

बिग डाटा और मापन : महंगाई से विभेद तक* रॉबर्टो रिगोबॉन

मैं अपनी बात द्वितीय सुरेश तेंदुलकर स्मृति व्याख्यानमाला के अंतर्गत भाषण देने हेतु आमंत्रित किए जाने के लिए भारतीय रिज़र्व बैंक के प्रति आभार प्रदर्शन के साथ प्रारंभ करना चाहूंगा। प्रोफेसर तेंदुलकर भारत में सर्वाधिक प्रभावशाली नीति निर्माताओं में से थे। उनके द्वारा किए गए अनेक कार्यों में से एक यह था कि उन्होंने भारतीय राष्ट्रीय सांख्यिकी आयोग के प्रमुख के रूप में कार्य किया। राष्ट्रीय सांख्यिकीय प्राधिकारियों द्वारा आंकड़े एकत्र करने के मार्ग में आ रही परेशानियों को कम करना इस तरह के आयोग के लक्ष्यों में से एक था। इसलिए यह उपयुक्त होगा कि मैं इस संपूर्ण व्याख्यान के दौरान मापन और आंकड़े एकत्र करने के बारे में ही बात करूं। जिन लोगों ने यह सोचा हो कि मैं इस विषय पर बात करने वाला हूँ कि अमेरिका में और अमेरिकी ब्याज दर के बारे में क्या होने जा रहा है, तो यह बात कहीं और की जा रही होगी। यहां पर हम लोगों में से उन लोगों के एकदम नीरस और अत्यधिक उकताने वाले जीवन के बारे में बात की जाएगी जो आंकड़े एकत्र करने के लिए समर्पित हैं। इस प्रसिद्ध व्याख्यान माला में भाषण देना मेरे लिए बहुत सम्मान की बात है। आपका आमंत्रण मैंने विनम्रता पूर्वक स्वीकार किया है।

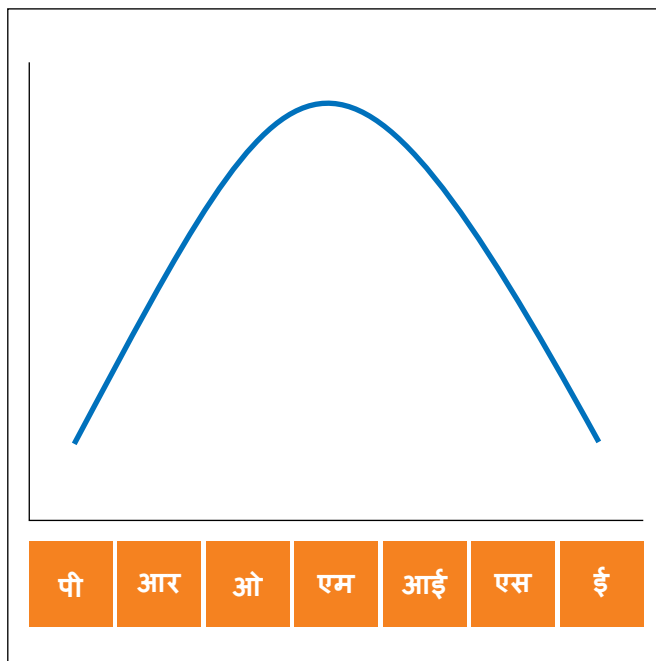
इसके अलावा, भारतीय रिज़र्व बैंक में आने का यह मेरे लिए तीसरा अवसर है, किंतु पहली बार है जब क्रिकेट का विश्व कप नहीं देखा जा रहा है और मैं इस बात से थोड़ा दुखी हूँ। इसलिए हमें यह सुनिश्चित करना होगा कि 2023 में, जब क्रिकेट विश्व कप सचमुच खेला जाएगा, मैं यहां आऊँ। और हां, इस बारे में कोई बहाना नहीं चलेगा।

अब, मैं मुझे अनुसंधानकर्ता के रूप में परिभाषित करने वाली विशेषताओं में से एक को मान्यता प्रदान करते हुए शुरुआत करना चाहूंगा। मुझ पर मापन का जुनून सवार है। मैंने अपनी जिंदगी के विगत 17 वर्ष मापन के बारे में विचार करते हुए गुजारे हैं। क्यों ? प्राथमिक रूप से इसका कारण यह है कि मैं यह महसूस करता हूँ कि हम यह कार्य बहुत खराब ढंग से करते हैं।

* द्वितीय सुरेश तेंदुलकर स्मृति व्याख्यानमाला के अंतर्गत भारतीय रिज़र्व बैंक, मुंबई में 02 अगस्त 2018 को प्रोफेसर रॉबर्टो रिगोबॉन, एमआईटी स्लोआन स्कूल ऑफ मैनेजमेंट, कैंब्रिज, मैसाचुएट्स, यूएसए का भाषण। इसमें व्यक्त मत एवं विचार सिर्फ लेखक के हैं।

मैं, यह बात और स्पष्टता के साथ कहना चाहूंगा। स्लोआन में मेरी कक्षाओं में हम सामाजिक भलाई तथा मापन संबंधी अपनी समस्याओं से जुड़े मुद्दों के पहलुओं पर विचार-विमर्श करने के लिए प्रणाली (फ्रेमवर्क) का उपयोग करते हैं। इस प्रणाली को हम पीआरओएमआईएस (प्रॉमिस) कहते हैं। इसमें प्रॉमिस से तात्पर्य यह होता है - व्यक्तिगत (पी), संबंध (आर), संगठन (ओ), बाजार (एम), संस्थान (आई), सामाजिक एवं राजनीतिक (एस) एवं पर्यावरण (ई)। जब आप इनके बारे में विचार करते हैं तो पाएंगे कि किसी भी समाज में भलाई का आकलन करने के वक्त ये सभी आयाम महत्वपूर्ण होते हैं। हम किसी को भी भूलकर यह बहाना नहीं कर सकते कि सब कुछ ठीक है।

हम किसी विशिष्ट पहलू को वहनीय बनाने वाले सिद्धांतों को समझने के फ्रेमवर्क का प्रयोग करते हैं, यह इसके अनुप्रयोगों में से एक है। उदाहरण के लिए, प्रदूषण के संदर्भ में हम प्रणाली में डाले गए प्रदूषण के उससे हटाए जा सकने वाले प्रदूषण की तुलना में कम होने की स्थिति को पसंद करेंगे। दूसरा महत्वपूर्ण अनुप्रयोग मापन विधि की गुणवत्ता को विशेषरूप से दर्शाना है।



पी	• व्यक्तिगत (पर्सनल)
आर	• संबंध (रिलेशनशिप)
ओ	• संगठन, फर्म एवं रोजगार (ऑर्गनाइजेशन, फर्म एंड जॉब्स)
एम	• बाजार एवं अर्थव्यवस्था (मार्केट्स एंड इकोनॉमी)
आई	• संस्थान (इंस्टीट्यूशन्स)
एस	• सामाजिक एवं राजनीतिक (सोसियल एंड पॉलिटिकल)
ई	• पर्यावरण (इन्वायरमेंट)

जैसा कि मैंने पहले कहा, मेरा मानना है कि मापन के संबंध में हम बहुत भयानक कार्य करते हैं। उदाहरण के लिए, हम संबंधों की गुणवत्ता की माप नहीं करते हैं। ऐसा अधिकांशतः उनके टूट जाने पर। हम अपनी श्रम शक्ति (अथवा उनके मनोबल) के संतुष्टि स्तर की माप नहीं करते हैं, अधिकांशतः उनके निराश होने या काम छोड़ने पर अथवा उनके श्रम अदालतों में शिकायतें दर्ज कराने की स्थिति में होता है। हम इसकी माप नहीं करते हैं हमारा समाज कितना निराश है, उनके विद्रोह करने पर ही इसका पता चलता है। हम किसी विशिष्ट झील पर पड़ने वाले प्रभाव की माप नहीं करते हैं, अधिकांशतः झील को खो देने पर। हम ड्रग्स की खपत की माप नहीं करते हैं, हम सिर्फ ड्रग्स की अधिक खुराक से परेशानी झेल रहे लोगों की गणना करते हैं। बेशक, आर्थिक परिणामों की गणना उनमें से है जिनकी माप हम बेहतरीन ढंग से करते हैं, किंतु जैसे ही हम विशुद्ध रूप से बाजार के परिणामों से इतर जाते हैं, माप की गुणवत्ता खराब होती जाती है।

मेरा मानना है कि "खराब" माप की पाँच विशेषताएँ होती हैं : हम मापन **विलंब से, विषम बारंबारता** पर करते हैं। तकरीबन हमेशा हम **पराकाष्ठा की स्थिति के व्यवहार** पर ध्यान केंद्रित करते हैं जिसमें हमारा आकलन अपनी **मान्यता** पर आधारित होता है, और गलत **सांख्यिकी** दिमाग पर हावी होती है। इसलिए, मैं मानता हूँ कि अगर हम सही निर्णय लेते हैं तो यह अधिकांशतः चमत्कार ही होता है - यह इस बात का प्रमाण है कि कोई दैवीय शक्ति हमारा संरक्षण कर रही है।

