

Ekonomisk kommentar

Webbinsamlade prisuppgifter och kortsiktiga inflationsprognoser

Isaiah Hull, Mårten Löf och Markus Tibblin

Författarna är verksamma vid Riksbankens avdelning för penningpolitik.¹

Kortsiktiga inflationsprognoser är viktiga för centralbankerna

Riksbanken har, precis som många andra centralbanker, som mål att upprätthålla prisstabilitet. Detta tolkas som att inflationen, mätt enligt konsumentprisindex (KPI), ska hållas låg och stabil.² KPI, som baseras på en korg av varor och tjänster som hushållen brukar konsumera, omfattar ett antal underkategorier. I Sverige är priserna på frukt och grönsaker, flygresor och bränsle exempel på underkategorier som ofta varierar kraftigt. Denna volatilitet kan ibland göra det svårt att förutse förändringar i KPI på kort sikt (de närmaste månaderna), trots att dessa kategorier tillsammans utgör endast en liten andel av det totala KPI. Man kan tycka att kortsiktiga prognosfel inte borde vara något större problem, eftersom penningpolitiska beslut baseras på inflationsutvecklingen sett över en längre period. Men för att kunna göra tillförlitliga prognoser över den framtida inflationen måste man ha en god kännedom om den nuvarande inflationen och en bra träffsäkerhet när det gäller inflationsutvecklingen på kort sikt. För såväl prognosarbete som penningpolitiskt beslutsfattande är det viktigt att kunna avgöra om ett prognosfel beror på tillfälliga eller mer permanenta faktorer. Eftersom inflationstakten uttrycks i årsförändring kan ett stort prognosfel, som antas bero på permanenta faktorer, påverka inflationsbanan de kommande tolv månaderna förutsatt att inga andra förändringar antas ske (se diagram 1).

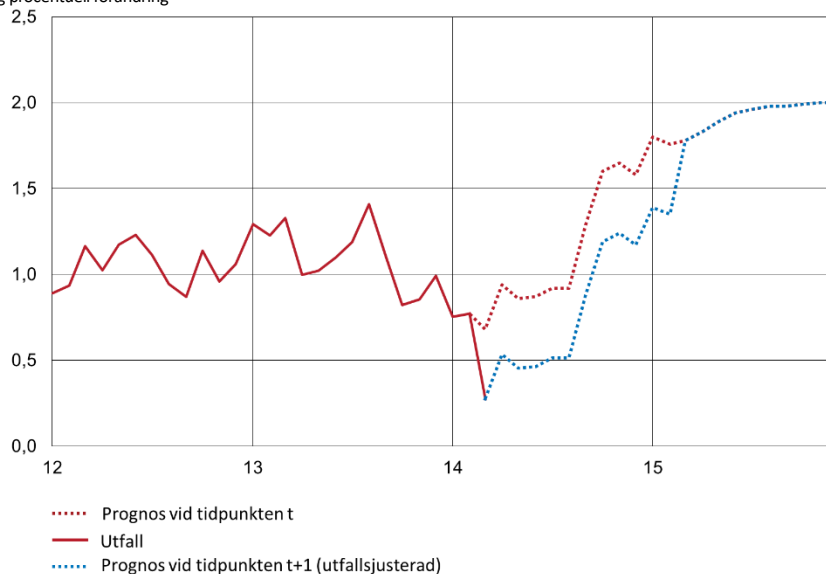
Prognoser för den kortsiktiga inflationsutvecklingen är viktiga för en centralbank. Vissa inslag i de publicerade inflationsciffrorna är volatila och därför svåra att förutse. Priser på frukt och grönsaker, energi och flygresor är exempel på sådana produktgrupper. En automatisk process har utvecklats som dagligen samlar in uppgifter från internet om priser på vissa frukter och grönsaker. Resultaten här tyder på att dessa uppgifter kan ge mer träffsäkra prognoser på kort sikt.

