

BETAM ÇALIŞMA TEBLİĞLERİ SERİSİ

TÜRKİYE HANEHALKI İŞGÜCÜ ANKETİ MİKROVERİLERİNDEN
İL KİMLİKLERİNİN GERİ KAZANILMASI

Luis Pinedo Caro

Çalışma Tebliği

#202603

Bahçeşehir Üniversitesi Ekonomik ve Toplumsal Araştırmalar Merkezi (BETAM)
Osmanpaşa Mektebi Sokak No:4-6 34353 Beşiktaş, İSTANBUL, TÜRKİYE
Telefon: +902123815962

E-mail:

betam@eas.bau.edu.tr

11 Eylül 2026

BETAM çalışma tebliğleri yalnızca tartışma ve görüş alma amacıyla yayımlanmaktadır. Bu çalışmalar hakem değerlendirmesinden geçmemiştir. Burada ifade edilen görüşler yazar(lar)a ait olup BETAM'ın görüşlerini yansıtmamaktadır.

Türkiye Hanehalkı İşgücü Anketi Mikroverilerinden İl Kimliklerinin Geri Kazanılması

Luis Pinedo Caro¹

Bahçeşehir Üniversitesi

Özet

Türkiye Hanehalkı İşgücü Anketi'nin (HLFS) resmi mikroverileri, coğrafi göstergeleri NUTS-2 düzeyinde (26 alt bölge) sunmaktadır. Verilerin bu şekilde toplulaştırılması, ülkenin 81 ilini gizleyerek, potansiyel olarak heterojen yerel ekonomileri aynı etiket altında bir araya getirmekte ve böylece politikaların yerel etkisinin değerlendirilmesini sınırlamaktadır. Bu makale, 2004-2013 dönemine ait mikroverilerde il düzeyinde kimliklerin geri kazanılmasına yönelik bir strateji sunmaktadır. Araç plaka il numarasına dayalı sıralı bir tanımlama algoritması uygulanmakta, buna il düzeyinde yıla özgü kentsel-kırsal nüfus yeniden ağırlıklandırması eklenmektedir. Tanımlama, mevcut çok-illi kümelerin arkasındaki mantık kullanılarak içsel olarak, Nüfus Sayımı'ndan, Adrese Dayalı Nüfus Kayıt Sistemi'nden (ADNKS) ve Sosyal Güvenlik Kurumu'ndan (SGK) elde edilen başlıca ekonomik faaliyetlere ilişkin idari kayıt verilerine karşı da dışsal olarak doğrulanmaktadır. Ampirik bir kanıt niteliğinde olarak, 81 ilin tamamı için il düzeyinde kayıt dışılık oranları tahmin edilmekte ve daha önce NUTS-2 raporlamasının altında gizli kalan bölgesel farklılıklar ortaya konmaktadır.

Anahtar Kelimeler: *İşgücü Anketi, Veri Yeniden Yapılandırma, Post-stratifikasyon, Türkiye İstatistik Kurumu (TÜİK).*

J.E.L. Sınıflandırması: *C-81, J-21, R-12*

¹ Luis Pinedo Caro, Bahçeşehir Üniversitesi Ekonomik ve Toplumsal Araştırmalar Merkezi'nde (BETAM) araştırmacı olarak görev yapmaktadır. E-posta: Luis.pinedo@bau.edu.tr.

1. Giriş

Türkiye'de sosyal hizmetler, ekonomik destek ve bölge planlaması gibi çeşitli yetkiler, büyükşehir belediyeleri için 5216 sayılı Kanun ve 6360 sayılı Kanun, diğer belediyeler için ise 5393 sayılı Kanun uyarınca, NUTS-3 düzeyinde 81 ili yöneten sözde büyükşehir ve büyükşehir olmayan belediyelere aittir. Ancak, Türkiye İstatistik Kurumu (TÜİK) tarafından sağlanan kamuya açık mikroyerler, ulusal Hanehalkı İşgücü Anketi'nde (HLFS) ve Gelir ve Yaşam Koşulları Araştırması'nda (SILC) mekânsal ayrıntı düzeyini 26 NUTS-2 alt bölgesiyle, Aile Yapısı Araştırması'nda ise yalnızca 14 "bölge" ile sınırlamaktadır; bunlar, sosyal konulara odaklanan başlıca araştırmalardan sadece birkaçıdır. Bu araştırmalar, ücretler, meslekler, işsizlik, işgücü piyasası dinamikleri, yoksulluk, yaşam memnuniyeti ve aile dinamikleri gibi hem yerel hem de Ankara merkezli karar alıcılar için önemli olan konularda başlıca bilgi kaynağını oluşturmaktadır. Bu konulara ilişkin il düzeyinde bilgi eksikliği, pratikte önemli bir analitik sürtünme yaratmaktadır: ağır sanayileşmiş üretim merkezleri ile büyük ölçüde tarıma dayalı komşu bölgeler gibi ekonomik olarak birbirinden farklı iller, tek bir istatistiki birim altında toplulaştırılmakta; bu durum yerleşmiş ekonomik şokları gizlemekte, yapısal ücret bozulmalarını saklamakta ve politika değerlendirmelerini daha gürültülü ve daha az anlamlı hale getirmektedir.²³

