

Dimpfl, Thomas; Langen, Tobias

**Conference Paper**

## A Cross-Country Analysis of Unemployment and Bonds with Long-Memory Relations

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: International Macroeconomics, No. A09-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Dimpfl, Thomas; Langen, Tobias (2015) : A Cross-Country Analysis of Unemployment and Bonds with Long-Memory Relations, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: International Macroeconomics, No. A09-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/112921>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A Cross-Country Analysis of Unemployment and Bonds with Long-Memory Relations\*

This version: February 10, 2015

---

**Abstract:** We analyze the relationship between unemployment rate changes and government bond yields during and after the most recent financial crisis across nine industrialized countries. The study is conducted on a weekly basis and we therefore nowcast unemployment data, which are only available once a month, on a weekly frequency using Google search query data. In order to account for the time series' long-memory components during the first-stage nowcasting and the second-stage modeling, we draw on Corsi's (2009) heterogeneous autoregressive time series model. In particular, we adapt this idea to a setting of mixed-frequency nowcasting. Our results indicate that Google searches greatly increase the nowcasting accuracy of unemployment rate changes. The impact of an idiosyncratic rise in unemployment on bond yields turns out to be positive for European countries while it is negative for the United States and Australia. The speed of the response also varies. Not unexpectedly, bond yields do not have an impact on unemployment. Our findings have interesting implications for the way shocks are absorbed in economic systems that differ, in particular, with respect to the central bank's core tasks.

---

*Keywords:* Nowcasting; Long-memory components; Heterogeneous VAR; Unemployment; Bond yields; Google searches

*JEL Classification:* G10, G12

---

\*Financial support from the German Research Foundation (DFG) is gratefully acknowledged. We retain responsibility for all remaining errors.

# 1 Introduction

Government bonds are by no means risk-free investments. In the aftermath of the financial crisis investors demanded considerable compensation for taking the risk of owning government bonds from nearly insolvent countries like Greece, Spain or Portugal. Economic theory suggests that a combination of a high level of government debt and low tax revenue due to high unemployment rates may increase the default risk of a country. Still, government bonds are affected by monetary policy which could halt or even reverse the effect. Earlier studies on the relationship between government debt and unemployment rates did not find a clear-cut pattern of interdependency. In particular, empirical studies are hampered by the fact that unemployment rates and government bond yields are observed on different frequencies. Bonds are continuously traded while unemployment rates are only announced once a month. The general solution is to align the data on a monthly frequency. Decreasing the frequency of the bond data obviously entails a substantial loss of information.

In this study, we examine the relationship between unemployment and government bond yields on a weekly basis, that is on a frequency which is higher than the publishing frequency of unemployment figures. To this end, we describe a general method to increase the resolution of a lower frequency time series when additional, related high-frequency data are available. The method implies a nowcasting of the low-frequency time series that feeds the supplementary data as a cascade of frequencies into the nowcast. The cascade structure accounts for long-run components of the higher frequency time series in a parsimonious way. We refer to this model as M-HAR model as it combines methodologies from the literature on mixed-frequency nowcasting (M; see, *inter alia*, Marcellino and Schumacher, 2010) and heterogeneous autoregressions (HAR; in particular following Corsi, 2009). We use Google search query data to increase the frequency of unemployment data from a monthly to a weekly level. For that purpose, the principal component of various search queries for information about unemployment is fed into the prediction model. In the process of nowcasting unemployment based on weekly Google data and monthly unemploy-

ment information, the M-HAR model can process both short-term and long-term relations between these time series. This is important because an individual's behavior that is indicative of becoming unemployed in the future can be observed far ahead of the date where the person actually becomes officially unemployed. Internet searches that were conducted one or multiple weeks before are used for the current unemployment nowcast. Therefore, our model can account for a general rise in searches that took place already during last three months, while current search intensity might be going down. We show that including Google search query data greatly increases the nowcasting accuracy compared to a pure autoregressive model. The resulting weekly unemployment time series is then used in a heterogeneous vector-autoregression (HVAR) of unemployment changes and bond yields. For a sample of seven European countries we consistently find that bond yields react positively to a rise in unemployment while for the United States and Australia this effect is negative. In contrast, there is virtually no impact of shocks in bond prices on unemployment.

Currently, the literature that directly analyzes the relationship between unemployment and bond prices is scarce. A notable exception is Bayoumi et al. (1995) who find that debt financing costs rise by nine basis points if unemployment rises by one percentage point. Still, there is quite a number of event studies that investigate how the announcement of unemployment figures impacts on government bond rates. Balduzzi et al. (2001) find that short-term (three-month T-bill, two-year note) and long-term bonds (ten-year note) in the US react positively to a surprise rise in jobless claims. Similarly, Fleming and Remolona (1999) report that employment announcements are the macroeconomic announcements that have the greatest effect on bond prices. In line with these findings, Afonso et al. (2011) document that the structural level of unemployment has a negative long-run impact on the credit rating of a country which may ultimately lead to rising refinancing costs, i.e. higher bond prices.

A possible explanation for why this kind of analysis is rare is a data issue. Unemployment figures and bond prices are not available at an identical frequency and one faces the question how to suitably aggregate bond data to monthly levels. Recently,

mixed data sampling (MIDAS) regression models (in particular Ghysels et al., 2004) have become popular as a means to alleviate this problem. For example, Ghysels et al. (2006) predict daily stock price volatility combining different intraday frequencies of returns. In this spirit, Marcellino and Schumacher (2010) combine the virtue of MIDAS regressions with factor models to reduce the dimensionality in macroeconomic models and to overcome the sampling frequency issue when now- and forecasting macroeconomic variables.

Another challenge in modeling government bond data is the long-memory property of this financial time series. Currently, there are two major ways to model stationary time series with long-memory properties. First, there is the class of ARFIMA (autoregressive fractionally integrated moving average) models. Sibbertsen et al. (2014) for example document that the bonds of France, Germany, Italy, and Spain are highly persistent and are close to unit-root behavior during the financial crisis. A second way to account for this high persistence is the cascading structure of heterogeneous autoregressive models as suggested by Corsi (2009). As the model is readily implemented it is now frequently and successfully applied (see, amongst others, Bauer and Vorkink, 2011; Dimpfl and Jung, 2012; Tseng et al., 2009). Chiriac and Voev (2011) compare the two methodologies and find a similar performance of the ARFIMA and the HAR model.

The forecast of a low-frequency time series at a higher frequency is termed nowcasting. This is of particular interest in macroeconomic modeling where GDP and other important economic indicators are only available on a quarterly or even yearly basis and are generally provided with a substantial time lag. Factors that influence the variable of interest are, however, available on a higher frequency and can be used to provide a nowcast of the still unobserved variable. Giannone et al. (2008) and Kuzin et al. (2011), for example, provide intra-quarter nowcasts of current GDP growth rates. On an even longer horizon Navicke et al. (2014) nowcast the at-risk-of-poverty rate in the EU which is only published with a delay of 2-3 years. This publication lag is obviously too long for policy recommendations.

In order to nowcast unemployment rates in France, Germany, Italy and Portugal, Barreira et al. (2013) draw on Google search query data. Google data have recently gained particular attention as they allow to forecast a broad range of (economic) data. The ground-breaking work in this context is Ginsberg et al. (2009) who use search query data to find early warning signs of an upcoming influenza epidemic. The basic idea behind using Google search queries is to find a measure for the relative importance of a subject for the individuals that are directly concerned. Da et al. (2011) use this relation to predict stock price returns while Bank et al. (2011) explain time-varying liquidity supply in the German stock market. Choi and Varian (2012) were the first to use Google search query data in predicting unemployment figures. They find the addition of Google data particularly useful to identify turning points in unemployment.

This paper exploits both the advances in the literature on modeling long-memory components and the literature on mixed-frequency nowcasting. Combining these methodologies we can conduct an analysis of the relationship between unemployment and government bond yields at a high frequency. Changes in unemployment are preceded by changes in relevant search activity several months earlier. Using the information available in search activity data significantly increases the nowcasting accuracy of unemployment for each of the nine considered countries. The reaction of government bond yields to a rise in unemployment is country-specific. In Europe, government bond yields increase in reaction to a shock in unemployment. This result is in line with economic theory which suggests that countries with higher unemployment face increased insolvency risk. Government bond yields of Australia and the United States decrease in reaction to a shock in unemployment. This might be explained by a policy of more aggressive monetary interventions, as pursued in particular by the US federal reserve system.

The remainder of this paper is structured as follows. Section 2 introduces the M-HAR method to nowcast unemployment and presents the heterogeneous VAR that is used to investigate the relationship of bonds and unemployment. Section 3 describes the data and extensively discusses the intricacies of working with Google's search volume data. Section 4 presents the empirical results and section 5 concludes.

## 2 Methodology

### 2.1 Nowcasting with Long-Run Relations

The M-HAR model can produce nowcasts of a low-frequency series by drawing on information from a related long-memory high-frequency series, where the timing of important informative events is not obvious ex-ante. Consider the case of Google searches and unemployment. How long does it take for an increased interest in unemployment benefits due to economic distress to translate into actual unemployment figures? How far in advance do individuals know they might be laid off? The relevant time frame could span a single week (e.g. in a hire-and-fire state like Louisiana in the US) or months (e.g. in Germany where the majority of workers have a three months notice period). To capture the entire, possibly important time structure in a vector-autoregression, a large number of lags is needed. Estimating such a model, however, suffers from the curse of dimensionality. In the M-HAR model this long lag structure is replaced by a much more parsimonious heterogeneous frequency cascade.

