Research Department **Bank of Israel**

Query Indices and a 2008 Downturn: Israeli Data

Tanya Suhoy*

Discussion Paper No. 2009.06
July 2009

* Research Department, Bank of Israel. http://www.boi.org.il
Tanya Suhoy – Phone: 972-2-655-2620; Email: tanya suhoy@boi.org.il

[1] This study was undertaken in 2008 at the suggestion of Prof. Stanley Fischer. I thank Prof. Hal Varian for help and access to the Google's experimental database, Jonathan Sidi for research assistance, and Ran Sharabani for his contribution to the discussion at the Research Department seminar.

חטיבת המחקר, בנק ישראל ת״ד 780 ירושלים 91007
Research Department, Bank of Israel, POB 780, 91007 Jerusalem, Israel

Abstract

"Predicting the Present with Google Trends" (Choi, Varian, 2008) offers a nowcasting model, suggesting that the popularity of web searches tracked by Google over time may be applied as an indicator of contemporaneous economic activity, before the official data become available and/or are revised. Such evidence was found in different countries with respect to automobile sales, home sales, retail trade and travel.

Using Google's Insights for Search application, which aggregates search terms by large socioeconomic categories, I test the predictive ability of Israeli query indices. At a monthly frequency, six predictive categories were found: Human Resources (Recruitment and Staffing), Home Appliances, Travel, Real Estate, Food and Drink and Beauty and Personal Care. According to the pair-wise Granger causality tests, the strongest predictor is in the Human Resources category (Recruitment and Staffing). The latter, taken at a quarterly frequency, is a leading indicator with regard to the job openings ratio, currently surveyed by the Ministry of Industry, Trade and Labor, and consequently may help in drawing monthly inferences about the unemployment rate. The large fraction of innovation variance explained by other selected categories is related to the employment rate. It suggests an attitudinal interpretation of the queries effect, which may be conveyed through the employment channel.

In-sample Bayesian probabilities of a downturn computed between 2004:2 and 2009:2, peak in the first half of 2007 and since March 2008, conforming official assessments of a slowdown in economic growth and an economic decline, respectively. Real-time simulations made since December 2007 signal a downturn likely to occur from April 2008.

For monitoring purposes, an index of Home Appliances queries has been incorporated as an instrumental variable in the State-of-the-Economy composite index, while current assessments of private consumption (trade and services revenue) have been improved, in terms of RMSE. The operative use of query data, however, may encounter problems that deserve attention. First, query indices may appear not to be stationary, due to alternative social search which is not tracked by the Google engine. Second, the predictive ability of query indices may vary over time.

"חיזוי ההווה" לפי מדדים של חיפוש ב"גוגל" והתפנית של 2008 : נתונים ישראליים[1]

תקציר

במאמר "Predicting the Present with Google Trends" (Choi, Varian, 2008) מוצע מודל לחיזוי אינדיקטורים בו-זמניים לפעילות הריאלית, אשר לפיו הפופולריות של מושאי-חיפוש מסוימים ב"גוגל" - המוגדרת כשיעור החיפוש של מושא מסוים מסך החיפושים במדינה או באזור - עשויה לשמש משתנה מסביר לפני שהנתונים הרשמיים נעשו זמינים או עברו עדכון. עדות לכך נמצאה לגבי מכירת מכוניות, מכירת בתים, מסחר קמעוני ונסיעות בארצות שונות.
באמצעות יישום של Google Insights for Search, המסכם חיפושים ב"גוגל" לפי קטגוריות סוציו-אקונומיות גדולות, אני בוחנת את יכולת החיזוי של מדדי חיפוש הרלבנטיים לפעילות הכלכלית - בישראל. נמצאו שש קטגוריות לחיזוי, כפי שהוגדרו על ידי "גוגל": Human Resources (Recruitment and Staffing), Home Appliances, Travel, Real Estate, Food and Drink and Beauty and Personal Care. על פי מבחני הסיבתיות הדו-משתניים של גריינג'ר, כוח החיזוי הרב ביותר נמצא בקטגוריה Human Resources. נתון זה, הנמדד בתדירות חודשית, מקדים את שיעור המשרות הפנויות שמשרד התמ"ת מפרסם לכל רביע, ולפיכך עשוי לסייע בהסקת מסקנות חודשיות לגבי שיעור האבטלה. המשתנה שהיכולת לחזות אותו באמצעות מדדי החיפוש של "גוגל" היא הגבוהה ביותר הוא השינוי במספר משרות השכיר במגזר העסקי.שני ממצאים אלו מאפשרים לפרש את כוח הניבוי של הקטגוריות האמורות כביטוי למידת אמון הצרכנים.
ההסתברויות הבייסיאניות להאטה בפעילות הריאלית, המחושבות על סמך מדדי החיפוש ב"גוגל" בתקופה פברואר 2004 עד פברואר 2009, מאותתות על האטה במחצית הראשונה של 2007, וכן מאז מארס 2008 - תוצאה המתיישבת עם ההערכות הרשמיות בדבר האטה בצמיחה הכלכלית במחצית הראשונה של 2007 ומיתון מאז המחצית השנייה של 2008.
לשם יישום הממצאים האמורים במעקב אחר התפתחות הפעילות הריאלית שולב מדד חיפוש למוצרים בני קיימא, כמשתנה עזר לחיזוי הרכיב "הפדיון מסחורות ושירותים" במדד המשולב למצב המשק.
עם כל זאת אפשר שניתוח נתוני החיפושים ייתקל במכשולים, שיחייבו תשומת לב: ראשית, מדדי חיפוש, המתבססים, כאמור, על שיעורי חיפושים מסוימים מתוך סך החיפושים, יכולים להיות לא-סטציונריים, בעוד שהמעקב אחר הנתונים הרשמיים הרלבנטיים על ההתפתחויות הריאליות מתנהל במונחים של שיעורי שינוי, שהם בדרך כלל סטציונריים; שנית, ייתכן שיכולת החיזוי של מדדי החיפוש משתנה על פני זמן.

[1] מחקר זה נערך בשנת 2008 על פי הצעתו של נגיד בנק ישראל, פרופ' סטנלי פישר. אני מודה לפרופ' האל ואריאן על ההסברים והגישה לבסיס הנתונים הניסיוני של "גוגל"; ליונתן סידי על עזרתו במחקר; לרן שהרבני, המתדיין בסמינר חטיבת המחקר של בנק ישראל.

1. Introduction

This work relates to the article "Predicting the Present with Google Trends" (Choi, Varian, 2008), which provides an overview of the recently launched Google application, Insights for Search, which enables web-search patterns (called web queries) to be compared over time, locations, industries and consumer items. The main idea underlying this tool is that query statistics are available in close to real time and may predict overall economic activity before the official data become available. They detect the predictive ability of queries with respect to automobile sales, home sales, retail trade and travel.

The present work aims to discover whether Israeli query indices can be applied as economic predictors, or more specifically, which of them might be helpful for economic monitoring purposes.

During the last few years the penetration rate of the internet in Israel increased rapidly, and the range of e-commercial operations has expanded greatly. In addition to the ability to make price comparisons and purchases via virtual stores, auctions and tenders, the internet has many other commercial uses such as on-line banking and payment of bills, intensive use of on-line job advertisements, open access to government services, companies' financial reports and press releases, closing deals via Web portals, including the transfer of business and legal documents. With regard to households' on-line expenditure, home electrical equipment, electronic goods and travel services constitute the major part, constituting between 5 percent and 8 percent of total household expenditure per item. At the same time, the Israeli e-market is relatively small: the average customer is wary of executing online transactions due to reservations concerning their security, and tends to gather pre-purchase information rather than buying online. These facts, in addition to the quick accessibility of Google data, make the latter an attractive source of data for economic monitoring.

This work follows a concept of growth cycles,[1] in which aggregate economic growth alternates between low and high rates. When the rate of economic activity declines from its long-run trend, the probability of recession increases. The current cycle may be predicted by past known rates, until it comes to a turning point. In real-time evaluation, there is a high degree of uncertainty because of the delay in official data and further revisions. A question is whether scanning web-queries for their popularity provides timely exogenous information from which inferences can be drawn. If popularity dynamics coincide with economic dynamics, they should exhibit the same downturns.[2]

Popularity is the likelihood of a random user being interested in a particular category of searches on Google, measured as a fraction of the search volume in this category in the total volume of queries over the same time interval (a week). Query indices track popularity as time-series. In the long run this metric appears not to be

---

[1] This approach, pioneered by Stock and Watson (1989) in their seminal model of coincident economic indicators, was further developed by Kim and Nelson as a regime-switching model and applied in the Bank of Israel composite state-of-the-economy index and the probability of recession (Menashe et al., 2003) .

[2] Query indices have been stored by Google application since 2004, when a major economic upturn was over. As a time span is still short, no upturns were tested.

stationary; the trend is driven by the changing social profile of the internet, which has emerged recently as a social and entertainment utility; an enterprise search utilizes social search tools and knowledge-based sites and becomes less tracked by the Google engine.

In the presence of this non-stationary component, I consider first differences of query indices rather than their levels. The short-term predictive ability of query indices is tested with regard to monthly rates of real growth of industrial production, retail trade, trade and services revenue, consumer imports and services exports, as well as employment rates in the business sector, while controlling for their past known rates.

This work resulted in six query categories, which are of importance for growth-cycle assessments, namely human resources (recruiting and staffing), home appliances, travel, real estate, food and drink, and beauty and personal care.

The result involves two issues. First, the possibility of a monthly unemployment projection using the recruiting and staffing query index, given the fact that the unemployment rate and the preceding job openings ratio[3] move in opposite directions (Figure 1).

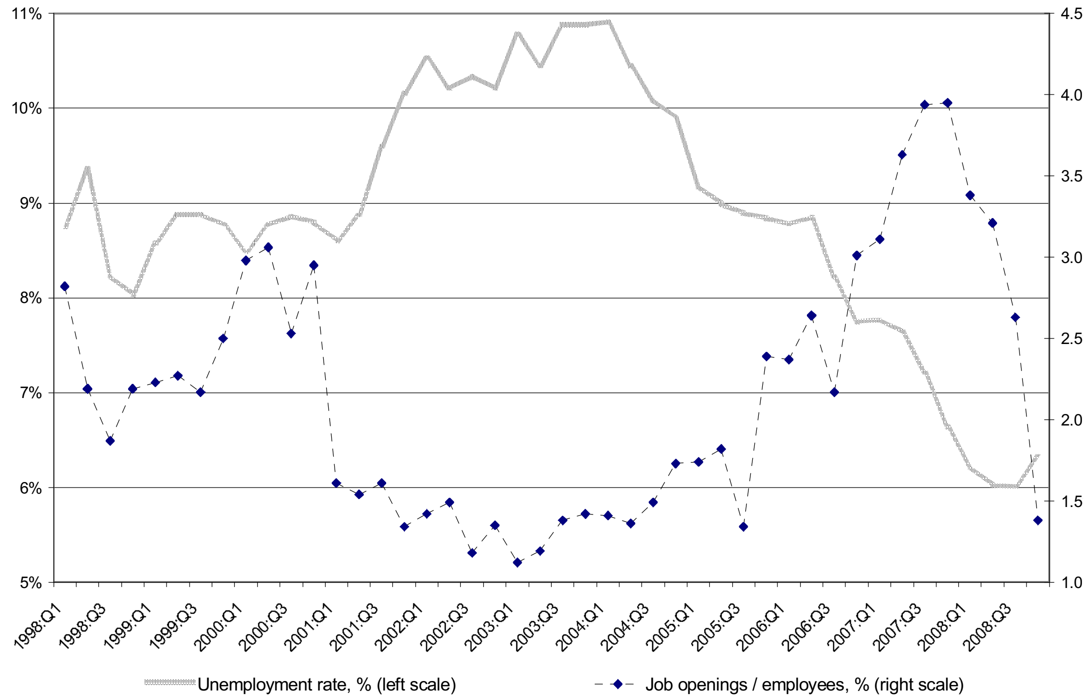

Figure 1. Unemployment rate (%, lagged) vs. job openings ratio (%) in business sector, seasonally adjusted quarterly data, Israel, 1998-2008.
Source: CBS Labor Force Survey, Ministry of Industry, Trade and Labor Survey

The second is a possible interpretation of five remaining categories in the sense of next-term consumer confidence. Thus, particular attention was paid to the large fraction of innovation variance of employment rates, explainable by selected queries. The current employment position, evaluation of job market opportunities and

[3] Job openings relative to employees number; accordingly to the findings of the quarterly surveys of Ministry of Industry, Trade and Labor, running since 1998 (available in Hebrew)

expectations affect both consumer spending and consumer mood. The slowdown in job growth curbs consumers' confidence and purchasing intentions; until the pace of hiring picks up, the cautious attitude will prevail, resulting in reduced popularity of inquiries for apartments, home equipments, restaurants, cosmetics and travel. The link between the two issues is straightforward, illustrated by Figures 2.1 and 2.2.

