

बिग डेटा तकनीकों से आवास मूल्य सूचकांक का संकलन*

इस अध्ययन का लक्ष्य बिग डेटा के जरिये भारत में आवासीय संपत्ति की कीमतों का एक वैकल्पिक सूचकांक प्राप्त करना है। आवासीय मूल्य सूचकांक तैयार करने के लिए चुनिंदा वेब पोर्टलों से आवासीय संपत्ति की कीमतों पर आँकड़े लिए गए हैं और बिग डेटा के औजारों से उनको प्रोसेस किया गया है। यह सूचकांक प्रत्येक तिमाही समाप्त होते ही लगभग तुरंत बन सकता है। यह पाया गया कि वेब आँकड़ों पर आधारित आवासीय घरों की कीमतों और पंजीकरण आँकड़ों (सूचकांक व वृद्धि दोनों) में अहम सहसंबंध हैं, यद्यपि कीमतों के स्तर में अंतर है। पंजीकरण डेटा और वेब डेटा पर आधारित सूचकांकों में दोतरफा संबंध है।

आवास मूल्य अर्थव्यवस्था का एक महत्वपूर्ण सूचक है। सकल देशी उत्पाद (जीडीपी), उपभोग, निवेश, मुद्रास्फीति, चालू खाता शेष, और उत्पादन अंतराल के साथ इसकी सह-गति है (अहियर्ने व अन्य, 2005)। विश्व भर में अधिकांश हाउसहोल्ड्स के लिए आवासीय संपत्ति सबसे बड़ी एकमात्र आस्ति है। आवासीय संपत्ति की कीमत में अंतर हाउसहोल्ड्स की दीर्घावधि निवेश रणनीति और उनके खर्च व उधार के पैटर्न को प्रभावित करते हैं। बैंक उधार व बंधक के रास्ते संपत्ति कीमत में परिवर्तन अर्थव्यवस्था के बैंकिंग व वित्तीय क्षेत्रों पर असर डालते हैं।

इस दृष्टि से नीति निर्माताओं, विशेषतः केंद्रीय बैंक के लिए आवासीय संपत्ति मूल्य एक महत्वपूर्ण सूचना है। भारतीय संदर्भ में दो स्थापित आवास मूल्य सूचकांक हैं, यथा, (ए) भारतीय रिज़र्व बैंक द्वारा तैयार आवास मूल्य सूचकांक (एचपीआई) और (बी) रेसिडेक्स जो राष्ट्रीय आवास बैंक (एनएचबी)¹ संकलित

* यह आलेख संख्यिकी और सूचना प्रबंध विभाग के बिग डेटा एनलिटिक्स डिविज़न के इंद्रनील गायेन और शशांक शेखर माइती ने तैयार किया है। वयक्त विचार लेखकों के हैं और यह आवश्यक नहीं कि रिज़र्व बैंक के विचारों का प्रतिनिधित्व करें।

¹ इस अध्ययन के लिए हमने भारतीय रिज़र्व बैंक (आरबीआई) द्वारा तैयार एचपीआई पर ध्यान दिया है क्योंकि तुलना के लिए एक विस्तारित समय शृंखला होगी। आगे, आरबीआई अंतरराष्ट्रीय मंचों पर एचपीआई को रिपोर्ट करता है, जबकि एनएचबी रेसिडेक्स के पास अखिल भारतीय स्तर का सूचकांक नहीं है (और उसका कवरेज अपरस्पर अनन्य है)।

करता है। ये सूचकांक सामान्यतः (कम से कम) एक तिमाही के अंतराल से उपलब्ध हैं।

आवासीय घरों की कीमतों पर प्रारंभिक संकेत पाने के लिए, भारत के चुनिंदा रीयल इस्टेट एजेंसियों के वेब पोर्टलों पर उपलब्ध डेटा के आधार पर एक वैकल्पिक आवासीय घर मूल्य सूचकांक तैयार करने का प्रयास किया गया है। इन आँकड़ों को एकत्र, प्रोसेस और स्टोर करने में बिग डेटा के औजार लगाए गए हैं।

आलेख के शेष भाग की संरचना इस प्रकार है : खंड 2 इस क्षेत्र में उपलब्ध साहित्य की एक संक्षिप्त समीक्षा करता है। खंड 3 गतिशील चार्ट स्क्रेपिंग विधि से डेटा का संग्रह और बिग डेटा के औजारों से अनस्ट्रक्चर्ड डेटा की प्रोसेसिंग पर प्रकाश डालता है। खंड 4 वेब डेटा के आधार पर सूचकांक तैयार करने की विधि बताता है। रिज़र्व बैंक की एचपीआई के आकलन हेतु सूचकांक की उपयोगिता का अध्ययन करने के लिए सरलीकृत तथ्य और आँकड़ा-आधारित विधियां खंड 5 में दी गई हैं। खंड 6 अनुभवजन्य परिणामों पर केंद्रित है और खंड 7 में निष्कर्षात्मक टिप्पणियां दी गई हैं।