Let  $y_{t+1}$  be the low-frequency time series for which a nowcast is needed. Assume that  $y_{t+1}$  follows its own autoregressive structure of order  $p$  such that

$$\begin{aligned} y_{t+1} &= \phi_1^{(y)} y_t + \phi_2^{(y)} y_{t-1} + \dots + \phi_p^{(y)} y_{t-p+1} + \varepsilon_t \\ &= \sum_{j=1}^p \phi_j^{(y)} y_{t-j+1} + \varepsilon_t, \end{aligned} \tag{1}$$

where  $\varepsilon_t$  follows some distribution with finite first and second moments. Now let  $x_{t,i}$  be a related higher frequency time series that can be used to produce nowcasts of  $y_{t+1}$ . Introducing a suitable structure of  $x$  into equation (1) results in

$$y_{t+1} = \sum_{j=1}^q \sum_{i=1}^{\tau} \phi_{j,i}^{(x)} x_{t-j+1,i} + \nu_{t,i} + \sum_{j=1}^p \phi_j^{(y)} y_{t-j+1} + \varepsilon_t, \tag{2}$$

where  $\tau$  is the number of partitions of the interval  $[t, t+1]$  and  $q$  is the number of low-frequency lags of  $x_{t,i}$ . In principle, the parameters in equation (2) could be directly estimated. However, equation (2) potentially includes a large number of lags

(namely  $q \cdot \tau$ ) of the higher frequency series  $x$ , which might ultimately render estimation infeasible. Instead let us replace the homogeneous autoregressive structure of  $x$  with a parsimonious heterogeneous frequency cascade  $\mathbb{C}(S)$  that captures the original autoregressive properties.  $S$  is a set of frequencies (e.g. weekly, monthly, quarterly, ...) and  $\mathbb{C}$  is the corresponding operator that aggregates  $x$  to the respective frequencies in  $S$ . Replacing the first term on the right hand side of equation (2) by this frequency cascade leads to

$$y_{t+1} = \mathbb{C}(S)x_{t,i} + \nu_{t,i} + \sum_{j=1}^p \phi_j^{(y)} y_{t-j+1} + \varepsilon_t \quad \forall i \in [1, \tau]. \quad (3)$$

In our application, unemployment data is available on a monthly basis while Google search volume is available on a weekly basis. In order to nowcast the monthly unemployment  $u_{t+1}$  we propose a frequency cascade structure of weekly data supplied with monthly, quarterly and yearly aggregates to cover the long-memory property of search query data:

$$u_{t+1} = c_w g_{t,i}^{(w)} + c_m g_{t,i}^{(m)} + c_q g_{t,i}^{(q)} + c_y g_{t,i}^{(y)} + \nu_{t,i} + \sum_{j=1}^p \phi_j^{(u)} u_{t-j+1} + \varepsilon_t \quad \forall i \in [1, 4], \quad (4)$$

where  $g_{t,i}^{(w)}$  is the weekly search volume series as provided by Google, and  $g_{t,i}^{(m)}$ ,  $g_{t,i}^{(q)}$  and  $g_{t,i}^{(y)}$  are series aggregated to monthly, quarterly and yearly averages, respectively. If we wanted to cover a similar time span without using the frequency cascade, we would have to include  $12 \times 4 = 48$  lags of  $g_{t,i}^{(w)}$ . The higher frequency Google series  $g_{t,i}^{(w)}$  has an additional amount of  $\tau = 4$  observations for each time period  $t$ , allowing for four consecutive nowcasts before observing  $u_{t+1}$ :

$$\begin{aligned} \hat{u}_{t+1,1} &= c_w g_{t,1}^{(w)} + c_m g_{t,1}^{(m)} + c_q g_{t,1}^{(q)} + c_y g_{t,1}^{(y)} + \sum_{j=1}^p \phi_j^{(u)} u_{t-j+1} \\ \hat{u}_{t+1,2} &= c_w g_{t,2}^{(w)} + c_m g_{t,2}^{(m)} + c_q g_{t,2}^{(q)} + c_y g_{t,2}^{(y)} + \sum_{j=1}^p \phi_j^{(u)} u_{t-j+1} \\ \hat{u}_{t+1,3} &= c_w g_{t,3}^{(w)} + c_m g_{t,3}^{(m)} + c_q g_{t,3}^{(q)} + c_y g_{t,3}^{(y)} + \sum_{j=1}^p \phi_j^{(u)} u_{t-j+1} \\ \hat{u}_{t+1,4} &= c_w g_{t,4}^{(w)} + c_m g_{t,4}^{(m)} + c_q g_{t,4}^{(q)} + c_y g_{t,4}^{(y)} + \sum_{j=1}^p \phi_j^{(u)} u_{t-j+1} . \end{aligned} \quad (5)$$



The parameters  $c_i$  and  $\phi_j^{(u)}$  in equation system (5) are estimated using quasi maximum likelihood (QML).

## 2.2 Unemployment and Bonds – Heterogeneous VAR

In order to analyze the relationship between unemployment and government bonds we draw on the well-established framework of vector autoregressive (VAR) models. The VAR model is implemented on a weekly basis. The nowcasts of unemployment based on the Google search volume data are matched with the quarterly observed data to construct a weekly time series. We then implement a heterogeneous VAR which accounts for past observations up to three years. Denote the weekly unemployment changes by  $\Delta u_t$  and the weekly bond yields by  $b_t$ . The set  $S$  includes the frequencies of one week, one, three and six months, and one, two and three years. The model reads as follows:

$$\begin{aligned} b_t &= a_{10} + \mathbb{C}_{11}(S)b_{t-1} + \mathbb{C}_{12}(S)\Delta u_{t-1} + \epsilon_{1,t} \\ \Delta u_t &= a_{20} + \mathbb{C}_{21}(S)b_{t-1} + \mathbb{C}_{22}(S)\Delta u_{t-1} + \epsilon_{2,t}. \end{aligned} \tag{6}$$

As we only have a reduced-form structure, the model can be estimated linewise with OLS.

## 3 Data Description

### 3.1 Unemployment Rates and Bond Yields

We analyze the relationship between unemployment rate changes and bond yields for a panel of nine countries: seven European countries (Austria, France, Germany, Portugal, Spain, Switzerland, and the UK) and two non-European countries (Australia and the US). All monthly unemployment series are obtained from Eurostat,<sup>1</sup> except for Australia and Switzerland for which the series are available on the Aus-

---

<sup>1</sup>[ec.europa.eu/eurostat](http://ec.europa.eu/eurostat)

tralian Bureau of Statistics<sup>2</sup> and the Swiss Amstat<sup>3</sup> websites, respectively. The time period considered is January 2004 (when Google search volume data became available) until April 2014. Daily government bond series are obtained from Datastream. Ten-year government bond yields are computed by subtracting the respective short-term bond rates. For the European countries for which short-term bond data are not available the EURIBOR is used instead.

[Figure 1 about here.]

Figure 1 depicts time series plots of the unemployment rates (upper graph) and bond yields (lower graph) for all countries in the panel. At first glance we can already observe some co-movement between rises in unemployment and bond yields. Furthermore, the effect of the European debt crisis is clearly visible for Spain and Portugal: both countries experienced a sharp increase in unemployment, starting during the financial crisis in 2008, which is accompanied by a similar rise in long-term government bond yields.

For the analysis, all unemployment series are seasonally adjusted and transformed into first differences. Subsequently, they are tested for stationarity using augmented Dickey-Fuller (DF) tests and are all found to be  $I(0)$ . Following Campbell et al. (1997) we calculate the yield spread using the theoretical cointegrating vector  $(1, -1)$  and perform a cointegration analysis. We generally find  $p$ -values that reject the cointegration hypothesis. This finding shows that the yield still has a long-memory property which we will account for using the cascading lag structure as outlined in section 2. The fact that the DF test fails to reject the null hypothesis of a unit root might well be due to the lacking power of this test. Furthermore, the literature on ARFIMA models documents that the DF test fails to distinguish unit root behavior (with  $d = 1$ ) from near unit root behavior ( $d < 1$ ). We therefore also calculate the fractional difference parameter  $d$  using the method of Geweke and Porter-Hudak (1983) and find it to be close to one, depending on the bandwidth selection criteria. Again no clear answer can be provided by this test. Ultimately, we rely on the

---

<sup>2</sup>[www.abs.gov.au](http://www.abs.gov.au)

<sup>3</sup>[www.amstat.ch](http://www.amstat.ch)

theoretical result in Campbell et al. (1997) and assume stationarity for the yield spreads while being aware that there is still a long-memory property which we have to account for.

Table 1 provides descriptive statistics. On average, only Germany and Switzerland could reduce the unemployment rate between 2004 and 2014. In Austria, Australia and the US, the unemployment rate remained roughly stable while it grew slightly in France and in the UK. In Spain and Portugal the sharp rise in unemployment after the financial crisis is reflected by the higher monthly growth rates: for Spain, this rate is ten times higher (on average) than for France or the UK. Similarly, these two countries exhibit the highest weekly bond yields, followed by the US and the UK. The lowest bond yields are observed for Australia and Switzerland; in case of the latter it is even negative. We also find that unemployment changes are generally left-skewed while the skewness of bond yields is not clearly drawn to left or right.

[Table 1 about here.]

## 3.2 Google Search Volume

Ideally we would like to know how many people engage in preparations to file for unemployment and how far in advance of becoming unemployed they do so. With this information we could easily nowcast unemployment figures. This would also solve the drawback that even month-end data, which correspond to the reporting frequency, are not immediately available on the first day of the following month, but only with a lag. Of course we do not have this exact data, but we can try to approximate how many people prepare facing unemployment with the help of Google search volume data (see Choi and Varian, 2012; Barreira et al., 2013).

The data is obtained from Google Trends,<sup>4</sup> where normalized weekly search volume data is provided for any search query that surpasses a certain volume threshold. These series are available starting in January 2004. The raw search volume data

---

<sup>4</sup>[www.google.com/trends](http://www.google.com/trends)

is corrected for multiple searches from one IP address and normalized by the total search volume at that time:

$$v_{s,t} = \frac{V_{s,t}}{V_t}, \quad (7)$$

where  $v_{s,t}$  is the normalized search volume at time  $t$  for search query  $s$ ,  $V_{s,t}$  is the actual search volume, and  $V_t$  the total search volume. To better be able to compare changes in search terms with different volume levels, the normalized volumes  $v_{s,t}$  are rescaled by their maximum historical value. The normalized search volume index  $g_{s,t}$  is thus given as

$$g_{s,t} = \frac{v_{s,t}}{\max_t \{v_{s,t}\}} \cdot 100. \quad (8)$$

This preparation and pre-selection of the data by Google causes several complications. Firstly, the volume threshold criterion restricts the set of keywords we can use for our analysis. This is problematic especially for smaller countries. Secondly, the rescaling of the normalized volume series  $v_{s,t}$  by its historical maximum causes the whole series to change when a new maximum occurs.

When working with Google’s search data, it is vital to keep in mind what is measured: The number of times a keyword is entered into the search engine. To make assumptions about the intentions that an individual had when she entered a query, the choice of keywords is crucial. They need to be specific to the intention and at best unique. However, most of them will also appear in other possibly related or – worse – competing contexts. The famous prediction of influenza epidemics with search query data by Ginsberg et al. (2009) is well-known to suffer from this keyword ambiguity. Some symptoms like fever or headache are not specific to having the flu, but occur in the course of numerous illnesses. Entering the keyword “fever” is, thus, not necessarily related to the illness of interest. When a fever causing epidemic unrelated to the flu spreads, the flu index rises erroneously.

Besides the choice of keywords, the context in which search volume data is used to draw conclusions about intentions can favor or hinder the application. Search queries are usually performed to acquire information on a specific topic. When the topic is only of interest to individuals when they are in a certain situation, then we

can infer that changes in search volume indicate a higher amount of individuals in that situation (being sick, being unemployed). Still, this only works if the situation occurs rarely, so that individuals are not yet informed from previous experience and do not need to search for information anymore. When an individual suffers from flu symptoms in the flu season, he might not search for information on it anymore. For rare symptoms no such learning can occur.

These insights are very important for the application to unemployment. The context is favorable to the application because when an individual faces unemployment it is likely a rare event, unlike getting the flu. The situation is likely to be unfamiliar and new information needs to be retrieved. To make the keywords we use as situationally specific as possible, we focus on the process of filing for unemployment and receiving unemployment benefits. We sort all keywords into four categories, each linked to the search for information on the process of filing for unemployment. These categories guide the choice of keywords in the different languages and countries and ensure comparability. The category “Benefits” relates to queries about unemployment benefits while “Process or Agency” subsumes queries on how and where to file for unemployment. These two categories are very specific. The categories “Unemployed” and “Unemployment” are more general and as such more prone to be used by individuals not faced with unemployment. Still they are very specific to the context of unemployment. Out of these two categories, “Unemployed” is more likely to be part of individuals’ sentence queries like “I am unemployed”, while the category “Unemployment” more likely captures general attention. Table 2 provides an overview of the keywords used and the categories they represent. Queries marked with a star (\*) are combined queries of several similar words and include different (mis-)spellings. Not all categories are represented for each country due to limited availability as a result of low search volume.