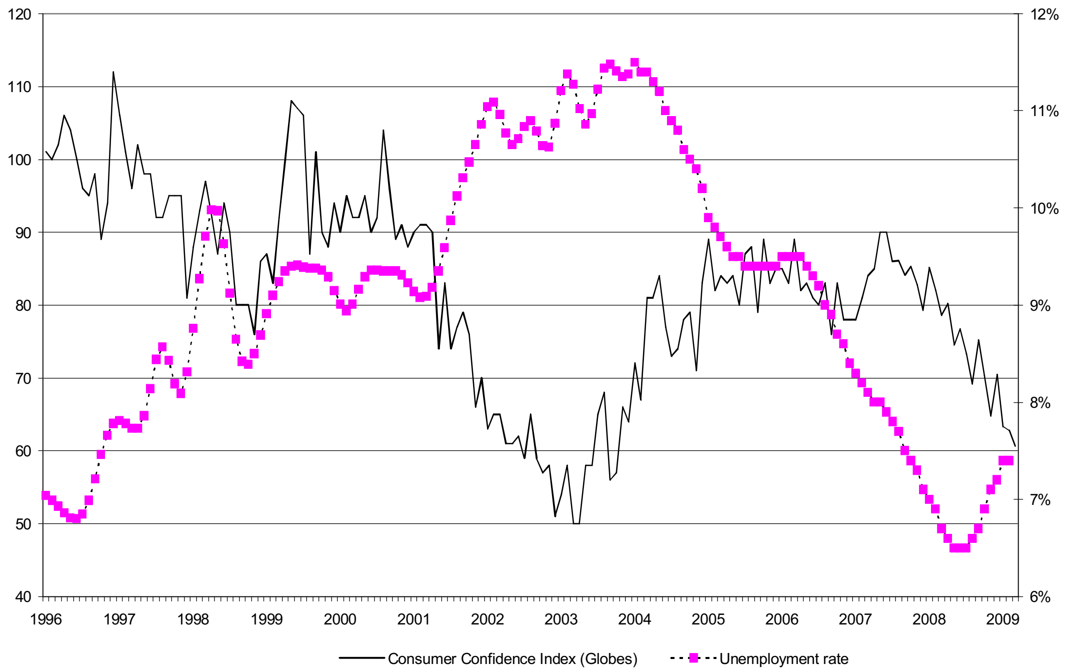

Figure 2.1 Unemployment rate vs. Consumer Confidence (Israel, 1996-2008, monthly)
Sources: CBS; Globes-Smith' Index of Consumer Confidence

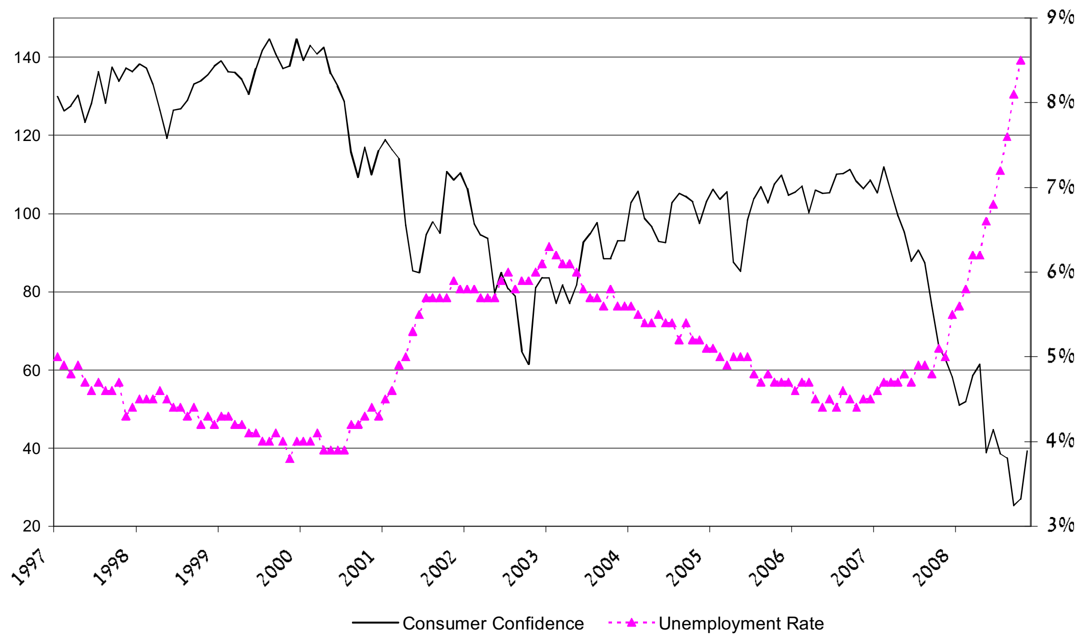

Figure 2.2 Unemployment rate vs. Consumer Confidence (U.S., 1997-2009, monthly)
Sources: U.S. Bureau of Labor Statistics; The Conference Board.

Empirical findings suggest incremental flexibility of consumer confidence to economic uncertainty, negative shocks and recession (Haugh, 2005; Ludvigson, 2004;

Braude and Friedman, 2003; Globes-Smith's Consumer Index of Israel, 2006–09). Then, based only on query data, I test six selected categories for the timing of downturn signals. Assuming that popularity changes may be described by two states, up and down changes relative to the long-run, I calculate the probabilities of the lower state, using Bayesian estimation technique fully described in Hamilton (1989), Chauvet and Hamilton (2005), and Kim and Nelson (1999).

Special investigation is required to combine individual signals into a synthetic one. Mostaghimi (1996, 2008) suggests a practical solution of a weighting scheme depending (inversely) on “relative entropy” of information, shared by components in the likelihood function. In our case, it means incorporating in the likelihood function the “true” probability distributions of a downturn and an upturn, evaluated exogenously from the state-of-the-economy/probability of recession monthly reports. So far, this problem has not been fully investigated. Thus, for comparison with official assessments of the state of economic growth equally weighted composed probability was used.

This paper is structured as follows: the next section provides some background information. The third deals with a methodology of testing query data for predictive ability. The fourth section describes the relevant query categories and reference data of official statistics. Section 5 discusses the empirical results, grouped by application: first, pairs of official cyclical indicators are tabulated, with corresponding query predictors; second, Bayesian probabilities of a downturn found upon leading queries are presented; third, two models of monthly projection, namely, of private consumption (trade and services revenue) and of the unemployment rate, are suggested. The last section concludes.

2. Background

Marketing studies concerned with high-involved products and services reveal that the frequency of a consumer’s visits to a website can be explained by purchasing intentions, the desire to gather information and consequently choose between online and offline channels. They justify the underlying behavioral rationale (Pickering, 1984), and find that intentions are transformed into purchases within a short time horizon (Morwitz et al., 2006).

During the last few years e-commerce has expanded worldwide faster than total economic activity: accordingly to 2004–07 data'[4] the e-markets of Britain, France, Germany and the U.S. grew at annual rates of between 25 percent and 39 percent. The market size of e-commerce reached almost 10 percent of total retail sales in Britain, 9 percent in the U.S.,[5] 5 percent in Germany and 4 percent in France.[6]

The size of the Israeli e-market is relatively small: accordingly to the non-official statistics reported by Global Technology Forum, the value of e-commerce consumer transactions was about NIS 1.5 billion at the end of 2005, only 1 percent of total retail

[4] See in: "The E-Commerce Market: Size and Trends"—http://www.gspay.com/the-e-commerce-market.php .

[5] According to U.S. Census Bureau E-Stats Report, 2008 May: www.census.gov/estats .

[6] Selected data from the Forrester Research have been reported by European E-commerce, Electronic Retailer Magazine, overview, February 2009.

trade and accommodation services revenue. Israeli consumers are less experienced in shopping on-line[7] for historical reasons: the relatively high price of personal computers in Israel in the late 1990s; high internet charges and payment restrictions resulted in a low penetration rate, with Israeli households lagging behind their European or American counterparts in their use of the internet. After the recession of 2001–03 internet use started accelerating rapidly: computer prices became more attractive, internet providers restructured their services due to convergence of telecom, internet and broadcast services (since 2005), on-line banking, investment and bill payments were permitted and electronic access to government services was made available to the public (by mid-2006); all this resulted in a rapid increase in the number of internet subscribers. Companies exploited web-portals intensively to exchange documents and catalogues and to close deals; on-line job advertisements by employment agencies became a powerful tool for both job-seeking and recruitment.

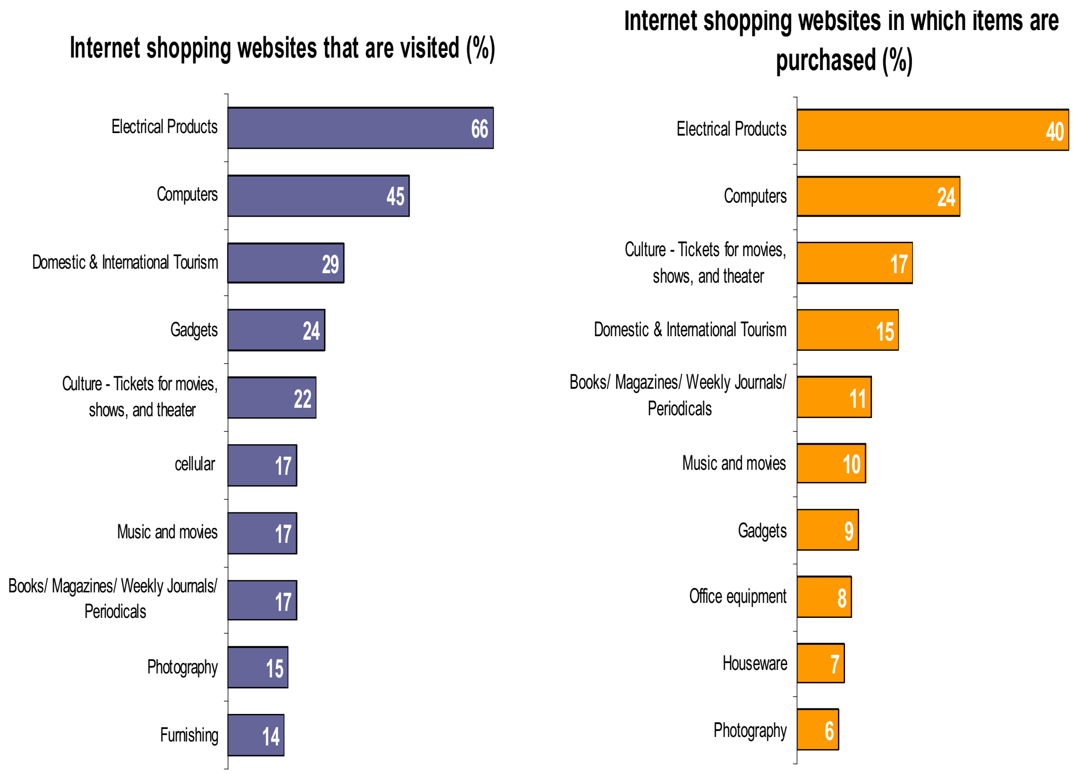

Figure 3. Distribution of Web visits alongside on-line purchases via Internet sites in Israel

According to the Central Bureau of Statistics and other surveys, 47 percent of the Israeli population used the internet in 2004, 67 percent in 2005 and 71 percent in 2006.[8] Based on the Household Expenditure Surveys 2003–04, 82 percent of household-subscribers use internet on a daily basis and 42 percent purchase goods or

[7] E-commerce analysts recall Vogel research which states that there is significant correlation between users' experience on the internet and their propensity to purchase online, explaining why the nations with the most experienced internet users have a higher percentage of online purchasers.

[8] The last two figures relate to Jewish households only, according to Taylor Nelson Sofres Teleseker.

services on a regular basis; 47 percent purchase electronic goods, 22 percent books, 17 percent hardware, and 14 percent pay for travel.[9,10] The Household Expenditure Survey of 2007[11] reports that among household expenditure on durable goods, internet purchases of electric home appliances (kitchen equipment, washing machines and dryers), durable entertainment goods, computers and accessories reached between 5 percent and 8 percent of total expenditure per item. The distribution of web visits and on-line purchases look similar, as revealed by the Shiluv Institute[12] survey of Israeli surfing habits (Figure 3).