¹ Vi vill gärna tacka kollegor på Riksbanken för synpunkter på tidigare utkast. De åsikter som uttrycks i denna kommentar är våra egna och skall inte uppfattas som Riksbankens ståndpunkt.

² <http://www.riksbank.se/sv/Monetary-policy/>

Diagram 1. Inflation utfall och prognosrevidering till följd av ett prognosfel på kort sikt

Årlig procentuell förändring



Anm. Diagrammet illustrerar hur ett prognosfel vid tid t (skillnaden mellan den heldragna och den streckade röda linjen) kan påverka den nya prognosbanan för inflationen under det kommande året (blå streckad linje).

Ny data och nya analysmetoder

Mängden information som publiceras på internet ökar mycket snabbt. Det har diskuterats flitigt hur centralbanker ska tillämpa nya metoder för att samla in och analysera den här typen av uppgifter på bästa sätt. I september 2015 anordnade Riksbanken en workshop om Big data, där ett antal centralbanker, forskare och privata företag visade exempel på hur sådan information kan användas till stöd för centralbankernas analyser och beslut.³ Ett område som centralbankerna intresserar sig allt mer för är den ökade tillgången på information om försäljningspriser på internet. Ledande forskning inom området har bedrivits inom projektet "Billion Prices" som genomförts på initiativ av Cavallo och Rigobon. De har i ett antal undersökningar kunnat visa att prisindex som konstruerats med hjälp av uppgifter inhämtade från internet följer de officiella mätningarna av konsumentprisindex, t.ex. KPI, ganska väl (se Cavallo och Rigobon, 2014). Samtal med handelsföretag i Riksbankens företagsundersökning visar också att skillnaderna är förhållandevis små mellan priserna i fysisk butik och priserna som publiceras på internet i Sverige.⁴ Detta resultat stämmer överens med en stor internationell jämförelse som utförts av Cavallo (2016). Det har också visats att prisuppgifter från internet går bra att använda för att prognostisera den officiella KPI. I vissa länder ger det bättre resultat än modeller som inkluderar prisuppgifter från fysiska butiker (Aparicio och Bertolotto, 2016).

Pilotprojektet – insamling av priser på frukt och grönsaker från internet

Om centralbankerna använde prisuppgifter från internet i prognosarbetet skulle några problem med kortsiktiga inflationsprognoser lösas. För det första uppstår det ingen tidsfördröjning i insamlingen, eftersom informationen kan samlas in i realtid, dvs. den finns

³ <http://www.riksbank.se/sv/Press-and-published/Notices/2015/The-Riksbank-organises-a-workshop-on-big-data/>

⁴ http://www.riksbank.se/Documents/Rapporter/Foretagsintervjuer/2016/rap_foretagsundersokning_160615_sve.pdf

tillgänglig innan det officiella KPI publiceras. Med hjälp av internet går det dessutom att samla in prisuppgifter på väldigt detaljerad nivå, vilket betyder att det går att göra mer disaggregerade analyser. Det skulle till exempel kunna göra det enklare att avgöra om en ovanlig prisutveckling är tillfällig eller inte. Eftersom insamlingen av uppgifter på internet kan automatiseras kan man även producera dagliga prisindex och inkludera dessa i modellerna till låga kostnader och med mycket begränsade resurser.

