HEDG
HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY of York

WP 13/02

Exploring the relationship between macroeconomic conditions and problem drinking as captured by Google searches in the US

Paul Frijters; David W. Johnston; Grace Lordan & Michael Shields

February 2013

york.ac.uk/res/herc/hedgwp

Exploring the relationship between macroeconomic conditions and problem drinking as captured by Google searches in the US

Paul Frijters[a], David W. Johnston[b], Grace Lordan*[c] and Michael A Shields[b]

*** Grace Lordan, corresponding author. Email: G.lordan@lse.ac.uk.**

a Department of Economics, University of Queensland

b Centre of Health Economics, Monash University

c Department of Social Policy, LSE

This paper has been accepted and is forthcoming in Social Science and Medicine and a link to the article once published will be found here http://gracelordan.com/peer-reviewed-publications-2

Abstract

There is considerable policy interest in the impact of macroeconomic conditions on health-related behaviours and outcomes. This paper sheds new light on this issue by exploring the relationship between macroeconomic conditions and an indicator of problem drinking derived from state-level data on alcoholism-related Google searches conducted in the US over the period 2004-2011. We find the current recessionary period coincided with an almost 20% increase in alcoholism-related searches. Controlling for state and time-effects, a 5% rise in unemployment is followed in the next 12 months by an approximate 15% increase in searches. The use of Internet searches to inform on health-related behaviours and outcomes is in its infancy; but we suggest that the data provides important real-time information for policy-makers and can help overcome the under-reporting in surveys of sensitive information.

Keywords: United States, Recession; alcoholism; global financial crisis; Google Insights

1. Introduction

There is considerable speculation about the potential health consequences of the current Great Recession. This is exacerbated by the long lag in the collection and release of many relevant data, making 'real-time' analysis limited. One reason for this speculation is that, following the influential study by Ruhm (2000), the possibility that recessions may actually be good for some aspects of health and health-related behaviours has become serious proposition. A study of the Great Depression of the 1930s by Granadosa and Diez Roux (2009) found that population health in the US did not decline, and they provide some evidence of health improvements with reduced mortality observed over the period 1930-33. The only exception was an increase in suicide, but this accounted for less than 2% of all deaths. Turning to the current recessionary period, Cotti and Tefft (2011), using US quarterly state level data for 2003-2009, found that unemployment is associated with reduced fatal road accidents, with the number of fatal accidents falling by over 17% after the beginning of the current recessionary period. Similar results have been found in different contexts (Granados (2008) and Kondo et al (2008)). However, European evidence generally highlights a different pattern. For example, Stuckler et al. (2009) examined the aggregate relationship between unemployment and mortality for 26 European Union countries over the period 1970 to 2007, and find no consistent evidence that increased unemployment was associated with a change in all-cause mortality. They do find that large rises in unemployment were associated with an increase in suicide, and also a short-term reduction in road fatalities. Similar findings are reported by Neumayer (2004) for Germany.

It is argued that a limitation of these studies is that they consider only death rates, thus excluding morbidity. In addition, some of the included outcomes in these studies suggest a biologically implausible reaction to increased unemployment rates (Stuckler et al., 2009). That is, some of the included outcomes, such as deaths due to cancer and heart disease, can take a long time to develop to the stage of death. Therefore, a finding that deaths attributed to these causes increase in a period of high unemployment may be spurious. If anything, the health outcomes that can be impacted by changes in unemployment rates are likely to be those where stress triggers a quick manifestation. Examples, of such outcomes include alcoholism and alcohol abuse.

In this paper we will look at the relationship between unemployment and the relative frequency of Internet search for alcoholism at the level of US states during the last 5 years, under the running assumption that the volume of Internet searches reflect the severity of alcohol abuse. The issue of problem drinking in the US is an important issue. Specifically, there are more than 30 ICD-

10 three digit or four digit codes that include alcohol in their name and definition (WHO, 2004) and perhaps most notably the risk of cardiovascular disease, cancer, liver cirrhosis and injury (Rehm et al., 2009) is greatly increased. Estimates of the economic cost associated with alcohol consumption suggest that it absorbs more than 1% of GDP in high and middle-income countries (Casswell and Thamarangsi, 2009). In the US, alcohol use and misuse is the third leading cause of preventable deaths (Mokdad et al., 2000) and the national costs attributable to alcohol are almost $150 billion (The National Centre for Addiction and Substance Abuse, 2009).

Yet, as Dávalos et al. (2011) recently note, "research on the effects of macroeconomic conditions on alcohol consumption is scant" and the results in the literature so far have been inconclusive. In the US context, the early evidence pointed to falling levels of alcohol consumption in bad times, with the largest reductions associated with heavy drinking (Ruhm, 1995; Freeman, 1999; Ruhm and Black, 2002). This is consistent with the premise that during a recession people are less likely to engage in affluent activities (Ruhm, 2000), in this case excess alcohol consumption. It is also consistent with the hypothesis that in the absence of employment related stress individuals may 'self-medicate' less using alcohol (Ruhm, 2005). If this is true for the current Great Recession also, then we would expect more people would search for advice on alcohol abuse on the Internet right after the recession starts, be it for themselves or a loved one, as they seek to curtail their drinking.

This hypothesis is however contradicted by Dávalos et al. (2011), who use 2001-2005 individual-level survey data for the US and find that recessions are associated with increased problem drinking. Notably, Davlalos et al. (2011) differs from the other fore-mentioned studies by using micro level data. This is important, as even if alcoholic drinks are a normal good and we expect aggregate consumption to decline in recessions, this does not tell us about the important right-tail of the distribution (i.e. problem drinking or alcoholism). This fits with the suggestion by Stuckler et al. (2005) that alcohol consumption may increase during spells of higher unemployment. If this is true, then we should not see an immediate increase in the number of people who seek help to curb their problem drinking after the recession hits, but rather a delayed increase in search behaviour as the increased extent of problem drinking works it way through to more search for help. With respect to this, alcohol abuse screening tests frequently include questions on an individual's behaviour that suggest a lag of at least one year with respect to getting help. For example, the U.S. Alcoholics Anonymous 'Is A.A. for you?' test asks "Have you had problems connected with alcohol in the last year" and "Have you had to have an eye-opener upon awakening during the past year". Therefore people may turn to the Internet for help at a substantial lag after they develop their problem.