Bu makale, 2004-2013 dönemi için HLFS'te illeri (NUTS-3) tanımlayarak ve il düzeyinde anket ağırlıklarına imkân tanıyan bir post-stratifikasyon düzeltmesi sunarak bu bilgi boşluğunu doldurmaya çalışmaktadır. Bununla birlikte, kamu araştırmalarının gizli boyutlarını ortaya çıkarmaya yönelik bu tür bir girişim yeni değildir; anket verilerindeki mekânsal veya nüfusa dayalı gizlemeleri aşmak, giderek adli veri yeniden yapılandırma ve tersine mühendislik tekniklerine dayanan ve uluslararası literatürde rastlanan bir uygulamadır.

Böyle bir örnek, Hindistan'ın resmi Periyodik İşgücü Anketi'nde (PLFS) kritik bir veri kısıtını da ele alan Enevoldsen (2023) tarafından sunulmaktadır; burada değiştirilmiş coğrafi ve örnekleme birimi kodları, araştırmacıların kentsel hanehalklarını ardışık yıllık paneller boyunca izlemesini engellemiştir. Yazar, ilk yıl anketindeki tanımlama kodlarının, veri anonimleştirme sırasında basit bir yerine koyma şifresiyle sistematik biçimde karıştırıldığını ortaya koymaktadır. Değişmez hanehalkı özelliklerine dayanan bir optimizasyon algoritması geliştirilmiş, bu sayede araştırmacılar ilk yıl kayıtlarını sonraki dalgalarla doğru biçimde eşleştirebilmiş ve anketin işgücü piyasası analizi için tam boyamsal potansiyelini ortaya çıkarabilmiştir.

Farklı bir veri geri kazanım türü ise, tam sokak adreslerinin gizlilik amacıyla saklandığı veya değiştirildiği durumlarda konumları haritalamaya yönelik yeni bir yöntem sunan Helderop, Nelson ve Grubestic (2023) tarafından gösterilmektedir. Standart haritalama yazılımları genellikle bu bulanıklaştırılmış konum verileriyle başa çıkmakta zorlanmakta ve noktaları bir sokağın veya mahallenin ortasına hatalı biçimde yerleştirmektedir. Bunu düzeltmek için araştırmacılar, izin verilen alan içinde en doğru ve gerçekçi noktayı bulmak amacıyla istatistiksel

² NUTS 3, İstatistiki Bölge Birimleri Sınıflandırması'nın (Nomenclature of Territorial Units for Statistics) üçüncü düzeyidir. Avrupa Birliği, istatistiki veri toplama amacıyla bölgeleri ayırmak için bu sistemi kullanmaktadır.

³ Türkçede Türkiye İstatistik Kurumu (TÜİK).

gruplama kullanan bir bilgisayar tekniği geliřtirmiş ve böylece arařtırmacılara hassas verileri haritalamak için çok daha güvenilir bir yöntem sunmuřtur.

Bazen veriler toplulařtırılmış olsa dahi mikroverilerin kendisi bile yeniden yapılandırılabilir. Nitekim bilgisayar bilimleri ve istatistiksel ifřa kontrolü alanındaki bazı çalışmalar, kamuya açık řekilde yayımlanan tabloların, altta yatan bireysel mikroveri kayıtlarını tam olarak yeniden yapılandırmaya yetecek matematiksel kısıtlar içerdini göstermiştir. Temel bir örnek, veritabanı yeniden yapılandırma saldırılarının teorik temelini sunan Dinur ve Nissim'dir (2003). İstatistiksel bir veritabanını kiři başına bir bit olacak řekilde basit bir bit dizisi olarak modelleyen yazarlar, gizlilięi korumak amacıyla bir miktar gürültü eklenmiş olsa dahi, yeterli sayıda toplulařtırılmış sorgu (kayıt alt kümeleri üzerindeki toplamlar) yanıtlandığında, hızlı bir algoritmanın orijinal veritabanının neredeyse tamamını yeniden yapılandırabildiğini göstermektedir. Başka bir deyiřle, katı bir matematiksel sınır olduęunu kanıtlamaktadırlar: belirli bir noktadan sonra, bir veri kümesi hem istatistiksel açıdan yararlı hem de gerçek anlamda gizli olamaz.