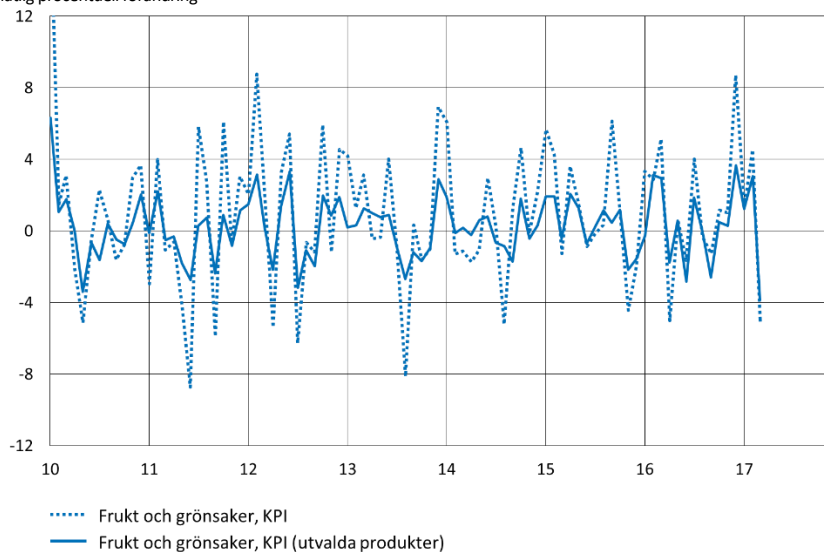
En mindre pilotstudie inleddes för att undersöka om priser på frukt och grönsaker, som är tillgängliga på internet, kan användas för att öka träffsäkerheten för kortsiktiga inflationsprognoser i Sverige. Frukt- och grönsakspriserna utgör endast 3 procent av KPI-korgen, men kraftiga prisrörelser i detta delindex kan få tydliga effekter på de aggregerade siffrorna. Syftet med pilotprojektet var följande:

1. Skapa en automatisk process för att dagligen samla in prisuppgifter på internet för några utvalda frukter och grönsaker från svenska handlare som bedriver e-handel. Internetpriser har samlats in för apelsiner, bananer, paprikor, äpplen, gurkor, vitkål, vindruvor, blomkål, päron, purjolök och tomater. Korrelationen mellan delmängden utvalda frukt- och grönsakspriser och det totala indexet för frukt och grönsaker i KPI är förhållandevis hög (se diagram 2). Därför antogs det också att internetuppgifter för dessa utvalda produkter skulle vara användbara för att göra prognoser på delaggregatet frukt och grönt i KPI.
2. Skapa ett viktat månadsprisindex baserat på prisinformation från internet för de utvalda frukter och grönsaker som anges i punkt 1.
3. Testa om det konstruerade indexet med internetpriser genererar mer träffsäkra prognoser än Riksbankens befintliga modeller för priser på frukt och grönt.

Resten av kommentaren är disponerad på följande sätt: Först ger vi en kortfattad beskrivning av den metod som har använts för att samla in uppgifterna och sedan beskriver vi hur vi konstruerar de index som används. Slutligen diskuterar vi resultatet och presenterar några övergripande slutsatser från pilotprojektet.

Diagram 2. Priser på utvalda produkter i KPI och aggregatet av frukt och grönsaker i KPI