What evidence is there for the running assumption that changes in alcohol-related searches reflect real changes in problem drinking? Choi and Varian (2009) provide a general discussion about predicting trends for Google search data. They find considerable evidence to suggest that the Internet is a key information source for information on health and health-relates issues (Baker et al., 2003; Polgreen et al., 2008; Lam-Po-Tang and McKay, 2010, Weaver et al., 2010; and Teft, 2011), with Google being the most used search engine. Askitas and Zimmermann (2011) provide a comprehensive discussion of the use of Internet search data to inform on changes in health and wellbeing in crisis periods, particularly focusing on search terms relating to 'symptoms' and 'side-effects', capturing aspects of self-diagnosis and treatment. They find that in the US, Germany and the G8, both types of Internet searches vary with recent macroeconomic conditions.

An important advantage of using search engine data, apart from the fact that it can provide information in real-time, is that it likely lessens the well-known problem of under-reporting of stigmatised outcomes (alcohol abuse, mental illness) in survey data (Feunekes et al., 1999; Stockwell et al., 2004), and of alcohol-related mortality in death records (Pollock et al., 1987; Wise et al., 2008; and Haemstrom, 2002). However, the data currently available from search engines does have the weakness, in contrast to survey data, that specific groups of individuals who might be most at risk cannot be identified nor can the timing of the problem. Additionally, it cannot be determined whether it is the individuals themselves who search the Internet, or if it is their concerned family or friends. A concerned relative is perhaps just as likely as the individual to search for information, given that family members are often involved with the decision to address the problem (Merikle et al., 2001; Simpson and Sells, 2000). Yet, even if a large number of searches are not undertaken by the problem abuser but rather by concerned family and friends, the frequency of Internet searches on alcoholism-related terms should still be strongly correlated with the prevalence of problem drinking. As such, an increase in people seeking help for alcohol-related problems during periods of higher unemployment warrants a policy response given that heavy drinking adversely affects many health outcomes.

Consistent with the study of Dávalos et al. (2011), our results support the view that increased unemployment is associated with increased Google searches associated with problem drinking six to twelve months later. This result is robust to a large variety of measures of macroeconomic activity; an allowance for state-specific quadratic time trends; controlling for the macroeconomic trends in general Internet search use and in health-related Internet search use; and separate analyses for different groups of US states and time periods. The peak of these searches lag unemployment by around nine months.

2. Definitions, Data and Methods

Google Search Terms

The level of problem drinking in the United Sates by state and month is approximated using data on Internet searches performed through Google. The data is accessed through Google Insights for Search (www.google.com/insights/search), which samples a portion of Google web searches to form an estimate of the quantity of searches containing a particular phrase, relative to the total number of searches. Therefore, the data provides a measure of the likelihood a random user will search for a particular phrase during a certain time period. The data is standardised by Google Insights such that each data point is divided by the maximum value for the time series:

$$y_t = 100 \cdot \frac{[number\ of\ searches\ for\ "phrase"/total\ \ searches]_t}{[number\ of\ searches\ for\ "phrase"/total\ \ searches]_{max}}$$

where y_t is the Google Insight data point for period *t*. In addition, data is provided only if the volume of search traffic passes an unspecified minimum threshold, and depending on the volume of search traffic, it is provided at a weekly or monthly frequency.

Data was collected by entering the phrase “alcohol+alcoholic+alcoholics+alcoholism+aa” into Google Insights for Search, and by applying the restriction that only those Google searches in the “Health” category should be considered (the data was downloaded on May 29, 2011 and can be accessed by contacting the corresponding author). We use this broad search to ensure that the minimum search volume was passed for the smaller US states and we restrict searches to those relating only to Health to eliminate searches unrelated to alcoholism. This approach provides us with a measure of the relative likelihood that a random user in a particular US state in a particular month between January 2004 (first point in which this data is actually available) and April 2011 will complete a health-related Google search for “alcohol” or “alcoholic” or “alcoholics” or “alcoholism” or “AA”.

To better understand what is measured by our search terms, Table 1 shows the top 10 related searches for our search phrase, and encouragingly suggests that individuals searching for these terms are predominantly searching for information on alcoholism treatment options and for information related to the symptoms of alcoholism. For example, the most common related search terms (used before and/or after one of our five search terms) in California, New York and Texas are "alcoholics anonymous", "AA meetings", "alcohol effects", and "alcohol abuse".

Throughout our empirical analyses we use a log transformation of the downloaded series, such that our measure of alcoholism for a particular state in period *t* equals:

$$alc_t = log\left(\frac{y_t}{y_{12.10}}\right) = 100 \cdot log\left(\frac{[number\ of\ searches\ for\ "phrase"]_t}{[number\ of\ searches\ for\ "phrase"]_{12/10}}\right)$$

where *y* is defined as above, and the subscript "12/10" refers to the December 2010 time period. This transformation, coupled with the use of state-fixed effects, implies that variation in our problem drinking measure represents the within-state percentage difference between the popularity of the search phrase in month *t* and in December 2010. In other words, a value of 10 in month *t* implies that problem-drinking searches in month *t* are 10% higher than in December 2010. Across all months and states, the mean and standard deviation of this standardised index equals 4.17 and 19.66, respectively.

To control for the possibility that general Internet search activity increases or decreases during recessions (e.g. due to changes in access or leisure time), we include additional search variables as covariates in all subsequent analyses. The three search variables we include are based on searches for "Yahoo", "Hotmail" and "Cancer". The first two variables are included to control for general increases in search activity, and are chosen because of their high popularity across the whole sample range (i.e. from 2004-2011). The cancer search variable is included to control for increases in health-related searches that are not caused by changes in true levels of health. The reasoning is that cancer rates should not plausibly change with short-term changes in unemployment rates, and that changes in searches for cancer rates are thus a useful proxy for changes in the overall degree to which individuals search online for health-related information. In Ruhm's (2000) influential study, cancer fatalities are found to be insignificantly related to the unemployment rate. His explanation of this finding is that "it would be surprising to observe significant variation in cancer fatalities, since deaths from this source are unlikely to respond much to short-tern changes in medical care or lifestyles" (p.632).