उदाहरणार्थ, नारी सशक्तिकरण की माप हम कैसे करते हैं ? हम नारियों की संख्या गिनते हैं या यौन उत्पीड़न के मामलों की संख्या की गणना करते हैं। पहला, दर्ज किए गए "यौन उत्पीड़न के मामले"? लंबे समय से यही परिभाषा है। दूसरा, उनकी गणना अदालत के मामले के सार्वजनिक हो जाने के बाद (यदाकदा) करना। तीसरा, यौन उत्पीड़न शत्रुतापूर्ण माहौल के पराकाष्ठापूर्ण स्थिति होती है। चौथा, हम सिर्फ उन मामलों पर ध्यान देते हैं, जो सुर्खियाँ बनते हैं - क्या आपने महाविद्यालयों के उन मामलों के बारे में सुना है जो 2015 में अनियंत्रित संख्या में थे ? क्या आपने हाल ही में किसी मामले के बारे में सुना है ? 2018 का कोई भी मामला मेरे ध्यान में नहीं है। क्या इसका मतलब यह है कि अब कोई समस्या है ही नहीं अथवा बात यह है कि समाचार पत्रों का रुख हॉलीवुड के अनैतिक आचरणों की तरफ हो गया है ? अंततः, सिर्फ अदालती मामलों पर ध्यान केंद्रित करने से हम यह सुनिश्चित करने वाले हैं कि अपेक्षाकृत कम मामले घटित हों, किंतु अधिकांश नारियों को जिस शत्रुतापूर्ण माहौल में जीवन यापन करना पड़ता है उसे परिवर्तित करने के लिए हम कुछ भी नहीं कर रहे होंगे।

मेरा अनुसंधान यह समझने की कोशिश है कि मापन की समस्या को प्रॉमिस के सभी पहलुओं को ध्यान में रखते

हुए कैसे कम करें। यह एक व्यापक एवं अस्पष्ट कार्य है, और संभवतः जीवन काल में यह समाप्त भी नहीं होगा। यहां पर, मैं कम से कम उन बातों का प्रदर्शन करना चाहूंगा जो हम कर सकते हैं।

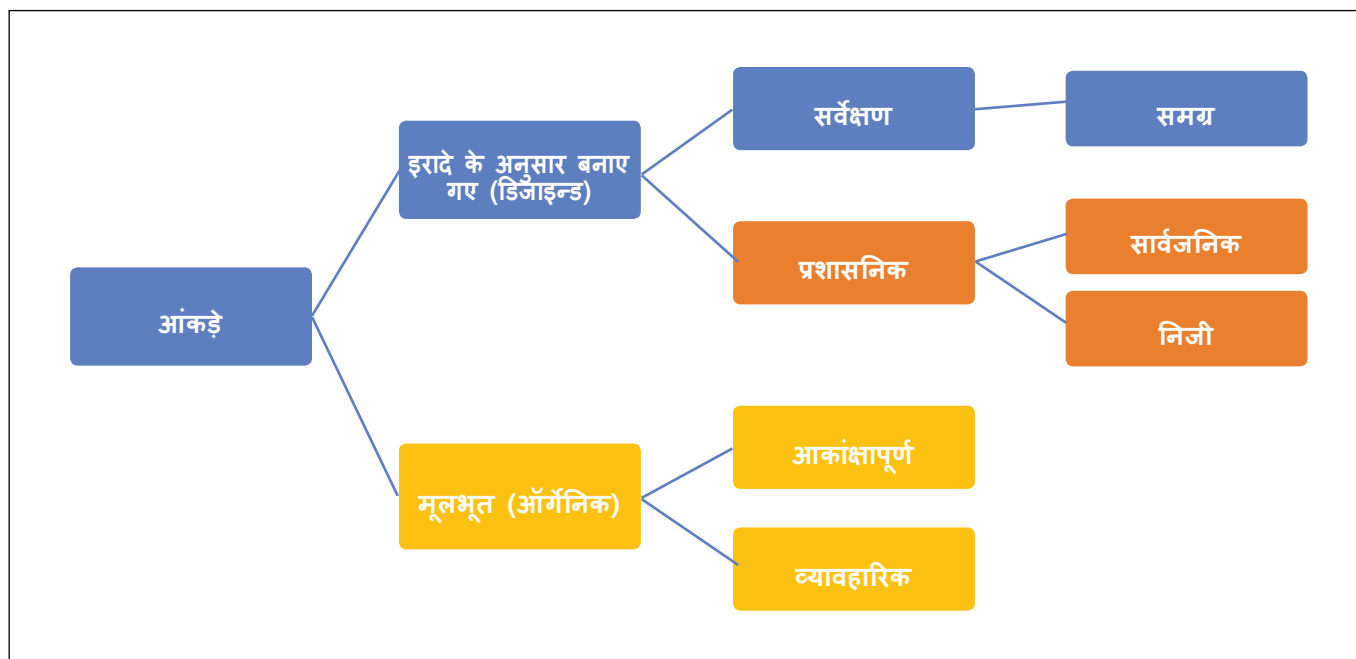
संबद्ध का अपूर्ण मापन

मैं उन सिद्धांतों को स्पष्ट करते हुए शुरुआत करना चाहूंगा जिनकी हम मैसाचुएट्स इंस्टीट्यूट ऑफ टेक्नोलॉजी (एमआईटी) में मापन के लिए तलाश कर रहे हैं - मैं इन्हें 'आधुनिक मापन' की आधारशिला (पिल्लर) कहता हूँ। पहली बात कि आंकड़ों के संग्रह के बारे में दो मार्गदर्शक सिद्धांत हैं : प्रथम सिद्धांत यह है कि यथा समय सूचक तैयार करने के लिए हमें निरंतर मापन की जरूरत होती है। दूसरा, आंकड़ा संग्रह अनुचित हस्तक्षेप रहित, या कम से कम यथासंभव अनुचित हस्तक्षेप रहित होना चाहिए। इसका तात्पर्य यह है कि आंकड़ों का संग्रह इलेक्ट्रॉनिक रूप में होने की बहुत संभावना है। आंकड़ों के संग्रह का यह एक मात्र रूप है जिसमें हम न्यूनतम हस्तक्षेप के साथ तेजी से आंकड़े एकत्र कर सकते हैं। अगले दो सिद्धांतों का संबंध सूचकों के प्रसार से है : पहला, कार्यप्रणालियाँ खुले स्रोत के रूप में होनी चाहिए, और माप किए जाने वाले व्यक्तियों, फर्मों तथा समुदायों को इसमें होने वाली गड़बड़ियों या इस संबंध में अपील करने की स्पष्ट और खुली प्रक्रिया उपलब्ध होनी चाहिए। दूसरा, आंकड़े जारी करने में निजता के संरक्षण की जरूरत होती है। चूंकि अभी हमारी सांख्यिकी काफी विलंब से तैयार की जाती है हमें यदाकदा ही महसूस होता है कि आंकड़ों का समेकन निजता के संरक्षण के लिए पर्याप्त है। हालांकि, जैसे ही आंकड़े लगभग वास्तविक समय पर जारी किया जाने लगेगा, निजता संरक्षण के लिए समेकन पर्याप्त नहीं होगा। निजता संरक्षण के लिए अन्य बेहतर प्रक्रियाओं को अपनाया जाना होगा। ये चार सिद्धांत हमें आधुनिक मापन की अंतिम विशेषता की ओर ले जाते हैं, जो यह है कि अपनी संरचना के अनुसार वे अपूर्ण ही रहने वाले हैं।

मैं परेशान नहीं हूँ। मैं सचमुच यह मानता हूँ कि किसी संबद्ध की अपूर्णता के साथ माप करना असंबद्ध की पूर्णता के साथ माप करते रहने से बेहतर है।

माप के इस नए सिद्धांत को साकार रूप देने के क्रम में हमें आंकड़ों के नए स्रोतों की ओर संभवतः नई प्रक्रियाओं की जरूरत है। बॉब ग्रोव्स (अमेरिकी जनगणना के पूर्व प्रमुख) के अनुसार आंकड़ों के दो व्यापक स्रोत होते हैं : इरादे के अनुसार बनाए गए (डिजाइन्ड) आंकड़े और मूलभूत (ऑर्गेनिक) आंकड़े। इनमें से दूसरे को अधिकांश लोग 'बिग डाटा' कहते हैं। इसलिए मेरा विचार है कि पहले वाले को स्माल डाटा कहा जाना चाहिए।