Bu sonuç yalnızca teorik deęildir. Garfinkel, Abowd ve Martindale (2019), yayımlanmış bir dizi istatistiki tabloyu, altta yatan mikroveriyi geri kazanmak için çözülebilecek bir matematiksel kısıtlar sistemi olarak ele alarak, bu tür yeniden yapılandırma saldırılarının pratikte nasıl gerçekleştirilebileceğini göstermektedir. ABD Nüfus Sayım Bürosu'nun kendisi de bu güvenlik açığının boyutunu daha sonra doğrudan doğrulamıştır: toplamda 150 milyardan fazla yayımlanmış istatistik arasından, 2010 Nüfus Sayımı'ndan yayımlanan 180 tablo setinin yalnızca 34'ünü kullanan içsel bir yeniden yapılandırma çalışması, bireysel kayıtlar için beř deęiřkeni (sayım bloęu, cinsiyet, yař, ırk ve etnik köken) geri kazanmış ve yeniden yapılandırılan kayıtların ülkenin sayım bloklarının %70'inde, yaklaşık 97 milyon kiřiye karşılık gelecek řekilde, gizli orijinleriyle tam olarak eřleřtiğini doğrulamıştır (Vilhuber ve dięerleri, 2023). Sade bir ifadeyle: hiçbir mikroveri yayımlanmadan yalnızca toplulařtırılmış tabloların yayımlanması dahi, ABD nüfusunun büyük bölümü için neredeyse eksiksiz bireysel düzeyde veri yayımlamakla matematiksel olarak eřdeęer olduęu ortaya çıkmıştır. Bu bulgu, Nüfus Sayım Bürosu'nun 2020 Nüfus Sayımı için geleneksel ifřa sınırlama yöntemlerini terk edip diferansiyel gizlilik yöntemini benimseme kararında merkezi bir rol oynamış ve bu makalenin kendi yönteminin de yararlandığı daha geniş dersi vurgulamaktadır: istatistiki yayınlar, toplulařtırmalarının izin verdięi görünenden çok daha fazla yapısal bilgi sızdırabilir.

Türkiye'deki uygulamalı mikro-ekonometri literatüründe veri geri kazanımına iliřkin örnekler dikkat çekici biçimde azdır; bu durum, TÜİK'in gizleme uygulamalarının, benzer ülkelerdeki uygulamalara kıyasla ne kadar az ilgi gördüğünü yansıtmaktadır. Tunalı (2023), Hanehalkı İřgücü Anketi'ne iliřkin metodolojik rehberinde bu kısıtları ayrıntılı biçimde ele almaktadır: TÜİK'in dönüşümlü örneklem çerçevesi, aynı hanehalkının ardışık dönemlerde dört kez ziyaret edilmesini içermekte ve iřgücü piyasası geçiřlerinin incelenmesi açısından açık bir araştırma deęeri taşıyan kısa bir panel yapısı oluşturmaktadır; ancak yazarın belirttięi üzere, bu panel boyutu "TÜİK'in veri yayma politikası nedeniyle mikroveri kullanıcılarının erişimi dışında" kalmaktadır. Yazarın deęerlendirmesi, yayımlanan mikroverilerin dayattığı bir dizi yapısal kısıt -birincil örnekleme birimi (PSU) tanımlayıcılarının bulunmayışı, dalgalar arası panel baęlantısının gizlenmesi ve bölgesel tahmin için elveriřsiz bir hedef nüfus olan kurumsal