2. साहित्य की समीक्षा

बिग डेटा एनलिटिक्स के क्षेत्र में, किसी निर्दिष्ट वेबसाइट से अपेक्षित डेटा को प्रोग्राम की मदद से जुटाने में लगने वाली तकनीक को वेब स्क्रेपिंग कहा जाता है। डेटा संग्रह के तरीके के रूप में वेब स्क्रेपिंग पहले अपनाया गया था। वैसे, इसका उपयोग मुख्य आर्थिक गतिविधि में बिलियन प्राइसेज प्रोजेक्ट (बीपीपी) में हुआ था जो कि एमआईटी स्लोअन और हारवर्ड बिजनेस स्कूल (कवालो एंड रिगोबॉन, 2016) का एक अकादमिक प्रयास था। यह कार्य बताता है कि किस प्रकार वैकल्पिक स्रोत, जैसे बीपीपी में ऑनलाइन ई-कॉमर्स, मूल्य सूचकांक के संकलन के लिए संभावित उम्मीदवार हो सकते हैं। उनके कार्य में, ऑनलाइन ई-कॉमर्स पोर्टलों से डेटा प्राप्त करने, साफ़ करने तथा प्रोसेस करने के लिए बिग डेटा तकनीकों का प्रयोग किया गया है। उनका तर्क था कि ये स्रोत प्रायः सस्ते हैं और उपयोगकर्ता इससे लगभग तत्काल मूल्य सूचकांक संख्याएं पा सकते हैं। यही

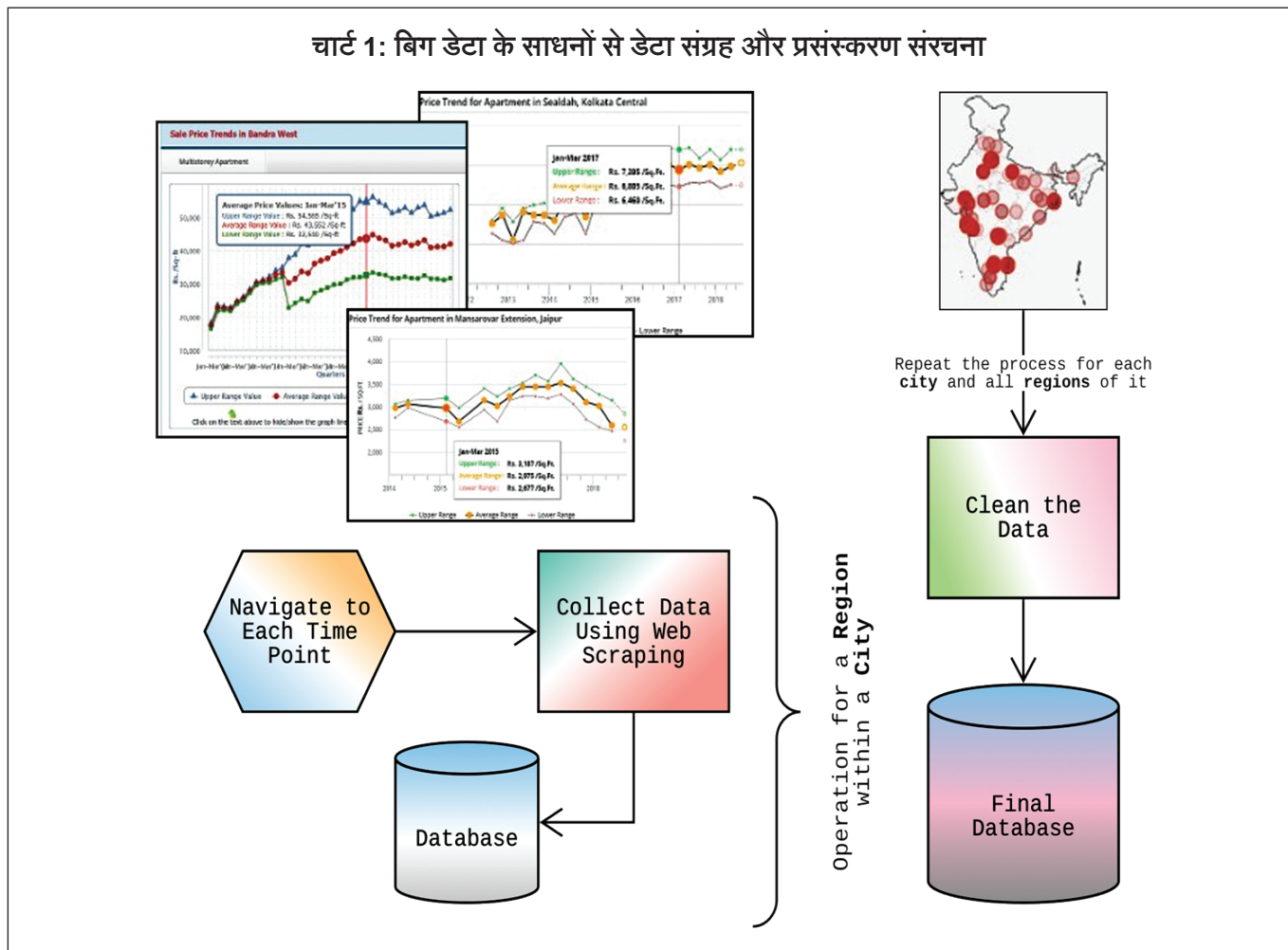
अवधारणा भारतीय संदर्भ में अपनाई गई थी जहाँ यह तर्क दिया गया था कि ऑनलाइन कीमतों की सीमाओं के कारण, वेब-डेटा आधारित सूचकांक भारतीय आधिकारिक उपभोक्ता मूल्य सूचकांक (सीपीआई) की दिशा एवं आकार दोनों को सफलतापूर्वक ट्रैक करता है (बनर्जी एवं अन्य, 2018)।