Despite the growing familiarity of the Israeli consumer with the internet, he tends to gather information and compare prices rather than buy on-line. A survey reveals that 49 percent of Israeli web-users in 2007[13] reported that they were still wary about fraud when giving particulars of credit cards over the internet; they were concerned that the item purchased may not be supplied in time, may be defective, may not be what they expected, and that they may encounter difficulties in canceling the deal, etc. While the average Israeli becomes familiar with web activities, on-line purchases are still more likely to be made by the higher-percentile income groups. By the end of 2005 the Association of Banks reported that only 20 percent of new subscribers who had joined the banks' online services had actually executed on-line transactions. These findings are also supported by research of the Ministry of Industry, Trade and Labor (2007).[14]

Actually, Google Trends opens up a whole new field of market intelligence. At early stages, the tool was focused on microeconomic benefits: a particular search term becoming more and more popular would resonate best as an advertisement message; related and rising new search topics would help better understand competitors' offers and create brand associations; seasonal searches could anticipate demand and prove useful for business plans, budgets and allocations; geographic distribution of searches could be important in identifying new markets. Aggregation of queries into large socioeconomic categories,[15] compatible with official statistical divisions, encouraged macroeconomic checks for monitoring purposes.

[9] These data have been reported by Finkel, Y. and Yiftach, M. (2003, 2006) in their presentations dedicated to the weight of e-commerce in the Israeli CPI.

[10] By the Nielsen Company's 2008 survey, the most popular and purchased items globally over the internet are books (41 percent), clothing/accessories/shoes (36 percent), videos / DVDs / games (24 percent), airline tickets (24 percent) and electronic equipment (23 percent). In Europe travel leads the way; in Britain the popularity of online grocery shopping heads the list.

[11] CBS press release dated September 8, 2008.

[12] The data have been published by Globes on 10/09/2009; see in:
http://www.globes.co.il/news/docView.aspx?did=1000252490&fid=598

[13] According to the InterSight company's survey data, published by Ynet portal on 08.05.2007.

[14] Bar Zuri, R. (2007), (in Hebrew).

[15] In aggregation, particular terms are weighted by probabilities, calculated by categorization engine; for example, "apple" has a 70 percent chance of being in Computers and Electronics and a 30 percent chance of being in Food and Drink.

Are Israeli query indices pro-cyclical? At the top categories level, a link between popularity and economic dynamics is not so obvious. Despite the persistent economic expansion of 2003–07, query data collected since 2004 have shown an ongoing decline of interest in the automotive, computer, and electronic industries categories. on the other hand, the popularity of entertainment, news and current events and online communities in Israel, the U.S. and worldwide is rising (Table 1). Figure 4 depicts Israeli weekly query indices of consumer electronics and programming alongside those of "youtube" and "facebook".

Table 1. Interest over time by selected query categories - annual average indices[1] of Israel, the U.S. and worldwide , 2004-08 [2]

	Worldwide			US			Israel		
	Automotive	Computers & Electronics	Industries	Automotive	Computers & Electronics	Industries	Automotive	Computers & Electronics	Industries
2004	2.8%	-4.6%	-1.7%	2.2%	-5.0%	-4.7%	5.8%	-9.8%	-18.3%
2005	1.1%	-17.4%	-5.5%	-2.3%	-16.4%	-7.6%	4.9%	-28.6%	-29.5%
2006	-1.3%	-28.2%	-11.8%	-5.8%	-26.4%	-12.2%	-1.2%	-39.4%	-38.7%
2007	-1.5%	-33.2%	-13.9%	-7.1%	-31.3%	-14.4%	-5.5%	-46.0%	-43.9%
2008	-2.3%	-36.4%	-13.9%	-5.1%	-32.3%	-11.8%	-4.7%	-46.4%	-39.1%
	Entertainment	News & Current Events	Online Communities	Entertainment	News & Current Events	Online Communities	Entertainment	News & Current Events	Online Communitie
2004	0.9%	14.1%	10.7%	1.3%	22.7%	1.1%	3%	11%	18%
2005	4.5%	6.1%	33.1%	5.0%	-8.2%	10.1%	6%	23%	59%
2006	7.3%	12.6%	68.4%	6.3%	-3.9%	24.8%	11%	40%	98%
2007	10.5%	19.0%	117.7%	8.3%	-3.1%	50.5%	8%	38%	135%
2008	18.2%	44.0%	219.5%	13.9%	38.3%	119.7%	11%	50%	179%

[1] Each index represents average change of popularity (in %) relatively to the first week of 2004.

[2] Source: http://www.google.com/insights/search , weekly data.

These long-term trends seem to be driven by global web-technology innovations, development of free-access social networks and expanding areas of web-activity. Emerging quickly as an entertainment and social connectivity facility, the internet has created a new profile of user involved in a huge variety of web-activities. The Stanford internet study[16] suggests that self-selection no longer plays a role; the longer people have been web users the more activities they report engaging in.

Other surveys also report rapidly increasing numbers of e-mailers and Facebook-subscribers, of connections among people who share hobbies, political interests, institutionalized religion and professional associations.

What has happened in last few years is that the search has moved into social channels such as Twitter or Facebook, or enterprise social engines. Seeking specific business information, people do not evaluate their chances of success as high as that

[16] http://www.stanford.edu/group/siqss/Press_Release/press_detail.htm .

of a search for general information; failures occur if query terms do not suit the Google approach, or if the hits on the first page of results are not ranked as expected, or if Google does not recognize the professional jargon used in the search. Surfers prefer to get a quick answer rather than analyze hyperlinks for relevance, and trust co-workers more than random web-surfers. Social searching provides a kind of solution, where instead of extracting information from websites, the search engine utilizes its users' knowledge. Enterprise search is the first to become less Google-based.

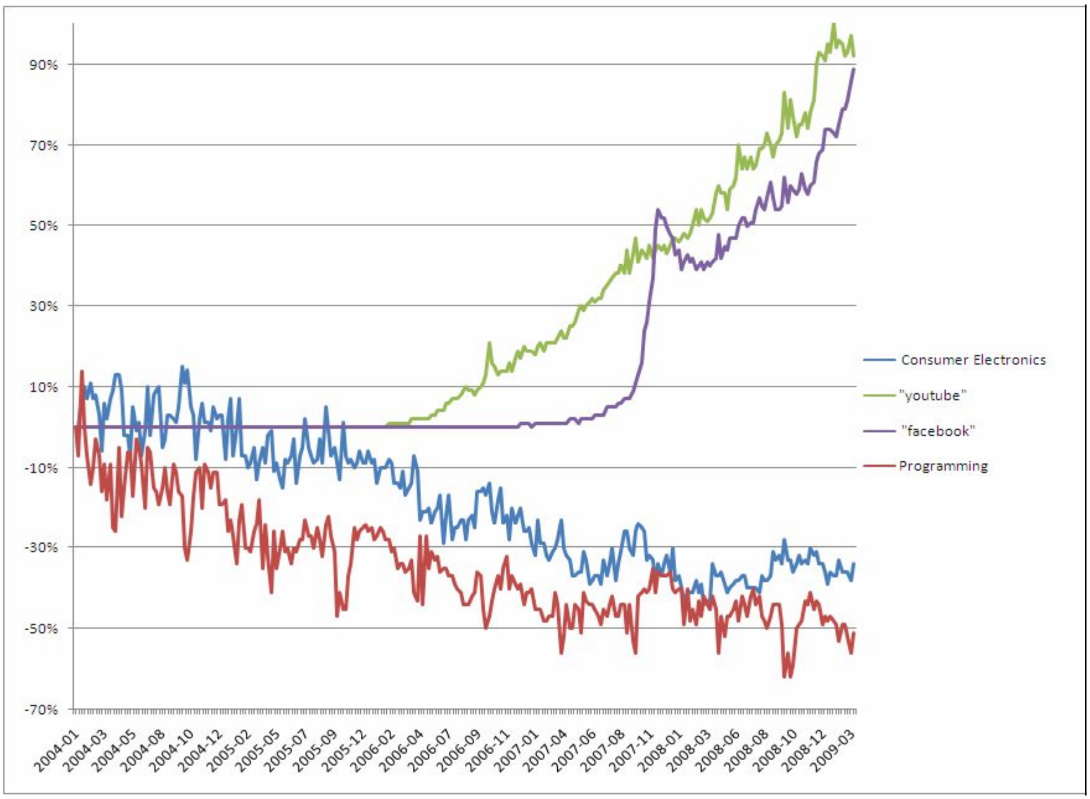

Figure 4. Selected Israeli query indices, original weekly data; sample 2004:1-2009:3; normalized to the first week of 2004.

To eliminate the trend, I consider two states of popularity dynamics: change up and down compared to the long-run change. For illustration purposes, Figure 5 depicts kernel densities of first differenced monthly query indices[17] of the flowers gifts and greetings category (id=70), computed from two sub-periods: first, between 2004:2 and 2007:11, that is, a period of economic expansion, and second, between 2007:12 and 2009:2, passing through a slowdown and recession. The sample mean of the first sub-period is approximately zero, with standard deviation 0.02 percent, indicating stable interest over time; the sample mean of the second sub-period is -0.02 percent with standard deviation 0.01 percent, indicating decreasing popularity of gift inquiries. Furthermore, considering the *t*-statistic, the two sample means differ at the 95 percent level of significance.

To select predictive queries, tests were conducted on the simple correlations and via pair-wise Granger causality. The first correlation checks were run in May, 2008

[17] Adjusted for seasonal and calendar effects and smoothed by 3-month moving average.

through the period of economic expansion between 2004:1 and 2008:2. The checks were repeated in February–April 2009, after the recessionary data have been received. If the correlation was lost, the category was filtered out.

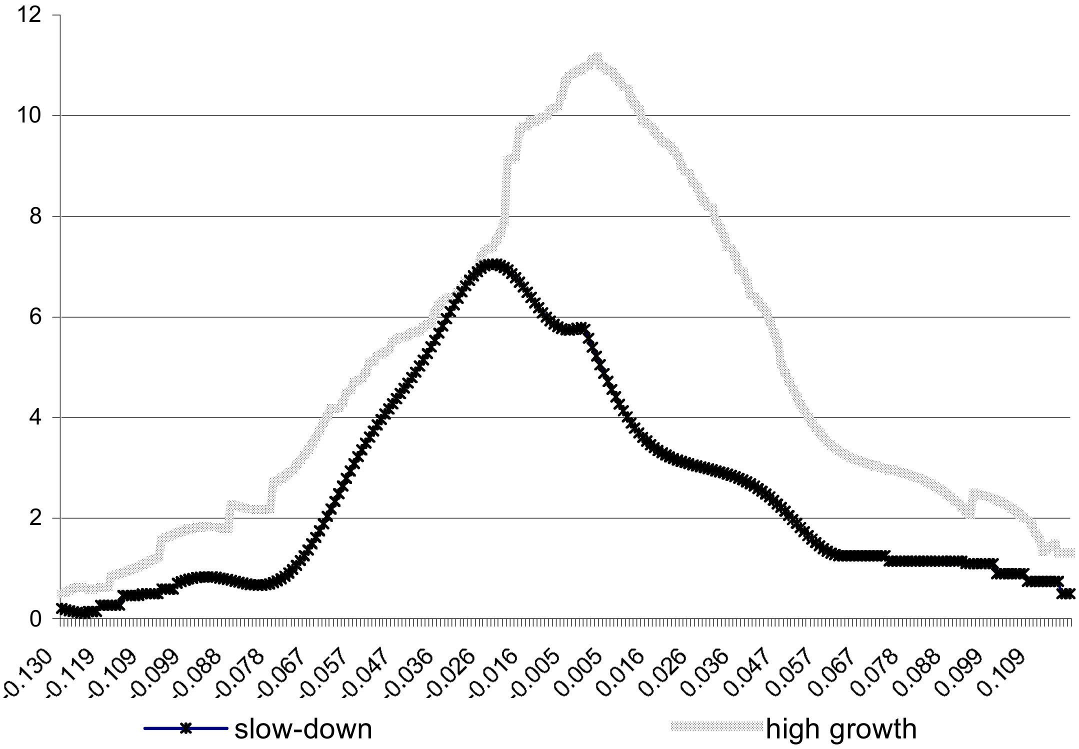

Figure 5. Kernel densities of monthly changes of query index in the category of Flowers Gifts & Greetings (id=70); the high growth curve (dashed) refers to the 2004:2 – 2007:11 sub-period; the slow-down curve (solid) refers to the 2007:12-2009:2 sub-period; month-to-month changes of popularity are evaluated upon two first weeks' averages, adjusted for seasonal and calendar effects and smoothed by 3-month moving average.

There are several reasons for Granger causality failures. The first is a new sample effect: Once the economy entered a recession, correlations weakened unless queries anticipated decline. Next, correlations turned out to be flexible to partial monthly information,[18] used to obtain more leads. Also notice that repeated data retrievals differ due to sampling error of Google's algorithm. Assessment of the importance of this noise has not been addressed in this article.

