Sovereign Default Risk and Bank Fragility in Financially Integrated Economies. *IMF Economic Review* 59: 162–194.

Breckenfelder, J., and B. Schwaab. 2018. Bank to Sovereign Risk Spillovers across Borders: Evidence from the ECB's Comprehensive Assessment. *Journal of Empirical Finance* 49: 247–262.

Brunnermeier, M., L. Garicano, P. Lane, M. Pagano, R. Reis, T. Santos, D. Thesmar, S. V. Nieuwerburgh, and D. Vayanos. 2016. The Sovereign-Bank Diabolic Loop and ESBies. *American Economic Review* 106: 508–512.

Calvo, G. 1988. Servicing the Public Debt: The Role of Expectations? *American Economic Review* 78: 647–661.

Chiarella, C., S. ter Ellen, X. Z. He, and E. Wu. 2015. Fear or Fundamentals? Heterogeneous Beliefs in the European Sovereign CDS Market. *Journal of Empirical Finance* 32: 19–34.

Chinco, A., A. D. Clark-Joseph, and M. Ye. 2019. Sparse Signals in the Cross-Section of Returns. *The Journal of Finance* 74: 449–492.

Corradin, S., N. Grimm, and B. Schwaab. 2020. "Euro Area Sovereign Bond Risk Premia during the Covid-19 Pandemic." *Working papers series 2021*. ECB.

Corsetti, G., and L. Dedola. 2016. The Mystery of the Printing Press: Self-Fulfilling Debt Crises and Monetary Sovereignty. *Journal of the European Economic Association* 14: 1329–1371.

De Grauwe, P., and Y. Ji. 2012. Mispricing of Sovereign Risk and Macroeconomic Stability in the Eurozone. *JCMS: Journal of Common Market Studies* 50: 866–880.

De Grauwe, P., and Y. Ji. 2013. Panic-driven austerity in the Eurozone and its implications. *VOX*.

De Santis, R. 2014. The Euro Area Sovereign Debt Crisis: Identifying Flight-to-Liquidity and the Spillover Mechanisms. *Journal of Empirical Finance* 26: 150–170.

De Santis, R. 2019. Redenomination Risk. *Journal of Money Credit and Banking* 51: 2173–2206.

Delatte, A. L., J. Fouquau, and R. Portes. 2017. Regime-Dependent Sovereign Risk Pricing during the Euro Crisis. *Review of Finance* 21: 363–385.

Dewachter, H., L. Iania, M. Lyriox, and M. de Sola Perea. 2015. A Macro-Financial Analysis of the Euro Area Sovereign Bond Market. *Journal of Banking and Finance* 50: 308–325.

Di Cesare, A.G. Grande, M. Manna, and M. Taboga. 2015. "Recent estimates of sovereign risk premia for Euro-area countries." Banca D'Italia Occasional Papers No. 128.

Ehrmann, M., and M. Fratzscher. 2005. Exchange Rates and Fundamentals: New Evidence from Real-Time Data. *Journal of International Money and Finance* 24: 317–341.

European Commission. 2015. Highlight: Revisiting the relation between inventories and production using survey data. *European Business Cycle Indicators*, 1st quarter, 7–11.

Fama, E. F., and J. D. MacBeth. 1973. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy* 3: 607–636.

Favero, C. 2013. Modelling and Forecasting Government Bond Spreads in the Euro Area: A GVAR Model. *Journal of Econometrics* 177: 343–356.

Feng, G. S. Giglio, and D. Xiu. 2019. "Taming the Factor Zoo: A Test of New Factors." Working paper 25481. National Bureau of Economic Research.

Freyberger, J., A. Neuhierl, and M. Weber. 2020. Dissecting Characteristics Nonparametrically. *The Review of Financial Studies* 33: 2326–2377.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33: 1–22.

Gabaix, X. 2014. A Sparsity-Based Model of Bounded Rationality. *The Quarterly Journal of Economics* 129: 1661–1710.

Ghysels, E., C. Horan, and E. Moench. 2018. Forecasting through the Rearview Mirror: Data Revisions and Bond Return Predictability. *The Review of Financial Studies* 31: 678–714.

Giannone, D., J. Henry, M. Lalik, and M. Modugno. 2012. An Area-Wide Real-Time Database for the Euro Area. *Review of Economics and Statistics* 94: 1000–1013.

Giannone, D., L. Reichlin, and D. Small. 2008. Nowcasting: The Real Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics* 55: 665–676.

Hastie, T.R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*. 2nd edn. New York, NY: Springer-Verlag.

Kaufman, L., and P. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: Wiley.

Kehoe, T., and H. Cole. 2000. Self-Fulfilling Debt Crises. *Review of Economic Studies* 67: 91–116.

Kim, S. J., L. Salem, and E. Wu. 2015. The Role of Macroeconomic News in Sovereign Cds Markets: Domestic and Spillover News Effects from the US, the Eurozone and China. *Journal of Financial Stability* 18: 208–224.

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor. 2016. Exact Post-Selection Inference, with Application to the Lasso. *The Annals of Statistics* 44: 907–927.

Lindgren, F., B. Hansen, W. Karcher, M. Sjöström, and L. Eriksson. 1996. Model Validation by Permutation Tests: Applications to Variable Selection. *Journal of Chemometrics* 10: 521–532.

Longstaff, F., J. Pan, L. Pedersen, and K. Singleton. 2011. How Sovereign is Sovereign Credit Risk? *American Economic Journal: Macroeconomics* 3: 75–103.

Manasse, P., and N. Roubini. 2009. Rules of Thumb for Sovereign Debt Crises. *Journal of International Economics* 78: 192–205.

Sims, C. A. 2003. Implications of Rational Inattention. *Journal of Monetary Economics* 50: 665–690.

Taylor, J., and R. Tibshirani. 2018. Post-Selection Inference for -Penalized Likelihood Models. *The Canadian Journal of Statistics = Revue Canadienne de Statistique* 46: 41–61.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society* 58: 267–288.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight. 2005. Sparsity and Smoothness via the Fused Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67: 91–108.

Tibshirani, R.R. Tibshirani, J. Taylor, J. Loftus, S. Reid, and J. Markovic. 2019. Tools for post-selection inference (manual for the r package 'selectiveinference'). *CRAN Depository.*

Yuan, M., and Y. Lin. 2006. Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68: 49–67.

Zhang, C. H., and S. S. Zhang. 2014. Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76: 217–242.

Zou, H. 2006. The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* 101: 1418–1429.

Zou, H., and T. Hastie. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67: 301–320.