आवासीय संपत्ति कीमत के संदर्भ में बिग डेटा तकनीकों और वेब स्क्रेपिंग का प्रयोग न्यूनतम रहा है। तथापि इस क्षेत्र में, गूगल ट्रेंड डेटा का प्रयोग करते हुए कई अध्ययन हुए हैं – प्रेडिक्टिंग इकॉनॉमिक एकटिविटीज़ (चोई एंड वैरियन, 2012); एसेसमेंट ऑफ़ रीयल इस्टेट मार्केट एंड हाउसिंग प्राइस (वू एंड ब्यांजजोल्फ़सॉ इन द इंडियन कॉन्टेक्स्ट, 2015); एंड हाउसिंग प्राइस इन द इंडियन कॉन्टेक्स्ट (मित्रा व अन्य., 2017)। इस

संदर्भ में ऐसा ही अध्ययन बैंक ऑफ़ इंडोनेशिया में किया गया था। यह बताया गया कि इंडोनेशिया का ऑनलाइन विज्ञापन डेटा प्रॉपर्टी की कीमतों को सफलतापूर्वक ट्रैक कर रहा है (क्रिष्टियावर्दानी एंड साम्पे, 2017)।

3. डेटा संग्रह और प्रोसेसिंग

भारत की चुनिंदा रीयल इस्टेट विज्ञापन एजेंसियों की वेबसाइट² से डेटा संग्रहित करने के लिए बिग डेटा के साधनों का प्रयोग हुआ है। आवासीय घर कीमत पर ऐतिहासिक डेटा प्राप्त करने के लिए, एक विशिष्ट वेब स्क्रेपिंग तकनीक विकसित की गई है जिसे 'डायनामिक चार्ट स्क्रेपिंग' कहा जा सकता है (चार्ट 1)।



² भारतीय बाजार में अग्रणी रीयल इस्टेट विज्ञापन वेबसाइट्स ।

स्रोत वेबसाइट में कीमत ट्रेंड को एक डायनामिक चार्ट के रूप में प्रस्तुत किया गया है। वेब स्क्रेपिंग तकनीक का विकास किया गया और वेब पोर्टलों में इन डायनामिक चार्टों पर उनका प्रयोग वांछित आवास कीमत सूचना पाने के लिए किया गया। स्रोत वेबसाइट, शहर (सिटी), शहर के अंदर अवस्थिति (क्षेत्र के नाम से), संपत्ति का प्रकार, संपत्ति की कीमत, कीमत समुच्चयन का प्रकार, कीमत की इकाई और कीमत का समय बिंदु वो डेटा फील्ड्स हैं जो डायनामिक चार्ट स्क्रेपिंग के प्रयोग से प्राप्त हुए हैं।

प्राप्त डेटासेट में कुछ वैल्यू अनुपस्थित पाए गए हैं। इन अनुपस्थित वैल्यू का आकलन स्थानीय बहुपदों (लोकल पॉलिनोमियल्स) के प्रक्षेप (इंटरपोलेशन) द्वारा किया गया है³

कभी कभी डेटासेट में कीमत अंतर बहुत अधिक पाया गया। ये अंतर इन विधियों से एडजस्ट किए गए हैं : (ए) अपवाद (आउटलायर) पहचान नियम जैसा कि X-13एरिमा-सीट्स प्रोग्राम में निहित है; (बी) योगज (एडिटिव) अपवाद पहचान नियम; और (सी) गतिमान अंतश्चतुर्थक परिसर (इंटरक्वार्टाइल रेंज) वाली अनुमानी पद्धति। आवास मूल्य का सूचकांक प्रोसेस किए हुए अंतिम (फाइनल) डेटा के आधार पर निकाला गया है।

4. आवास मूल्य सूचकांक का संकलन

संकेत P_{ijkt}^a , a प्रकार के कीमत समुच्चयन के साथ i शहर के j क्षेत्र में t तिमाही और संपत्ति के k वर्ग के लिए। नोट करें कि जहाँ भी P_{ijkt}^a उपलब्ध नहीं है, उसे \bar{P}_{ijkt}^a से भरा गया है (जहाँ (i, j, k, t, a) उपलब्ध नहीं, वहाँ आकलित)। आकलन विभिन्न मशीन लर्निंग और समय शृंखला अन्वारोप पद्धतियों द्वारा किया गया है। सूचकांक के संकलन में निम्नलिखित चरण शामिल हैं:

(ए) एक निर्दिष्ट समय बिंदु को आधार-अवधि मानते हुए (मान लिया द्वारा इसे $\pi_{ijkt}^a = \frac{P_{ijkt}^a}{P_{ijkt_0}^a}$ से दर्शाते हैं), इनमें से (π_{ijkt}^a) प्रत्येक वर्ग हेतु मूल्य सापेक्ष की गणना इस प्रकार की जाती है। प्रत्येक मूल्य सापेक्ष के भार को निम्नलिखित अंकन (नोटेशन) से दर्शाया जाता है?