3. Methodology

The model of Choi and Varian (2008) suggests that the monthly prediction of economic indicator (e.g., sales) y_t, depending on its past level y_{t-1} and seasonal lag y_{t-12}, may be improved, while the contemporaneous query index x_t is taken into account. The query index is defined as a fraction of queries relating to a selected category and performed on Google from Israel every week in the corresponding total,

[18] At monthly frequency, query indices have been computed as two first weeks' averages. Spectral analysis shows well-distinguished harmonics of 4 week. Due to intra-monthly fluctuations, monthly indices, calculated upon partial or full monthly information may have different predictive power.

subsequently normalized to the first week of 2004. As soon as a category is selected and its index x_t is known, the "nowcast" of y_t is calculated as follows:

$$\log(y_t) = \alpha_0 + \alpha_1 \log(y_{t-1}) + \alpha_{12} \log(y_{t-12}) + \beta x_t + e_t\,, \quad e_t \sim IID(0,\delta^2) \quad (1)$$

where the estimated positive slope β matches intuition and the sum of parameters α_1 and α_{12} is close to 0.9. While run on seasonally adjusted data of a dependent variable, the first lag parameter was also approximately 0.9.

Thus, we would expect $\beta > 0$ while (1) is reformulated in terms of log differenced dependant variable, where both dependant and exogenous (query) variables are seasonally adjusted (denoted as $\tilde{y}_t$ and $\tilde{x}_t$, respectively), that is α_1 is restricted to be unit:

$$\log(\tilde{y}_t) - \log(\tilde{y}_{t-1}) = \alpha + \tilde{\beta}\tilde{x}_t + \zeta_t \qquad \tilde{\beta} > 0\,, \quad \zeta_t \sim IID(0,\delta_\zeta{}^2) \quad (2)$$

It is possible that $\tilde{x}_t$ is not stationary. Then, by applying first differenced query indices, the short-time predictive relationship could be written as:

$$\Delta y_t = \rho_0 + \sum_{l=1}^{L} \rho_l \Delta y_{t-l} + \sum_{l=0}^{L} \beta_l \Delta x_{t-l} + \zeta_t\,, \quad \zeta_t \sim IID(0,\delta_\zeta{}^2) \quad (3)$$

where $\Delta y_t = \log(\tilde{y}_t) - \log(\tilde{y}_{t-1})$, $\Delta x_t = \tilde{x}_t - \tilde{x}_{t-1}$, $L = 1$.

To test Δx_t for extra information which is not contained in the past known rates of official indices, the pair-wise Granger causality rule was applied. Under the null hypothesis $H_o : \beta_0 = \ldots = \beta_l = 0 \quad (l = 1,2)$ an alternative autoregressive model has to be evaluated, which does not include query information:

$$\Delta y_t = \tilde{\rho}_0 + \sum_{l=1}^{L} \tilde{\rho}_l \Delta y_{t-l} + \eta_t\,, \quad \eta_t \sim IID(0,\delta_\eta{}^2) \quad (4)$$

Then, respective sums of squared residuals of the unrestrictive (3) and the restrictive (4) equations are compared via the F-test:

$$\frac{(\sum \eta_t^2 - \sum \zeta_t^2)/L}{\sum \zeta_t^2/(T-2L-1)} \sim F_{L,T-2L-1} \qquad (5)$$

If the test-statistic $F_{L,T-2L-1}(L = 1,2)$ exceeds its critical value, we reject the null hypothesis and conclude that the selected queries category is a leading one.

The pair-wise Granger causality tests have been applied sequentially to filter query categories with respect to their predictive power.

Given selected categories, Bayesian probabilities of downturn have been calculated according to Hamilton's (1989) two-state AR(0) model. Specifically, values of popularity changes $\Delta\tilde{x}_t$, adjusted for seasonal effects, may be considered as drawn from a mixture of two distributions, one, corresponding to the rising popularity relative to its long-run change, and the other, to the popularity decline. These states are defined by a binary random variable $S_t = \{0,1\}$ receiving values 1 or 0 accordingly to rising or falling popularity. As the degree of data variability (coefficient of variation) is high and the time span is short, these states are hard to recognize. Thus, smoothed query indices are required. Another problem that arises is that of the well-known end point effect, when the T th observation, which is the month in question for the current monitoring, is likely to be revised later.

To deal with the end-point problem, a weekly frequency of Google indices is helpful. By accounting only for two first weeks' indices, one can get monthly averages, which predate official data by two months. Thus, the T-month smoothed estimate, which is of interest, can be conditioned on the $(T+1)$-month observation, available from the query time series. Such a treatment has the advantage, first, of a more robust T-month smoothed estimate,[19] conditional on subsequent data availability; secondly, of enabling a smoothed probability of a downturn, which will be considered below, also conditional on subsequent observations.

Let $z_1^{(2w)},...z_T^{(2w)} \mid \Omega_{T+1}$ be smoothed (trend) estimates of a query index $x_t^{(2w)}$, calculated as the first two weeks' average. For a particular smoothed change of popularity, calculated as $\Delta z_t^{(2w)} = z_t^{(2w)} - z_{t-1}^{(2w)}$, $t = 1,...T$, a Bayes' rule can be applied to obtain the in-sample conditional probability of downturn:

$$\Pr(S_t = 0 \mid \Delta z_t^{(2w)}, \Omega_{T+1}) = \frac{f(\Delta z_t^{(2w)} \mid S_t = 0)\Pr(S_t = 0)}{f(\Delta z_t^{(2w)} \mid S_t = 0)\Pr(S_t = 0) + f(\Delta z_t^{(2w)} \mid S_t = 1)\Pr(S_t = 1)} \quad (6)$$

where $f(\Delta z_t^{(2w)} \mid S_t = 1, \Omega_{T+1})$ and $f(\Delta z_t^{(2w)} \mid S_t = 0, \Omega_{T+1})$ are query densities, conditional on the state of rising or decreasing popularity, respectively;
$\Pr(S_t = 1)$ and $\Pr(S_t = 0)$ are unconditional (transition) probabilities of increasing or decreasing popularity, respectively.

Since the queries span is short, the parameterization of (6) is the most simple: only the shift of the mean is allowed, while the variance is kept constant. The mean value μ_0 corresponds to the state of down-changes $(S_t = 0)$ and the mean value $\mu_0 + \mu_1, \mu_1 > 0$ corresponds to the state of up-changes $(S_t = 1)$. The current smoothed change $\Delta z_t^{(2w)}$ depending on a state S_t is modeled as an AR(0) Markov-switching process:

$$\begin{aligned} &\Delta z_t^{(2w)} = \mu_{S_t} + \nu_t, \quad \nu_t \sim N(0, \delta_\nu^2) \qquad (7) \\ &\mu_{S_t} = \mu_0 + \mu_1 S_t, \quad \mu_U > 0, \quad S_t \in \{0,1\} \end{aligned}$$

where S_t is subject to a first-order Markov chain with transition probabilities:

$$\Pr[S_t = 0 \mid S_{t-1} = 0] = q, \quad \Pr[S_t = 1 \mid S_{t-1} = 1] = p$$

Having evaluated (7), one gets densities of popularity changes in each state, that is:

$$f(\Delta z_t^{(2w)} \mid S_t = 0) = \frac{1}{\sqrt{2\pi\delta_\nu}} \exp(\frac{-(\Delta z_t^{(2w)} - \mu_0)^2}{2\delta_\nu^2} \text{ for down-changes;}$$

[19] Calculated here by the X-12-Arima procedure as a Henderson curve with asymmetric end-point weights; prior to smoothing, the series is extrapolated via an appropriate seasonal ARIMA process; thus, each additional end observation matters.

$$f(\Delta z_t^{(2w)} \mid S_t = 1) = \frac{1}{\sqrt{2\pi\delta_v}} \exp(\frac{-(\Delta z_t^{(2w)} - \mu_0 - \mu_1)^2}{2\delta_v^2} \text{ for up-changes}$$

and proceed to (6).

Gibbs-sampling scheme is employed to obtain the joint distribution of all unknown parameters, through iterative draws of each unknown parameter from its appropriate distribution, conditioned on other parameters drawn earlier. Using arbitrary starting values and appropriate distributions described in (Kim and Nelson, 1999), at the first stage I generate states, at the second, transition probabilities, at the third stage, mean changes, at the last, the standard deviation; I then proceed to a new iteration. This sequence of multiple draws from the respective conditional distributions has been proved to converge to the joint distribution after a sufficiently large number of iterations. To eliminate initial values effect, I discard the first portion of 500 draws and use the remaining 1000 to calculate posterior means and the standard deviation of query changes. No evidence was found concerning flexibility of two posterior means to their starting values.

On the contrary, the results appear to be flexible to the initial values of the transition probabilities. To start iterating, I adopt transition probabilities from the State-of-the-Economy Index monthly report and set standard deviation of popularity change equal to the 2004:2 – 2007:11 sample value; I assign non-positive initial value to μ_0 and derive initial value of μ_1 under the restriction that $\mu_0 + \mu_1$ is positive.

Once (6) is evaluated, state recognition becomes easier using smoothed probabilities, depending on next-month probability estimates, i.e., $\Pr[S_t = 0, S_{t+1} = s \mid \Omega_{T+1}] \quad (s = 0,1 \quad t = 1,..T)$. The latter have been calculated with Kim's smoothing algorithm (Kim and Nelson, 1999, pp.68-70).

Thus, for m selected categories vector of individual downturn probabilities $\Pr_t = \{P_t^{(1)},...P_t^{(m)}\}$ has been evaluated for each month t , conditional on query information till $t+1$. To make state assessment, some synthetic composition of individual probabilities should be constructed. Until a more accurate weighting scheme is investigated, the most intuitive average (i.e., equal-weighted) composition has been used.

4. Data description

As accepted, the determinants of Israeli growth cycle are monthly percent changes of real values of industrial production, retail trade, revenue of trade and services, consumer imports, exports of services and the employment rate. Corresponding official indices, seasonally adjusted, are taken here as reference series (Appendix 1).

As mentioned before, matching query categories to reference series proceeds from Google's division into 30 categories at the top level and a further division into subcategories, excluding books, entertainment, education, health and sport (Appendix 2).

To avoid overestimation, query indices series were tested for seasonality, although the time span is too short for Jewish calendar shift to be definitely detected. I assume it exists, if moving seasonality is detected at the 90 percent significance level; all seasonal/calendar effects have been removed by X-12-Arima routine.

A brief overview of the major considered categories is given below.

The Automotive category includes searches of vehicle brands, auto parts, auto insurance, leasing etc. Like worldwide and in some cases more acutely, the interest in this category in Israel has declined since the middle of 2005, excluding leasing, insurance and auto parts (predominately tires), which have exhibited stable or even rising interest, at least till 2007. The most popular search is the Yitzhak Levi price list of used cars. Searches for garages were rising faster than the total search volume of this category until the end of 2007. Based on the sub-period of economic expansion between 2004:2 and 2007:11, positive correlations of the automotive query index were recorded with the trade and services revenue and industrial production, notably of high and medium technologies, characterized by intensive use of leasing services. Queries for rent-a-car, which relate to the auto-financing sub-category, shared a positive correlation with the exports of services monthly percent changes. Also, some predictive ability relating to consumer imports was detected, due to imports of vehicle brands to consumers and leasing firms. Based on the whole period between 2004:2 and 2009:2, the correlations appear to have weakened or even been lost.

The business categories searches worldwide, as well as in the U.S. are mostly concentrated on office equipment, marketing and pay scales; the top searches contain these keywords: "calendar,,"[20] "staples," "marketing" and "salary." There is remarkable rising popularity, also in Israel, of the "pay pal" search—an e-commercial tool of payment and money transfer. However, the main inquiries do not refer to salary, but are mostly concerned with office equipment (the total category), job seeking, recruiting and management consulting. Among searches rising faster than the total category are Pilat (a psychometric test for job candidates), Machon Noam (an Israeli preliminary assessment center, that also provides services of personal preparation for tests and job interviews), and “jobmaster” (a Israeli portal of job advertising). The overall index of the business category has gradually declined in Israel, as well as worldwide, while the popularity of human resources and recruitment and staffing sub-categories is stationary; the latter started declining since the middle of 2008. Two important features of the recruitment and staffing index have been exploited in this work. First, it exhibits high stable positive association with the employment rates. Second, at a quarterly frequency it is likely to conform to the percentage changes in job openings, available from the Ministry of Industry, Trade and Labor quarterly survey.