Månatlig procentuell förändring



Skrapning

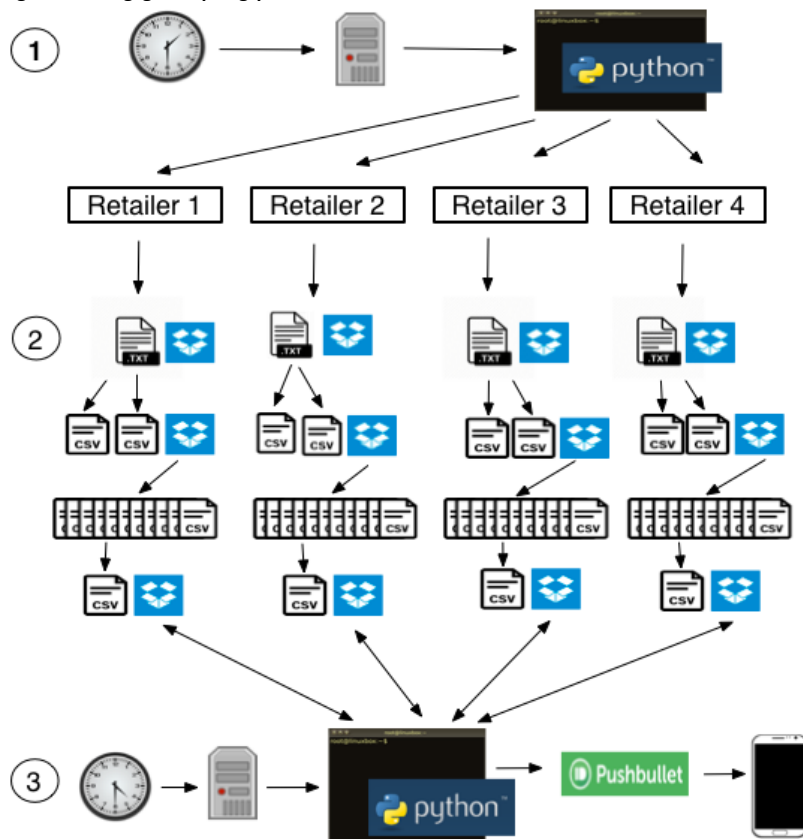
För att samla in prisuppgifterna i det här projektet använder vi en teknik som kallas skrapning. Den innebär att programmatiska sökfrågor skickas till servern för en webbplats. Servern svarar genom att skicka tillbaka den underliggande kod som körs i en webbläsare när en internetanvändare besöker en viss sida på webbplatsen. Kodens struktur kan sedan användas för att identifiera den information man vill få fram, t.ex. produktnamn och priser. Vi använder ytterligare verktyg för att schemalägga skrapningarna så att de utförs automatiskt på de berörda webbplatserna vid samma tidpunkt varje dag. Vi begränsar också hastigheten på förfrågningarna till varje webbplats för att inte belasta serverna.

Alla aktiviteter utförs på en virtuell server (Linus VPS). Servern kör tre skript i följd vid samma tidpunkt varje dag (se figur 1). Det första skriptet besöker livsmedelshandlars webbplats. Skriptet hämtar kod från alla sidor som behandlar priser på frukt och grönsaker. Från koden identifieras en lista över alla produktpriser och produktnamn och denna sparas i en .csv-fil för varje plats och dag. Den obearbetade koden sparas också i .txt-format i 90 dagar. På så sätt kan fel som upptäcks i efterhand korrigeras. Därefter identifierar skriptet mönster i texten, för att filtrera uppgifterna ytterligare så att nya .csv-filer som enbart innehåller information om de utvalda produkterna kan skapas.

När skrapningen är klar kör servern igång ett nytt skript, som fogar samman de nya uppgifterna med historiska data. Först identifierar skriptet matchningar mellan den nya informationen och specifika produkter som har identifierats vid tidigare skrapningar. Produkter som inte kan matchas tilldelas nya ID-nummer, och produkter som redan har ID-nummer slås ihop med redan befintliga tidsserier.

Slutligen kör servern igång det tredje skriptet som söker efter fel. Skriptet identifierar antalet filer som har producerats, storleken på filerna och de uppgifter som har använts. Därefter skickar servern en rapport via sms, vilket gör att eventuella fel snabbt kan identifieras och korrigeras.