³ समय शृंखला (टाइम सिरीज) डेटा पर स्लाइडिंग विंडों में लैंगरेंज पॉलीनोमियल (https://en.wikipedia.org/wiki/Polynomial_interpolation) का प्रयोग हुआ है।

$$\omega_{ijk}^a = \text{Weight for } \begin{cases} i^{\text{th}} \text{ city, } j^{\text{th}} \text{ region, where } (i, j) \in L \\ k^{\text{th}} \text{ category of property, where } k \in K \\ a^{\text{th}} \text{ type of price aggregation, where } a \in A \\ \text{It is independent of quarter (time point) } t \end{cases}$$

(बी) निर्दिष्ट प्रॉपर्टी प्रकार और अवस्थिति (अवस्थिति / लोकेशन i और j को मिलाकर बना है) के लिए, विभिन्न प्रकार के समुच्चय के लिए भारित मूल्य $(\pi_{ijkt} = \frac{\sum_a \pi_{ijkt}^a \omega_{ijk}^a}{\sum_a \omega_{ijk}^a})$ सापेक्ष निकाले गए हैं। समुच्चय स्तर⁴ औसत समुच्चय प्रकार (ω_{ijk}^a) को उच्चतम भार और अन्य दो $(\omega_{ijk}^{\text{upper}} < \omega_{ijk}^{\text{average}} > \omega_{ijk}^{\text{lower}})$ पर कमतर भार निर्धारित करता है। इसे $\omega_{ijk}^a = e^{-|\bar{P}_{t_0}^a - \bar{P}_{t_0}^{\text{average}}|}$ के रूप में परिभाषित किया गया है। (यहाँ $\bar{P}_{t_0}^a$ आधार अवधि में सामान्यीकृत यूनिट फ्री प्राइस का सही औसत दर्शाता है।) (अंतिम भार औसत के लिए 1, ऊपर के लिए 0.913 आया।)

(सी) इसके बाद इन सभी भारित मूल्य सापेक्षों को प्रत्येक तिमाही के लिए $\pi_{ijt} = \frac{\sum_k \pi_{ijkt} \omega_{ijk}}{\sum_k \omega_{ijk}}$ के रूप में अवस्थिति स्तर पर आगे और समुच्चयित किया जाता है। प्रत्येक संपत्ति प्रकारों $\{\omega_{ijk}\}$ के अनुसार भारों का चयन आधार अवधि (t_0) में शहर (i) में विशिष्ट किस्म की संपत्ति ऑफर करने वाले क्षेत्रों (j) की गणना पर आधारित है।

(डी) प्रत्येक तिमाही के लिए, प्रत्येक अवस्थिति से संबंधित भारित मूल्य सापेक्षों (π_{ijt}) को क्षेत्र भारों (ω_{ij}) के प्रयोग से पुनः समुच्चयित कर शहर स्तरीय तिमाही भारित मूल्य सापेक्ष $(\pi_{it} = \frac{\sum_j \pi_{ijt} \omega_{ij}}{\sum_j \omega_{ij}})$ प्राप्त किए जाते हैं। क्षेत्रीय अवस्थितियों $\{\omega_{ij}\}$ के लिए भार $\omega_{ij} = e^{-\bar{P}_{ijt_0}^5}$ बनाया गया है।

⁴ सेट ए (सभी संभावित समुच्चयन प्रकारों का) में तीन तत्व हैं जो ऊपरी, औसत एवं निम्नतर हैं।

⁵ यह आधार अवधि में, किसी क्षेत्र की औसत सामान्यीकृत कीमत से व्युत्क्रमानुपाती रूप से जुड़ा है, जो किसी निर्दिष्ट अवस्थिति में विज्ञापनों की संख्या का प्रतिनिधित्व करता है। डेटा से यह देखा गया कि उच्चतर सापेक्ष मूल्य वाले विज्ञापन निम्नतर सापेक्ष मूल्य की तुलना में कम बार आते हैं।

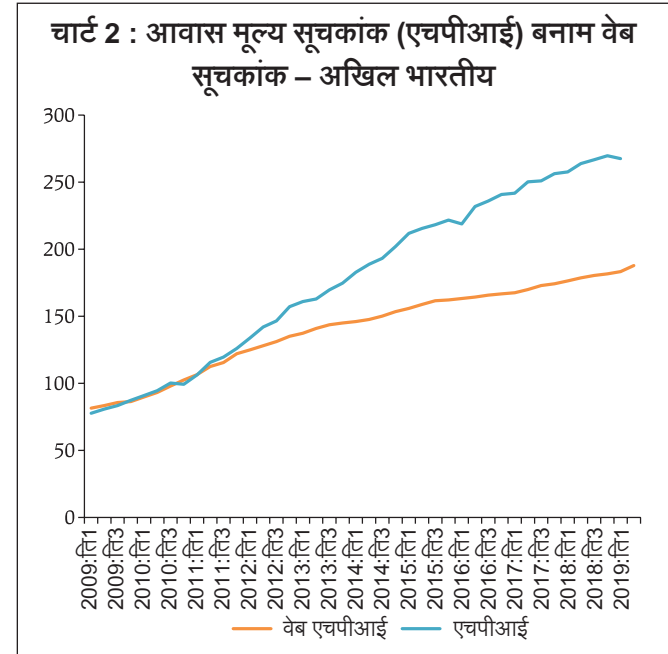
(ई) अंततः, तिमाही अखिल भारतीय आँकड़े प्राप्त करने के लिए, यह मानकर कि शहर और क्षेत्र का प्रतिनिधि कवरेज है, शहर भार को ω_i के रूप में लेते हुए ये मूल्य सापेक्ष (π_{it}) भारत औसत की गणना के लिए प्रयुक्त होते हैं। यह अंक $\pi_t = \frac{\sum_i \pi_{it} \omega_i}{\sum_i \omega_i}$ के रूप में दर्शाया गया है। आवास मूल्य सूचकांक (एचपीआई) बनाने में रिज़र्व बैंक की पद्धति का अनुसरण करते हुए, 2011 की जनगणना के डेटा के अनुसार प्रत्येक शहर के जनसंख्या आकार को तत्संबंधी शहर भार के रूप में लिया गया गया है।⁶