The popularity of home and garden is stationary, like its worldwide counterpart; the trend of the corresponding U.S. index has a mild negative slope. The top searches are “furniture,” “ikea,” “doors,” “lighting,” “tambour,” “ace,” “design,” the top searches reported worldwide are quite similar. According to the first sub-sample, i.e., 2004:2-2007:11, the corresponding index was found well-correlated with retail trade (durables goods) rates, trade and services revenue and consumer imports rates. It is not surprising, that the home appliances sub-category’s fit is as good; the top searches (all in Hebrew) are “washing machine,” “refrigerator,” “tadiran,” “electra,” “air conditioner,” “oven” refer to the most popular household on-line purchases. Through

[20] Here and later I have translated inquiries originally entered in Hebrew into English. Google, of course, differentiates between them.

the whole available period 2004:2–2009:2 this index has outperformed the overall category index, especially when data of half-month coverage have been used.

Beauty and personal care embodies searches for fitness clubs, hair, face and body care, cosmetic surgery, weight loss, spa, beauty magazines; regarding the Israeli queries, the top searches are “hair," “fitness," “spa," “Holmes place," “studio-c," “yoga”. As appears from Israeli data, this index seems to be a litmus paper of consumers' mood. Having a high correlation with the employment rates, this remains informative after the span has been extended, pointing at slowing down since the middle of 2008; it performs as well through the data of partial monthly coverage.

It is possible that the food and drink popularity works similarly. The top searches are "recipe" or "recipes," "café,"[21] "restaurants," "chocolate," "chicken" (meaning recipes of cooking as revealed by related terms), "pizza," "cake" (baking recipes and decorating). While eating out directly affects consumer spending, the link of cooking searches is vague. We could interpret it by different ways: first, people are sharing culinary tips, chefs' cookbooks and cooking career information when activities are rather high; second, most people do not use recipes when making dinner, as confirmed by surveys; they are more relevant for healthy or gourmet food and could also be associated with consumer confidence. Third, a high correlation of the corresponding index with the employment rate gives rise to the possibility, for instance, that people are looking for recipes while in the workplace.

The travel category performs well, as expected. The top searches are "hotel" or "hotels," "tzimmer," "flights," "holiday" (meaning vacation packages), "sea," "el-al," "issta". Beyond seasonality, this index has cyclical component, predictive about trade and accommodation services revenue. The hotel and accommodation index is well associated with the exports of services (through travel services) and the employment posts rates, supporting our assessment above. Nevertheless, any clear preference could be given to the sub-categories "air travel," "hotels and accommodation," "car rental and taxi services" over the total, until additional checks are performed.

The shopping queries differ substantially from their U.S. counterparts: in the top are "greetings" (for greeting cards download) and "zap" (the most popular site for price comparisons) contents; the others -"sport," "ikea" and "ebay," which relate also to paypal, Amazon and other shopping sites, localized in different countries, contribute less. On the contrary, the "ebay" search is the most popular one within this category both in the U.S. and worldwide. This contrast provides further evidence that the on-line shopping activity in Israel is relatively minor. Based on the first sub-period of economic growth from 2004:2 till 2008:4, various query indices from the shopping category were found positively correlated with retail trade, revenue and employment rates; however, after the sample has been extended, all but flowers, gifts and greetings category failed to signal a downturn of 2008.

The real estate category fits the monthly changes in dwelling supply and trade and services revenue, through queries for rental and sales listings, real estate agencies and land registry documentation. The top searches are "apartments," "homeless" and "yad2" (popular web-sites which contain rent, sale and short-term apartments), "for

[21] In the U.S. and worldwide this search does not belong to the top, while "restaurants" do.

rent," "for sale," "tabu" (entry in the land registry). Compared to the U.S., the "mortgage" term is not on the top of popular Israeli queries via internet.

The queries for finance and insurance are rising in Israel faster than in the U.S., due to higher rates of new subscribers for online banking services; the top searches are in Hebrew, and relate to the small number of large Israeli banks. After the current crisis erupted, the popularity index started slowing down. Although the downturn is very apparent ex post, it is not clear in advance what motivates the query. In the first sub-period, no significant correlation with the revenue of financial services was obtained. The positive correlation with the employment rate strengthened since the start of the recession.

As in the earlier discussion, popularity of the Computers and Electronics category is declining in Israel, the U.S. and worldwide; since 2004, the corresponding indices have fallen by 46 percent, 32 percent and 36 percent respectively. Among the top searches, the common are programming terms like "download," "hp," "vista," "linux," "java," "microsoft". The Israeli-specific top searches are "crack" and "serial" indicating the fast-developing industry of pirated software, key generators and crack tools, downloaded from websites. We can only guess which motivation is prevalent in this activity: hacking for fun, seeking confidential information assets or the investigation of security risks. Among the sub-categories, only the hardware popularity exhibits cyclical movement around the steeply declining trend; the corresponding first differenced index correlates positively with percentage monthly changes in industrial production (namely, of high and medium technologies), employment and imports of consumer durables. The top searches of the Hardware sub-category are similar everywhere: “hp," “driver," “asus," “usb," “dell," “computer” and “laptop.”

In the last few years, the popularity of the industry queries has also declined everywhere. Taken at first differences, Israeli indices provide some evidence of positive association with official data, but not at the top level: for example, building materials, which is a further sub-division of the construction and maintenance sub-category, appears to be positively correlated with the industrial production sub-division, namely, transport equipment and machinery and equipment. Despite the fact that the top Israeli queries about construction equipment and materials have been entered in English, like “screw," “concrete," “wood," “doors” no significant correlations with the imports of investment goods were found. Neither correlation with building starts was obtained. These findings, or more precisely, their absence, make sense if Google search appears not to be a main tool of the Enterprise search.

5. Estimation results

5.1. Granger causality tests

Table 2 lists query categories positively associated with reference series. The last two columns of this table provide details of two mapping procedures, performed in May 2008 and February-April 2009.

Based upon month-to-month changes of query indices between 2004:2 and 2008:2, column (6) presents the number of positive correlations relating to different reference subjects[22] found for each category. Column (7) presents number of positive correlates, found for each category upon the whole period 2004:2 - 2009:2, available by February-March 2009.

Categories with no positive correlations found are not presented.

Table 2. Predictive characteristics of selected query categories, by the sub-period of 2004:2–2008:2 and the whole period of 2004:2–2009:2

ID	Category	Subcategory	ADF (p-values)[1]	Seasonal and calendar effects [2]	Number of positive correlates[3] (satisfying Granger-test)	
					(2004:2 - 2008:2)	(2004:2 - 2009:2)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
47	Automotive	ND	0.16	*	3	2
89	Automotive	Auto Parts	0.09 *	no	2	-
815	Automotive	Vehicle Brands	0.17	no	2	-
4	Business	ND	0.00 ***	**	3	2
157	Business	Human Resources	0.00 ***	*	3	3
330	Business	Human Resources (Recruitment & Staffing)	0.00 ***	***	2	2
30	Computers & Electronics	Hardware	0.32	*	3	3
11	Home & Garden	ND	0.00 ***	***	2	2
271	Home & Garden	Home Appliances	0.00 ***	***	3	3
48	Industries	Construction & Maintenance	0.31		2	-
650	Industries	Construction & Maintenance	0.12		3	1
18	Shopping	ND	0.00 ***	***	2	1
73	Shopping	Mass Merchants & Department Stores	0.59	***	2	-
323	Shopping	Flowers Gifts & Greetings	0.00 ***	***	2	1
696	Shopping	Luxury Goods	0.02 **		2	-
7	Financing	ND	0.74		1	1
29	Real Estate		0.45	**	3	3
44	Beauty & Personal Care	ND	0.00 ***	**	1	1
71	Food & Drink	ND	0.66	**	4	5
67	Travel	ND	0.00 ***	***	2	2
179	Travel	Hotels & Accommodations	0.00 ***	***	3	3
203	Travel	Air Travel	0.00 ***	***	2	-

[1] The null hypothesis is non-stationarity (presence of unit root); ***, **, * correspond to 1%, 5% and 10% significance levels, respectively

[2] The null hypothesis is no seasonal effects; ***, **, * correspond to 1%, 5% and 10% significance levels, respectively

[3] Only reference series which refer to different reference subjects (Appendix 1) have been counted

[22] For example, overall industrial production index and one of high-medium technologies relate to the same reference subject.

Table 3. Pair-wise values of F-test statistics and corresponding p-values

Industrial production	Employment posts in business sector	Trade and services revenue	Of which: Accommodation services and restaurants	Retail trade - durables	Consumer imports[2]	Dwellings supply - private sector	Services exports
Category ID = 157,330: "Business (Human Resources(Recruitment & Staffing))"							
8.636	29.707						5.182
(0.003)	(0.000)						(0.022)
Category ID = 11: "Home & Garden (ND)"							
	10.882			8.034			
	(0.001)			(0.005)			
Category ID = 271: "Home & Garden (Home Appliances)"							
	3.596	4.216		33.696			
	(0.057)	(0.040)		(0.000)			
Category ID = 47: "Automotive (ND)" [1]							
		4.722			2.879		
		(0.029)			(0.089)		
Category ID = 29: "Real Estate (ND)"							
		5.153				3.660	2.580
		(0.023)				(0.056)	(0.100)
Category ID = 30: "Computers & Electronics (Hardware)" [2]							
5.788	9.309				12..408		
(0.016)	(0.002)				(0.000)		
Category ID =44: "Beauty & Personal Care (ND)"							
	21.889						
	(0.000)						
Category ID = 67: "Travel (ND)"							
		3.113	3.204				
		(0.077)	(0.073)				
Category ID = 179: "Travel (Hotels & Accommodation)" [3,4]							
	6.344	3.565					6.794
	(0.011)	(0.059)					(0.009)
Category ID = 71: "Food & Drink (ND)" [4]							
	3.468	5.411	4.206	2.733			3.977
	(0.060)	(0.020)	(0.040)	(0.098)			(0.046)
Category ID =323: "Shopping (Flowers Gifts & Greetings)" [4]							
	2.918						
	(0.080)						
Category ID = 650: "Industries(Construction & Maintenance(Building materials))" [5]							
16.759							
(0.000)							

[1] For industrial production reference, "high-medium technologies" sub-division index has been used; the *t*-statistic for the overall index is not significant, although the F-test holds. For consumer imports, a "vehicles" sub-division index has been used.
[2] For consumer imports reference, a durable goods sub-index has been used.
[3] For exports of services reference, a tourism services sub-index has been used.
[4] For trade and services revenue reference, a commerce sub-index has been used.
[5] For industrial production reference, transport equipment and machinery and equipment sub-divisions have been used; due to similar results the latter has been omitted.

To assess stationarity of query indices, column (4) shows *p*-values of Augmented Dickey-Fuller statistics, compared to their critical values; a null hypothesis of unit root is not rejected if corresponding *p*-value exceeds the 10 percent level. According to the results, query indices, unlike confidence indices, may be non-stationary, and first differencing is reasonable. Column (5) marks with asterisks categories with significant seasonal and calendar effects, adjusted appropriately.

Table 3 presents F-test values of pair-wise Granger causality; only categories of predictive ability approved at least at the 10 percent significance level are reported. This table is constructed as a guide, suggesting the best query predictor for each reference series. In most cases a weak to moderate predictive ability of queries has been recorded; in some cases no predictive ability was recorded with regard to overall reference index, but only at the sub-index level, as, for example, high-medium technologies, transport and machinery equipment or consumer imports of durable goods.

With that, two features attract attention. First, the human resources category appears to be the most predictive one; apart from a natural correlation with employment rates, no other query category explains innovation variance of overall industrial production growth. Second, a large fraction of innovation variance, explained by queries lies in the employment rate. It suggests that the interpretation of queries effect is attitudinal and may be conveyed through the employment channel.

5.2. Bayesian probabilities of downturn

Table 4 presents estimated parameters of a Markov-switching AR(0) model, by selected categories. All parameters of mean changes are statistically significant, suggesting that popularity changes may be treated as state dependant.