इरादे के अनुसार बनाए गए (डिजाइन्ड) आंकड़े वे आंकड़े होते हैं जो सर्वेक्षणों और प्रशासनिक रिकॉर्डों से लिए जाते हैं। सर्वेक्षणों का कोई उद्देश्य होता है या उसके प्रश्न



होते हैं, और प्रश्नावली इस प्रकार तैयार की जाती है कि इन प्रश्नों के उत्तर से आंकड़े जुटाए जाते हैं। इस परिप्रेक्ष्य में, प्रशासनिक रिकॉर्ड सर्वेक्षण का एक रूप होते हैं। जनगणना एक सर्वेक्षण है, वित्तीय विवरण सर्वेक्षण होते हैं, कर प्रपत्र सर्वेक्षण होते हैं, इत्यादि। उन सभी के अलग-अलग उद्देश्य होते हैं, जो क्रमशः इस प्रकार हैं : देश में व्यक्तियों की संख्या की गणना करना, फर्मों की वित्तीय स्थिति की तुलना करना और उसे समझना, तथा किसी व्यक्ति या फर्म की कर देयता समझना। हम व्यक्तियों और फर्मों को एक विशिष्ट तरीके से जानकारी भरने को कहते हैं, और तब राष्ट्रीय सांख्यिकी कार्यालय जानकारी तैयार करता है, जिसे देश में साझा किया जाता है। अंततः, इरादे के अनुसार बनाए गए (डिजाइन्ड) आंकड़े के निम्नलिखित लाभ होते हैं - ये प्रतिनिधि आंकड़े होते हैं, छोटे प्रतिदर्श चयन की समस्याएं होती हैं, इन्हें बेहतर ढंग से समझा जाता है, और इसलिए यह स्पष्ट होता है कि इनमें सुधार कैसे करें। इसकी बारंबारता नहीं होने के कारण इसमें निजता के संरक्षण की प्रवृत्ति होती है। दूसरी ओर, नुकसानपूर्ण स्थिति यह होती है कि आंकड़ों का संकलन महंगा पड़ता है और यह कि सर्वेक्षण बहुत अधिक अनुचित हस्तक्षेपकारी होते हैं - हस्तक्षेप इतना अधिक होता है कि हाल के दशकों में जवाब देने की दर घट रही है और लोग इन सर्वेक्षणों को इस कदर बोझ मानने लगे हैं कि आवश्यक नहीं है वे सच्चाई पूर्वक जवाब दे रहे हों। जनगणना में अभी-अभी शामिल किया गया नागरिकता से जुड़ा प्रश्न इसका एक उदाहरण है। वाणिज्य विभाग से एक चुक हुई और उसके अनुपात में हमें 2020 की जनगणना में अधिक संख्या में बुरे जवाबों का सामना करना पड़ेगा। समय और प्रयासों का यह कैसा अपव्यय है ?

मूलभूत (ऑर्गेनिक) आंकड़े वे आंकड़े होते हैं जो लोगों को इस बात का भान कराए बिना तैयार किए जाते हैं कि वे सर्वेक्षण का हिस्सा हैं। इनका संबंध निम्नलिखित से हो सकता है - आपके फोन में जीपीएस, वेब में आपकी सर्च, आपके नेटवर्क में दोस्त इत्यादि। प्रत्येक बार, जब भी कोई व्यक्ति इनमें से किसी माध्यम से कार्य करता है, वह व्यक्ति स्वयं के बारे में जानकारी उपलब्ध कराता है। उदाहरण के लिए, ऐसा यदा कदा ही होता है कि लोग विकिपीडिया में ऐसे ही कोई प्रश्न पूछते हों। आमतौर पर लोगों की रुचि जवाब में होती है। इसका मतलब यह है कि विकिपीडिया में की जाने वाली प्रत्येक सर्च, सर्च करने वाले व्यक्ति के बारे में विकिपीडिया कुछ बता देता है। उनके व्यक्तित्व के किसी छोटे पहलू के बारे में कुछ जानकारी देता है। जब आप किसी मूवी की पूरी देखते हैं, तो नेटफ्लिक्स जान जाता है कि आप या आपके घर में कोई उस तरह की मूवी में रुचि रखते हैं। पुनः, आपके रुझानों के बारे में छोटी कितु उपयुक्त जानकारी के बारे में पता चलता है।

दूसरे शब्दों में, जैसा कि सैंडी पेंटलैंड कहते हैं, इन संपर्कों के साथ हम अपनी पसंद के बारे में अवशेष छोड़ते जाते हैं। इस प्रकार से, अवशेषों को एकत्र करते हुए हम व्यक्ति के बारे में जानकारी जुटा सकते हैं। अंत में उस व्यक्ति के बारे में थोड़ी बातें जान सकते हैं। यह एक सर्वेक्षण की तरह है, किंतु व्यक्ति को यह पता नहीं चलता की उस पर सर्वेक्षण किया जा रहा है !

वस्तुतः, मूलभूत (ऑर्गेनिक) आंकड़े का सबसे बड़ा लाभ यह है कि वे गैर-बाधाकारी हैं और आंकड़े तैयार करने के संबंध में व्यक्ति की सच्चाई बचायी जाती है। हम

अपने जीपीएस से या गूगल से झूठ नहीं बोलते हैं, या नेटफ्लिक्स से गड़बड़ी की कोशिश नहीं करते हैं। हम अपने जीवन से संबंधित सच्चे अवशेष छोड़ जाते हैं। आंकड़ों के सभी स्रोतों के मामले में ऐसा नहीं होता है। उदाहरण के लिए, फेसबुक में लोग अपनी पहचान नहीं दर्शाते हैं, बल्कि यह दर्शाते हैं कि वे कैसे होना चाहेंगे। फिर भी, यह व्यक्ति की महत्वाकांक्षाओं के बारे में जानकारी तो देता है, किंतु यह नहीं पता चलता कि वे क्या हैं। अतः, सत्यता में अंतर होता है जो उनकी आकांक्षाओं की तीव्रता बनाम आंकड़ों के स्रोत में बदलाव करने के अनुसार बदलती रहती है। हालांकि, मूलभूत आंकड़ों की मुख्य समस्या उनके प्रतिनिधि होने और प्रतिदर्श चयन से संबंधित है। वेब पेजों में सभी लोग बराबर से सहभाग नहीं देते हैं। इसका अर्थ यह होता है कि आंकड़ों के अनुमानों और निष्कर्ष निकालना मुश्किल होता है। इसके अलावा, अनुमानों के भेदभावपूर्ण होने की संभावना है तथा ये मॉडल अस्थिर एवं अनिश्चित होंगे।

मेरा मानना है कि परिमाण, वेग एवं विविधता अपने आप में अप्रासंगिक होते हैं, इसे नोट किया जाए। वे तभी सार्थक होते हैं जब किसी प्रासंगिक प्रश्न का उत्तर ढूँढने के लिए उनका प्रयोग किया जा सके। दूसरे शब्दों में, आंकड़ों के आकार से अनुमान की त्रुटि कम हो जाती है, किंतु अभिन्न त्रुटि को कम नहीं किया जा सकता। अतः, अनुसंधानकर्ता गलत बातों का अधिक सटीकतापूर्वक अनुमान करते रहते हैं। पुनः, प्रतिनिधिक होने अथवा प्रतिदर्श चयन की कमी की समस्या आदर्श अनिश्चितता को प्रथम श्रेणी का मुद्दा बना देती है। इसलिए, “बिग डाटा” - या मूलभूत डाटा जैसा कि मैं इसे कहना पसंद करता हूँ, के संबंध में कार्य करना सिर्फ सटीकता की समस्या को सुलझाता है। गलत विवरण, आदर्श अनिश्चितता, और आदर्श अस्थिरता के कारण पूर्वाग्रह और अधिक बढ़ जाता है।

दो निष्कर्ष : पहला, किसी भी प्रकार का आंकड़ा पूर्ण नहीं होता है, इसलिए मैं मानता हूँ कि भविष्य में आंकड़ा संग्रह की कार्यप्रणाली एक संकर उपागम को अपरिहार्य बनाती है। हमें डिजाइन किए हुए आंकड़ों को लेते हुए इसे ऑर्गेनिक आंकड़ों से एकीकार करने की जरूरत है, ताकि स्माल डाटा के साथ बिग डाटा में बेहतर लाई जा सके। दूसरे शब्दों में, बिग डाटा से उत्पन्न पूर्वाग्रह को दूर करने के लिए स्माल डाटा का प्रयोग करें। दूसरा, जैसा कि मैंने पहले कहा, किसी संबद्ध बात की अपूर्णतापूर्वक माप करना किसी असंबद्ध बात की पूर्ण माप करने से कहीं बेहतर है।