Main Measures of Economic Activity

Two main measures of unemployment are considered in this study. First, state level monthly unemployment rates (UR) were retrieved from the Bureau of Labour Statistics (BLS); and second, insured unemployment rates were constructed from monthly state level unemployment insurance (UI) claims data, which were downloaded from the Employment and Training Administration. The insured-UR represents the quantity of continued UI claims divided by the number of workers who qualify for UI. This second measure is also used in Tefft (2011). While the first measure is closer to the economic definition of unemployment, the second measure is less afflicted by variations due to small samples in the labour surveys and thus less noisy.

Descriptive Figures

Figure 1 presents kernel regression estimates of the unemployment rate (UR) across time and our problem-drinking index across time, where an Epanechnikov kernel function and a rule-of-thumb bandwidth are used in the calculations. It is clear that the index is positively correlated with the UR, though it appears to lag unemployment by 6-12 months. Both series are falling between 2003 and 2007, before sharply increasing in 2008. In fact, the near doubling of the unemployment rate during 2008 coincided with an increase in searches of near 20% (relative to the search volume in December 2010). Results from Levin-Lin-Chu panel unit-root tests (Levin et al., 2002) show that the Google search alcoholism index and the two unemployment measures (UR and insured-UR) are both stationary, with tests of the null hypothesis of non-stationarity having p-values <0.0001.

Figure 2 presents the estimated cross-sectional relationship between the UR and the problem-drinking index using between-state variation (an Epanechnikov kernel function and a rule-of-thumb bandwidth are again used). It shows that the relationship is relatively flat for UR values less than 9%, and strongly positive for higher UR values. Though it is difficult to draw any firm conclusions from cross-sectional correlations, this suggests it may be important to test for nonlinearity in the underlying relationship.

Statistical Methods

To model the effects of macroeconomic conditions on problem drinking, we use a linear regression model with state-specific intercepts and state-specific quadratic time trends:

$$
\begin{aligned}
(1)\quad alc_{jt} &= \delta_1 ur_{jt} + \delta_2 ur_{jt-3} + \delta_3 ur_{jt-6} + \delta_4 ur_{jt-9} + \delta_5 ur_{jt-12} \\
&\quad + yr_t + mth_t + s_j + s_j t + s_j t^2 + \boldsymbol{X}_{jt}\boldsymbol{\beta} + \varepsilon_{jt}
\end{aligned}
$$

where alc_{jt} is our measure of problem drinking in state j in month t, ur_{jt} to ur_{jt-12} are contemporaneous and lagged measures of unemployment, yr_t and mth_t are year and month specific effects that control for country-wide changes across time and for seasonality, s_j is a state-specific intercept, $s_j t$ and $s_j t^2$ represent a state-specific linear and quadratic time trend, $\boldsymbol{X}_{jt}$ represents additional control variables, and ε_{jt} is the regression disturbance term. Standard errors are clustered by state to allow for correlation between disturbances across time within states.

Lagged unemployment terms at 3 months (ur_{jt-3}), 6 months (ur_{jt-6}), 9 months (ur_{jt-9}) and 12 months (ur_{jt-12}) are included in equation (1) to allow for the possibility that the unemployment effects on problem drinking searches are not instantaneous. The state-specific intercepts and time trends in equation (1) capture differences in the levels of problem drinking across states and in the growth rates of such drinking across states – this approach amounts to the inclusion of 50 dummy variables, 50 linear time variables and 50 quadratic time variables. This approach is more conservative than random-effects or OLS regression models, but the additional terms limit the possibility that our estimates are driven by omitted-variable bias. Given the inclusion of these terms, the effect of the unemployment (or unemployment insurance claims) rate on problem drinking is captured by within-state variations in unemployment rates in relation to within-state variation in problem drinking around its trend. It does mean that we become more conservative in terms of what 'counts' as real variation due to unemployment: any effect that looks highly quadratic will be partially absorbed by the state-specific time-trends rather than the unemployment changes.

3. Results

Main Results

Estimates of the relationship between our two main unemployment measures and our problem-drinking index are presented in Table 2. The estimates in row 1 suggest that contemporaneous unemployment is unrelated to problem drinking: the estimated coefficient on ur_{jt} is not statistically different from zero in column (1) or (2). However, the remaining estimates indicate that it takes several months following an increase in unemployment for a rise in problem drinking to be evident. For both unemployment measures, unemployment lagged nine months has the largest estimated effect: a 1 percentage-point increase in the UR (or insured-UR) is associated with an increase in the problem drinking index by 1.76% (or 2.4% for insured-UR). The cumulative effect estimates presented in the bottom row give the cumulative effect of the five unemployment rate terms, and can be interpreted as the total effect of a 1 percentage-point increase in the UR 12 months ago. Thus, the cumulative effect estimates suggest that a 5 percentage-point increase in the unemployment rate (as seen in the Great Recession) is associated with around a 14% increase in the problem drinking index after 12 months. The equivalent figure for the insured-UR is around 28%. We also tested for potential nonlinearity in the relationship between UR and problem drinking by estimating a model with four unemployment rate dummy variables ($4 < ur_{jt-9} \leq 6$, $6 < ur_{jt-9} \leq 8$, $8 < ur_{jt-9} \leq 10$, $ur_{jt-9} > 10$). The results from this specification suggest that the marginal effect of UR on problem drinking is larger at higher UR values. For example, relative to $ur_{jt-9} \leq 4$ the estimated effect of $4 < ur_{jt-9} \leq 6$ equals 2.9%, whereas the estimated effect of $ur_{jt-9} > 10$ equals 10.1%.

A direct comparison with the results presented in Tefft (2011) is complicated by the use of a different model specification (we include lags in our specification) and a different standardisation of the search data. Nevertheless, he finds that a one percent increase in the contemporaneous unemployment rate increases his depression index (constructed using a standard normal transformation) by roughly 5% of a standard deviation - see Column 1, Table 2 in Tefft (2011). Similarly, when we estimate a model that only controls for contemporaneous unemployment, our estimated unemployment rate effect equals 0.922, indicating that a one percent increase in the unemployment rate increases our problem drinking index by roughly 5% of a standard deviation.