5. सरलीकृत तथ्य और आँकड़ा आधारित पद्धतियां

रिज़र्व बैंक की एचपीआई⁷ हर तिमाही बनाई जाती है जिसमें 1आधार वर्ष 2010-11 (आरबीआई, 2014) होता है तथा प्रमुख शहरों यथा मुंबई, दिल्ली, चेन्नई, कोलकाता, बंगालुरु, लखनऊ, अहमदाबाद, जयपुर, कानपुर और कोच्ची के राज्य सरकारों के पंजीकरण प्राधिकारियों से प्रॉपर्टी लेन-देन संबंधी सरकारी आँकड़ों को आधार बनाया जाता है। कोच्ची शहर को इस एचपीआई के विश्लेषण से बाहर रखा गया है। वेब से प्राप्त आँकड़ों पर आधारित यह सूचकांक कोच्ची को छोड़कर इन सभी शहरों को कवर करता है। इसे वेब (डबल्यूईबी) एचपीआई कह गया है (सरलता के लिए हमने वेब का प्रयोग अदल-बदल कर एक ही अर्थ में किया है)। डबल्यूईबी का आधार वर्ष वही है जो एचपीआई का है।

समग्रतः, वेब (डबल्यूईबी) सूचकांक एचपीआई से कम पाया गया, यानी पंजीकरण डेटा पर आधारित कीमत स्तर विज्ञापनों में दर्शाए गए कीमत स्तर से अधिक है (चार्ट 2)।

आगे, वेब इंडेक्स ग्राफ्स एचपीआई से सहजतर है। अर्थात् विज्ञापन कीमतें बार-बार नहीं बदली जाती, जबकि वास्तविक



पंजीकरण (रजिस्ट्रेशन) डेटा पर आधारित एचपीआई में समय के साथ कीमतों में अधिक भिन्नता है (क्योंकि वास्तविक विक्रय में बहुत अंतर होता है)। इसके अतिरिक्त रजिस्ट्रेशन कीमतों में प्रॉपर्टी के बहुत से कारक जुड़े होते हैं, जैसे मंजिल/तल (फ्लोर) वृद्धि, पोजीशन, मुखी/अग्रभाग (फेसिंग) और अन्य बातें जो विज्ञापन में जाहिर नहीं की जाती। एचपीआई और वेब में पार्थक्य का आँकड़ापरक विश्लेषण इस अध्ययन के क्षेत्र से बाहर है। वैसे, इसके निम्नलिखित कारण हो सकते हैं:

- विज्ञापित कीमतें दीर्घतर अवधि में बदलती हैं, जबकि प्रॉपर्टी की वास्तविक कीमतें काफ़ी ज्यादा होती हैं क्योंकि उसमें कई बातें शामिल होती हैं;
- ग्राहकों को लुभाने के लिए कीमतें छूट (डिस्काउंट) पर विज्ञापित की जाती हैं जबकि अंतिम लेन-देन उच्चतर लागत पर होता है;
- रीयल इस्टेट बाजार अनुसंधान पर कई रिपोर्टें बताती हैं कि 2013-14 से⁸ आवासीय कीमतों की बिक्री में रुझान ढलान का है। फलतः, संभव है कि विज्ञापन में दर्शाई गई कीमतें वास्तव में बिकी संपत्तियों का प्रतिनिधित्व न करें; और

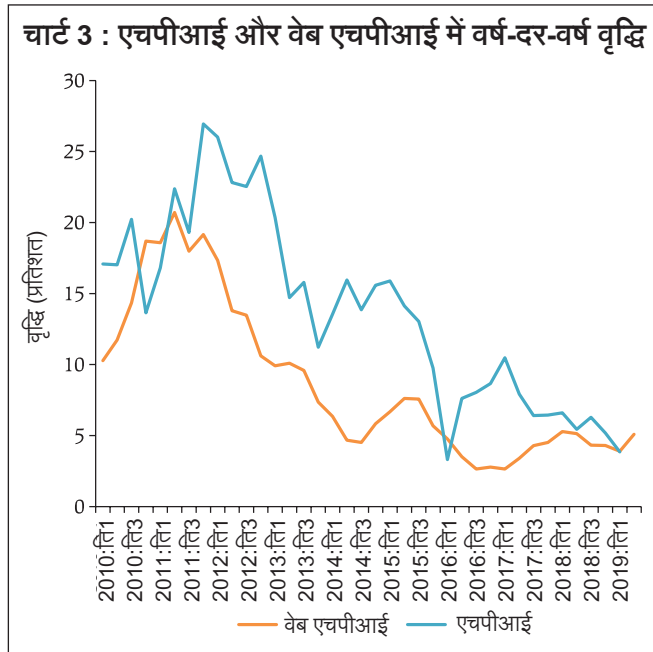
⁶ भार निर्धारण के सभी स्वरूप (अखिल भारतीय स्तर के समुच्चयन को छोड़कर) लेखकों के अनुमान वाली पद्धति पर आधारित है। इस प्रकार के डेटा पर अनुसंधान कार्य या तो सीमित हैं या इन पर कोई भी अनुसंधान कार्य नहीं है जिसके विवरण इस अध्ययन के दायरे से बाहर रखे गए हैं।