मैं इस चर्चा में ऐसे तीन उदाहरण प्रस्तुत करूंगा जो इस संबंध में होंगे कि हम मापन को कैसे बेहतर बना सकते हैं, और ऐसा करने के लिए हम आंकड़ों के स्रोतों को एकाकार कर रहे हैं।

बिलियन मूल्यों की परियोजना

बिलियन मूल्य की परियोजना का उद्देश्य महंगाई की वैकल्पिक माप की गणना करना है। दैनिक महंगाई दरों की गणना करने के लिए हम इंटरनेट से प्राप्त मूल्य के आंकड़ों का प्रयोग करते हैं। ‘वेबस्क्रेपिंग’ नामक प्रक्रिया का इस्तेमाल करना हमने 14 वर्ष पहले प्रारंभ किया था। सामान्य शब्दों में, वेबस्क्रेपिंग का उद्देश्य मानक वेब पेज में दर्शाई गई जानकारी को लेते हुए उसे डाटाबेस के रूप में व्यवस्थित करना है। कृपया जेईपी में हमारे लेख (अलबर्टो कैवाल्लो के साथ) का अवलोकन करें, जिसमें हमने सूचकांकों के तैयार किए जाने के विवरण का उल्लेख किया है।

हम लोगों ने खाद्य स्फीति पर कार्य 2006 में प्रारंभ किया। आज की स्थिति में वेब पर बहुत से क्षेत्रों का बहुत बढ़िया प्रतिनिधित्व है और हमारे कार्यक्षेत्र में बहुत अधिक विस्तार हुआ है। वस्तुतः, 2016 में प्राइस वाटर हाउस कूपर्स द्वारा जिन 26 देशों में यह सर्वेक्षण किया जा रहा था, उन्होंने ऑनलाइन खरीद पर दिलचस्पी दिखाई है। परिमाण की दृष्टि से आज की स्थिति में इसकी दखल 10 वर्ष पूर्व की हमारी स्थिति के तुल्य है। नीचे दी गई तालिका में सभी देशों में प्राप्त उत्तरों का सारांश प्रस्तुत किया गया है, और प्रश्न यह है कि वे खरीद किस जगह पर करना पसंद करेंगे। मैंने इस जानकारी के ढंग को बहुत रोचक पाया। उनके पास जानकारी देश वार उपलब्ध है।

जानकारी की गणना हेतु हमें दो प्रकार की सूचना की आवश्यकता होती है : मूल्य और बास्केट। बास्केट उन भारों (वेट्स) का संग्रह है जो देश के उपभोग समूहों (बंडल) का प्रतिनिधित्व करने के मन्तव्य के साथ तैयार किया जाता

ऑनलाइन बनाम इन-स्टोर खरीद को प्राथमिकता

ऑनलाइन		इन-स्टोर
60%	पुस्तकें, संगीत, फिल्में एवं वीडियो गेम	28%
39%	खिलौने	37%
43%	उपभोग की इलेक्ट्रॉनिक वस्तुएं एवं कंप्यूटर	51%
36%	खेल-कूद के उपकरण/आउटडोर	44%
37%	स्वास्थ्य एवं सौंदर्य (प्रसाधन सामग्री)	47%
40%	पहनने के कपड़े एवं फुटवियर	51%
32%	आभूषण/घड़ियां	49%
33%	हाउसहोल्ड उपकरण	56%
30%	डीआईवाई/भवन सुधार	52%
30%	फर्नीचर एवं घरेलू सामग्री	59%
23%	किराने का सामान	70%

है। सांख्यिकी कार्यालय भारों का एकत्रण उपभोक्ता सर्वेक्षणों के माध्यम से करता है ताकि यह निर्धारित किया जा सके कि टमाटरों, प्याजों, आवास, शिक्षा, आईफोनों पर कितना व्यय किया जाता है। उनके यह समझने के बाद कि उन मूल्यों को भार किस तरह से प्रदान किया जाना चाहिए, वे प्रत्येक माह (अधिकांश देशों में) आंकड़े एकत्र करते हैं और बास्केट की लागत में परिवर्तन की गणना करते हैं।

कीमत तो ऑफलाइन और ऑनलाइन स्टोर्स से एकत्र की जा सकती हैं। हालांकि, ऑनलाइन डाटा संग्रह के प्ररप्रेक्ष्य में तीन प्रकार के क्षेत्र हैं - वे क्षेत्र जिनके बारे में उपलब्ध ऑनलाइन जानकारी भयावह है; वे क्षेत्र जिनके बारे में ऑनलाइन उपलब्ध जानकारी प्रतिनिधिक है और सांख्यिकी कार्यालय द्वारा एकत्र भार उपभोक्ता बास्केट के अनुमान के बारे में बहुत अच्छी जानकारी उपलब्ध कराते हैं; और वे क्षेत्र जिनके बारे में ऑनलाइन उपलब्ध आंकड़े ऑफलाइन की तुलना में काफी बेहतर हैं, ऐसा विशेष रूप से खरीदे गए उत्पादों की संरचना के संबंध में जानकारी के मामले में है।

उदाहरण के लिए, शिक्षा के मामले में ऑनलाइन एकत्र की गई जानकारी भयावह है। सिर्फ निजी ट्यूशन एवं शिक्षा की लागत को प्रभावित करने वाले कुछ कारक, यथा गणवेश एवं पुस्तकों की जानकारी ही ऑनलाइन एकत्र की जा सकती है। खाद्य क्षेत्र दूसरे प्रकार के क्षेत्र का उदाहरण है। टमाटरों और प्याजों के बीच हिस्से का एकत्रण सांख्यिकी कार्यालय द्वारा काफी अच्छा किया जाता है। अतः, इनमें भार का अनुमान उच्च गुणवत्ता वाले आंकड़ों के आधार पर लगाया जाता है। खाद्य क्षेत्र में, बहुत से देशों में ऑनलाइन उपलब्ध कीमतें बहुत प्रतिनिधिक होती हैं। मदवार मुद्रास्फीति की गणना करने के लिए हम वेब से दैनिक कीमतें ले सकते हैं और आधिकारिक भारों का प्रयोग करते हुए उनका समूहन कर सकते हैं। अंततः, इलेक्ट्रॉनिक्स तौसरे प्रकार का उदाहरण है। आईफोन एक्स और 8 पहले से बाजार में उपलब्ध हैं और अधिकांश सांख्यिकी कार्यालय अभी भी आईफोन 7 की जानकारी जुटाने में लगे हुए हैं। हमें इन क्षेत्रों में बास्केटों की गणना करने और ऑनलाइन स्रोतों से मूल्य एकत्र करने की आवश्यकता है।

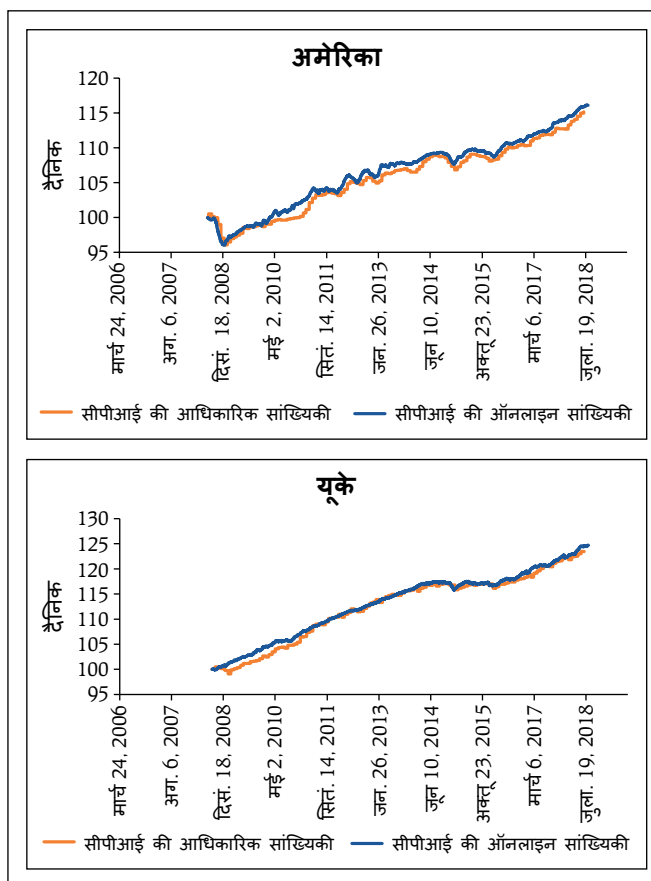
जिन क्षेत्रों के बारे में हमारे पास बैड डाटा ही उपलब्ध हैं, हम दैनिक आधार पर अनुमान लगाते हैं कि सांख्यिकी कार्यालय से क्या रिपोर्ट किए जाने की संभावना है - इसलिए, यह वेब में उपलब्ध दैनिक मूल्यों का प्रयोग करते हुए अनुमान लगाने की गतिविधि है। दूसरे प्रकार के क्षेत्रों के लिए हम कीमतें एकत्र करते हैं और उसके बाद आधिकारिक भारों का प्रयोग करते हैं (ये सबसे सरल होते हैं)। तीसरे प्रकार के क्षेत्र के लिए हम वेबपेजों में दर्शाई गई सूचना का प्रयोग करते हुए भारांक का अनुमान लगाते हैं - निश्चितरूप से यहां पर विशिष्ट प्रकार की त्रुटि की शुरुआत होती है।