Comparing the results across Table 2 we see that the measure of unemployment used is not particularly important for the main conclusion that unemployment is positively associated with problem drinking, and that that the relationship is largest at a lag of nine months. Note, if we include UR and insured-UR variables in the same regression model we find that both measures have statistically significant effects. In this joint model, UR lagged 6 months is the largest and most significant UR term (the estimate equals 1.358 with a *t*-statistic equal to 2.4), and insured-UR lagged 9 months is the largest and most significant insured-UR term (the estimate equals 2.672 with a *t*-statistic equal to 3.5).

The bottom row of Table 2 shows that the insured-UR has a much larger overall effect on problem drinking than the UR (5.6 versus 2.7). One explanation for this difference is that the insured-UR measure is less afflicted by variations due to small samples in the labour surveys and thus the estimated effects are less attenuated. Another explanation is that the insured-UR is more representative of individuals that are negatively affected by unemployment. For example, only individuals who became unemployed involuntarily are eligible for unemployment insurance. In the robustness section below we investigate this second explanation and find that it is credible.

Robustness Results

One issue with respect to the interpretation of our results concerns the change in the composition of unemployment over the sample period – for example, Mayer (2010) notes that one of the main characteristics of the recent recession in the US was the unprecedented rise in long-term unemployment. In order to provide some information on the robustness of our results, we use various measures of local-area labour underutilization rates provided by the BLS. Importantly, these data are only available quarterly, rather than monthly as for the UR and insured-UR measures. (see measures U-1 to U-6 described at http://www.bls.gov/lau/stalt.htm). Note that the BLS suggests using these quarterly state-level measures of labour underutilization with caution, stating that: "The sample of the CPS (*Current Population Survey*) is designed to be able to reliably estimate total annual unemployment on a state-by-state basis. State subsamples of the CPS generally cannot support reliable estimation of even total employed and unemployed on a quarterly basis, much less for the sort of detailed economic characteristic data that underlie the alternative measure calculations" (personal correspondence from the BLS).

The results for the 6 alternative definitions are shown in Table 3, with (1) 'Persons unemployed 15 weeks or longer', being a measure of longer-term unemployment, and (6) 'Total unemployed + marginally attached workers + total employed part time for economic reasons', being the broadest measure of labour underutilization that includes both discouraged workers, and marginally-attached workers. Our working hypothesis is that increased long-term unemployment will be more strongly associated with increased alcohol-related Google searches, with an expected lesser association with changes in the measure that includes marginally attached and discouraged workers. Looking down the columns of Table 3 gives a clear and consistent picture in support of this hypothesis. We see a decreasing association between the economic activity and the problem-drinking index as we incrementally broaden the measure of labor underutilization, and the results are consistent with our main results presented in Table 2, with the largest association consistently found for the lagged 6-12 month measures of economic inactivity.

In Table 4 we explore the estimated unemployment effects presented in Table 2 by re-estimating equation (1) for a post-recession time period and for three different sub-samples of states. These sub-sample analyses allow for the possibility that the effect of unemployment on problem drinking differs across states due to certain policies or characteristics of the population. Importantly, we implicitly control for the ‘main’ effect of all time-invariant state-level characteristics by including state fixed-effects in all regression models. We also control for time-varying state characteristics by including state-specific linear and quadratic time trends.

In column (1) the sample is limited to the 40 months post December 2007. As Tefft (2011) notes, the recession from 2008 onwards represented a substantial change in the United States’ economic and political landscapes, which may mean the unemployment-problem drinking relationship during this period differed from the relationship in the preceding period. The estimates provide evidence in support of this hypothesis: the estimated cumulative effect is almost double the size of the equivalent estimate in Table 2, indicating that an increase in UR post 2008 had a larger total effect than an increase in UR pre 2008. This result may also be driven by nonlinearity in the relationship between UR and problem drinking, such that increases in UR from high UR levels (as seen post 2008) have a larger effect on problem drinking than do increases in UR from low UR levels (as seen pre 2008). This pattern of nonlinearity is reflected in the cross-sectional relationship shown in Figure 2.

Following Ruhm (2007), in column (2) of Table 4 we present estimates from a regression estimated with the 20 most populous states, given that the unemployment rate and problem drinking index variables are likely to suffer from less measurement error in larger states. The 20 states used in the Table 4, column (2) regression models are: California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, Michigan, Georgia, North Carolina, New Jersey, Virginia, Washington, Massachusetts, Indiana, Arizona, Tennessee, Missouri, Maryland and Wisconsin. We find that omitting the less populous states increases the contemporaneous effect, suggesting that the effects of unemployment are more immediate in larger states. We also find that the cumulative effect is larger, which supports the hypothesis that measurement error is attenuating the estimates presented in Table 2.

In column (3) the sample is restricted to the 21 states with state-legislated restrictions on alcohol sales. The 21 states used in the Table 4, column (3) regression models are: Alabama, Idaho, Iowa, Maine, Maryland, Michigan, Minnesota, Mississippi, Montana, New Hampshire, North Carolina, Ohio, Oregon, Pennsylvania, South Carolina, Utah, Vermont, Virginia, Washington, West Virginia and Wyoming. The estimates are similar to those in Table 2, suggesting that alcohol access does not mediate the unemployment-problem drinking relationship.

In column (4) of Table 4, we test whether average religiosity within a state influences the estimated unemployment effect. We hypothesise that the unemployment-problem drinking relationship is weaker for individuals with high religiosity, as social norms in religious groups may tend to encourage moderation when it comes to intoxication. Thus, people who frequently attend church or religious services may be more likely to avoid negative health behaviours, such as alcohol. In addition, church attendance can provide informal help rather than the help that can be found via the Internet. Such local informal insurance was found to be important in the case of income: Deheija et al. (2007) find that religious participation reduces the impact of income changes on consumption by roughly 40 percent, strongly suggesting religious communities serve as informal insurance networks.