olmayan nüfus- sıralamakta, ancak bunlardan hiçbirini için bir çözüm önermemektedir. Nitekim, bahsi geçen kısıtlardan bazılarının yalnızca dile getirildiği araştırmalar da görülmektedir; örneğin Bağır, Küçükbayrak ve Torun (2021), Türkiye'de ücret eksik bildirimine ilişkin araştırmaların, gerekli ayrıştırılmış verinin bulunmaması nedeniyle kısıtlı kaldığını belirtmektedir. Başka durumlarda ise, HLFS'teki belirli değişkenlerin eksikliğini telafi etmek için başka kaynaklardan yararlanılmaktadır; ana ücret analizi HLFS'in eğitim değişkenine dayanan ancak bağımsız bir kontrol olarak Türkiye Nüfus ve Sağlık Araştırması'nın sürekli eğitim yılı ölçütünü kullanan Aydemir ve Kırdar (2017) buna örnektir. Bizim yaklaşımımız, tam olarak HLFS'in kendi örnekleme mekaniğinden yararlanarak ve onu daha bilgilendirici hale getirerek bu örüntüden ayrılmaktadır.

Bu genel kıtlığın ortasında, iki çalışma istisna olarak öne çıkmakta ve Türkiye HLFS'i bağlamında burada geliştirilen yaklaşıma en yakın mevcut örneği oluşturmaktadır. Pinedo Caro (2020), HLFS dalgaları içinde açıklanmayan Suriyeli mülteci alt örneklemlerini ayırt etmek için artık (residual) bir tanımlama stratejisi geliştirmiştir: HLFS doğum yerini kaydettiği ancak vatandaşlığı kaydetmediği için, Suriyeli mülteciler Türkiye'nin olağan göçmen akışından ayırt edilememektedir. Pinedo Caro, "standart" göç örüntüsünü karakterize etmek için savaş öncesi (2004-2010) geliş kohortundan yararlanmakta, bunu 2011-2017 yurt dışında doğan nüfustan çıkarmakta ve elde edilen Suriyeli alt örneklemini, ki-kare testi kullanarak Türkiye Göç İdaresi Genel Müdürlüğü'nün resmi yaş dağılımı verileriyle doğrulamaktadır. Pinedo Caro (2021), benzer şekilde, COVID-19 pandemisi sırasındaki işgücü piyasası aksamalarını öngörmek amacıyla, TÜİK'in kamuya açık olarak yayımladığı toplulaştırmadan daha ayrıntılı, aylık sıklıkta gözlemler çıkarmaktadır. Ancak her iki çalışma da mekânsal tanımlamayı ele almamaktadır. Bu uygulamalı ölçüm geleneği üzerine doğrudan inşa edilen bu makale, örnekleme bütünlüğünden ödün vermeksizin, 2004-2013 dönemine ait tüm HLFS mikroveri arşivinde 81 il için il göstergelerini tam olarak yeniden yapılandıran ilk sistematik yöntemi geliştirmektedir.

Bu makale, yukarıda tartışılan ilk literatür akımının içinde yer almakta -gizlenmiş kimlikleri geri kazanmak için deterministik örnekleme protokollerinden ve anket uygulama mekaniğinden yararlanmakta- ve üç temel katkı sunmaktadır. Birincisi, anonimleştirilmiş mikroverilerden 81 il kimliğini geri kazanmak için deterministik anket taramalarından ve demografik dağılımlardan yararlanan, tekrarlanabilir bir Stata/Mata algoritması sunmakta ve nüfusa dayalı orantılı atama kuralı yoluyla deterministik olarak ayrıştırılamayan altı il çiftini çözüme kavuşturmaktadır. İkincisi, alt bölge düzeyindeki örneklem tasarımı sapmalarını düzeltmek amacıyla resmi demografik sayımlara göre kalibre edilmiş, çok yıllık bir nüfus yeniden ağırlıklandırma matrisi oluşturmakta ve yeniden yapılandırılan il panelinin her yıl TÜİK'in kendi yayımladığı toplulaştırılmış verilerle makro düzeyde tutarlı kalmasını sağlamaktadır. Üçüncüsü, yeniden yapılandırılan tahminleri idari Sosyal Güvenlik Kurumu (SGK) kayıtlarına karşı doğrulamakta -il büyüklükleri ve sektör sınıflandırmaları genelinde yüksek mekânsal güvenilirlik ortaya koymakta- ve standart NUTS-2 raporlaması altında gizlenen bölge içi kayıt dışılık farklarını ortaya çıkararak NUTS-3 ayrıntı düzeyinin ampirik gerekliliğini göstermektedir.