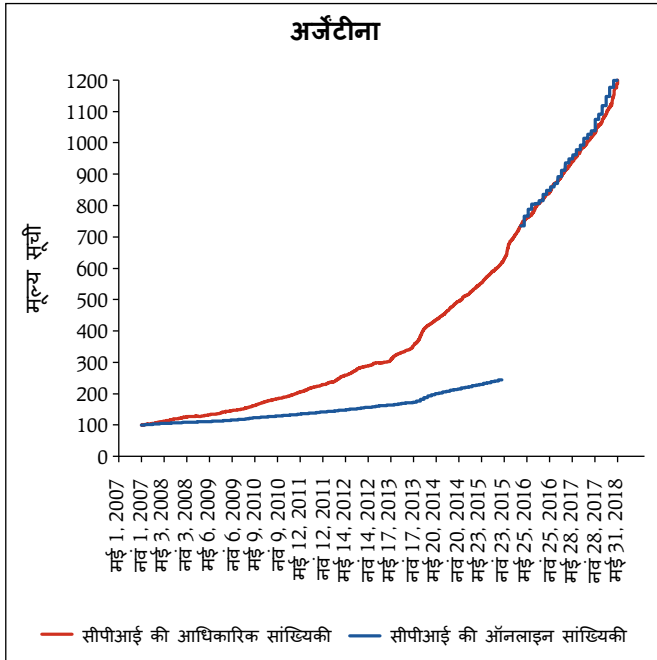
हमारी मुद्रास्फीति सूचियां आधिकारिक सूचियों के समान नहीं होती हैं। उनके अलग-अलग होने के बहुत से कारण हैं। दो बहुत अंतर्दर्शी स्रोत हैं : पहला, ऑनलाइन

और ऑफलाइन बाजार भिन्न-भिन्न हैं और इसलिए अल्प समयान्तराल में मुद्रास्फीति एक समान नहीं होना चाहिए। संभव है कि दीर्घवधि में ये समान हों, किंतु महीना दर महीना की बारंबारता में तो निश्चितरूप से ऐसा नहीं होगा। ऑनलाइन उपभोक्ताओं के अधिक धनवान और कम धैर्यवान (कम उम्र वाले) होने की प्रवृत्ति होती है, और ऑनलाइन बाजार में अधिक प्रतिस्पर्धा की प्रवृत्ति भी होती है। दूसरा, हमारी प्रक्रिया भी भिन्न होती है - इसलिए मुद्रास्फीति दरें अलग-अलग होना आश्चर्य की बात नहीं होना चाहिए। हालांकि, 14 वर्षों से हम दैनिक मुद्रास्फीति दरों की गणना करते आए हैं, किंतु उनका आधिकारिक दरों के बहुत करीब होना बहुत उल्लेखनीय रहा है।

अगले तीन ग्राफों में, नारंगी रेखा आधिकारिक सांख्यिकी उपभोक्ता मूल्य सूचकांक (सीपीआई) को और नीली रेखा ऑनलाइन दैनिक उपभोक्ता मूल्य सूचकांक को दर्शाती है। यूएस और यूके को देखें तो हम पाएंगे कि ऑनलाइन और ऑफलाइन मुद्रास्फीति एक दूसरे के बहुत करीब हैं।

सभी देशों में ऐसा नहीं दिखाई देता है। हम लोगों ने अर्जेंटीना में 2007 में शुरुआत की क्योंकि सांख्यिकी कार्यालय में 2006 में हस्तक्षेप किया गया। उस समय के दौरान हमारी विसंगतियां बहुत अधिक थीं। 9 वर्षों की आधिकारिक संचयी मुद्रास्फीति 120 प्रतिशत है, किंतु हमारे आंकड़े इसे 640 प्रतिशत दर्शाते हैं। रोचक बात है कि मई





2016 से, जब से सांख्यिकी कार्यालय ने बिना किसी जोड़-तोड़ के मुद्रास्फीति की गणना प्रारंभ की है, आधिकारिक और ऑनलाइन मुद्रास्फीति काफी करीब रही हैं (इस ग्राफ में आधिकारिक सोपीआई को हमें पुनः सामान्य बनाने की आवश्यकता पड़ी ताकि मई 2016 में इसे ऑनलाइन स्फीति के अनुरूप बनाया जा सके - हालांकि, पुनः सामान्य बनाए जाने के बाद दोनों एक दूसरे के अनुरूप हैं)। 120 बनाम 640 प्रतिशत ! यह जोड़-तोड़ का वह आकार है जिसके अंतर्गत किर्चनर सरकार ने उनके नागरिकों पर करारोपड़ किया। मजदूरी को आधिकारिक मुद्रास्फीति के साथ सूचीबद्ध किया गया है, पेंशन को आधिकारिक मुद्रास्फीति के साथ सूचीबद्ध किया गया है किंतु जीवन यापन के मानक वास्तविक महंगाई से प्रभावित होते हैं। 500 प्रतिशत का स्वामित्वहरण पूर्णतः झूठ पर आधारित है। मुझे खुशी है कि नई सरकार ने केंद्र बिंदु पर ध्यान केंद्रित किया।

अर्थव्यवस्था के हालात तीव्रतापूर्वक तक काफी परिशुद्धता से जानने के लिए ऑनलाइन कीमतों का प्रयोग किया जा सकता है। इसके अलावा, ऑनलाइन कीमतों में ऑफलाइन कीमतों की तुलना में तेजी से परिवर्तन की प्रवृत्ति होती है (बाजार अधिक प्रतिस्पर्धी है) और इसलिए ऑनलाइन माध्यम को आधिकारिक सूचनाओं के लिए प्रयोग में लाया जा सकता है। विकसित देशों में 2 से 3 महीनों का पूर्वानुमान लगाया जाता है, जबकि उभरते बाजारों में लगभग 1 से 2 महीने का ही अनुमान लगाया जाता है।

हम लोग लगभग 20 देशों के लिए मुद्रास्फीति के आंकड़े तैयार करते आए हैं। और यह घोषित करते हुए मुझे बहुत खुशी हो रही है कि इस वर्ष के अंत तक हम भारत की मुद्रास्फीति की गणना भी करने लगेंगे।

थाउजेंड बिग मैक इन्डैक्स

मैं दूसरी परियोजना का रुख करना चाहूंगा : वास्तविक विनिमय दर या क्रय शक्ति साम्यता (पीपीपी) के मापन की कोशिश करना।

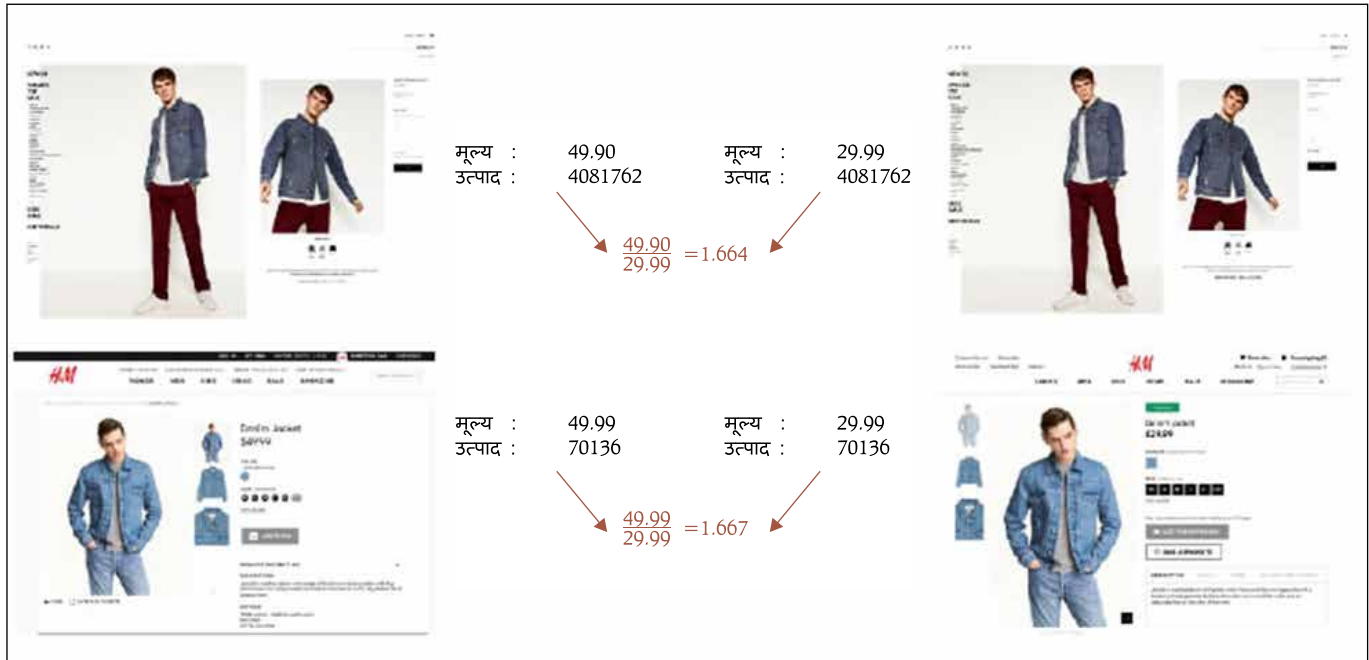
पीपीपी के मामले के दो मानक हैं। एक तो विश्व बैंक (डब्लूबी) का बड़ा प्रयास है जो पीपीपी के अनुसार समायोजित माप तैयार करने के लिए 105 सांख्यिकीय अधिकारियों को एक साथ काम पर लगाए हुए है। और हमारे पास द इकोनॉमिस्ट का बिग मैक इन्डैक्स है जो संभवतः अब तक तैयार की गई सबसे आसान सांख्यिकी है। निराशा की बात है कि विश्व बैंक की नज़र में द बिग मैक महज पीपीपी के समान ही है।