The estimated effects using the 27 states with the highest rates of religiosity (column 4) weakly support this hypothesis. At a 9-month lag, a 1 percentage-point increase in UR is associated with an increase in problem drinking of around 1.55%, which is slightly smaller than the estimate of 1.76% for the full sample. Moreover, the estimated cumulative effect is substantially lower (1.547 versus 2.745). The 27 states used in the Table 3, column (4) regression models are: Alabama, Louisiana, South Carolina, Mississippi, Utah, Arkansas, Nebraska, North Carolina, Tennessee, Georgia, Oklahoma, Texas, Kentucky, Kansas, West Virginia, Indiana, Missouri, Iowa, South Dakota, Virginia, Minnesota, Delaware, Wisconsin, Idaho, Pennsylvania, North Dakota and Ohio.

As a final robustness check, we examine the importance of our particular problem-drinking index by experimenting with an alternative index formed with the search phrase "addict+addiction+abuse" in the health domain. The top related searches to this search phrase are "drug abuse", "substance abuse", "alcohol abuse" and "drug addiction", and so is sensitive to changes in rates of drug addiction in addition to changes in problem drinking. The estimated unemployment effects when using this alternative index were also significantly positive, with the largest effect occurring at a nine-month lag (results available upon request).

Before we conclude, we note a number of caveats of our study. First, while we have undertaken what we believe to be the first exploratory study focusing on the relationship between changing macroeconomic conditions and a Google-search derived measure of problem drinking, the main limitation (or data trade-off) of the analysis is the lack of individual-level data. This means that we are not able to identify the characteristics of individuals or groups who use the Internet to gain information. In particular, we are not able to say whether it is the individual with alcohol problems who undertakes the search, or a concerned relative, friend or even colleague. Second, we are not able to verify the external validity of our problem-drinking index by comparing it with other state-level information. Comparing the index with, for example, alcohol sales data or alcohol-related traffic fatalities, is unlikely to be useful in such a validation exercise, as these measures do not clearly capture the right-hand tail (problem drinking) of alcohol consumption. This would even be the case for alcohol-related hospital admissions, which to a large extent is likely to capture the short-term consequences of binge drinking, rather than a chronic alcohol problem. Moreover, Alcoholics Anonymous is not a member organisation and does not collect any accurate attendance data. Clearly, being able to identifying a measure with which to validate our problem-drinking index (and other health indices) is an important question for future research.

Third, there is the issue of the timing of searches: we have found a significant association between changes in eight measures of unemployment and labour underutilisation, and changes in problem-drinking searches. Our results consistently suggest that it takes around 9 months for the effect of increased unemployment to be fully felt in terms of problem drinking searches. However, while the analysis we have undertaken is a useful starting point in understanding the dynamics of this relationship, we would suggest that more research (including qualitative research) is needed to more fully understanding why, and at what stage in the development of a drinking problem do individuals turn to using the Internet to access for health and health-care related information. It is clearly important to know whether the peak of the problem coincides with the peak of the search, or whether the problem peaks before or after the search.

Finally, in order to make a causal statement, which is that unemployment causes an increase in problem drinking search behaviour, we have adopted a conservative modelling approach. In particular, the models control for state fixed effects, and state-specific linear and quadratic time trends. However, while we believe it is safe to assume that problem drinking did not cause the Great Recession (no reverse causality), it is impossible to fully rule-out the existence of time-varying unobservables co-moving with both unemployment and Internet search behaviour. Ideally, in order to address this set of caveats we would need a large longitudinal dataset following individuals (and families) that identifies their (changing) socioeconomic circumstances, provides their detailed Internet search behaviour and records, and also includes information on their alcohol consumption and symptoms of alcohol abuse (measured in such a way as to avoid under-reporting).

4. Conclusion

In this paper we have contributed to the understanding of how macroeconomic conditions are related to health, by investigating if the current recessionary period in the US is associated with increased problem drinking (as measured by alcoholism-related Internet searches) over the period 2004 to 2011. This is an important exercise because, in contrast to general alcohol consumption, alcohol-related traffic accidents and longer-term alcohol-related mortality, there is only scant evidence relating macroeconomic conditions to problem drinking or abuse (Davalos et al., 2011). This is the case even though alcohol use and misuse is the third leading cause of preventable deaths in the US (Mokdad et al., 2005) and imposes substantial costs on society (Casswell and Tharmarangsi, 2009). It is also an important exercise because the health effects of the current recessionary period are largely unknown due to the unavailability of 'traditional' data sources. Our work can be seen as complementary to Tefft (2011) who found evidence that mental health problems, as measured by Google search activity for the terms "Depression" and "Anxiety", increased as the state of the US economy worsened over the period 2004-2010.

Focusing on within-state variation, we have found that increased unemployment is associated with a substantive increase in problem drinking searches. The roughly 5% increase in unemployment rates during this period is predicted to increase such internet searches by around 14% after 12 months, with most of this effect occurring after a nine month lag. Moreover, this result is robust to a wide-variety of measures of macroeconomic activity; an allowance for state-specific linear and quadratic time trends; controlling for the macroeconomic trends in general Internet search use and in health-related Internet search use; and separate analyses for different groups of US states and time periods. These results, seen in conjunction with recent research that uses individual-level survey data (Davalos et al., 2011), provide additional evidence pointing to the potential adverse effects of the current recessionary period on problem drinking.

While our study and a number of others, suggests that the Internet is used to find help to deal with alcohol abuse, to date there is incomplete evidence as to whether the Internet can actually help in improving health outcomes. Some suggest it has great promise (Tang, 2006 and Kiley, 2002), but others suggest poorer outcomes (Hainer et al., 2000; Gardner, 2006) or at least urge caution (Ryan and Wilson, 2008). Given the potential though, there is a role to play for policy to ensure that the websites that come up first when a person searches for a particular illness are trusted and accredited.