[Table 2 about here.]

We intentionally restrict ourselves to a very precise set of queries closely related to unemployment. Other studies like Askitas and Zimmermann (2009) and D’Amuri and Marcucci (2010) include different search terms that individuals who face unem-

ployment might enter. The intention behind these queries is not to seek information about receiving unemployment benefits, but finding employment. Such queries range from searches for “jobs” or names of specific job search engines (like “indeed” or “job24”). We find that the volume of these queries behaves differently in every country.

[Figure 2 about here.]

An example is presented in Figure 2. The graphs depict the first principal component of changes in normalized search volume for queries related to jobs and unemployment benefits. For the United States the sharp increase in search volume of both series occurs almost at the same time during the crisis. For the United Kingdom however this does not hold. Here, the sharpest increase in search volume for jobs is observed years after the crisis. Searches for specific job websites have the additional problem that fluctuations in popularity lead to changes in search volume which are completely unrelated to unemployment. Moreover, the keywords would have to be frequently adjusted as new websites emerge. In our sample, only unemployment-related searches consistently occurred ahead of changes in unemployment for all nine countries. A possible explanation is that job search related queries are not necessarily specific to individuals in the situation of facing unemployment. They can just as well be a sign of a growing job market. The series just happen to be very similar for the US, which may be due to specifics of the social security system coupled with a high pressure to find a new job quickly.

All search queries listed in Table 2 are different measures of the same underlying quantity, namely impending unemployment. We perform a principal component analysis to reduce the dimensionality of search volume data and produce a single indicator time series. Before computing the principal components, each normalized search volume series is smoothed and first differences are formed. These steps are necessary to ensure that the variance-maximizing principal component analysis does not simply pick the noisiest series as the most important variance contributor. To evaluate the fit of the first identified principal component with the search volume data of the different categories we compute the average correlation per category.

Overall, the principal component represents all categories well, i.e. a high correlation is observed for all nine countries. The weakest relationship is found for the “Process” category and the principal component for Austria where the correlation is still 0.457.

Figure 3 illustrates the results.

[Figure 3 about here.]

## 4 Results and Discussion

The analysis of the relation between changes in unemployment and government bond yields is separated into two steps. In a first step we use weekly Google search volume data to nowcast monthly unemployment changes on a weekly frequency employing the M-HAR method. In a second step we use the resulting higher frequency unemployment series to analyze the relationship with government bond yields in a heterogeneous VAR framework.

### 4.1 Nowcasting Unemployment – M-HAR Estimation

The M-HAR model requires the identification of a suitable cascading structure. As outlined in Subsection 2.1, the number of relevant lags might differ across countries. Therefore, we use cross-correlations between unemployment changes and the principal component of normalized search volume changes to determine an appropriate structure. Figure 4 depicts cross-correlations at different lags of search volume changes for France and the United States. The search volume changes feature a high persistence as correlations change slowly between lags. One can observe that the correlation peaks at different lags: -2 months for the US, -5 months for France. The strength of the correlation also varies. This may be yet another indicator that the timing when people search for information about unemployment is country-specific.

[Figure 4 about here.]

Figure 5 provides an overview of the two key characteristics – time and strength – of the cross-correlation functions for all nine countries in the panel. The scatter plot depicts the maximum correlations between unemployment changes and search volume changes alongside the lag at which these occur. The blue points mark the two non-European countries, crisis countries are labeled by red squares, and the remaining European countries are tagged by green triangles. Again, we observe that there are large differences in the timing of the maximum correlation peak and the size of the correlation. There also seems to be a pattern that the closer the timing of changes in either time series, the higher is the correlation between them. This should be intuitively clear since a larger distance in time allows for more unrelated innovations to dilute the relation.

[Figure 5 about here.]

We use the M-HAR model based on equation (5) to produce four weekly nowcasts of unemployment each month. Figure 5 shows that the peak of the cross-correlation for Austria occurs at a lag of eight months. This is the longest distance observed, so a frequency cascade structure that covers one year should be able to capture a sufficient amount of lags. The 48 weekly lags are represented in the cascade by five frequency factors (week, month, three months, six months, year). The benefit of including the search volume frequency cascade in the estimation is evaluated by comparing the nowcasts to a forecast based on a pure autoregressive (AR) model of unemployment. The number of monthly lags of the autoregressive component for unemployment in both models is chosen in a way to guarantee the best model fit of the pure AR model. Selection criteria are forecasting performance, lag significance and BIC. The maximum of the optimal lag length  $p$  turns out to be three. Lag length is determined for each country individually. Parameters are estimated via QML. Table 3 summarizes the results.

[Table 3 about here.]

The magnitude of the cascade coefficient estimates, as well as their significance, closely mirrors the time at which the respective cross-correlation functions peak in



Figure 5. This result highlights that the cascade coefficients pick up the information occurring at these lags. Countries where the peak in the cross-correlation function occurs less than four months ahead have insignificant year coefficients in the M-HAR estimation. For countries with a peak more than four months ahead the year coefficients are significant. The pattern is more pronounced when the size of the coefficients is considered as well. The two exceptions are Portugal and the United Kingdom. For Portugal the overall high variance of the parameter estimates might indicate a data quality issue. The overall search volume for unemployment-related topics is so low that many queries that are used for other countries are below the Google-imposed threshold and thus unavailable. For the United Kingdom most coefficients have relatively low  $p$ -values but the largest absolute values are still observed before the four-months mark. The coefficient estimates for Spain are particularly large due to the big changes in unemployment during the financial and debt crises.

Mean squared forecasting error (MSE) and adjusted  $R^2$  are reported in the last two columns of Table 3 to measure the benefit of including the Google search volume frequency cascade in addition to a pure autoregressive model of unemployment. Overall, we find a significant reduction in MSE and a higher adjusted  $R^2$  for all countries.<sup>5</sup> The degree of improvement, however, varies greatly. Figure 6 depicts the changes as a scatter plot of MSE and  $R^2$  for the pure AR and the M-HAR model.

[Figure 6 about here.]

Interestingly, the largest improvements in both MSE and  $R^2$  can be observed for the United States, even though none of the frequency cascade estimates are statistically significant. Large improvements in both criteria can also be observed for Australia. France and Switzerland, which already feature a low MSE in the pure model, display a large increase in  $R^2$ . For the solvent European countries the unemployment forecasts based on the AR model result already in a comparatively low MSE. Therefore, the inclusion of the search query data has no large impact. The  $R^2$

---

<sup>5</sup>Differences in MSE are tested for significance using the Diebold-Mariano test (Diebold and Mariano, 1995).

values indicate that while the search volume data does not greatly improve the average forecasting error in this case, it nevertheless helps explaining a larger fraction of the total variation in unemployment. For the non-European countries and the European countries hit by the solvency crisis the inclusion of search data distinctly improves the prediction of unemployment changes by both decreasing the initially high forecast error as well as explaining a larger part of total variation.

[Figure 7 about here.]

The plots in Figure 7 present a comparison of the M-HAR model and the autoregressive model based forecasts during the crisis for the United States and France. The vertical lines represent the observed changes in the official unemployment figures, while the solid line represents the nowcasts of the M-HAR model and the dashed line is the AR-based nowcast. Due to the additional search volume data, the M-HAR forecast can anticipate movements in unemployment much better than the autoregressive model. For the United States the large changes during the crisis cannot be inferred from its past changes alone, so the pure autoregressive model forecasts deviate by a large margin from actual data, which results in a high MSE for this model. The M-HAR forecast utilizes the increase in unemployment-related search activity that takes place one month ahead (cp. Figure 4), improving the forecast noticeably. But even for France, where the autoregressive model already performs well, the inclusion of unemployment related search activity improves the forecasts.

## 4.2 Unemployment and Bonds – Heterogeneous VAR Estimation

To estimate the relationship between unemployment changes and bond yields we employ a heterogeneous VAR as described in equation (6) using the higher frequency unemployment series provided by the M-HAR model. The frequency cascades used in the heterogeneous VAR have to cover a longer time span than in the M-HAR model in the previous subsection to account for the extreme long-memory feature of the government bond yield time series (see also Diebold and Li, 2006). We find

that a period of three years captures a sufficiently long lag structure, such that the impact of innovations is not permanent but slowly decreases over time. To cover this lag structure we use seven heterogeneous AR-terms, i.e. two frequency terms more than in the previously used cascade. Like in the standard VAR context, increasing the number of parameters in the HVAR leads to larger variances of the estimates, which would in turn call for a parsimonious modeling strategy. Instead of using the same frequency cascade for each country, we could use cascades that are tailored to fit the individual countries best, i.e. leaving out insignificant frequencies or adding additional ones where necessary. This decreases the variance of the remaining estimates but it also leaves ample room for ambiguity. Instead, we opt for keeping the estimation as general as possible and comparable across countries, thereby accepting the fact that parameter estimates suffer from higher variance.

Figure 8 visualizes the estimation results of the bond equation, only displaying significant frequency estimates.<sup>6</sup> The shading indicates the  $p$ -value of the parameter estimates on a scale from  $< 0.001$  (blue) to  $> 0.2$  (white).

[Figure 8 about here.]

For Switzerland, Spain, and the US we find a rather timely response of bond yields to unemployment changes. For the remaining countries, it takes longer for the bond yields to react. However, the impact seems stronger in the latter case where the largest values of the estimates can be observed between the frequencies of six month and two years. The order of magnitude of the majority of the insignificant estimates is considerably lower than the reported values. Bond yield estimates are around one for the one week frequency which is due to the high persistence of the time series. This effect is countered by subsequent negative estimates on lower frequencies. The insignificant estimates of the bond yields cascade structure are again small in magnitude, similar to the insignificant unemployment cascade estimates. While for the impact of unemployment a more country-specific pattern can be observed, the autoregressive structure of the bond yields is rather similar across all countries.

---

<sup>6</sup>Detailed estimation results are reported in the appendix.

Figure 9 provides impulse-response functions of bond yields to an idiosyncratic shock of one standard deviation in unemployment for all countries. Contemporaneous effects are identified via Cholesky decomposition. The ordering is not crucial in the present context as the important influences are not contemporaneous but lagged.

[Figure 9 about here.]

The information contained in the impulse-response functions is condensed in Figure 10. The scatter plot depicts the peaks of the impulse-response functions and the lag at which they occur.

[Figure 10 about here.]