बिग मैक इन्डैक्स की आलोचना करने के बारे में बहुत सी बातें हैं। इसके बारे में आपकी राय से इतर, इसकी सरलता बहुत बेहतरीन है। यह पूरे संसार के लिए एक समान है। मुझे पता है कि अब इसका इस्तेमाल कोई नहीं करता, किंतु यह वही पुराना इन्डैक्स है। डब्लूबी डाटा में एक समान मर्दों का मिलान करना बड़ी समस्या रही है। डब्लूबी का लाभ यह है कि इसमें एक से अधिक मर्दें होती हैं। परन्तु इसकी लागत और जटिलता के कारण इसको समझना और कठिन हो जाता है, और इससे भी खराब बात यह है कि इन कारणों से यह अप्रासंगिक भी हो सकते हैं।

अलबर्टो कैवालो और मैं ने इन दो सूचकांकों के बीच किसी सूचकांक तैयार करने का निर्णय लिया। हमारा उद्देश्य दुनिया भर के उन एक समान उत्पादों की तुलना करना, जिनकी सक्रियतापूर्वक खरीद की जा रही हो, और सूचकांक का आधार एक से अधिक मद को बनाना था।

यह हम कैसे करते हैं ? वेबपेजों में उपलब्ध एक बहुत महत्वपूर्ण जानकारी आईडी की मद से संबंधित है। उसी खुदरा विक्रेता (वेबपेज) के लिए आईडी की मद से संबंधित उत्पाद की विशिष्ट जानकारी सभी विभिन्न भौगोलिक स्थानों, अर्थात् देशों के संबंध में उपलब्ध होती है। यह यूपीसी या बारकोड कोड से बेहतर है क्योंकि वे कोड देश विशेष से संबंधित हैं। वस्तुतः ये भाषा विशेष से संबंधित हैं - कनाडा में कोक की किसी बोतल का यूपीसी भिन्न : या तो फ्रांसीसी या अंग्रेजी में, होता है जो पैकेजिंग पर निर्भर होता है। स्टोर्स इन मर्दों की आईडी का प्रयोग अपनी आपूर्ति शृंखला के प्रबंधन में करते हैं, और इसलिए एक समान मर्दों को ऐसा ही - एक समान- मानते हैं, चाहे किसी में क्रिसमस का लेबल लगा हो या दूसरे में नहीं लगा हो। हम इन मर्दों की आईडी से दुनिया भर की हजारों मर्दों का मिलान और उनकी तुलना कर सकते हैं।

उदाहरण के लिए, पारदर्शिता के संदर्भ में धरती पर उपलब्ध दो सबसे बेहतरीन वेबपेज ज़ारा और एचएंडएम हैं। मैं उनके उत्पाद पसंद भी करता हूँ। पुरुषों के लिए, फास्ट

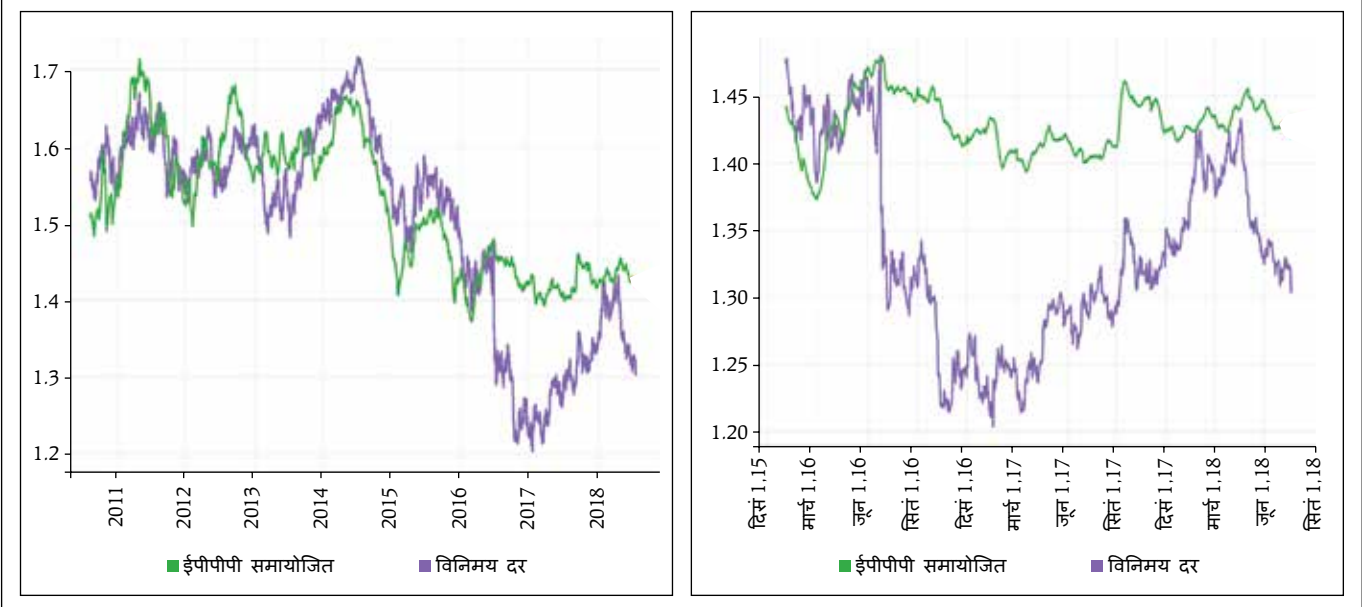


फैशन, यंग, जैकेट, डेनिम जैसी चीजों के लिए हम सामानों की खोज दो अलग-अलग देशों के स्टोर्स में भी कर सकते हैं (मैंने यहां पर यूएस को बाईं ओर तथा यूके को दाईं ओर दिखाया है)। हम लोगों ने कर को अलग किया और मूल्यों के अनुपात की गणना की। यह एचएंडएम जैकेट की अंतर्निहित सांकेतिक विनिमय दर होती है। प्रत्येक मद के संबंध में ऐसा करने के बाद हम उन हजारों उत्पादों पर काम करते हैं, प्रत्येक श्रेणी के अंतर्गत उनका भार तय करते हैं और देश के संबंध में सूची तैयार करते हैं। देखिए, द बिग मैक इंडेक्स का नया वर्धित रूप आपके समक्ष है।

इस सूचकांक की बहुत सी विशेषताएं हैं। इसका कोई रखरखाव नहीं करना पड़ता, इसमें एक भी गैर-व्यापार योग्य उत्पाद नहीं है, और लगभग सभी सामानों का व्यापार अंतरराष्ट्रीय स्तर पर होता है। इसके अलावा, ये मदें वे हैं जो लोग सचमुच में खरीदते हैं। इलेक्ट्रॉनिक्स, कपड़ों, निजी देखभाल की वस्तुओं, गैसोलिन इत्यादि पर विचार करें। हमारी सूची में इन मदों को शामिल किया गया है।

मैं, नीचे दिए गए आरेख में यूके-यूएस के सूचकांकों को दर्शा रहा हूं। बैंगनी रेखा सांकेतिक विनिमय दर (वित्तीय