Table 4. Parameters of Markov switching AR(0) model, by leading categories

ID	Category	Subcategory	Average change of popularity[1] (%), by state		Transition probabilities	
			Increasing (s=1)	Decreasing (s=0)	p	q
157	Business	Human Resources	1.4	-0.9	0.96	0.87
			(0.08)			
271	Home & Garden	Home Appliances	0.6	-0.3	0.85	0.80
			(0.02)			
29	Real Estate	ND	3.7	-1.5	0.95	0.78
			(0.58)			
44	Beauty & Personal Care	ND	1.2	-0.6	0.94	0.88
			(0.06)			
71	Food & Drink	ND	1.0	-0.3	0.96	0.82
			(0.08)			
67	Travel	ND	1.5	-1.1	0.95	0.90
			(0.04)			

[1] Standard deviations (%) are given in parentheses

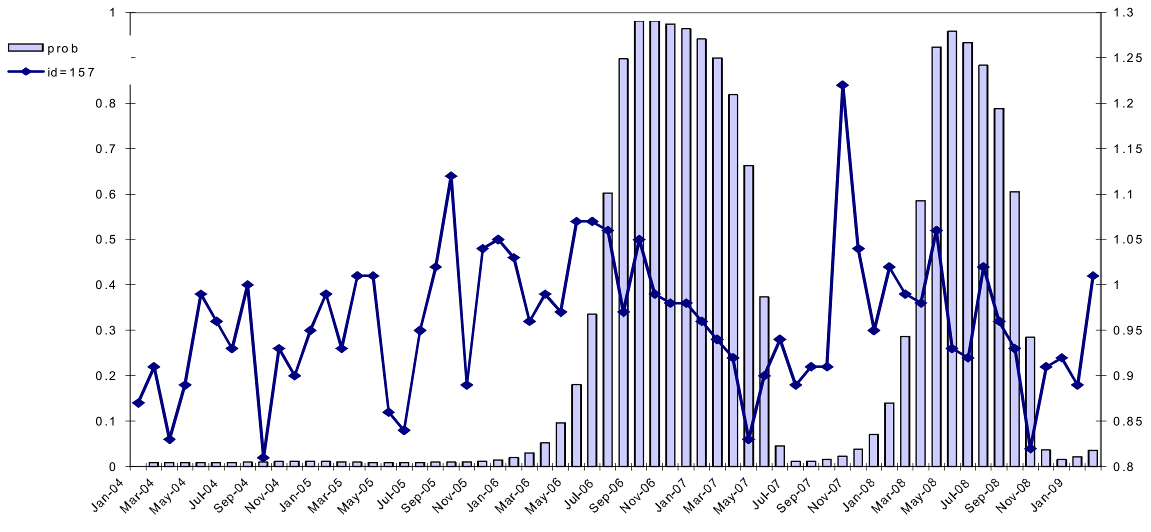

Figure 6.1. Human resources (ID=157): in-sample probability of a downturn (bars) and the seasonally adjusted query index.

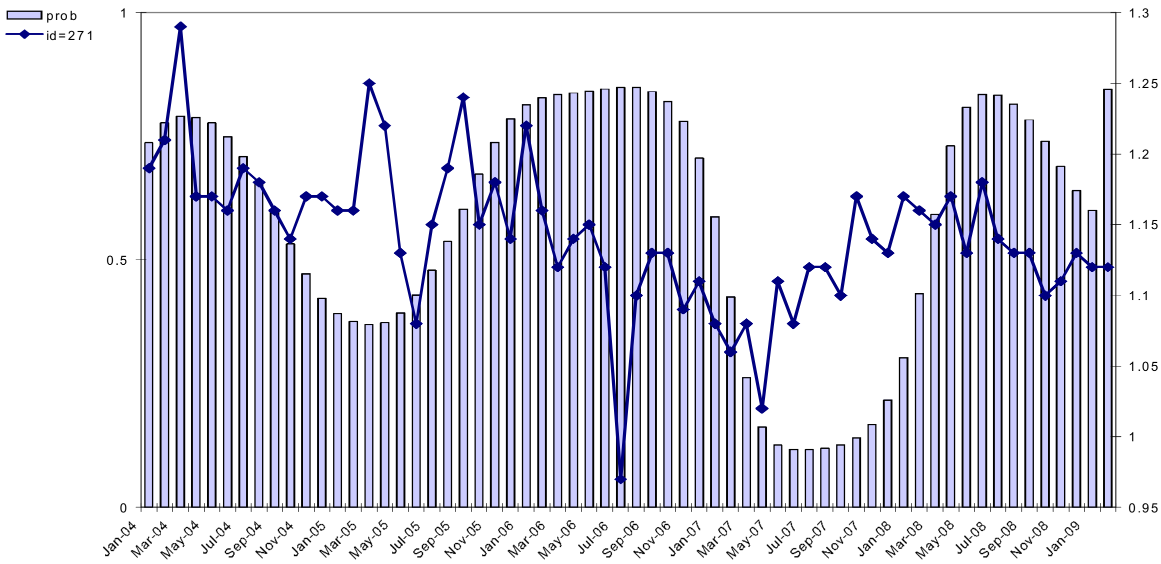

Figure 6.2. Home appliances (ID=271): in-sample probability of a downturn (bars) and the seasonally adjusted query index.

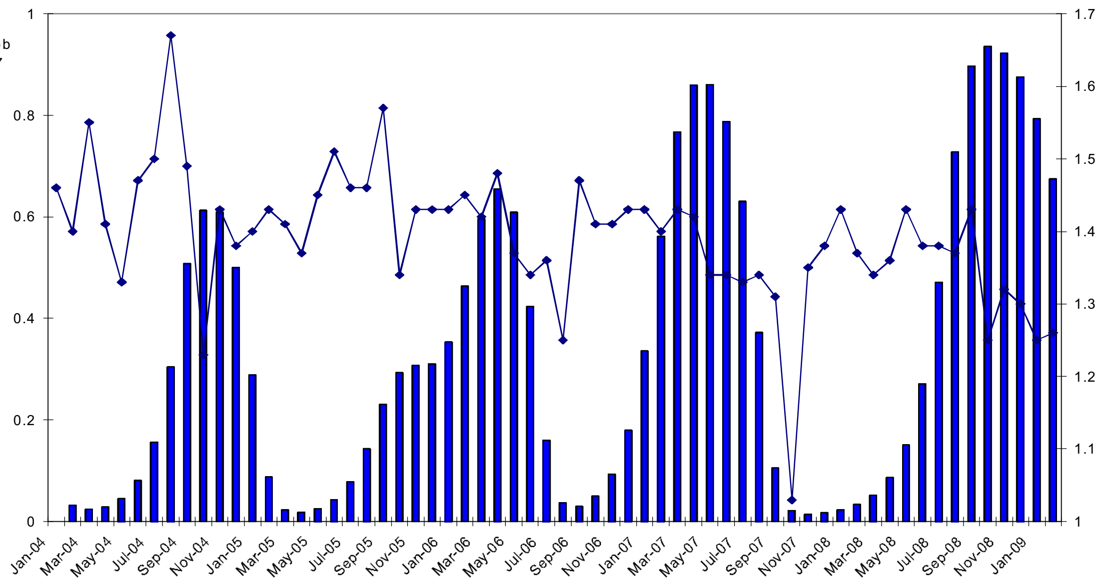

Figure 6.3. Travel (ID=67): in-sample probability of a downturn (bars) and the seasonally adjusted index (line).

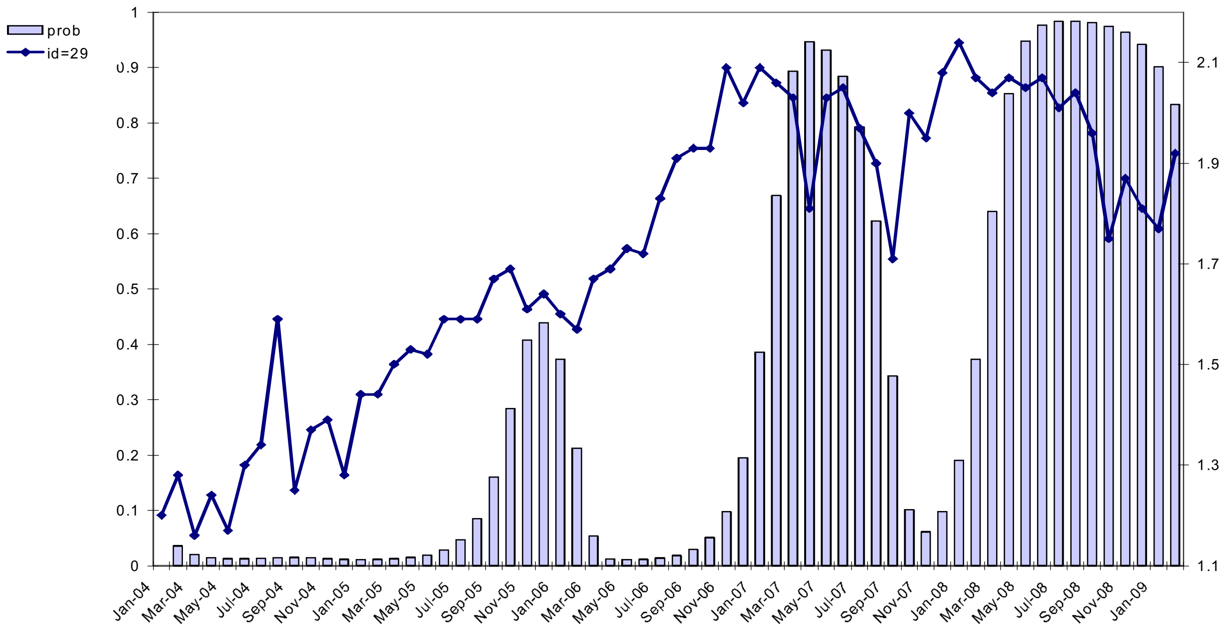

Figure 6.4. Real estate (ID=29): in-sample probability of a downturn (bars) and the seasonally adjusted index (line).

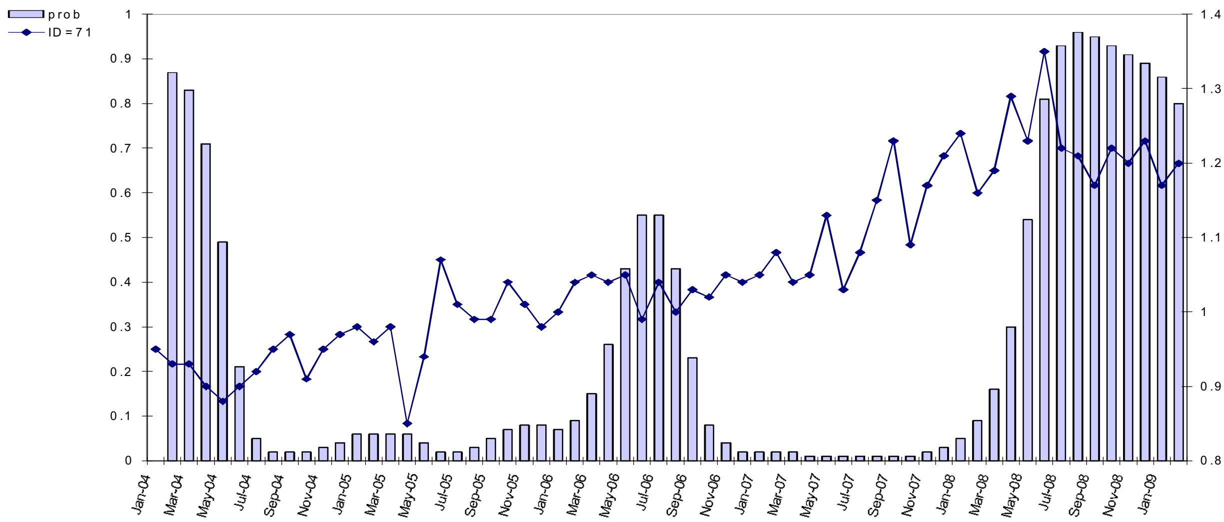

Figure 6.5. Food and drink (ID=71): in-sample probability of a downturn (bars) and seasonally adjusted query index (line).

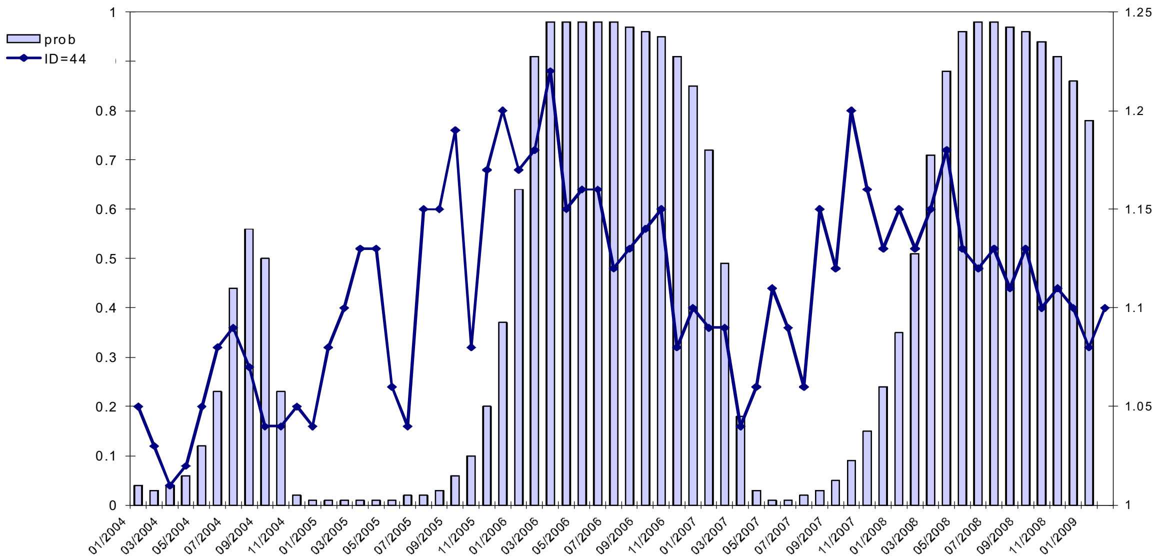

Figure 6.6. Beauty and personal care (ID=44): in-sample probability of a downturn (bars) and seasonally adjusted query index.