The graph illustrates an interesting pattern. Government bond yields of all European countries react positively to idiosyncratic shocks in unemployment. This matches the intuition that a higher unemployment rate increases future debt of the country and decreases tax income revenue. This negatively impacts future solvency and increases the risk involved in holding long-term government bonds. For the European crisis countries the impact peaks at a much later stage, implying that the impact is much more persistent. This may be due to differences in the local economic systems and the way shocks to unemployment can be absorbed quickly. The German labor market was arguably the most flexible one during the crisis, which is reflected in this graph as well. The bond markets in the United States and Australia react differently to idiosyncratic shocks in unemployment. Here, shocks to unemployment decrease government bond yields. In the United States the shock is absorbed relatively quickly while the Australian system adjusts more slowly. Differences in labor market efficiency may serve as an explanation.

The fact that the response of bond yields is negative may be an indication of a different monetary policy in the United States and Australia compared to that of the European Union. The most likely explanation is that the European Central Bank's core task "shall be to maintain price stability" (EU, 2012). Therefore, rising unemployment could be interpreted solely as a deterioration of the country's future

solvency which might lead to a rising bond yield as in particular long-term refinancing costs should rise. In contrast, the central mission of the Federal Reserve Board in the United States entails the “pursuit of maximum employment, stable prices, and moderate long-term interest rates”.<sup>7</sup> Therefore the Fed might react immediately to rising unemployment rates by lowering interest rates to stimulate public and private investment. The excess supply of money might then lead to lower bond yields. The same holds true for Australia where the Reserve Bank of Australia’s core task is also to “contribute to the stability of the currency, full employment, and the economic prosperity”.<sup>8</sup>

[Figure 11 about here.]

The impact of idiosyncratic shocks in bond yields on unemployment changes is negligible. Figure 11 presents the corresponding impulse-response functions. As can be seen, the impact graphs lie close to zero for all nine countries. This result is in line with economic reasoning. Interest rates might have an impact on both public and private investment, which in turn determines the number of available jobs in the long run. A direct impact has, to the best of our knowledge, never been derived.

## 5 Conclusion

With this work we set out to investigate the relationship between government bond yields and unemployment. Government-issued bonds are priced on the assumption that the country as a whole will be able to pay the mortgage back in the future. This ability is entirely determined by the labor force and unemployment is therefore a great risk factor in this bet. Besides the short-run negative effects of unemployment, we show that at least in Europe, rising unemployment leads to an increase in government bond yields, indicating that financing costs for the country rise. In a worst-case scenario one might imagine that the country fails to refinance mature bonds because investors deem the risk as abundant, and require extremely high risk

---

<sup>7</sup>[www.federalreserve.gov/aboutthefed/mission.htm](http://www.federalreserve.gov/aboutthefed/mission.htm)

<sup>8</sup>[www.rba.gov.au/about-rba/index.html](http://www.rba.gov.au/about-rba/index.html)

compensation (as happened in Greece or Spain) or refrain from investing at all, leading to a state bankruptcy. The current European sovereign debt crisis is such a situation. And even if it seems obvious, the reported result has an important policy implication: unemployment reduction has to be a major goal of government activity. Reducing unemployment not only raises the tax basis and, thus, alleviates the debt burden, but in the long run also leads to lower interest rates as the risk compensation requested by investors is reduced. A notable exception are the United States for which the trust of investors in the country's ability to meet its debt obligations is seemingly endless.

The investigation of the relationship between unemployment and bond yields requires a suitable preparation of the time series because bond and unemployment data are only available on different frequencies. Our methodological contribution consists of a proposition how to provide weekly data for unemployment using a mixed-frequency heterogeneous autoregressive model. We show that to this end Google search queries can successfully be used to nowcast unemployment changes. The application of the M-HAR model is of course not limited to unemployment nowcasting but can generally be applied to any context where a nowcast that employs a highly persistent time series is needed.

## Appendix

[Table 4 about here.]

## References

- AFONSO, A., P. GOMES, AND P. ROTHER (2011): “Short- and long-run determinants of sovereign debt credit ratings,” *International Journal of Finance & Economics*, 16, 1–15.
- ASKITAS, N. AND K. F. ZIMMERMANN (2009): “Google Econometrics and Unemployment Forecasting,” *Applied Economics Quarterly*, 55, 107–120.
- BALDUZZI, P., E. J. ELTON, AND T. C. GREEN (2001): “Economic News and Bond Prices: Evidence from the U.S. Treasury Market,” *Journal of Financial and Quantitative Analysis*, 36, 523–543.
- BANK, M., M. LARCH, AND G. PETER (2011): “Google search volume and its influence on liquidity and returns of German stocks,” *Financial Markets and Portfolio Management*, 25, 239–264.
- BARREIRA, N., P. GODINHO, AND P. MELO (2013): “Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends,” *NET-NOMICS: Economic Research and Electronic Networking*, 14, 129–165.
- BAUER, G. H. AND K. VORKINK (2011): “Forecasting multivariate realized stock market volatility,” *Journal of Econometrics*, 160, 93–101.
- BAYOUMI, T., M. GOLDSTEIN, AND G. WOGLOM (1995): “Do Credit Markets Discipline Sovereign Borrowers? Evidence from U.S. States,” *Journal of Money, Credit and Banking*, 27, 1046–1059.
- CAMPBELL, J. Y., A. W. LO, AND A. C. MACKINLEY (1997): *The Econometrics of Financial Markets*, Princeton, New Jersey: Princeton University Press.
- CHIRIAC, R. AND V. VOEV (2011): “Modelling and forecasting multivariate realized volatility,” *Journal of Applied Econometrics*, 26, 922–947.
- CHOI, H. AND H. VARIAN (2012): “Predicting the present with Google trends,” *Economic Record*, 88, 2–9.

- CORSI, F. (2009): “A Simple Approximate Long-Memory Model of Realized Volatility,” *Journal of Financial Econometrics*, 7, 174–196.
- DA, Z., J. ENGELBERG, AND P. GAO (2011): “In Search of Attention,” *The Journal of Finance*, 66, 1461–1499.
- D’AMURI, F. AND J. MARCUCCI (2010): “Google It! Forecasting the US Unemployment Rate with A Google Job Search Index,” FEEM Working Paper No. 31.
- DIEBOLD, F. AND R. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics*, 13, 253–263.
- DIEBOLD, F. X. AND C. LI (2006): “Forecasting the term structure of government bond yields,” *Journal of Econometrics*, 130, 337–364.
- DIMPFL, T. AND R. C. JUNG (2012): “Financial Market Spillovers Around the Globe,” *Applied Financial Economics*, 22, 45–57.
- EU (2012): “Consolidated Versions of the Treaty on European Union and the Treaty on the Functioning of the European Union,” *Official Journal of the European Union*, 55, 1–390.
- FLEMING, M. J. . AND E. M. . REMOLONA (1999): “What Moves Bond Prices?” *The Journal of Portfolio Management*, 25, 28–38.
- GEWEKE, J. AND S. PORTER-HUDAK (1983): “The Estimation and Application of Long Memory Time Series Models,” *Journal of Time Series Analysis*, 4, 221–238.
- GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2004): “The MIDAS Touch: Mixed Data Sampling Regression Models,” Anderson Graduate School of Management – Finance Working Paper Series.
- (2006): “Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies,” *Journal of Econometrics*, 1-2, 59–95.



- GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): “Nowcasting: The real-time informational content of macroeconomic data,” *Journal of Monetary Economics*, 55, 665–676.
- GINSBERG, J., M. H. MOHEBBI, R. S. PATEL, L. BRAMMER, M. S. SMOLINSKI, AND L. BRILLIANT (2009): “Detecting influenza epidemics using search engine query data,” *Nature*, 457, 1012–1014.
- KUZIN, V., M. MARCELLINO, AND C. SCHUMACHER (2011): “MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area,” *International Journal of Forecasting*, 27, 529–542.
- MARCELLINO, M. AND C. SCHUMACHER (2010): “Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP,” *Oxford Bulletin of Economics and Statistics*, 72, 518–550.
- NAVICKE, J., O. RASTRIGINA, AND H. SUTHERLAND (2014): “Nowcasting Indicators of Poverty Risk in the European Union: A Microsimulation Approach,” *Social Indicators Research*, 119, 101–119.
- SIBBERTSEN, P., C. WEGENER, AND T. BASSE (2014): “Testing for a break in the persistence in yield spreads of EMU government bonds,” *Journal of Banking & Finance*, 41, 109–118.
- TSENG, T.-C., H. CHUNG, AND C.-S. HUANG (2009): “Modeling Jump and Continuous Components in the Volatility of Oil Futures,” *Studies in Nonlinear Dynamics & Econometrics*, 13, Article 5.

### Descriptive Statistics – Unemployment Changes and Bond Yields

**Table 1:** The table provides the mean (denoted by  $\bar{x}$ ), the standard deviation  $\sigma$ , and the skewness coefficient  $\varsigma$  of monthly unemployment changes  $\Delta u$  and weekly ten-year government bond yields  $b$  for the nine countries listed in the first column.

	$\overline{\Delta u}$	$\sigma_{\Delta u}$	$\varsigma_{\Delta u}$	$\bar{b}$	$\sigma_b$	$\varsigma_b$
Austria	0.00	0.14	0.26	1.43	1.00	-0.50
France	0.01	0.08	-0.10	1.46	0.99	-0.69
Germany	-0.04	0.09	0.18	1.11	0.90	-0.50
Portugal	0.06	0.18	0.14	3.69	3.50	0.86
Spain	0.11	0.24	1.04	2.56	1.58	0.07
Switzerland	-0.01	0.07	0.78	-0.07	0.92	-0.64
UK	0.01	0.10	0.59	1.23	1.44	0.10
Australia	0.00	0.14	0.29	0.11	0.91	0.45
USA	0.00	0.18	0.57	1.96	1.15	-0.46

## List of Search Terms and Descriptive Statistics

**Table 2:** The table presents the keywords used to extract Google search volume data and descriptive statistics thereof. Keywords marked with a \* account for possible misspelling.  $v_{rel}$  is the relative average search volume normalized by the series denoted with  $v_{rel} = 100$ . For example  $v_{rel} = 10$  means that the average search volume of that series is ten times lower than for the base series.  $\sigma$  is the standard deviation of the respective search volume over time.

Country	Benefits	$v_{rel}$	$\sigma_{\Delta v}$	Keyword Categories		$v_{rel}$	$\sigma_{\Delta v}$	Unemployed	$v_{rel}$	$\sigma_{\Delta v}$	Unemployment	$v_{rel}$	$\sigma_{\Delta v}$
				Process or Agency									
Austria				ams		100	0.08	arbeitslos*	1	0.17			
France	allocation chômage*	14	0.28					au chômage*	14	0.21	chômage*	100	0.20
	calcul chômage*	7	0.23										
Germany	arbeitslosengeld	61	0.10	hartz iv		89	0.21	arbeitslos*	29	0.11	arbeitslosigkeit*	39	0.10
	arbeitslosengeld*	96	0.09	hartz iv*		100	0.20						
	arbeitslosengeld berechnung*	11	0.26										
Portugal	subsídio desemprego	35	0.17					desemprego	100	0.26			
	subsídio desemprego*	43	0.19										
Spain	subsídio desempleo*	8	0.13	inem*		100	0.21	en paro	8	0.22	paro	33	0.18
Switzerland	arbeitslosengeld*	6	0.33	arbeitslosenkasse*		14	0.19	arbeitslos*	9	0.19	arbeitslosigkeit*	11	0.31
				rav		100	0.16						
UK	jobseeker's allowance*	100	0.15					unemployed*	34	0.19	unemployment*	62	0.15
	jobseeker's allowance rate*	2	0.34										
Australia	unemployment benefits*	38	0.10					unemployed*	27	0.16	unemployment*	100	0.15
USA	unemployment benefits	6	0.23	file for unemployment*		2	0.24	unemployed*	4	0.25	unemployment	100	0.21
	unemployment benefits*	13	0.27								unemployment*	100	
	unemployment benefits rate*	2	0.28										

## Nowcasting Unemployment – Estimation Results

**Table 3:** The model presents the coefficient estimates for the model used to nowcast unemployment.  $p$ -values are given in parentheses. Model fit is evaluated by the mean squared error (MSE) and the adjusted  $R^2$  in the last two columns. The first row of each country is the pure autoregressive model, the third row is the M-HAR model. \* (\*\*) indicates that the reduction of the MSE is significant on the 5% (1%) level.