2. Veri ve tanımlama yöntemi

Temel veri kümesi, 2004-2013 dönemini kapsayan, Türkiye Hanehalkı İşgücü Anketi'nden (HLFS) elde edilen uyumlaştırılmış mikroverilerden oluşmaktadır. TÜİK, bu on yıl boyunca anket örnekleme çerçevesini, her alt bölgenin (NUTS-2 düzeyi) kırsal ve kentsel alanları için istatistiksel olarak temsil edici tahminler sağlayacak şekilde tasarlamıştır. Ancak il düzeyindeki (NUTS-3) bilgiler, yalnızca üç istisna dışında -İstanbul (TR10), Ankara (TR51) ve İzmir (TR31)- kamuya açık mikroveri dosyalarından açıkça çıkarılmıştır. Bu üç büyükşehir yalnızca, ilgili NUTS-2 alt bölgelerinin tek illi olması nedeniyle tanımlanabilmektedir. Sonuç olarak, araştırmacılar idari kimliği bulunmayan 78 ille baş başa kalmaktadır. Türkiye'nin, tarihsel olarak Alman Länder'leri veya İspanyol Özerk Toplulukları gibi federal yapılardan ziyade Fransız üniter devlet modelini örnek alan, oldukça merkeziyetçi idari mimarisinde, NUTS-2 birimleri işlevsel siyasi veya ekonomik yargı alanları olarak değil, kesinlikle yapay istatistiki toplulaştırmalar olarak var olmaktadır.⁴

HLFS'in NUTS-2 düzeyindeki temsil gücüne ilişkin sorun, bu toplulaştırılmış sınırlar içinde önemli ölçüde farklılık gösteren yerel işgücü piyasası dinamiklerinden ve idari yetkiden kaynaklanmaktadır. Örneğin, TR41 alt bölgesi, büyük bir sanayi ve otomotiv üretim merkezi olan Bursa'yı, ağırlıklı olarak seramik ve hammadde çıkarımına dayalı, seyrek nüfuslu bir ekonomiye sahip Bilecik ile birleştirmektedir. TR41'i homojen bir ekonomik birim olarak ele almak, yerelleşmiş kayıt dışı istihdam dinamiklerini, belediye politikası müdahalelerini ve hedefli işgücü piyasası denetimlerini gizleyerek ciddi bir toplulaştırma yanlılığına yol açmaktadır.

2.1 Tanımlama stratejisi

Bu yapısal kısıtların üstesinden gelmek ve daha titiz bir politika değerlendirmesine olanak sağlamak amacıyla, 81 ilin kimliğinin tamamını yeniden yapılandırmak üzere tasarlanmış deterministik bir tanımlama ve yeniden ağırlıklandırma yöntemi öneriyorum. İl düzeyinde ayrıntıyı geri kazandırmak, hem bir veri boşluğunu hem de kamusal bir ihtiyacı gidermekte; araştırmacılara ve karar alıcılara gerçek idari yargı alanlarına dayanan, yüksek çözünürlüklü ampirik kanıtlar sunmaktadır.

Tanımlama stratejisi, ham HLFS mikroverisinin yapısal bir özelliğinden yararlanmaktadır: örnekleme tasarımı, her yıllık dalga içinde, 12 düzenli döngü halinde yürütülen deterministik, artan yönlü bir mekânsal tarama uygulamaktadır. Hanehalkı düzeyindeki mikroveri gözlemlerini rastgele hale getirmek yerine, alt bölgeler nihai veri kümesine, Türkiye illerinin resmi sayısal plaka sırasını (01'den 81'e kadar) kesinlikle yansıtan sistematik bir düzende yüklenmektedir. Bu çoklu döngüler boyunca, neredeyse tüm iller istikrarlı ve öngörülebilir bir sırada temsil edilmektedir. Ek A, illerin değişmez konumunu ve plaka numarasına göre ardışık illerin yalnızca bir kez görüldüğü durumları gösteren, tam bir döngünün eksiksiz yapısal doğrulamasını sunmaktadır.⁵

⁴ Örnekleme yöntemi 2014 yılında değişmiş, TÜİK sabit referans haftalarından, yılın 52 haftasına eşit biçimde dağıtılmış sürekli haftalık saha çalışmasına geçmiştir. 2004-2013 dönemini karakterize eden artan yönlü sıralı mekânsal örüntü bu değişiklikte gizlenmiş olup, 2014 sonrası il kimlikleri önerilen yöntemle geri kazanılamamaktadır.

⁵ 2004'ten 2013'e kadar tüm anket turlarını kapsayan tam döngü matrisleri talep üzerine sunulabilir.