Figur 1 Den dagliga skrapningsprocessen



Sammanställa index

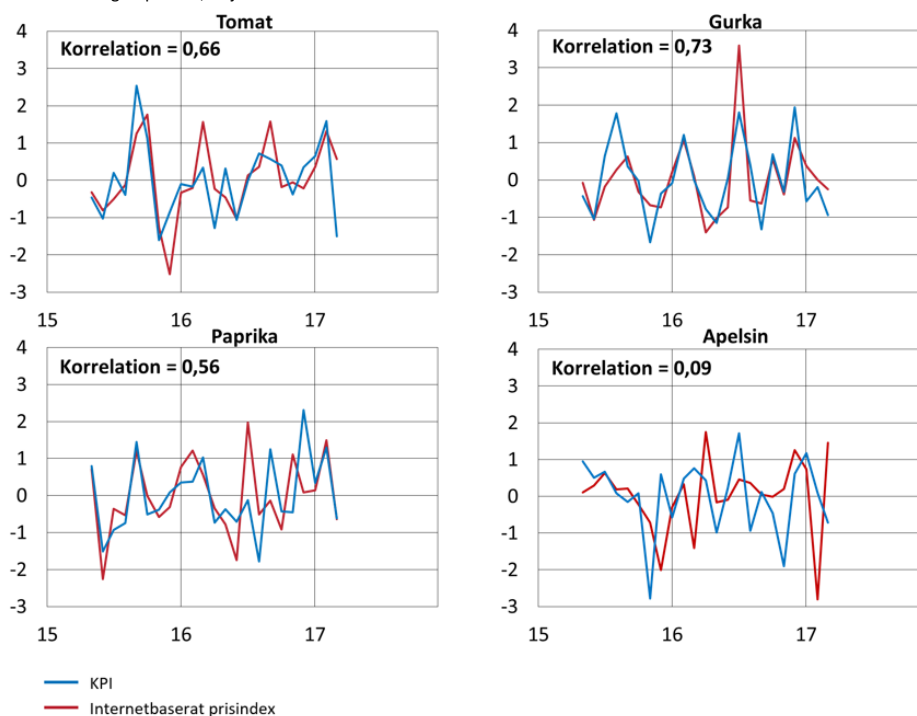
Först sorteras de insamlade internetpriserna in i olika grupper – apelsinpriser för sig, äppelpriser för sig och så vidare. Därefter anpassas data för att motsvara mätveckorna i den svenska KPI-undersökningen.⁵ I nästa steg beräknas ett geometriskt medelvärde för varje produkt och månad, dvs. ett genomsnittspris för apelsiner i januari och ett för februari, och så vidare.

I diagram 3 visas den procentuella månadsförändringen av priserna på apelsiner, gurkor, paprikor och tomater från maj 2015 till mars 2017. De blå linjerna visar priserna enligt KPI, medan de röda linjerna visar motsvarande förändringar baserat på uppgifter från internet. Korrelationen mellan KPI-priserna och de priser som finns tillgängliga på nätet är förhållandevis hög, och allra högst för priset på gurkor och tomater. Lägst korrelation uppmäts för priser på apelsiner.

⁵ Prisuppgifter för frukt och grönsaker i KPI samlades in under tre veckor i mitten av varje månad (sammanlagt 21 dagar) fram till och med december 2016. Därefter har Statistiska Centralbyrån gjorts vissa förändringar.

Diagram 3. Prisförändringar enligt internetdata och motsvarande prisförändringar enligt KPI

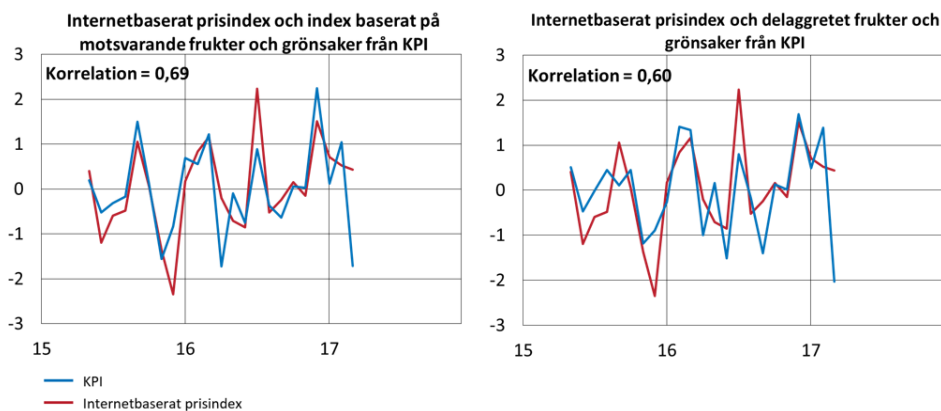
Månadsförändringar i procent, maj 2015–mars 2017



Anm. Diagrammen visar några exempel på jämförelser mellan index baserat på insamlad data från internet (röd linje) och motsvarande data från KPI (blå linje).