Country	$\phi_1^{(u)}$	$\phi_2^{(u)}$	$\phi_3^{(u)}$	$c_w$	$c_m$	$c_q$	$c_h$	$c_y$	$\sigma_\varepsilon$	MSE	$R^2$
Austria	0.28 (0.00)	-0.31 (0.00)							0.13 (0.00)	1.74	13.9
	0.21 (0.00)	-0.37 (0.00)		0.03 (0.96)	0.00 (1.00)	-0.03 (0.96)	-0.66 (0.11)	0.75 (0.00)	0.13 (0.00)	1.57**	22.1
France	0.39 (0.00)	0.17 (0.00)							0.07 (0.00)	0.55	23.9
	0.25 (0.00)	0.11 (0.01)		0.06 (0.70)	0.11 (0.64)	-0.43 (0.00)	0.44 (0.00)	-0.10 (0.02)	0.07 (0.00)	0.47**	38.4
Germany	0.53 (0.00)	0.24 (0.00)							0.07 (0.00)	0.43	32.9
	0.47 (0.00)	0.21 (0.00)		0.30 (0.05)	-0.26 (0.21)	0.02 (0.87)	-0.03 (0.77)	0.05 (0.16)	0.06 (0.00)	0.41*	35.4
Portugal	0.73 (0.00)								0.13 (0.00)	1.82	48.8
	0.61 (0.00)			-0.31 (0.26)	0.33 (0.38)	-0.17 (0.45)	0.26 (0.14)	0.05 (0.58)	0.13 (0.00)	1.67**	53.6

**Table 3: Nowcasting Unemployment – Estimation Results (cont.)**

Country	$\phi_1^{(u)}$	$\phi_2^{(u)}$	$\phi_3^{(u)}$	$c_w$	$c_m$	$c_q$	$c_h$	$c_y$	$\sigma_\varepsilon$	MSE	$R^2$
Spain	0.73 (0.00)	0.25 (0.00)	-0.07 (0.13)						0.12 (0.00)	1.36	75.0
	0.66 (0.00)	0.25 (0.00)	-0.07 (0.17)	1.38 (0.00)	-2.18 (0.00)	1.48 (0.00)	-0.71 (0.03)	0.13 (0.35)	0.11 (0.00)	1.25**	77.5
Switzerland	0.16 (0.00)	0.29 (0.00)	0.31 (0.00)						0.06 (0.00)	0.35	36.9
	-0.02 (0.62)	0.16 (0.00)	0.25 (0.00)	0.19 (0.04)	-0.29 (0.02)	0.28 (0.00)	-0.11 (0.08)	0.00 (0.95)	0.05 (0.00)	0.28**	48.9
UK	0.26 (0.00)	0.10 (0.03)	0.26 (0.00)						0.09 (0.00)	0.80	20.8
	0.17 (0.00)	0.05 (0.33)	0.18 (0.00)	-0.19 (0.38)	0.65 (0.03)	-0.31 (0.11)	-0.21 (0.14)	0.16 (0.01)	0.09 (0.00)	0.73**	28.6
Australia	-0.16 (0.00)								0.14 (0.00)	2.00	2.5
	-0.26 (0.00)			-0.50 (0.15)	0.69 (0.15)	-0.49 (0.07)	0.51 (0.01)	0.04 (0.67)	0.13 (0.00)	1.74**	15.4
USA	0.15 (0.00)	0.34 (0.00)	0.15 (0.00)						0.16 (0.00)	2.49	23.7
	-0.13 (0.01)	0.04 (0.42)	-0.07 (0.14)	-0.24 (0.31)	0.36 (0.27)	-0.06 (0.78)	0.09 (0.54)	0.06 (0.33)	0.14 (0.00)	1.90**	41.9

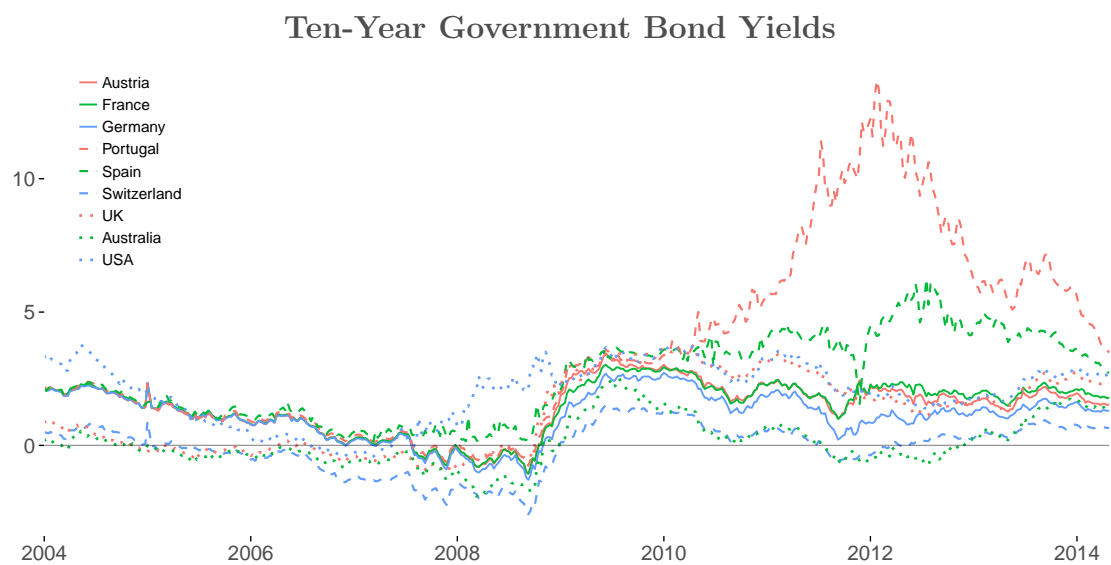
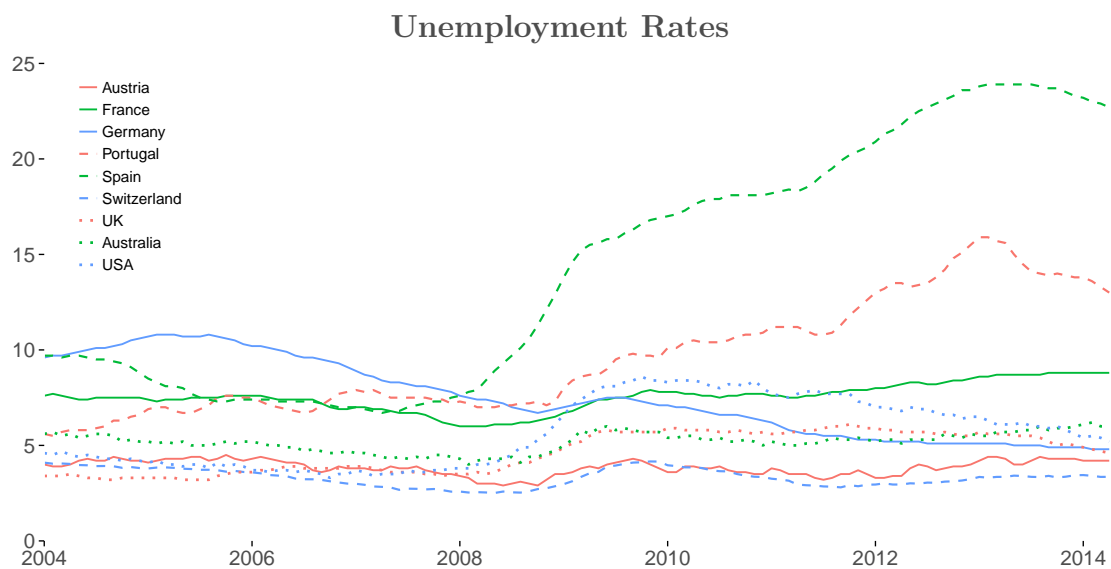
## HVAR – Estimation Results

**Table 4:** The table reports coefficient estimates for the HVAR model. DV marks the dependent variable bond yields  $b$  or unemployment changes  $u$ .  $p$ -values are in parentheses.

Country	DV	$c_{uw}$	$c_{um}$	$c_{u3m}$	$c_{u6m}$	$c_{uy}$	$c_{u2y}$	$c_{u3y}$	$c_{bw}$	$c_{bm}$	$c_{b3m}$	$c_{b6m}$	$c_{by}$	$c_{b2y}$	$c_{b3y}$
Austria	$b$	0.04 (0.83)	-0.25 (0.28)	0.01 (0.98)	0.79 (0.12)	1.20 (0.14)	0.85 (0.49)	0.15 (0.87)	1.14 (0.00)	-0.18 (0.05)	-0.02 (0.77)	-0.03 (0.60)	0.03 (0.54)	-0.12 (0.03)	0.16 (0.04)
	$u$	0.91 (0.00)	-0.27 (0.00)	0.01 (0.94)	0.34 (0.06)	-0.94 (0.00)	-0.87 (0.05)	0.42 (0.21)	0.02 (0.29)	-0.01 (0.77)	-0.01 (0.78)	0.01 (0.51)	0.01 (0.63)	0.05 (0.02)	-0.09 (0.00)
France	$b$	-0.86 (0.21)	0.91 (0.31)	-0.91 (0.22)	0.00 (1.00)	4.17 (0.02)	2.82 (0.02)	2.37 (0.15)	1.06 (0.00)	-0.16 (0.06)	0.03 (0.72)	-0.17 (0.01)	-0.08 (0.34)	0.20 (0.08)	-0.03 (0.69)
	$u$	1.02 (0.00)	-0.10 (0.22)	-0.17 (0.02)	-0.01 (0.90)	0.46 (0.01)	0.06 (0.60)	-0.10 (0.52)	0.00 (0.95)	-0.01 (0.27)	0.01 (0.14)	-0.01 (0.04)	-0.02 (0.03)	0.03 (0.00)	-0.01 (0.06)
Germany	$b$	-0.05 (0.91)	0.70 (0.28)	-0.67 (0.25)	0.19 (0.76)	1.68 (0.08)	1.25 (0.14)	1.20 (0.53)	1.10 (0.00)	-0.15 (0.09)	-0.01 (0.90)	-0.06 (0.38)	-0.04 (0.57)	0.10 (0.16)	-0.08 (0.05)
	$u$	0.89 (0.00)	0.03 (0.71)	-0.07 (0.37)	-0.03 (0.75)	-0.49 (0.00)	0.18 (0.12)	0.40 (0.13)	0.01 (0.20)	0.00 (0.82)	0.00 (0.66)	0.02 (0.03)	-0.02 (0.07)	-0.04 (0.00)	0.02 (0.00)
Portugal	$b$	0.03 (0.94)	0.11 (0.88)	-0.11 (0.87)	-0.38 (0.59)	-0.25 (0.76)	-2.07 (0.26)	4.42 (0.13)	1.00 (0.00)	-0.20 (0.03)	0.28 (0.02)	-0.07 (0.59)	-0.08 (0.42)	0.21 (0.18)	-0.21 (0.10)
	$u$	0.92 (0.00)	-0.08 (0.37)	0.05 (0.58)	-0.01 (0.93)	0.13 (0.24)	-0.37 (0.13)	-1.02 (0.01)	0.01 (0.08)	-0.01 (0.35)	0.01 (0.74)	0.00 (0.78)	0.01 (0.56)	0.03 (0.20)	-0.03 (0.06)

Table 4: HVAR – Estimation Results (cont.)