⁷ आँकड़े <https://dbie.rbi.org.in/> (Statistics > Real Sector > Prices & Wages > Quarterly) से लिए गए हैं। एचपीआई श्रृंखला उभय अवधि के दो भिन्न आधार वर्ष में उपलब्ध है। 2010-दूसरी तिमाही (2010-11 की पहली तिमाही की समतुल्यता से) के पहले के सूचकांक आँकड़े पुराने आधार वर्ष (2008-09) के सूचकांकों के साथ मिलाकर प्राप्त किए गए हैं।

(iv) एचपीआई की संकलन पद्धति के अनुसार, आवासीय संपत्तियों का समूह एक तिमाही से दूसरी तिमाही में अलग होता है, जो तत्संबंधी अवधि में हुई वास्तविक बिक्री पर निर्भर होता है।

जनवरी 2010- सितंबर 2010, अक्टूबर 2011 – दिसंबर 2015 और अप्रैल 2016- जून 2017 के दौरान पंजीकरण डेटा पर आधारित वर्ष-दर-वर्ष वृद्धि के अंक विज्ञापन आधारित वृद्धि से काफी अधिक थे। तथापि, दोनों शृंखलाओं में कीमतों में वर्ष-दर-वर्ष वृद्धि की सांख्यिकीय सह-गति है। जनवरी-मार्च 2019 में वृद्धि के जो अंक 5 प्रतिशत के नीचे चले गए, उसका कारण संभवतः भू-संपदा (विनियमन और विकास) अधिनियम [रीयल इस्टेट रेग्यूलेशन एंड डेवलपमेंट (रेरा) ऐक्ट], 2016 का कार्यान्वयन रहा हो (चार्ट 3)।

एचपीआई और वेब एचपीआई में आँकड़ापरक संबंध का अध्ययन करने के लिए स्तरों एवं वर्ष-दर-वर्ष वृद्धि पर सहसंबंध विश्लेषण किया गया है।



⁸ 2017 में घरों की बिक्री 2013/14 के स्तर से 40 प्रतिशत नीचे: एनएआरओसीके (स्रोत: <https://economictimes.indiatimes.com/industry/services/property-cstruction/housing-sales-down-40-per-cent-in-2017-from-2013/14-levels-anarock/articleshow/63352292.cms>)

आवास कीमतों की वर्ष-दर-वर्ष वृद्धि की गणना में निम्नलिखित सूत्र का प्रयोग⁹

$$Y_t^{HPI} = \left(\frac{I_t^{HPI}}{I_{t-4}^{HPI}} - 1 \right) * 100 \quad \dots \quad (1)$$

$$Y_t^{WEB} = \left(\frac{I_t^{WEB}}{I_{t-4}^{WEB}} - 1 \right) * 100 \quad \dots \quad (2)$$

एचपीआई के आकलन में वेब एचपीआई डेटा की उपयोगिता का विश्लेषण करने के लिए, निम्नलिखित चरणों का प्रयोग हुआ है:

1. चरों (वेरिएबल्स) Y_t^{HPI} और Y_t^{WEB} का अवर्धमान परीक्षण (स्टेशनरी टेस्ट) करें।
2. यदि तत्संबंधी चर अवर्धमान (नॉनस्टेशनरी) पाए जाते हैं तो यह देखने के लिए एक परीक्षण किया जाता है कि क्या ये चर सहसमेकित (कोइंटीग्रेटेड) हैं।

यदि चर सहसमेकित हैं, तो चरों Y_t^{HPI} और Y_t^{WEB} के आधार पर एक मॉडल बनाएं अन्यथा, दोनों चरों पर अंतरों की एक उपयुक्त संख्या लेते हुए चरों Y_t^{HPI} और Y_t^{WEB} को अवर्धमान शृंखला में बदलने के बाद एक वेक्टर आटो रिग्रेसन (वीएआर) मॉडल बनाएं (एस्टीरियो एंड हॉल, 2011)।

देखने में आया कि चर Y_t^{HPI} और Y_t^{WEB} अस्थिर हैं और सहसमेकित नहीं हैं। पहले क्रम अंतर में चर Y_t^{HPI} और Y_t^{WEB} स्थिर पाए गए। अतः तत्संबंधी वीएआर मॉडल ये हैं:

$$\begin{pmatrix} \Delta Y_t^{HPI} \\ \Delta Y_t^{WEB} \end{pmatrix} = \begin{pmatrix} \alpha^{HPI} \\ \alpha^{WEB} \end{pmatrix} + \sum_{j=1}^p \begin{pmatrix} \gamma_j^{HPI-HPI} & \gamma_j^{WEB-HPI} \\ \gamma_j^{HPI-WEB} & \gamma_j^{WEB-WEB} \end{pmatrix} \begin{pmatrix} \Delta Y_{t-j}^{HPI} \\ \Delta Y_{t-j}^{WEB} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{HPI} \\ \epsilon_t^{WEB} \end{pmatrix} \quad \dots \quad (3)$$