Slutligen aggregeras prisförändringarna som har samlats in från internet till ett sammanfattande index (nedan kallat *internetbaserat prisindex*) med hjälp av de vikter som används i KPI. I den vänstra figuren i diagram 4 jämförs det indexet (röd linje) med ett KPI-index för motsvarande frukt- och grönsakspriser (blå linje). I figuren till höger visas samma internetbaserade prisindex tillsammans med det totala prisindexet för frukt och grönsaker i KPI (dvs. inte enbart för de frukter och grönsaker som omfattas av pilotstudien). Av naturliga skäl sjunker korrelationen när det internetbaserade prisindexet jämförs med det totala indexet för frukt och grönsaker (se den högra figuren i diagram 4). Korrelationen är dock fortfarande relativt hög och det internetbaserade prisindexet fångar upp det mesta av variationen i det officiella frukt- och grönsaksindexet.

Diagram 4. Jämförelse mellan ett aggregerat index baserat på internetdata och två olika index med KPI-data
Månadsförändringar i procent av priser, maj 2015–mars 2017.



Anm. Diagrammen visar aggregerade prisförändringar enligt data från internet (röd linje i båda diagrammen). Blå linjer visar prisförändringar enligt KPI. Den blå linjen i det vänstra diagrammet visar prisförändringar då samma frukter och grönsaker beaktas som internetinsamlingen. Den blå linjen i det högra diagrammet visar månadsförändringar enligt hela delaggregatet frukt och grönt.

Prognosutvärdering

I det här avsnittet undersöks mer formellt om det går att använda internetpriser för att göra prognoser på frukt- och grönsakspriser i KPI.

Här jämförs prognoser från Riksbankens nuvarande modeller för priser på frukt och grönsaker med de prognoser som baseras på internetpriser. De modeller som Riksbanken använder i dag för prognoser på kort sikt beaktar ett stort antal indikatorer. Modellerna för frukt- och grönsakspriser inkluderar en indikator åt gången tillsammans med ett glidande medelvärde av växelkursindex (KIX). Dessutom ingår tidsfördröjda värden av den beroende variabeln, dummys för extrema värden och i vissa fall även glidande medelvärdestermerna (MA-termerna). I jämförelsen nedan används en medelvärdesprognos från alla dessa specifikationer och den metoden förkortas MEAN. Principalkomponentsanalys används också för att sammanfatta informationen från indikatorerna till ett index i steg 1. Ett eller flera sådana sammanfattande index ingår sedan i stället för de individuella indikatorerna i modellerna. Medelvärdesprognosen från dessa specifikationer kallas PC nedan. Dessa modellbaserade prognoser (MEAN och PC) jämförs med en metod där det internetbaserade prisindexet användes direkt som en prognos på prisaggregatet för frukt och grönsaker i KPI. Den metoden kallas OP.

Utvärderingsperioden är maj 2015 till mars 2017. Roten ur medelkvadratfelet, RMSE, används för att jämföra de olika metodernas träffsäkerhet. RMSE sammanfattar standardavvikelsen i prognosfelen och graden av systematik i felen (bias). Ju lägre RMSE, desto bättre prognosförmåga. En prognos som alltid är korrekt har ett RMSE lika med noll.

Resultatet sammanfattas i tabell 1. Under utvärderingsperioden har det internetbaserade prisindexet, gett bäst resultat. Den prisinformationen verkar alltså förbättra prognosförmågan för den allra närmaste tiden. Utvärderingsperioden är dock mycket kort och det krävs längre tidsserier för att man skall kunna dra tydligare slutsatser.