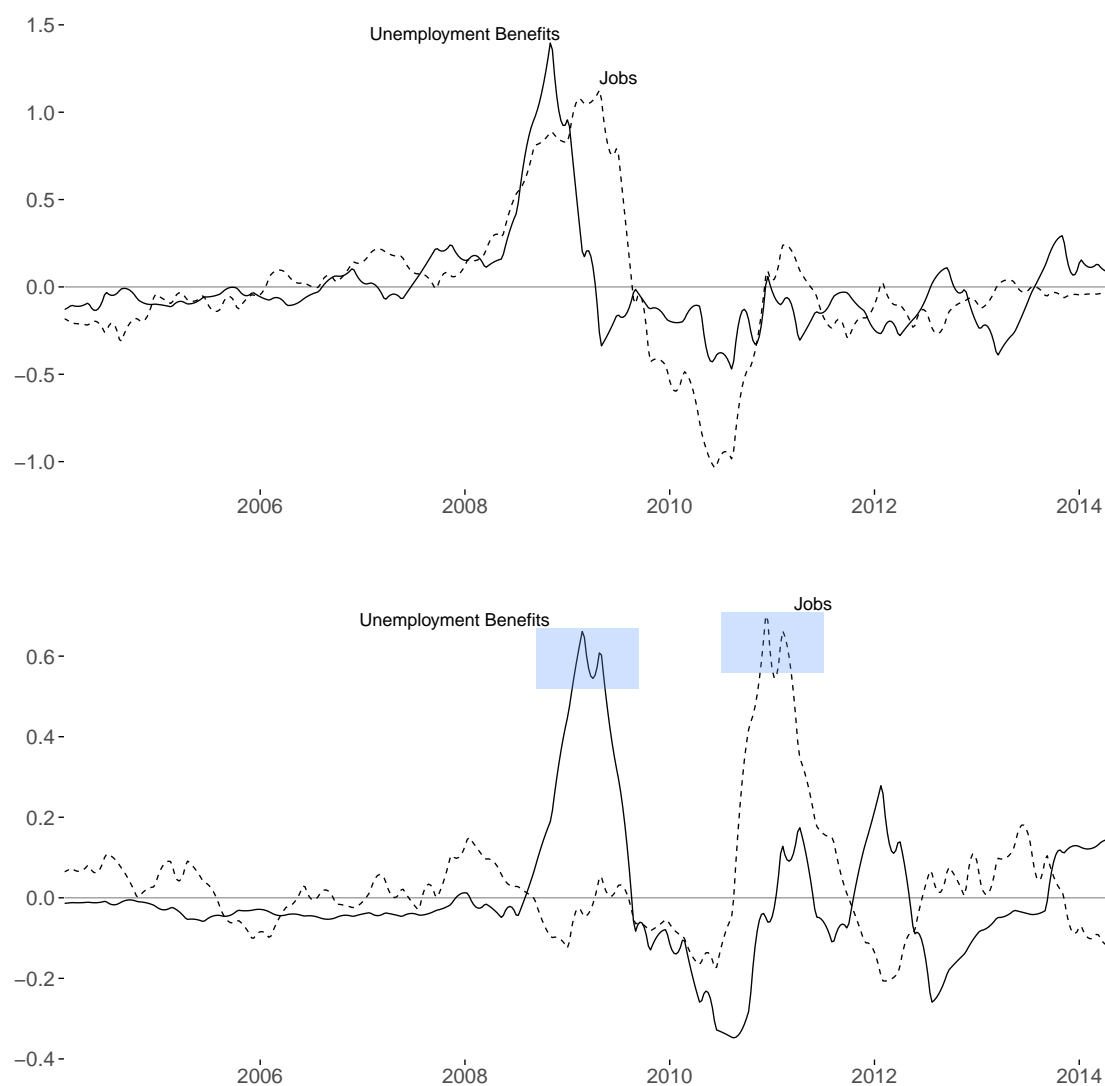
Country	DV	$c_{uw}$	$c_{um}$	$c_{u3m}$	$c_{u6m}$	$c_{uy}$	$c_{u2y}$	$c_{u3y}$	$c_{bw}$	$c_{bm}$	$c_{b3m}$	$c_{b6m}$	$c_{by}$	$c_{b2y}$	$c_{b3y}$
Spain	$b$	0.97 (0.02)	-0.74 (0.20)	-0.12 (0.81)	-0.27 (0.57)	1.31 (0.01)	0.44 (0.37)	0.21 (0.81)	0.74 (0.00)	0.02 (0.85)	0.23 (0.07)	-0.26 (0.03)	-0.03 (0.81)	0.50 (0.03)	-0.23 (0.21)
	$u$	0.88 (0.00)	0.15 (0.13)	-0.17 (0.05)	0.14 (0.09)	-0.03 (0.66)	-0.24 (0.00)	0.04 (0.78)	0.03 (0.03)	-0.03 (0.12)	0.01 (0.62)	-0.01 (0.51)	0.04 (0.04)	-0.04 (0.25)	0.00 (0.91)
Switzerland	$b$	0.76 (0.15)	-1.13 (0.11)	0.69 (0.40)	1.09 (0.29)	0.61 (0.60)	2.89 (0.00)	-1.44 (0.11)	1.11 (0.00)	-0.18 (0.03)	-0.01 (0.86)	-0.10 (0.14)	0.04 (0.55)	0.17 (0.15)	0.02 (0.78)
	$u$	1.03 (0.00)	-0.26 (0.00)	0.22 (0.02)	0.04 (0.77)	-0.32 (0.02)	-0.13 (0.26)	0.11 (0.30)	0.01 (0.36)	0.00 (0.90)	-0.01 (0.42)	0.01 (0.23)	0.00 (0.52)	-0.03 (0.03)	0.01 (0.60)
UK	$b$	-0.44 (0.40)	0.25 (0.72)	-0.10 (0.88)	1.65 (0.04)	-0.18 (0.84)	0.20 (0.81)	-3.59 (0.10)	1.01 (0.00)	0.01 (0.88)	-0.11 (0.15)	0.05 (0.52)	0.03 (0.67)	0.13 (0.26)	-0.12 (0.20)
	$u$	1.18 (0.00)	-0.36 (0.00)	0.04 (0.54)	0.00 (1.00)	-0.12 (0.22)	0.07 (0.42)	0.47 (0.05)	0.00 (0.91)	0.00 (0.90)	0.00 (0.96)	0.01 (0.16)	-0.02 (0.00)	-0.01 (0.60)	0.01 (0.34)
Australia	$b$	-0.11 (0.69)	-0.22 (0.58)	0.94 (0.17)	-0.50 (0.63)	-2.60 (0.15)	3.71 (0.08)	-0.58 (0.50)	1.06 (0.00)	-0.12 (0.17)	0.07 (0.41)	0.04 (0.64)	-0.11 (0.14)	-0.09 (0.38)	0.13 (0.27)
	$u$	0.95 (0.00)	-0.45 (0.00)	0.22 (0.12)	-0.07 (0.76)	-0.41 (0.27)	0.23 (0.61)	-0.10 (0.56)	0.01 (0.60)	-0.01 (0.69)	0.01 (0.49)	0.00 (0.85)	-0.01 (0.50)	-0.03 (0.11)	0.01 (0.72)
US	$b$	0.13 (0.77)	-1.57 (0.02)	1.25 (0.23)	-0.08 (0.95)	0.23 (0.80)	0.40 (0.49)	1.08 (0.18)	0.97 (0.00)	-0.02 (0.85)	-0.02 (0.80)	0.00 (0.97)	0.05 (0.55)	-0.38 (0.14)	0.34 (0.23)
	$u$	1.03 (0.00)	-0.60 (0.00)	0.21 (0.11)	0.53 (0.00)	-0.40 (0.00)	-0.08 (0.29)	-0.02 (0.82)	-0.01 (0.16)	0.02 (0.12)	-0.01 (0.51)	0.00 (0.71)	0.02 (0.09)	0.01 (0.87)	-0.05 (0.19)



**Figure 1:** The figure presents time series plots of monthly unemployment rates per country (upper graph) and daily ten-year government bond yields (lower graph). Both series are depicted in percent, from January 2004 to April 2014.