Figures 6.1 to 6.6 depict smoothed in-sample probabilities of a downturn for leading categories, alongside corresponding seasonally adjusted query indices, at monthly frequency. As it can be seen, the current downturn is well-captured by all leading indices, synchronized since April 2008. At this point, a composite probability of downturn, computed as the simple average of individual query probabilities, reached a value of 0.6, while the probability of recession, computed by state-of-the-economy index model basing upon retrospective official data, peaks first at 0.77 (Figure 7).

Besides, earlier episodes of decreasing popularity could be recognized in individual probabilities of a downturn, which mostly cancel each out, but which were synchronized in the first half of 2007. The latter coincides with the consumer confidence indices of Globes monthly reports on consumer expectations from May 2006 to January 2007, which diagnosed pessimism and intentions to reduce purchases as a result of economic uncertainty, inflationary pressure, the Second Lebanon War and its aftermath. A short episode of slowing-down economic growth was documented as well by the state-of-the-economy index in March–April 2007.

However, ex post comparisons might be misleading due to data revisions, which affect official and query statistics differently. The main sources of end-point revisions of official data are increased sample size between preliminary and subsequent data processing (industrial production index), repeated trend evaluation and seasonal adjustments when a new observation becomes available, and forecast errors of the state-of-the-economy model while missing current data are inferred by extrapolation. Thus, the cumulative effect of end-point official revisions is larger than that of query data, where only the seasonal component is slightly updated when a new observation is included.

To adjust our analysis for these effects, an ex ante comparison was simulated (Figure 8). On the one hand, a composite probability of downturn is presented, computed from real-time simulations of a Bayesian model for each leading query described above. Smoothed estimates of the individual downturn probabilities were obtained since 2007:12 going one month ahead each time. On the other hand, the first-vintage estimates of the probability of recession, derived from the state-of-the-economy index model have been used for comparison. It can be seen clearly that simulated composite probability obtained through query data would first signal a downturn in April 2008, two months before the first vintage of the probability of recession did. in fact, the reason lies in dramatic revisions of official data, performed repeatedly by Central Bureau of Statistics between April 2008 and September 2008, including downward revisions of trends of industrial production and trade and services revenue indices up to 0.8 percent.

5.3. Monthly projection of the trade and services revenue index rate

A monthly change of the trade and services revenue index is one of six coincident components of the state-of-the-economy index currently used for monitoring. Due to the delay until the actual current figure becomes available, data based on past rates and value added tax (VAT) data, available from the Ministry of Finance, were used in place of the missing data.

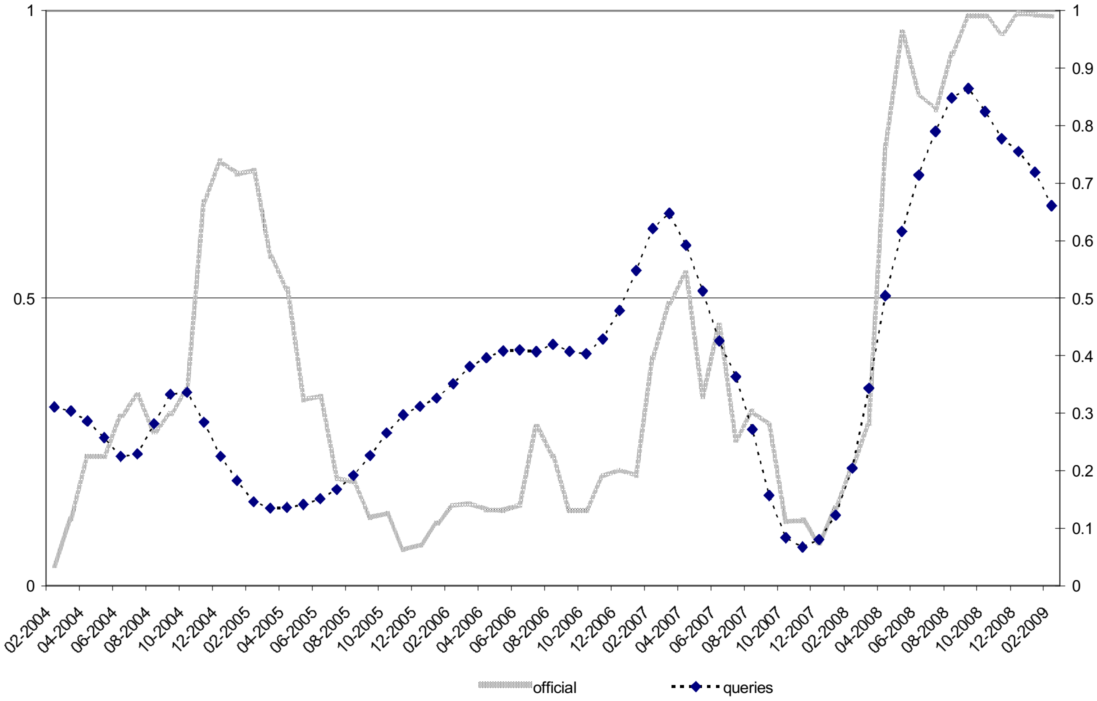

Figure 7. Overall in-sample probabilities of downturn, calculated upon official and query indices.

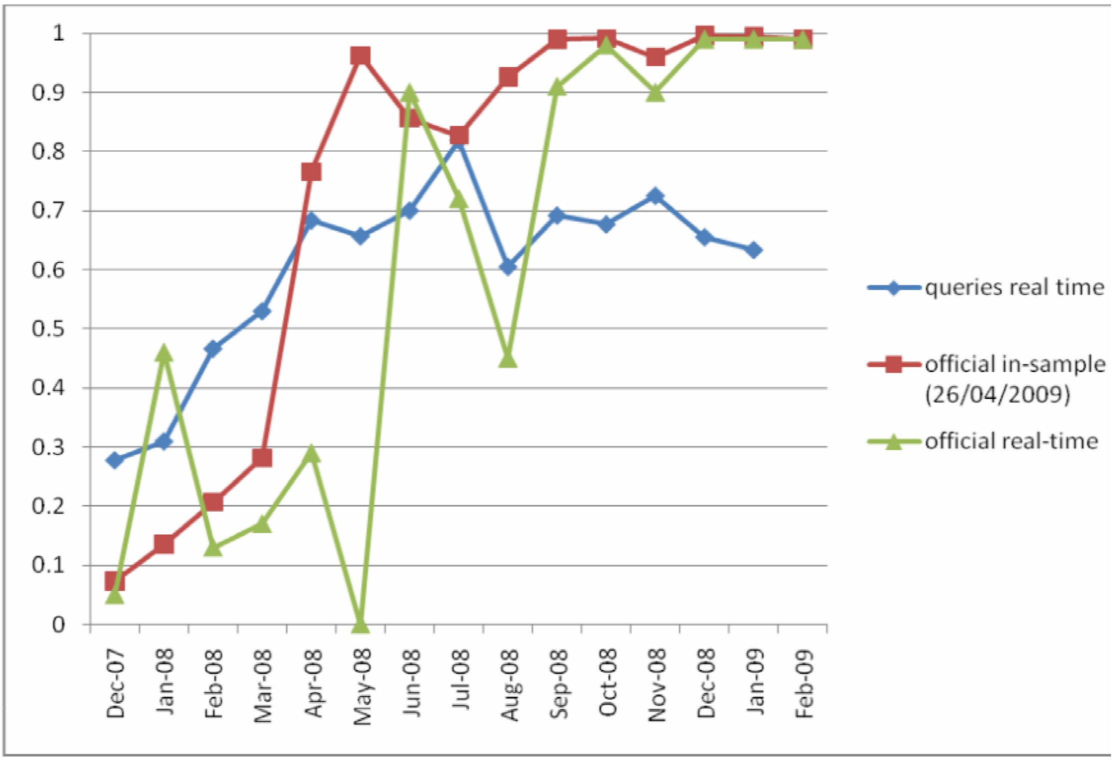

Figure 8. Composite real-time probability of downturn, simulated from query data, alongside first and last vintages of the Probability of Recession, evaluated by the State-of-the-Economy Index routine.

However, recent analysis has shown that retail trade information performs better in terms of mean squared forecast errors than the VAT data, which are subject to a bimonthly cycle and fluctuations due to postponed payments, subsequently smoothed by the CBS data processing.

Usually, no official retail trade index is available when the state-of-the-economy index is computed. Hence, using the query index of home appliances category as an instrumental variable, a monthly predictor of the percentage change in retail (durables) trade may be created which, in turn, refines the revenue component estimate.

Thus, the monthly projection of the trade and services revenue rate has been set as a two-stage least squares model:

$$rt\_dur_t = \alpha_0 + \alpha_1 rt\_subtot_{t-1} + \alpha_2 \Delta x_t^{(ID=271)} + \eta_{rt\_dur,t}$$
$$revenue\_st_t = \rho_1 r_{t-1}^{(trend)} + \rho_2 rt\_dur_t + \rho_3 maam_t + \eta_{r,t}$$

where

rt_dur_t and rt_subtot_t refer to monthly rates (in log-difference terms) of retail trade indices of durables goods and total, respectively[23];

$\Delta x_t^{(ID=271)}$ is the query index of the home appliances category (ID=271), first differenced monthly averages;

$revenue\_st_t$ and $revenue\_st_t^{(trend)}$ refer to the trade and services revenue index and its trend rates,[24] respectively;

$maam_t$ is the VAT monthly rate (in log-difference terms), available

Table 5. Nonlinear 2SLS parameter estimates and residual errors summary for the revenue of trade and services monthly projection model (sample: 2004:1–2009:3)

Dependant variable	Parameter	Estimate[1]	t-value	Pr > | t |	R^2-adj	RMSE
rt_dur_t	α_0	0.0094 (0.0057)	1.65	0.105	0.345	0.043
rt_dur_t	α_1	-0.7087 (0.2273)	-3.12	0.029		
rt_dur_t	α_2	0.6673 (0.1258)	5.31	< 0.0001		
r_t	ρ_1	0.7077 (0.3096)	2.29	0.0260	0.379	0.016
r_t	ρ_2	0.1567 (0.0650)	2.41	0.0192		
r_t	ρ_3	0.0754 (0.0404)	1.91	0.0472		

[1] Approximate standard errors are given into parentheses;

[23] This notation corresponds to that used in Appendix 1; all data used are seasonally adjusted.

[24] Trend figures are derived by the CBS based on the index levels and reported independently each month with a two-month delay; there is a warrant to use the corresponding rate (in log difference terms) as an instrumental variable, alongside query index changes, VAT rates and past rates of overall retail trade index.

contemporaneously from the Ministry of Finance;

$\alpha_0, \alpha_1, \alpha_2$ and ρ_1, ρ_2, ρ_3 are 2SLS parameters; an intercept is approved to be $\rho_0 = 0$;

and $\eta_{rt\_dur,t}$, $\eta_{r,t}$ are random disturbances.

Table 5 reports system parameters and depicts a notably improved fit, compared to a singular equation, which used to be: $revenue\_st_t = \hat{\rho}_1 revenue\_st_{t-1}^{(trend)} + \hat{\rho}_3 maam_t + \hat{\eta}_{r,t}$, having parameters (estimated over the same sample):

$\hat{\rho}_1 =$ 0.843 (0.416)**; $\hat{\rho}_1 =$ 0.093 (0.047)**; $R^2 - adj. =$ 0.107; $RMSE =$ 0.020.

5.4. Monthly projection of unemployment rate

The query index of the human resources (recruitment and staffing, ID=330) category is stationary over the time span from 2004:1 till 2009:2, according to Augmented Dickey-Fuller test results (table 2). At quarterly frequency, this is a leading index of the percentage change in job openings, currently surveyed by the Ministry of Industry, Trade and Labor[25] (Appendix 3, Table A3.1). In turn, the job openings ratio is negatively correlated with a subsequent unemployment rate[26] (Figure 2.1).