Finally, we have added to the broader recent literature that highlights the value that Internet search data can have for informing health-related issues. A detailed discussion of the use of such data can be found in Askitas and Zimmermann (2011). This potential seems particularly important for sensitive or socially stigmatised behaviour or health issues such as alcohol abuse, drug use and mental health, where survey data can suffer from substantial under-reporting by individuals, who in contrast, might seek the anonymity of the Internet for information and advice. Internet search data such as that from Google are also available in 'real-time' allowing researchers to inform on issues or trends with only a short time lag.

References

Askitas, N. and Zimmermann, KF. (2011). Health and well-being in the crisis. IZA Discussion Paper No. 5601, Bonn (March).

Breitung J, and Pesaran MH. (2008). Unit roots and cointegration in panels. In The econometrics of panel data: fundamentals and recent developments in theory and practice, Matyas L, Sevestre P (eds). Springer: Heidelberg.

Casswell, S. and Thamarangsi, T. (2009). Reducing harm from alcohol: call to action. The Lancet, 373 (9682), 2247-57.

Charles, K. and DeCicca, P. (2008). Local labour market fluctuations and health: Is there a connection and for whom? Journal of Health Economics, 27 (6), 1332-1350.

Choi, H. and Varian, H. (2009). Predicting the present with Google trends. Available from: http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf.

Baker, L., Wagner, TH., Singer, S. and Bundorf, MK. (2003). Use of the internet and e-mail for health care information: Results from a national survey. Journal of the American Medical Association, 289(18), 2400-6.

Cotti, C. and Tefft, N. (2011). Decomposing the relationship between macroeconomic conditions and fatal car crashes during the Great Recession: Alcohol- and non-alcohol-related accidents. B.E. Journal of Economic Analysis & Policy, 11 (1), Article 48.

Dávalos, ME., Fang, H. and French, MT. (2011). Easing the pain of an economic downturn: Macroeconomic conditions and excessive alcohol consumption. Health Economics. Article first published online: 12 SEP 2011, DOI: 10.1002/hec.1788.

Dee, T. (2001). Alcohol abuse and economic conditions: evidence from repeated cross-sections of individual-level data. Health Economics, 10 (3), 257–270.

Dehejia, R., DeLeire, T. and Luttmer, E. (2007). Insuring consumption and happiness through religious organizations. Journal of Public Economics, 91, 259-279.

Freeman DG (1999). A note on 'Economic conditions and alcohol problems'. Journal of Health Economics. 18(5), 661-70

Feunekes, GIJ., van't Veer, P., van Staveren, WA. And Kok, FJ. (1999). Alcohol intake assessment: The sober facts. American Journal of Epidemiology, 150 (1), 105-12.

Gardner M.(2006). Diagnosis using search engines - Probably heralds a much more sophisticated web resource. British Medical Journal. 333(7579),1131.

Gazeaux, J., E. Flaounas, P. Naveau, and A. Hannart (2011), Inferring change points and nonlinear trends in multivariate time series: Application to West African monsoon onset timings estimation, J. Geophys. Res., 116, D05101, doi:10.1029/2010JD014723.

Granados J. (2008) Macroeconomic fluctuations and mortality in postwar Japan. Demography. 45(2), 323-43.

Granadosa, J. and Diez Roux, A. (2009). Life and death during the Great Depression. Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.0904491106.

Haemstrom (2002). Alcohol-related deaths contribute to socioeconomic differentials in mortality in Sweden. European Journal of Public Health; 12 (4), 254-262.

Jiménez-Martín, S., Labeaga, JM., and Vilaplana Prieto, C. (2006). Further evidence about alcohol consumption and the business cycle. Working Paper No. 2006–06. Fundación de Estudios de Economía Aplicada

Johansson, E., Bockerman, P., Prattala, R. and Uutela, A. (2006). Alcohol-related mortality, drinking behavior, and business cycles: Are slumps really dry seasons? European Journal of Health Economics, 7(3), 215-220.

Hainer MI, Tsai N, Komura ST and Chiu CL (2000) Fatal hepatorenal failure associated with hydrazine sulfate. Annals of Internal Medicine. 133(11), 877-80.

Kiley R. Does the internet harm health? (2002). Some evidence exists that the internet does harm health. British Medical Journal. 324(7331), 238.

Koenig, H., George, L., Meador, K., Blazer, D. and Ford, S. (1994). The Relationship between religion and alcoholism in a sample of community-dwelling adults. Hospital and Community Psychiatry, 45, 225-231.

Kondo N, Subramanian SV, Kawachi I, Takeda Y, Yamagata Z. (2008). Economic recession and health inequalities in Japan: Analysis with a national sample, 1986-2001. Journal Epidemiology and Community Health. 62(10), 869-75.

Lam-Po-Tang, J. and McKay, D. (2010). Dr Google, MD: a survey of mental health-related internet use in a private practice sample. Australasian Psychiatry, 18 (2), 130-3.

Levin, A., Lin, C.F., and Chu, C.S.J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. Journal of Econometrics, 108, 1–24.

Marlowe, DB; Merikle EP; Kirby KC; Festinger DS and McLellan AT (2001). Multi-dimensional assessment of perceived treatment entry pressures amoung substance abusers. Psychology of addictive behaviours. 15 (2), 97-108.

Mayer, G. (2010). The trend in long-term unemployment and characteristics of workers unemployed for more than 99 weeks. Congressional Research Service, December 20th. www.crs.gov.

Mokdad, AH., Marks, JS., Stroup, DF. and Gerberding, JL. (2005). Actual causes of death in the United States, 2000. Journal of the American Medical Association, 291(10), 1238-45.

Neumayer E. (2004). Recessions lower (some) mortality rates: Evidence from Germany. Social Science and Medicine. 58(6), 1037-1047.

Pacula, RL. (2011). Substance use and recessions: What can be learned from economic analyses of alcohol? International Journal of Drug Policy, 22 (5), 326-334.

Polgreen, PM., Chen. Y., Pennock, DM., Nelson, FD. and Weinstein, RA. (2008). Using internet searches for influenza surveillance. Clinical Infectious Diseases, 47 (11), 1443-48.

Pollock, DA., Boyle, CA., Destefano, F., Moyer, LA. and Kirk ML. (1987). Underreporting of alcohol-related mortality on death certificates of young United States army veterans. Journal of the American Medical Association, 258 (3), 345-48.