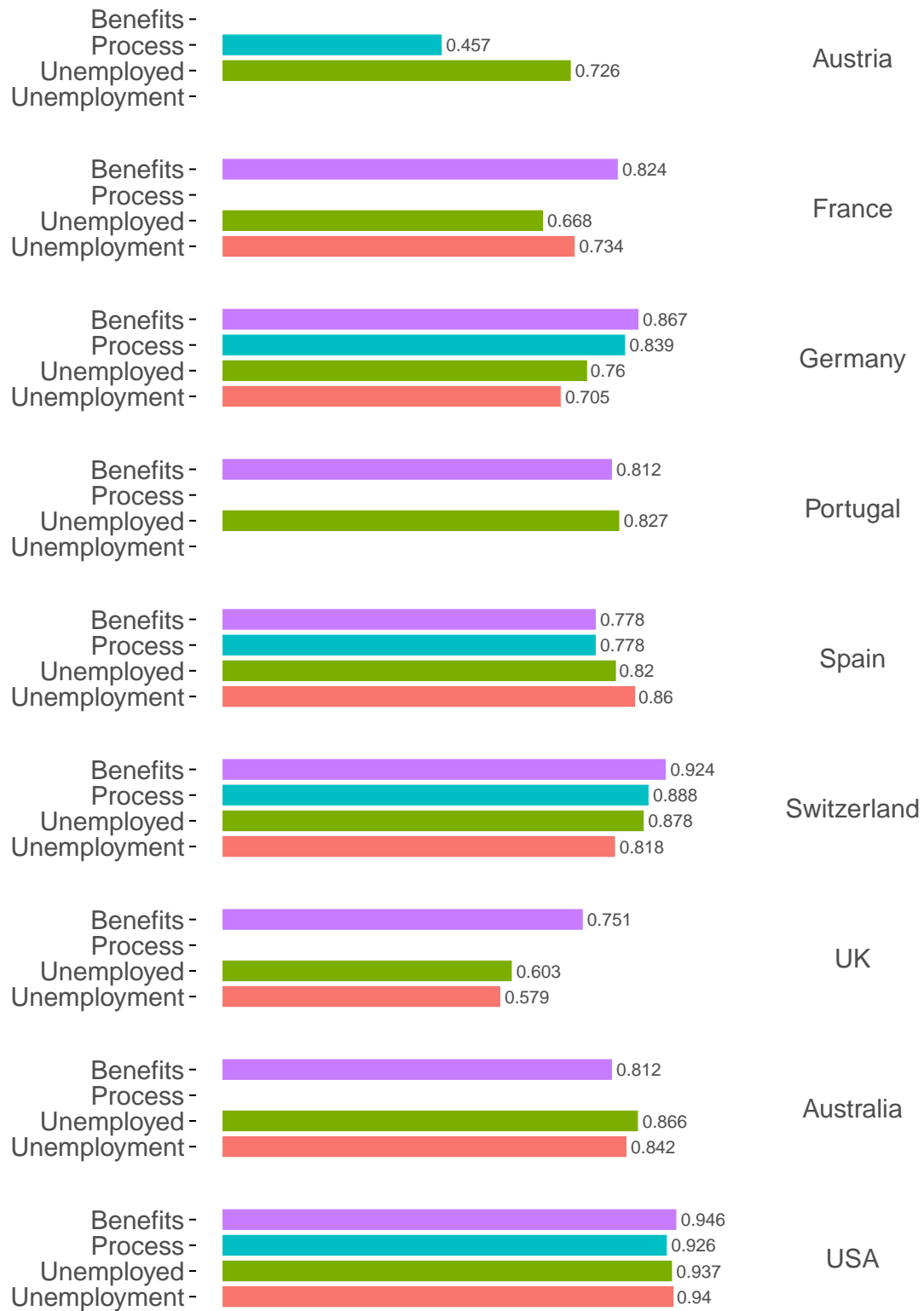


## Changes in Search Volume for "Jobs" and "Unemployment Benefits"



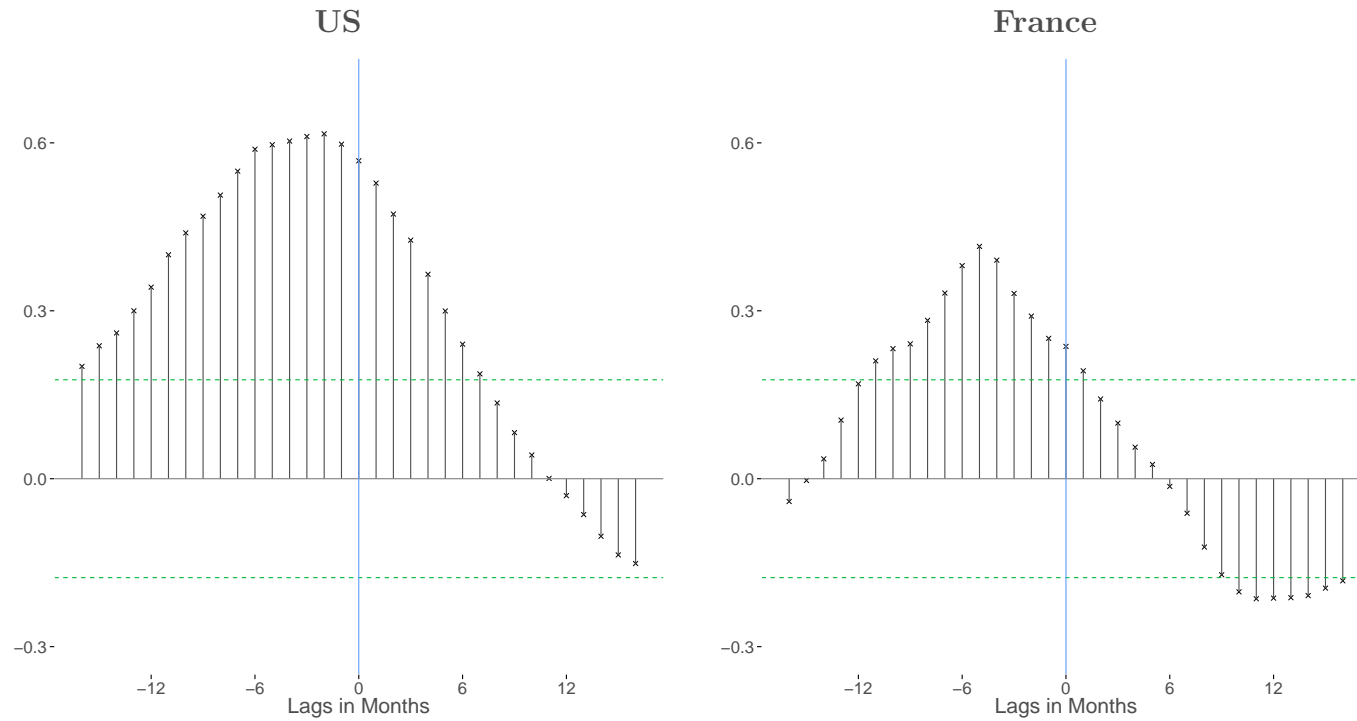
**Figure 2:** The figure presents time series plots illustrating the lagged co-movement of search volume for the keywords “jobs” and “unemployment benefits”. The upper graph presents data for searches conducted in the United States, the lower graph is UK data.

### Correlation of Principal Components and Keyword Categories



**Figure 3:** The figure illustrates the correlation of the extracted principal component with the search volume in the four categories “Benefits”, “Process”, “Unemployed” and “Unemployment” for the different countries.

## Cross-Correlation Functions of Unemployment Changes and Principal Component

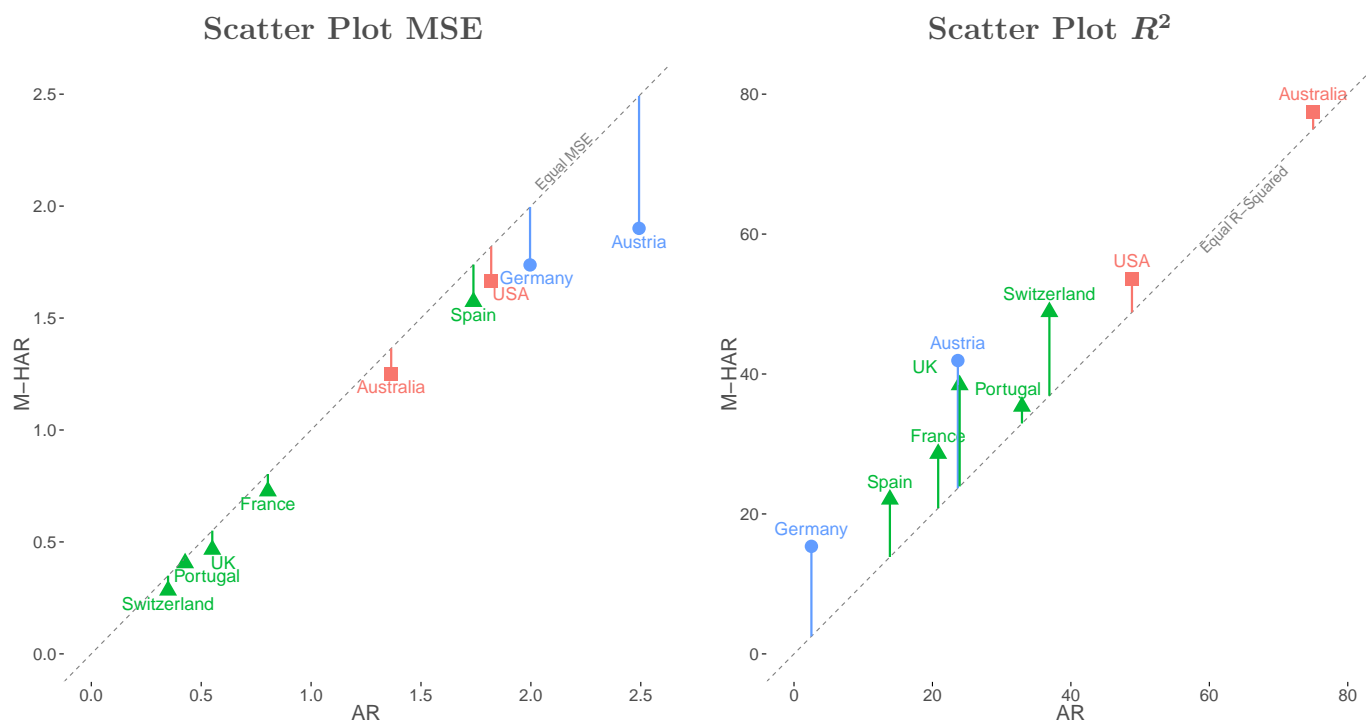


**Figure 4:** The figure presents the cross-correlation function of the current change in unemployment rates with different leads and lags of the principal component extracted from the search volume changes. The left graph depicts the situation for US data while the right graph presents data from France. The blue vertical line denotes lag 0. The green dotted horizontal lines represent 95% confidence intervals.

### Cross-Correlation Peak Size and Timing

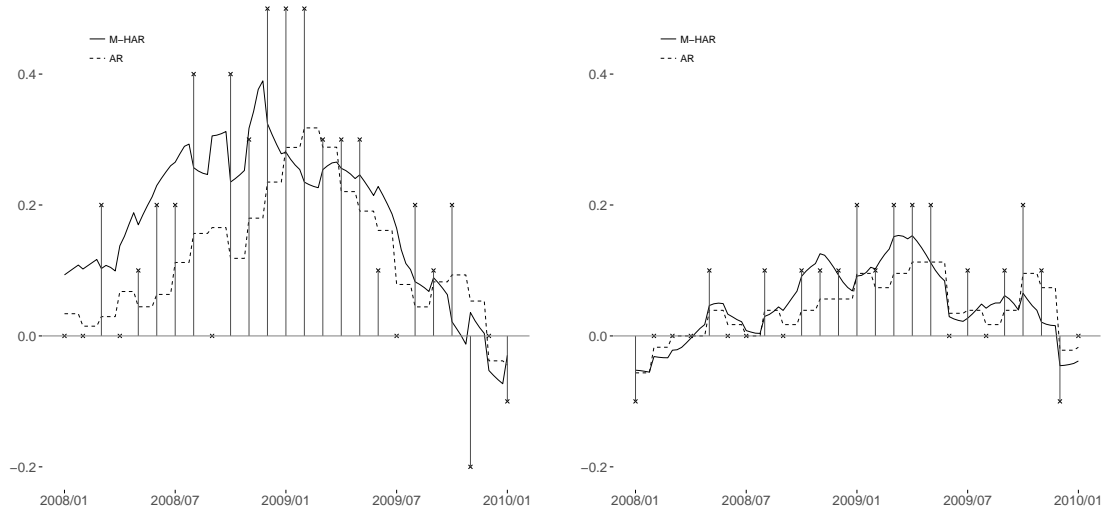


**Figure 5:** The figure shows how many months  $h$  ahead of a change in unemployment, the correlation with past search query data is highest, i.e. a scatter plot of  $h$  against  $\max_h \{\text{cor}(\Delta u_t, \Delta g_{t-h})\}$ .