नोट करें कि समीकरण के दाहिनी ओर गुणांक $\gamma_j^{HPI-WEB}$ होने का अर्थ है ΔY_t^{WEB} का पूर्वानुमान करते हुए ΔY_{t-j}^{HPI} के लिए गुणांक एवं अन्य सभी γ_j^{***} को उसी रेखा पर परिभाषित किया गया है। सरलता से दर्शाने के लिए हमने बाद में सारणी 3 प्रथम कॉलम में केवल γ_j^* का प्रयोग किया है (शेष कॉलम γ_j^{***} को पूरी तरह परिभाषित करते हैं)।

समीकरण (3) का आकलन अंतर्निहित डेटा के आधार पर किया गया है।

⁹ समय $t \in \{2, 3, \dots, T\}$ पर और जो हम I_t^{HPI} और I_t^{WEB} एचपीआई वेब पीआई के रूप में दर्शा रहे हैं।

6. प्रयोगिक परिणाम

जनवरी-मार्च 2009 से जनवरी-मार्च 2019 (41 तिमाहियों) तक उपलब्ध तत्संबंधी सूचकांक यथा एचपीआई और वेब एचपीआई पर तिमाही डेटा के आधार पर प्रायोगिक विश्लेषण किया गया है।

स्तर व वृद्धि (वर्ष-दर-वर्ष) पर एचपीआई और वेबएचपीआई के बीच आकलित सहसंबंध गुणांक 1 प्रतिशत के स्तर पर सांख्यिकीय रूप से महत्वपूर्ण पाए गए हैं (सारणी 1)।

तत्संबंधी चरों का यूनिट रूट टेस्ट फिलिप्स पेरोन टेस्ट के प्रयोग से किया गया है। देखने में आया कि दोनों चरों Y_t^{HPI} और Y_t^{WEB} का प्रथम क्रम अंतर (फर्स्ट ऑर्डर डिफरेंस) अवर्धमान (स्टेशनरी) पाया गया (सारणी 2)। Y_t^{HPI} और Y_t^{WEB} चरों के बीच जोहानसन सहसमेकन परीक्षण दर्शाता है कि ये चर सहसमेकित नहीं हैं। तदनुसार, सैंपल डेटा के आधार पर एक वीएआर मॉडल (समीकरण 3 की तर्ज पर) आकलित किया गया है।¹⁰ वीएआर मॉडल आवास कीमत की वृद्धिशील वृद्धि के लिए एक आकलन ढाँचा देता है।

आकलित वीएआर मॉडल दर्शाता है कि आवासीय मकान की कीमत की वृद्धिशील वृद्धि का आकलन उनके अंतराल मूल्यों (विगत चार तिमाही) से किया जा सकता है। विशेषतः, वेब पर उपलब्ध कीमत से, पंजीकरण आवासीय मकान कीमत का अनुमान लगाने में सहायता मिल सकती है। पंजीकरण कीमत और ऑनलाइन वेब कीमत के बीच द्विदिशात्मक कारण संबंध है (सारणी 3)।

टेबल 1: आकलित सहसंबंध गुणांक और सांख्यिकीय जाँच

| सहसंबंध गुणांक | आकलन | पी- वैल्यू | सैंपल डेटा अवधि |
|--|--------|------------|--------------------------------------|
| $\rho(I_t^{HPI}, I_t^{WEB})$ | 0.9869 | 0.00 (***) | जनवरी-मार्च 2009 से जनवरी-मार्च 2019 |
| $\rho(Y_t^{HPI}, Y_t^{WEB})$ | 0.7844 | 0.00 (***) | जनवरी-मार्च 2009 से जनवरी-मार्च 2019 |
| $\rho(\Delta Y_t^{HPI}, \Delta Y_t^{WEB})$ | 0.1296 | 0.23 | अप्रैल-जून 2010 से जनवरी-मार्च 2019 |

¹⁰ अंतिम अनुमान त्रुटि (एफपीई) सूचना मानदंड के आधार पर वीएआर मॉडल का अंतराल क्रम (लैग ऑर्डर) तय किया गया है। आकलित वीएआर मॉडल के लिए इष्टतम अंतराल क्रम (ऑप्टिमल लैग ऑर्डर) 4 है।

सारणी 2 : यूनिट रूट टेस्ट – सारांश निष्कर्ष

| निराकरण परिकल्पना (नल हाइपोथेसिस) | फिलिप्स पेरोन टेस्ट के सारांश निष्कर्ष | स्वीकार / रद्द करें |
|---------------------------------------|---|---------------------|
| Y_t^{HPI} का एक यूनिट रूट है | टेस्ट स्टैटिस्टिक -0.9246, पी वैल्यू 0.7688 | स्वीकार |
| ΔY_t^{HPI} का एक यूनिट रूट है | टेस्ट स्टैटिस्टिक -7.2818, पी वैल्यू 0.0000 | रद्द |
| Y_t^{WEB} का एक यूनिट रूट है | टेस्ट स्टैटिस्टिक -1.0897, पी वैल्यू 0.7096 | स्वीकार |
| ΔY_t^{WEB} का एक यूनिट रूट है | टेस्ट स्टैटिस्टिक -4.4201, पी वैल्यू 0.0012 | रद्द |