सांकेतिक विनिमय दर बनाम पीपीपी



बाजारों से) है, जिसमें गिरावट मूल्यहास को इंगित करती है। हरी रेखा तुलनात्मक मूल्यों (यूएस को यूके से विभाजित किया गया) को इंगित करती है। इसलिए, नीचे की ओर जाने का तात्पर्य इस एक समान व्यापार योग्य उत्पाद समूह के लिए यूके में महंगाई दर अमेरिका में महंगाई दर की तुलना में अधिक है।

दो बातें महसूस करने की हैं। पहली, ब्रेक्सिट के पहले यह उल्लेखनीय है कि दोनों रेखाओं में किस प्रकार समानता है। वस्तुतः, यह बताना मुश्किल है कि किस बात का कारण क्या है उस समय विनिमय दर में बहुत अधिक उतार-चढ़ाव (10 प्रतिशत से अधिक) हो रहे थे और खुदरा विक्रेताओं ने उन प्रवृत्तियों का करीब से अनुसरण किया। दूसरी, रोचक ढंग से वह पैटर्न ब्रेक्सिट के समय बदल गया।

दाईं ओर दर्शाया गया आरेख ब्रेक्सिट के मतदान के बाद की महज सूचियां हैं। काफी अधिक मूल्यहास हुआ है (जून में ब्रेक्सिट मतदान होने, और अक्टूबर में थेरसा के यह कहने के बाद कि वह धारा 50 को लागू करने जा रही हैं) किंतु यूके के खुदरा विक्रेता यूएस की तुलना में अधिक तेजी से मूल्य नहीं बढ़ाते हैं। आखिर क्यों ? इसकी दो व्याख्या संभव हैं : या तो उपभोक्ताओं में मुद्रास्फीति के संदर्भ में परिणामों की बहुत अधिक जागरूकता थी और वे कीमत वृद्धि (टोब्लेरोन का मामला इसका प्रमाण है) को स्वीकार करने के इच्छुक नहीं थे या मांग इतनी कम थी कि मूल्य वृद्धि होने पर मांग बनी नहीं रह सकती थी। दोनों ही मामलों में बड़ा अंतराल तैयार होता है। आयात योग्य सामान (इनमें से अधिकांश वस्तुएं आयातित हैं) की मांग

अपेक्षाकृत कम रहने के कारण चालू खाते में सुधार होने की जरूरत है (जो सचमुच में हुआ) और सांकेतिक विनिमय दर में सुधार होना ही चाहिए (जो स्पष्ट रूप से देखा जा सकता है कि ऐसा हुआ)। यहां पर हम कोई भविष्यवाणी नहीं कर रहे हैं। हम सिर्फ यूएस की तुलना में यूके में आयात के लिए समग्र मांग की स्थिति की माप कर रहे हैं। यह होता है, और अधिकांश आर्थिक मॉडल यह भविष्यवाणी करेंगे कि मांग में इस तरह के परिवर्तन का चालू खाते एवं विनिमय दर पर प्रभाव पड़ने की आवश्यकता होती है।

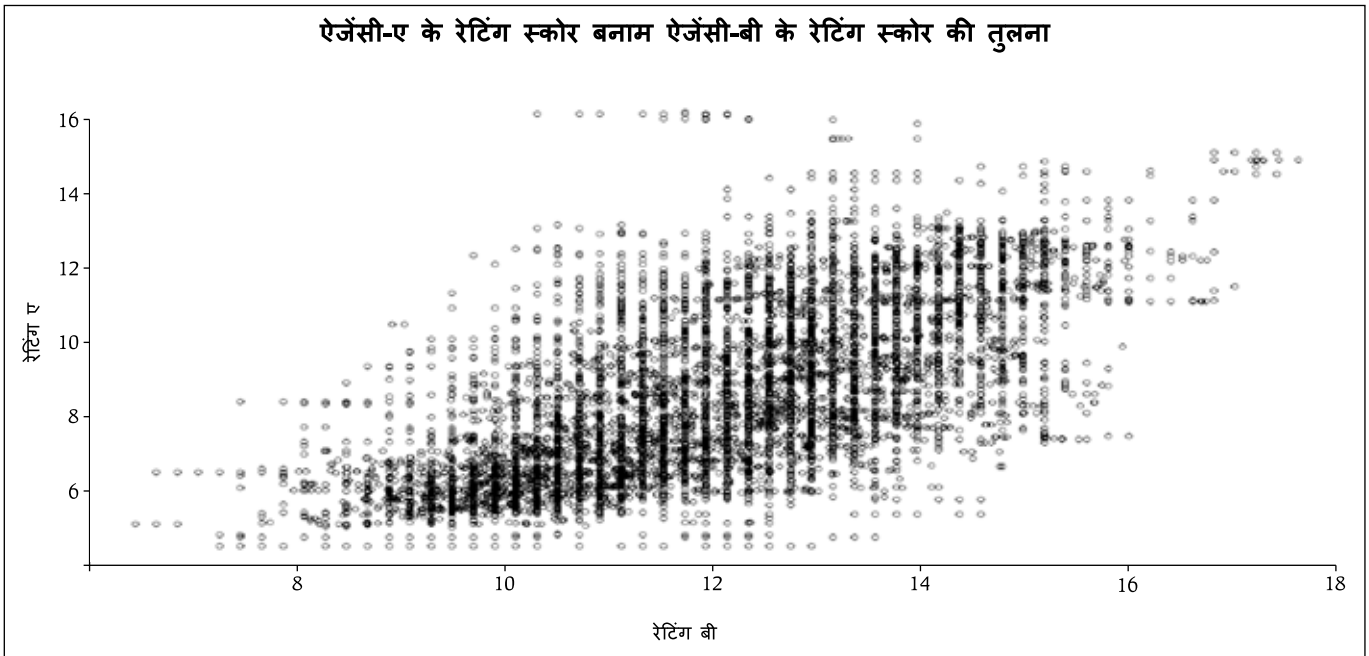
हमारे पास अभी तक सिर्फ नौ देश ही हैं, किंतु और देश इसमें शामिल होंगे।

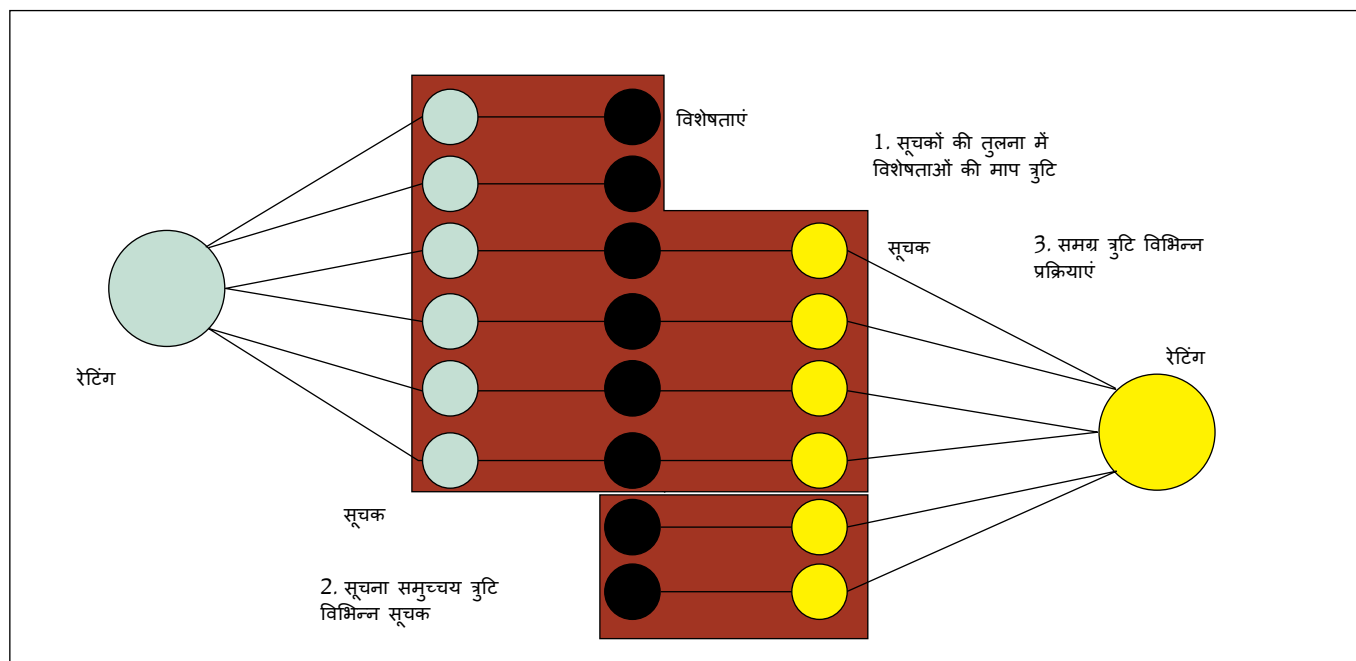
समग्र अव्यवस्था परियोजना (कन्फ्यूजन प्रोजेक्ट)