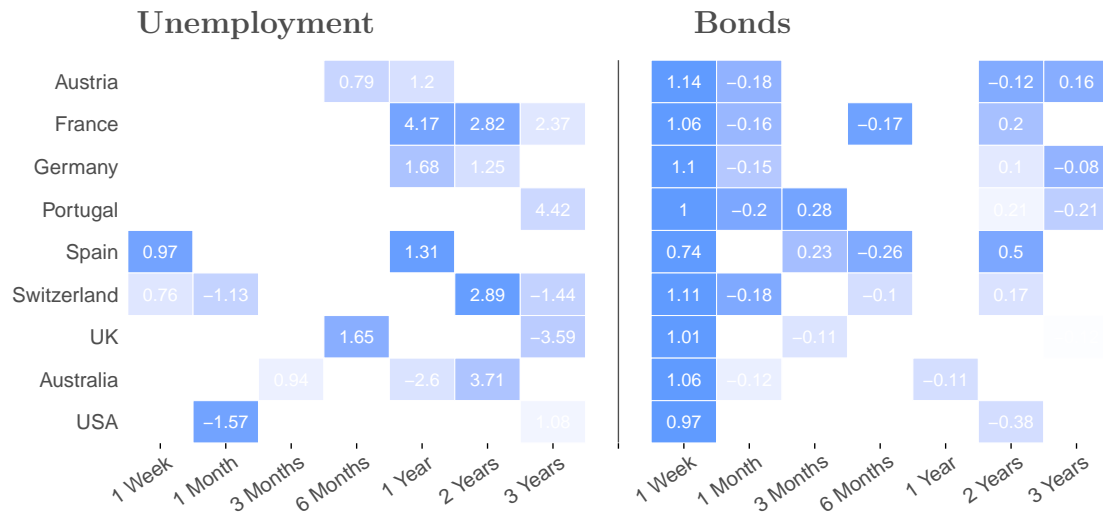


**Figure 6:** The figure compares the model fit of the pure autoregressive forecast model and the M-HAR model based on mean squared error (left graph) and adjusted  $R^2$  (right graph). The dashed 45-degree line indicates the reference line for equal performance of the two models.

## AR and M-HAR Nowcasts

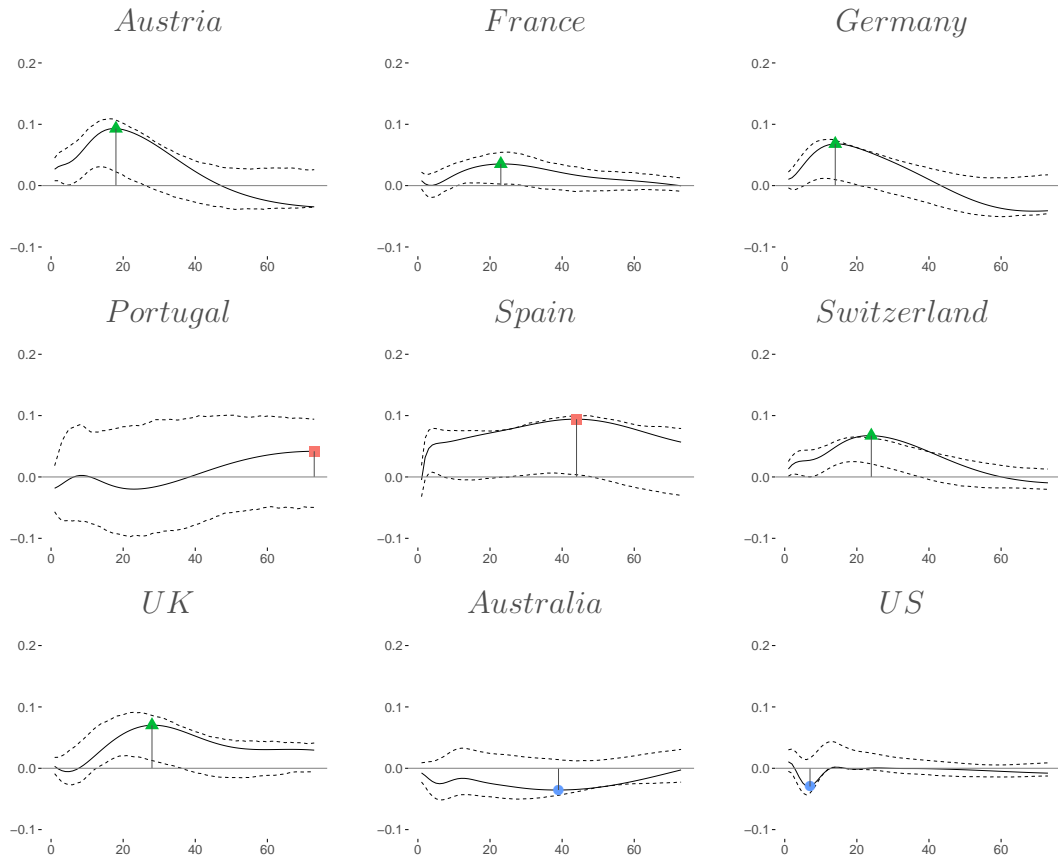


**Figure 7:** The figure shows a comparison of the unemployment change nowcasts based on the pure autoregressive model (dotted line) to the one based on the M-HAR model (solid line). The horizontal lines represent the actual changes. The left graph is for the US, the right graph for France.



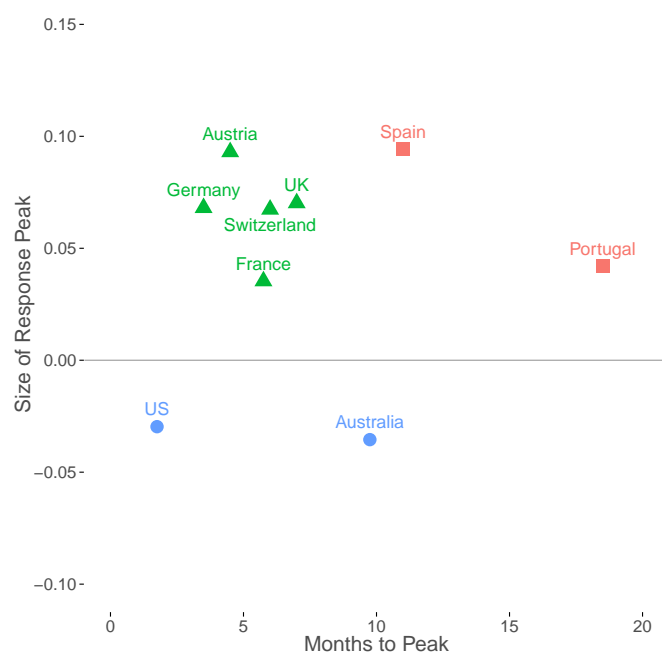
**Figure 8:** The figure illustrates the impact of the lag structure of unemployment rate changes and bond yields on government bond yields. Significance of the heterogeneous autoregressive parameters is illustrated through fading blue where dark blue is a  $p$ -value  $< 0.001$  and white is an insignificant parameter with a  $p$ -value  $> 0.2$ .

## Impulse-Response Functions of Bond Yields to a Shock in Unemployment



**Figure 9:** The figure presents the impulse-response functions of bond yields to a one standard deviation shock in unemployment (solid line). 95% confidence bounds (dotted lines) are based on a resampling bootstrap.

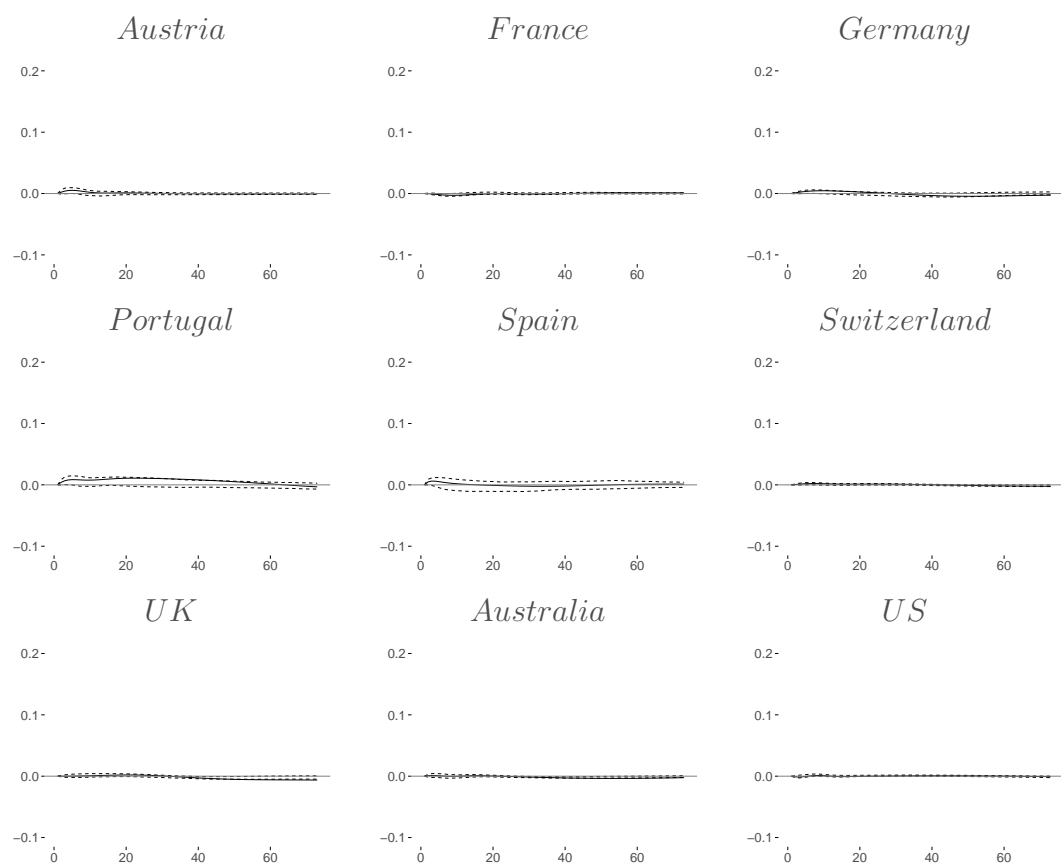
### Scatter Plot of Impulse-Response Peak and Timing



**Figure 10:** The figure illustrates the timing and magnitude of the impulse-response of government bond yields to an idiosyncratic shock in unemployment changes.



## Impulse-Response Functions of Unemployment to a Shock in Bond Yields



**Figure 11:** The figure presents the impulse-response functions of unemployment to a one standard deviation shock in bond yields (solid line). 95% confidence bounds (dotted lines) are based on a resampling bootstrap.