Tabell 1. RMSE för olika metoder, maj 2015–mars 2017.

Horizon	Nowcasting procedures		Online prices
	Mean	PC	OP
1 month	2.0	2.0	1.6

Diskussion

Arbetet med att utveckla processer för automatisk insamling (skrapning) och analys av internetbaserade prisuppgifter har krävt ett visst mått av trial-and-error. Den utvecklade processen måste vara skraddarsydd för att kunna hantera en rad olika och föränderliga webbplatser. Det har dock gått att bygga program som kan hantera dessa variationer, vilket gör att insamlingsprocessen och indexberäkningarna nu i princip är helt automatiserade.

Arbetet med att analysera internetbaserade mikrouppgifter har precis börjat i och med det här pilotprojektet. En del tid har även ägnats åt att undersöka om ytterligare bearbetning av de internetbaserade uppgifterna skulle kunna öka prognosförmågan. Exempelvis konstruerades ett index som baserades på internetpriser som historiskt uppvisat en hög korrelation med motsvarande officiella KPI-priser. Prisinformationen har också sammanfattats med hjälp av principalkomponentsanalys. Under kortare perioder har dessa index gett mycket goda prognosresultat, men generellt har de inte varit lika tillförlitliga som den enkla metoden vi beskriver ovan (OP). Uppgifterna bearbetas alltså inte för närvarande, men det är något som skulle kunna undersökas vidare.

De insamlade uppgifterna skulle också kunna analyseras med avseende på företagens prissättningsbeteende. Detaljerade mikrouppgifter kan vara en källa till att förstå när, hur och varför ett företag ändrar sina priser. Detta skulle också kunna undersökas vidare.

Slutsatser

Pilotprojektet har gett ett antal insikter om internetbaserad prisinsamling och analys av sådana uppgifter med inriktning mot prognoser.

För det första har projektet visat att det går att samla in internetbaserade försäljningspriser från e-handelsplatser som drivs av handlare i Sverige. Att skapa de skript och it-processer som krävs för en stabil uppgiftsinsamling tar visserligen tid och kräver programmeringskompetens som normalt sett inte brukar finnas hos centralbanker. Men när metoderna väl har införts krävs ett minimum av underhåll och utveckling.

Dessutom har de insamlade webbaserade uppgifterna lett till nya typer av utmaningar som skiljer sig från den vanliga analysen av och prognosmetoderna för inflation. Det krävs alltså mer analys inom det här området.

Resultatet hittills pekar på att internetbaserade prissättningsuppgifter tillför ett visst värde när man gör prisprognoser på frukt och grönsaker. Resultatet tyder också på att det kan finnas utrymme för att utöka insamlingen av priser som finns tillgängliga på internet och använda dem i kortsiktiga prognosmodeller. Det finns dock fortfarande många frågor som bör utredas, till exempel hur informationen kan användas på bästa sätt.

Källförteckning

Aparicio, Diego och Manuel Bertolotto (2016), "Forecasting Inflation with Online Prices", Working Paper – MIT.

Bernanke, B. S. och Boivin, J. (2003), "Monetary policy in a data-rich environment", *Journal of Monetary Economics* 50 (3), 525–546.

Bertolotto, M., Cavallo, A. och Rigobon, R. (2014), "Using Online Prices to Anticipate Official CPI Inflation", UTokyo Price Project Working Paper Series 031, University of Tokyo, Graduate School of Economics.

Cavallo, F. Alberto, (2016), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", NBER Working Papers 22142, National Bureau of Economic Research, Inc.

Stock, J. H. och Watson, M.W., (2002), "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association* 97:460, 1167–1179.

Stock, J. H. och M. W. Watson, (2004), "Combination forecasts of output growth in a seven-country data set", *Journal of Forecasting* 23 (Issue 6), 405–430.