Rehm, J., Mathers, C., Popova, S., Thavorncharoensap, M., Teerawattananon, Y. and Patra, J. (2009). Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. The Lancet, 373 (9682), 2223-33.

Ruhm, CJ. (1995). Economic conditions and alcohol problems. Journal of Health Economics, 14 (5), 583-603.

Ruhm, CJ. (2000) Are recessions good for your health? Quarterly Journal of Economics, 115 (2), 617-50.

Ruhm, CJ. and Black, WE. (2002). Does drinking really decrease in bad times? Journal of Health Economics, 21 (4), 659-78.

Ruhm CJ (2005). Healthy living in hard times. Journal of Health Economics. 24(2),341-63.

Ruhm, C. (2007). A healthy economy can break your heart. Demography, 44 (4), 829-848.

Ryan A and Wilson S. (2008). Internet healthcare: Do self-diagnosis sites do more harm than good? Expert Opinion Drug Safety. 7(3), 227-9.

Simpson DD and Sells SB (Eds). 1990. Opiod addiction and treatment: A 12 year follow up. Malabar, FL: Krieger.

Stockwell, T., Donath, S., Cooper-Stanbury, M., Chikritzhs, T., Catalano, P. and Mateo, C. (2004). Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduates-frequency and recent recall. Addiction, 99(8), 1024-33.

Stuckler, D., Basu, S., Suhrcke, M., Coutts, A. and McKee, M. (2011). Effects of the 2008 recession on health: a first look at European data. The Lancet. 378(9), 124-125.

Stuckler, D., Basu, S., Suhrcke, M., Coutts, A. and McKee, M. (2009). The public health effect of economic crises and alternative policy responses in Europe: An empirical analysis. The Lancet, 374 (9686), 315-23.

Tang H and Hwee Kwoon Ng J. (2006). Googling for a diagnosis - use of Google as a diagnostic aid: Internet based study. British Medical Journal. 333(1143).

Tefft, N. (2011) Insights on unemployment, unemployment insurance, and mental health. Journal of Health Economics, 30 (2), 258-64.

The National Centre for Addiction and Substance Abuse (CASA). Shoveling Up II: The Impact of Substance Abuse on Federal, State and Local Budgets. 2009 May Contract No.: NCJ 227065.

Weaver, JB., Mays, D., Weaver, SS., Hopkins, GL., Eroglu, D. and Bernhardt, JM. (2010). Health information seeking behaviors, health indicators, and health risks. American Journal of Public Health, 100 (8), 1520-5.

Wise M, Bialek S, Finelli L, Bell BP and Sorvillo F (2008). Changing trends in hepatitis C–related mortality in the United States, 1995-2004. Heapatology. 47 (4) 1128-1135.

World Health Organisation (2004). International classification of diseases and related health problems. 10th revision. Geneva; World Health Organisation

Figure 1: Unemployment and Problem Drinking Google Searches across Time

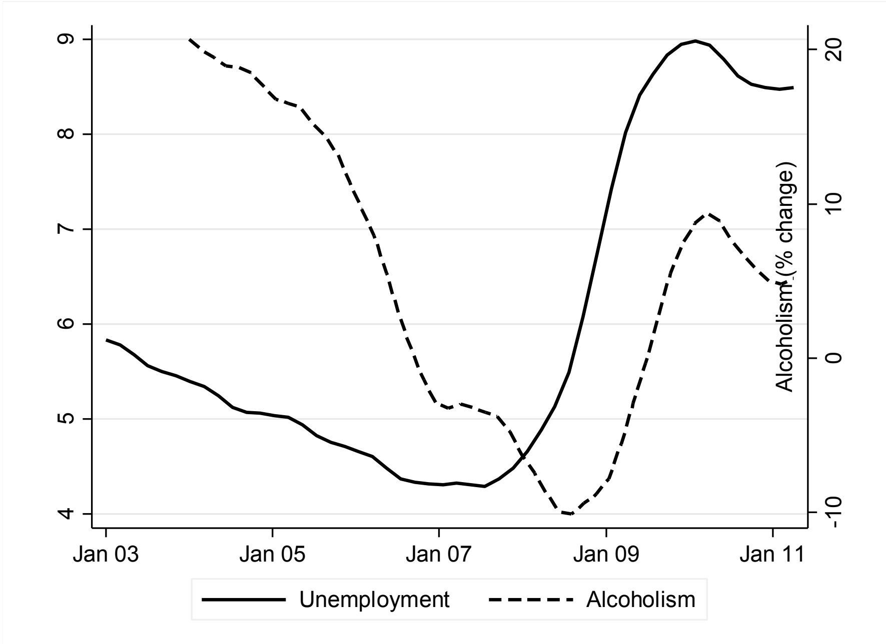

Figure 2: Estimated Non-Parametric Relationship between Unemployment and Problem Drinking Google Searches