7. निष्कर्ष

बिग डेटा के उपकरणों की सहायता से रीयल इस्टेट विज्ञापनदाताओं के वेब पोर्टलों में आवास कीमतों पर उपलब्ध डेटा को एकत्रित और प्रसंस्कृत (प्रोसेस) किया गया है। ऑनलाइन उपलब्ध विज्ञापन डेटा के आधार पर रिहाइशी मकान कीमत का एक वैकल्पिक संकेतक बनाया गया है। एचपीआई और वेब एचपीआई के कीमत स्तरों के बीच अंतराल है। विज्ञापन डेटा के आधार पर रिहाइशी मकान कीमत का स्तर पंजीकरण डेटा वाले स्तर से कम है। साथ ही, जनवरी 2010 -सितंबर 2010, अक्टूबर 2011 – दिसंबर 2015 और अप्रैल 2016 – जून 2017 के दौरान, पंजीकरण डेटा में मूल्य वृद्धि में बढ़ोतरी विज्ञापन के आधार पर जोड़ी गई बढ़ोतरी की तुलना में काफी अधिक थी। तथापि सूचकांकों (एचपीआई और वेब) और उनकी वर्ष-दर-वर्ष वृद्धि में सहसंबंध धनात्मक और सांख्यिकीय रूप से महत्वपूर्ण

सारणी 3 : आकलित वीएआर मॉडल

| सहगुणांक | अंतर्जात (एंडोजीनस) चर | |
|-------------------------------|------------------------|-----------------------|
| | ΔY_t^{HPI} | ΔY_t^{WEB} |
| $\alpha^{HPI} / \alpha^{WEB}$ | -0.3850 (0.44) | -0.2737 (0.18) |
| γ_1^{HPI} | -0.0189 (0.91) | -0.0103 (0.87) |
| γ_2^{HPI} | -0.1922 (0.21) | 0.0372 (0.53) |
| γ_3^{HPI} | 0.2007 (0.18) | 0.1073 (0.00) |
| γ_4^{HPI} | -0.4307 (0.01) | -0.1366 (0.04) |
| γ_1^{WEB} | 0.2033 (0.63) | 0.2101 (0.22) |
| γ_2^{WEB} | 0.0288 (0.93) | 0.4651 (0.00) |
| γ_3^{WEB} | -0.3672 (0.27) | -0.0591 (0.65) |
| γ_4^{WEB} | 0.9768 (0.00) | -0.3212 (0.02) |
| $R^2 (adj)$ | 0.32 (0.03) | 0.40 (0.01) |

ग्रेंगर कारण जाँच

| निराकरण परिकल्पना (नल हाइपोथिसिस) | सारांश परिणाम |
|--|--------------------------|
| $\Delta Y_t^{HPI} \Delta Y_t^{WEB}$ का ग्रेंगर कारण नहीं | एफ-टेस्ट = 2.9509 (0.03) |
| $\Delta Y_t^{WEB} \Delta Y_t^{HPI}$ का ग्रेंगर कारण नहीं | एफ-टेस्ट = 2.8186 (0.04) |

स्रोत: लेखकों द्वारा की गई गणना।
प्रथम कोष्ठकों के अंक पी- वैल्यू दर्शाते हैं।

पाया गया। अतः विज्ञापन डेटा पर आधारित रिहाइशी मकान का मूल्य एक महत्वपूर्ण संकेतक हो सकता है, और इसके प्रयोग पर और आगे अनुसंधान के प्रयास किए जा सकते हैं।

संदर्भ

Alan G. Ahearne, John Ammer, Brian M. Doyle, Linda S. Kole, and Robert F. Martin, (2005). "House Prices and Monetary Policy: A Cross-Country Study", *Board of Governors of the Federal Reserve System*.

Asteriou, D. & Hall, S. G., (2011). "Vector Autoregressive (VAR) Models and Causality Tests". *Applied Econometrics (Second ed.)*. London: Palgrave MacMillan, p. 319-333.

Banerjee, R., Singhal, N. & Subramanian, C., (2018). "Predicting Food Price Inflation through Online Prices in India". *Economic and Political Weekly, Volume LIII No 23, pp. 132-135*.

Cavallo, A., & Rigobon, R. (2016). "The Billion Prices Project: Using Online Prices for Measurement and Research". *Journal of Economic Perspectives, 30(2), 151-178*. <https://doi.org/10.1257/jep.30.2.151>

Choi, H., & Varian, H. (2012). "Predicting the Present with Google Trends". *Economic Record, 88(SUPPL.1), 2-9*. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>

Kristiawardani, K., & Sampe, I. (2017). "Using Online Property Advertisements Data as a Proxy for Property Market Indicators Bank Indonesia". *FC-Bank Indonesia Satellite Seminar on "Big Data" at the ISI Regional Statistics Conference 2017 Bali, Indonesia, March*.

Mitra, P., Sanyal, A., & Choudhury, S. (2017). "Nowcasting Real Estate Activity in India using Google Trend Data". *Reserve Bank of India - Occasional Papers, 38-No. 1*. https://rbi.org.in/Scripts/bs_viewcontent.aspx?Id=3516.

RBI, (2014). "House Price Index: 2010-11 to 2013-14". *RBI Monthly Bulletin, October, pp. 93-96*.

Wu, L., & Brynjolfsson, E. (2015). "The Future of Prediction : How Google Searches Foreshadow Housing Prices and Sales". *Economic Analysis of the Digital Economy: April 2015*. <https://doi.org/10.3386/w19549>