These relationships are likely to tackle the incompleteness of unemployment data available to policymakers during monthly economic monitoring.

Basically, the Labor Force Survey of CBS provides quarterly data. In addition, preliminary trend data of the unemployment rate are released at monthly frequency with a two-month delay.

A monthly application might be thought to incorporate the query index of human resources (recruitment and staffing), as well as the job openings ratio, derived from the Ministry of Industry, Trade and Labor Survey in a monthly "nowcasting" of the unemployment rate

In the context of the problem, this implies the following steps.

First, an interpolation of a quarterly variable with a monthly indicator, developed in (Mitchel et al., 2005) is applicable. This approach, written in a state-space form, yields monthly interpolands of a quarterly observed variable, namely, the job openings ratio, regressed by a monthly explanatory indicator, i.e., the query index of human resources (recruitment and staffing, ID=330).

Next, as soon as a leading explanatory monthly indicator of the job openings ratio is enabled, a linear projection of the monthly unemployment trend, recently available from the CBS, is completed, using ARIMA (p=(2), q=(2)[27]) specification and assuming that the labor force participation rate remains at its past known level.

---

[25] See also quarterly periodical overview "Demand for Labor in business sector" of the Ministry of Industry, Trade and Labor Research Division (available in Hebrew).

[26] Appendix 3 presents an example of an OLS regression, estimated at quarterly frequency through the sample 1998:I – 2008:IV.

[27] A moving average term is included only at the lag where the unemployment trend estimate is still available.

Table 6.1. Parameter of the state-space monthly interpolation of the job openings ratio using the query index of human resources (recruiting and staffing, ID – 330) (sample: 2004:1–2009:3)

Component	Parameter	Estimate[1]	t-value	Pr > | t |
Irregular	Error variance	0.28736 (0.0057)	5.43	< 0.0001
$\tilde{x}_t^{(ID=330)}$	Coefficient	0.32263 (0.06030)	5.35	< 0.0001
Quarterly lag[2]	AR(3) coefficient	0.84209 (0.07383)	11.41	< 0.0001
Fit statistics based on Residuals				
R^2-*adj*	0.627			
RMSE	0.545			

Table 6.2. Parameters of the ARIMA (p=(2),q=(2)) monthly projection of unemployment rate, using monthly interpoland of job openings ratio (sample: 2004:1–2009:3)

Parameter	Lag	Estimate[1]	t-value	Pr > | t |
Intercept	0	8.85191 (1.51394)	5.43	< 0.0001
AR	2	-0.78537 (0.09691)	5.35	< 0.0001
MA	2	0.98704 (0.01929)	51.16	< 0.0001
Monthly exogenous variable[3]	0	-0.11156 (0.05703)	-1.96	0.0501
Fit statistics based on Residuals				
R^2-*adj*	0.917			
RMSE	0.392			

[1] Approximate standard errors are given into parentheses;
[2] Assumed corresponding to the third monthly lag
[3] Monthly interpoland of job openings ratio

Table 6.1 presents final estimates of the free parameters of the state-space model, enabling monthly updates of the job openings ratio until the last corresponding query index is available.

Table 6.2 presents parameters of the monthly projection of the unemployment rate monthly trend, using monthly interpoland of job openings ratio obtained in the previous step.

6. Conclusions

Israeli data support the hypothesis that Google query indices may be helpful in drawing inferences about the state of current economic growth, given the fact that official data are released with a delay.

Six leading query categories were selected, namely human resources (recruiting and staffing), home appliances, travel, real estate, food and drink and beauty and personal care. Granger causality tests provide evidence that corresponding indices contain cyclical components which conform with cycles of economic growth.

The most predictive category is human resource (recruiting and staffing). A corresponding index may be applied for a monthly linear projection of the unemployment rate, whereas the current monthly assessments may be subject to a large bias due to lags in the data.

A large fraction of innovation variance, explained by query categories which satisfy pair-wise Granger causality, lies in the employment rate. It supports attitudinal interpretation of the queries effect in the sense of next-term consumer confidence.

According to Augmented Dickey-Fuller tests, the dynamics of query indices may be non-stationary. Downward trends detected in the automotive, industries, computer and electronics categories support the hypothesis that social search and knowledge-oriented documentation compete with Google searches. For short-term predictions, first-differencing of query indices facilitates the analysis.

Weekly frequency of Google's application is useful in real-time monthly monitoring, enabling query indices to precede official data by two months, by accounting for the first two weeks' information. This fact has been exploited in the treatment of the end-point problem to obtain smoothed/robust estimates for the relevant month.

Monthly changes of popularity, derived from smoothed query indices, may be modeled with a shifting mean, as suggested by the Bayesian two-state model. The lower state, when identified through the queries co-movement, helps to assess a slowdown in economic growth.

There is perhaps a problem of the varying predictive ability of query indices, which requires closer attention as the time span extends. In particular, a weighting scheme for individual probabilities of a downturn, derived from separate query indices, is worth further investigation.

Bibliography

Braude, Y., and Fiedman, A. (2005) Consumer Confidence and Private Consumption, Bank of Israel Survey No 78 (Hebrew)

Chauvet, M. and Hamilton, J. (2005) Dating Business Cycle Turning Points - NBER Working Paper No.11422

Choi, H. and Varian, H. (2008) Predicting the Present with Google Trends – Draft, November 19, 2008.

Finkel, Y. and Yiftach, M. (2006) E-commerce in the Israeli CPI – ECE/ILO Meeting on consumer Price Indices, May 10-12, 2006, Geneva

Finkel, Y. and Yiftach, M. (2003) E-commerce in the Israeli CPI – Seventh Meeting of the International Working group on Consumer price indices, Paris 2003.

Hamilton J. (1989), “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle” , Econometrica , 57 (2), 357-384.

Haugh, D., L. (2005) The influence of consumer confidence and stock prices on the US business cycle – CAMA Working Paper No 3/2005, Australian National University.

Ludvigson, S. (2004) Consumer confidence and consumer spending – Journal of Economic Perspectives, 18/2

Marom., A., Menashe, Y. and Suchoy, T. (2003) The State-of-the-Economy Index and the Probability of Recession: The Markov Regime-Switching Model – Discussion Paper No 2003.05, Bank of Israel

Mitchell, J., Smith, R.J., Weale, M.R., Wright, S. and Salazar, E.L. (2005) An indicator of monthly GDP and an early estimate of quarterly GDP growth. – The Economic Journal, V.115 (February), p.p.F108-F129

Kim C.J. and Nelson C.R. (1999) State-Space Models with Regime Switching. Cambridge, Mass.

Morwitz, V. et al. (2006) When do Purchase Intensions Predict Sales? – New York University, Working Paper Series.

Mostaghimi, M., M. (1996) Combining Ranked Mean Value-Forecasts – European Journal of Operational Research, No 94, 1996

Mostaghimi, M., M. (2008) Bayesian Probability Forecast of a 2007-2008 Economic Recession in the U.S. – http://ssrn.com/abstract=1138371

Pickering, J.F. (1984) Purchase expectations and the demand for consumer durables. - Journal of Economic Psychology 5 (1984), 341-352

Smith Institute's Consumer Index of Israel, 2006

Spink, A. and Jansen, B.J. (2004) A study of Web Search Trends – Webology, Vol.1, No 2, December, 2004

Stock, J.H. and Watson, M.W. (1989) New indices of coincident and leading economic indicators – NBER Macroeconomics Annual, p.p.351-94

רוני בר צורי (2007), דפוסי קניה מרחוק – מנהל מחקר וכלכלה, משרד התעשייה והמסחר, פברואר 2007.

Appendix 1. Reference series

Reference subject	Fame Database code	Reference series
Revenue of commerce and services	*revenue_st*	Total - Commerce and services
Revenue of commerce and services	*revenue_e*	Wholesale and retail trade, and repairs
Revenue of commerce and services	*revenue_f*	Accommodation services and restaurants
Revenue of commerce and services	*revenue_i*	Business activities
Revenue of commerce and services	*revenue_j*	Banking, insurance and other Financial institutions
Revenue of commerce and services	*revenue_k*	Education
Revenue of commerce and services	*revenue_m*	Community, social, personal and other services
Revenue of commerce and services	*credit_cards*	Private Consumers Credit Card Purchases
Retail trade	*rt_subtot*	Total excl. gas, fertilizers and petroleum
Retail trade	*rt_hokit*	Kitchen and house accessories
Retail trade	*rt_dur*	Durables
Retail trade	*rt_ misce*	Consumption goods - miscellaneous
Retail trade	*rt_textil*	Textile and clothing
Industrial production	*tpr_ high*	High technology
Industrial production	*tpr_ medium _high*	Medium-high technology
Industrial production	*tpr 29_31*	Basic metals and products
Industrial production	*tpr 27_29*	Metal and machinery
Industrial production	*tpr32*	Electronic components
Industrial production	*tpr 32T34*	Electronic equipment
Industrial production	*tpr 34*	Industrial equipment for control and supervision
Industrial production	*tpr 31_34*	Electronic motors, components etc.
Industrial production	*tpr 33*	Electronic communication equipment
Industrial production	*tpr 35*	Transport equipment
Employment posts	*ep_bs*	Total - business sector
Employment posts	*ep_i*	Business activities
Employment posts	*tmg*	Industry
Employment posts	*tmg_ medium _high*	Industry - Medium-high technology
Consumer imports	*im_c*	Total
Consumer imports	*im_c_dur*	Durables
Consumer imports	*im_c_veh*	Transport vehicles
Export of services	*ex_serv*	Total
Export of services	*ex_serv_tour*	Travel services
Construction	*Sup_flt*	Dwellings supply - private sector

Appendix 2. Google query categories

ID	Category	Subcategory
47	Automotive	ND
170	Automotive	Vehicle Licensing & Registration
468	Automotive	Auto Financing
473	Automotive	Vehicle Shopping
815	Automotive	Vehicle Brands
89	Automotive	Auto Parts
25	Business	Advertising & Marketing
157	Business	Human Resources
330	Business	Human Resources (Recruiting & Staffing)
11	Home & Garden	ND
270	Home & Garden	Home Furnishings
271	Home & Garden	Home Appliances
44	Beauty & Personal Care	ND
71	Food & Drink	ND
67	Travel	ND
179	Travel	Hotels & Accommodations
203	Travel	Air Travel
29	Real Estate	ND
18	Shopping	ND
70	Shopping	Flowers Gifts & Greetings
73	Shopping	Mass Merchants & Department Stores
696	Shopping	Luxury Goods
48	Industries	Construction & Maintenance
650	Industries	Construction & Maintenance (Building Materials)
5	Computers & Electronics	ND
30	Computers & Electronics	Hardware
78	Computers & Electronics	Consumer Electronics
434	Computers & Electronics	Electronics & Electrical
7	Finance & Insurance	ND
107	Finance & Insurance	Investing
279	Finance & Insurance	Credit & Lending

Appendix 3. OLS regression of unemployment rate by the job openings ratio (quarterly seasonally adjusted data, sample: 1998:I - :2008:Iv)

Quarterly projection of unemployment rate is as follows:

$$u_q = \gamma_0 + \gamma_1 u_{q-1} + \gamma_2 l_{q-1} + \gamma_3 \frac{V_{q-1}}{E_{q-1}} + \varsigma_q\,, \quad \varsigma_q \sim IID(0, \sigma_\varsigma^2)$$

where u_q is an unemployment rate at a quarterly frequency (CBS, Labor Force Survey);

l_q - a labor force participation rate (CBS, Labor Force Survey);

$\frac{V_q}{E_q}$ - job openings ratio (Ministry of Industry, Trade and Labor quarterly Survey).

$\gamma_0, \gamma_1, \gamma_2, \gamma_3$ - parameters, evaluated as (standard deviations into parentheses) :

$$\gamma_0 = 11.78 \quad (3.72)***$$
$$\gamma_1 = 0.84 \quad (0.06)***$$
$$\gamma_2 = -0.18 \quad (0.06)***$$
$$\gamma_3 = -0.25 \quad (0.09)**$$
$$R^2_{adj} = 0.93$$
$$DW = 1.80$$

Table A3.1. OLS regression of job openings[1] by query index of Human Resources (Recruiting & Staffing, ID=330) [2], sample: 2004:I – 2008:IV

Parameter	Value[3]
Constant	-28.92 (9.66) **
Coefficient	1.60 (0.44) **
DW	1.92
R^2-adj	0.39

[1] Percent quarterly change, seasonally adjusted
[2] One-quarter lead, seasonally adjusted
[3] Standard deviations are into parentheses;
** 95% significance level