अंततः, मेरी नई परियोजना फर्मा के नैतिक व्यवहार की माप के संबंध में है : इसे समग्र अव्यवस्था परियोजना कहा जाता है और यह कार्य जूनिअनकोइल्बेल एवं फ्लोरिअन बर्ग के साथ संयुक्त कार्य है। यह महज शुरुआत है (2 से 3 वर्ष हो रहे हैं) ।

एक दशक से अधिक समय से, वित्तीय बाजारों की रुचि पर्यावरणीय, सामाजिक एवं अभिशासन (ईएसजी) के बहुत से पहलुओं का फर्मा पर पड़ने वाले प्रभाव का समझने में रही है। कारण यह है कि निवेशक यह चाहेंगे कि उनके निवेशों का समाज पर सकारात्मक प्रभाव पड़े। इसके कारण फर्मा के बर्ताव की माप करने के मन्तव्य के साथ डाटा उपलब्ध कराने वालों के विस्तार की शुरुआत हुई है।

हालांकि, डाटा उपलब्ध कराने वालों ने अविश्वसनीय ढंग से एकरूप रेटिंग तैयार किया है। उदाहरण के लिए, नीचे





दिए गए रेखा-चित्र में हम यह दर्शाए हैं कि ईएसजी की शीर्ष दस रेटिंग एजेंसियों में से दो एजेंसियों द्वारा उसी फर्म को रेटिंग दी गई है। जैसा कि देखा जा सकता है, सहसंबंध धनात्मक है हालांकि यह बहुत कम है (30 प्रतिशत से कम)।

असंगति का स्रोत क्या है ? इस उद्देश्य के लिए, हम लोगों ने कच्चे आंकड़े बहुत सी रेटिंग एजेंसियों से लिया तथा उनकी रेटिंग प्रक्रिया पर विपरीत दिशा में कार्य किया। हम लोगों ने तीन स्रोतों में गलतियों को सुधारा। पहला, उन विशेषताओं का वर्गीकरण किया जिनकी माप की जानी चाहिए। उदाहरण के लिए, कार्बन डाइऑक्साइड का उत्सर्जन, जल प्रबंधन, श्रमिक प्रथाएं, महिलाओं का सशक्तिकरण, बोर्ड की संरचना इत्यादि / इसके बाद, हम लोगों ने श्रेणियों अथवा विशेषताओं के विभिन्न संकेतक निर्धारित किया। पहला कदम उन संकेतकों के बीच असंगति को समझना था, जो उन विशेषताओं को समाहित करते हैं। हम लोग इसे 'माप की त्रुटि' कहते हैं। एक बार साझा विशेषताओं की तुलना कर लेने के बाद हम लोगों ने यह नोट किया कि रेटिंग एजेंसियां विभिन्न विशेषताओं पर अपना ध्यान केंद्रित करती हैं। यह संभव है कि कोई रेटिंग एजेंसी किसी एक विशेषता की भी माप नहीं करती हो। बेशक, इसके कारण असंगति उत्पन्न होती है। इस स्रोत को हम लोग "सूचना समुच्चय त्रुटि" कहते हैं। अंततः, हम समेकन के विभिन्न नियमों का अवलोकन करते हैं और वे नियम भी असंगति उत्पन्न कर सकते हैं। इसकी पहचान "समेकन त्रुटि" के साथ की जाती है।

आंकड़ों का एक बार अवलोकन करने और चार रेटिंग एजेंसियों के बीच असंगतियों का विश्लेषण करने के बाद हम लोगों ने पाया कि असंगतियां लगभग 70 प्रतिशत माप, 25 प्रतिशत सूचना समुच्चय, और 5 प्रतिशत समेकन नियमों के कारण उत्पन्न हो रही थीं। ये अभी भी शुरुआती परिणाम हैं। अंतिम लक्ष्य ईएसजी की बेहतर माप करने में सक्षम होने का है, जो स्वतंत्र हों और सभी लोग उसका उपयोग कर सकें।

कुछ सीख

अंत में, मैं उन कुछ मिली सीख का संक्षेप में उल्लेख करना चाहूंगा जो इन वैकल्पिक आंकड़ों के स्रोत पर काम करते हुए प्राप्त हुईं। उम्मीद है कि ये सीख केंद्रीय बैंकों के लिए महत्वपूर्ण हैं।

पहली सीख, मूलभूत आंकड़े का सबसे बड़ा फायदा इनकी सत्यता होती है। यह फायदा इसके आकार में नहीं बल्कि गति, विविधता इत्यादि में निहित होता है। यह तथ्य है कि लोगों में सत्यतापूर्वक उत्तर देने की प्रवृत्ति होती है और मेरा मानना है कि इसके कारण ये आंकड़े अद्वितीय होते हैं। इस संबंध में, आकांक्षापूर्ण के स्थान पर व्यावहारिक आंकड़े एकत्र करने से प्रत्येक प्रेक्षण के सूचना-परक मूल्य में वृद्धि होती है। इसलिए, क्रेडिट कार्ड से होने वाले लेन-देन या जीपीएस लोकेशन से संबंधित आंकड़े ट्वीट किए जाने वाले अथवा फेसबुक स्टेटस के आंकड़ों की तुलना में अधिक सार्थक होते हैं।

दूसरी सीख, मूलभूत आंकड़े को भूगोल या सामाजिक-आर्थिक परिस्थितियों के अनुसार व्यवस्थित नहीं किया जाता

है। इसे बर्ताव के अनुसार व्यवस्थित किया जाता है। हम लोग भूगोल एवं सामाजिक परिस्थितियों पर ध्यान केंद्रित करते हुए सदियों से (वस्तुतः) सर्वेक्षण का संग्रह करते रहे हैं। सांख्यिकी कार्यालयों के लिए अर्थव्यवस्था और समाज को समझने के लिए यही प्रतिमान मार्गदर्शी ढांचे के रूप में उपलब्ध हैं। मेरी समझ से इस मत को चुनौती दी जाएगी।

तीसरी सीख, मूलभूत आंकड़े का सबसे नकारात्मक पहलू यह है कि ये आंकड़े प्रतिनिधिक नहीं होते हैं। व्यापक अनुमान लगाना असंभव है - हालांकि, आंकड़ों के इन स्रोतों पर आधारित गणना के सहसंबंधों के आधार पर व्यापक अनुमान लगाना लोगों को बहुत पसंद है। गैर-प्रतिनिधिकता तथा बहुत बड़े आंकड़ों के समुच्चय का संयोग एक खतरनाक घटना है। लोग बहुत अधिक परिशुद्धता पूर्वक भयानक समाश्रयण का अनुमान लगाते हैं। किसी तुच्छ बात पर जोर देना महत्वपूर्ण हो सकता है - किंतु यह बात आमतौर पर भुला दी जाती है। आंकड़े के आकार के कारण अनुमान की त्रुटि कम हो जाती है, किंतु पूर्वाग्रह समाप्त नहीं होते हैं !

चौथी सीख, आंकड़े के तात्कालिक होने की स्थिति में, उनके समेकित होने पर भी, निजता भंग होना अपरिहार्य

हो सकता है। हमें इन समस्याओं के बारे में जागरूक होने की जरूरत है क्योंकि हम लोगों ने कभी इनका सामना नहीं किया है। हम लोगों ने समुच्चयों के अपने तिमाही अनुमानों के आधार पर खुद को सुरक्षित महसूस किया है। इसके, उप-समुच्चय के दैनिक अनुमानों में परिवर्तित होने पर हमारी निजता संबंधी जरूरतों की चिंताएं अधिक होनी चाहिए।

अंततः, क्या आंकड़ों के नए स्रोत सांख्यिकी कार्यालयों को प्रतिस्थापित करने वाले हैं ? नहीं, कोई संभावना नहीं है। यह उनको परिवर्तित होने के लिए बाध्य करेगा, किंतु वे समाप्त नहीं होंगे। भविष्य में सांख्यिकी कार्यालयों को उन्नत उपागमों को अपनाना होगा। आंकड़ों का कुछ हिस्सा मूलभूत आंकड़ों के प्रयोग से जुटाया जाएगा और कुछ हिस्सा सर्वेक्षणों के माध्यम से। बहुत महत्वपूर्ण बात है कि मूलभूत आंकड़ों में अंतर्निहित पूर्वाग्रहों में सुधार करने के लिए सर्वेक्षणों का उपयोग किया जाने वाला है। सार रूप में, बड़े, तीव्र किंतु गैर-प्रतिनिधिक आंकड़ों की मरम्मत करने के लिए छोटे, धीमे किंतु प्रतिनिधिक आंकड़ों का प्रयोग किया जाएगा।