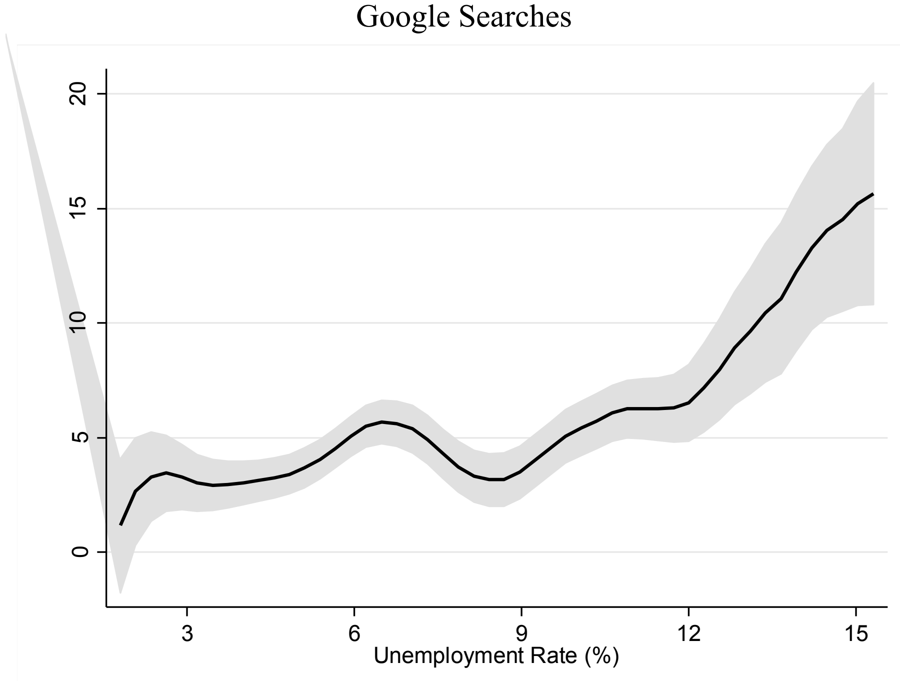

Table 1: Top 10 Related Search Terms for Three States

	California	Texas	New York
1	alcoholics anonymous	alcoholics anonymous	alcoholics anonymous
2	aa meetings	alcohol effects	alcohol effects
3	alcohol effects	aa meetings	aa meetings
4	alcohol abuse	alcohol abuse	alcohol abuse
5	effects of alcohol	effects of alcohol	effects of alcohol
6	alcohol treatment	alcohol treatment	alcohol withdrawal
7	aa meeting	alcohol withdrawal	drugs
8	alcohol withdrawal	Drugs	aa meeting
9	alcohol rehab	alcohol side effects	rehab
10	addiction	alcohol rehab	alcohol antibiotics

Notes: According to Google Insights, these terms are determined by "examining searches that have been conducted by a large group of users preceding the search term you've entered, as well as after".

Table 2: Estimated Effect of Unemployment Rate on Alcoholism-Related Google Searches

	UR		Insured UR	
	Estimate	*t*-statistic	Estimate	*t*-statistic
Contemporaneous rate	0.310	(0.723)	-0.176	(-0.420)
Rate 3 months ago	0.454	(1.185)	**1.129**	**(3.312)**
Rate 6 months ago	**1.235**	**(3.117)**	**1.462**	**(3.342)**
Rate 9 months ago	**1.760**	**(5.484)**	**2.410**	**(5.118)**
Rate 12 months ago	-1.013	(-1.806)	0.821	(1.644)
Cumulative effect	**2.745**	**(4.692)**	**5.646**	**(6.946)**

Notes: Estimates give the percent change in the dependent variable (searches) in relation to a 1 percentage-point increase in the (insured) UR. Standard errors are clustered by state. Statistically significant estimates are in bold font. Models also control for Yahoo searches, Hotmail searches, Cancer searches, month effects, years effects, state-specific intercepts and state-specific quadratic time trends. Number of state-months equals 4262.

Table 3: Estimated Effects of Labor Underutilization Rates on Alcoholism-Related Google Searches

	Rate 0-3 months ago	Rate 3-6 months ago	Rate 6-9 months ago	Rate 9-12 months ago	Cumulative effect
1. Persons unemployed 15 weeks or longer	**1.136** **(3.445)**	**0.768** **(2.501)**	**1.420** **(3.646)**	0.200 (0.383)	**3.524** **(5.491)**
2. Job losers and persons who completed temporary jobs	**0.838** **(2.668)**	0.379 (1.094)	**1.304** **(3.852)**	**1.501** **(3.826)**	**4.023** **(6.977)**
3. Total unemployed	**0.546** **(1.987)**	0.338 (1.120)	**0.892** **(3.118)**	**0.928** **(2.662)**	**2.704** **(6.124)**
4. Total unemployed + discouraged workers	**0.541** **(1.978)**	0.337 (1.139)	**0.813** **(3.028)**	**0.910** **(2.655)**	**2.602** **(5.821)**
5. Total unemployed + discouraged workers + all marginally attached workers	**0.572** **(2.280)**	0.287 (0.972)	**0.869** **(3.174)**	**0.707** **(2.174)**	**2.435** **(5.548)**
6. Total unemployed + marginally attached workers + total employed part time for economic reasons	**0.429** **(2.550)**	0.000 (0.001)	**0.670** **(3.317)**	**0.905** **(4.405)**	**2.004** **(6.491)**

Notes: Estimates in each row are from 5 different regression models. Each gives the estimated percent change in the dependent variable (searches) in relation to a 1 percentage-point increase in the underutilization rate. Standard errors are clustered by state. Statistically significant estimates are in bold font. Models also control for Yahoo searches, Hotmail searches, Cancer searches, month effects, years effects, state-specific intercepts and state-specific quadratic time trends. Number of state-months equals 4262.

Table 4: Estimated Effect of Unemployment Rate on Alcoholism-Related Google Searches by Sub-groups of States

	Post-Recession	20 Most Populous	Alcohol Controlled	Highest Religiosity
Contemporaneous rate	0.528	**1.640**	-0.025	-0.037
	(1.018)	**(5.020)**	(-0.049)	(-0.061)
Rate 3 months ago	**1.205**	0.104	1.327	-0.117
	(2.027)	(0.292)	(1.599)	(-0.341)
Rate 6 months ago	**1.295**	-0.149	0.885	**0.946**
	(2.164)	(-0.420)	(1.062)	**(2.104)**
Rate 9 months ago	**2.162**	**2.716**	**1.786**	**1.554**
	(3.408)	**(7.141)**	**(3.221)**	**(4.642)**
Rate 12 months ago	0.420	-0.512	-1.612	-0.799
	(0.602)	(-0.960)	(-1.585)	(-1.248)
Cumulative effect	**5.611**	**3.798**	**2.361**	**1.547**
	(7.219)	**(6.416)**	**(2.289)**	**(2.039)**
Number of states	51	20	21	27
Sample size (state months)	2090	1760	1736	2247

Notes: The estimates give the percent change in the dependent variable (searches) in relation to a 1 percentage-point increase in the UR. Standard errors are clustered by state. Statistically significant estimates are in bold font. Models also control for Yahoo searches, Hotmail searches, Cancer searches, month effects, years effects, state-specific intercepts and state-specific quadratic time trends.