Nitekim plaka sırası illerin büyük çoğunluğunu geri kazandırmakta, ancak ardışık plaka numaralarını paylaşan altı belirli il çifti, yalnızca mekânsal sıralama yoluyla tam bir ayrıştırmayı engellemektedir. Bu belirsizlikleri çözmek için, deterministik taramayı orantılı bir atama kuralıyla tamamlıyorum. Doğal plaka sırası izlenerek, belirsiz blok içindeki ilk hane halkı gözlemleri, ilgili ilin nüfusunun, eşleştirilen iki ilin toplam nüfusuna oranını tam olarak yansıtacak şekilde, sırayla ilk görünen ile atanmaktadır.

2.2 Post-stratifikasyon düzeltmesi ve il düzeyinde anket ağırlıkları

Her ilin tanımlanmasının ardından, NUTS-3 düzeyinde anket ağırlıkları oluşturan özel bir post-stratifikasyon düzeltmesi geliştiriyorum. Referans nüfus sayıları, 2000 Genel Nüfus Sayımı'ndan ve 2007-2014 dönemi için Adrese Dayalı Nüfus Kayıt Sistemi'nden (ADNKS) elde edilmektedir. Ara yıllar (2004-2006) için, il nüfusları, 2000 ve 2007 referans değerleri arasında sabit bir üstel büyüme oranı varsayılarak enterpole edilmektedir. Daha sonra, il ve alan bazında nüfus hedeflerini belirlemek için resmi kentsel-kırsal dağılım oranları uygulanmakta ve bu hedefler ile gösterilmekte, burada kentsel veya kırsal katmanı, yılı ve ise ili ifade etmektedir. Hanehalkı gözlemleri daha sonra, örneklem toplamalarının bu resmi hedeflerle eşleşmesi için il $\times$ kentsel/kırsal hücre düzeyinde yeniden ağırlıklandırılmaktadır: $N_{p,s,t}^{official} s \in \{R\}tp$

$$w_{i,p,t}^{temp} = w_{i,p,t} \cdot \frac{N_{p,s,t}^{official}}{\hat{N}_{p,s,t}}, \text{ burada , ilinde ve yılında bireyi için TÜİK'in orijinal tasarım ağırlığıdır}$$

ve

$$w_{i,p,t}^{ipt\hat{N}_{p,s,t}} = \sum_{i \in p,s,t} w_{i,p,t}$$

ilgili tasarım-ağırlıklı hücre nüfusedir. Bir ilin belirli bir yılda iki katmandan birine sahip olmadığı nadir durumlarda, il toplamı üzerinden benzer şekilde hesaplanan bir il düzeyi faktörü kullanılmaktadır (etkilenen iller ve yıllar için bkz. Ek B).

Elde edilen il düzeyindeki genişletilmiş nüfus, NUTS-2 düzeyinde toplulaştırıldığında, TÜİK'in kendi mikroyerisinde yayımladığı karşılık gelen NUTS-2 toplamını her zaman aşmaktadır. Bunun nedeni, TÜİK'in örnekleme tasarımının toplam yerleşik nüfusu değil, kurumsal olmayan nüfusu hedeflemesidir. Resmi istatistiki yayınlarla makro düzeyde tutarlılığı korumak için, il toplamaları, yayımlanan mikroyeride yer aldığı şekliyle ilgili NUTS-2 alt bölge nüfusuna tam olarak eşit olacak şekilde kısıtlanmaktadır. Kurumsal olmayan nüfus rakamları NUTS-3 düzeyinde gözlemlenemediğinden, her bir NUTS-2 alt bölgesinin kurumsal olmayan nüfus payının, kendisini oluşturan iller genelinde tek biçimli olarak uygulandığını varsayıyorum. Bu oran şu şekilde tanımlanmaktadır: r_g

$$r_{g,t} = \frac{\sum_{i \in g,t} w_{i,t}}{\sum_{p \in g} N_{p,t}^{official}},$$

burada pay, TÜİK'in orijinal tasarım ağırlıklarının ima ettiği NUTS-2 alt bölgesinin kendi toplulaştırılmış nüfusedir ve , o alt bölgeye ait iller üzerinden resmi yerleşik nüfusu

toplamaktadır. Her ilin anket ağırlıkları daha sonra sabit NUTS-2 faktörü ile ölçeklendirilerek nihai post-stratifikasyon ağırlıkları elde edilmektedir: $N_{g,t}^{official} r_{g,t}$

$$w_{i,p,t}^{province} = w_{i,p,t}^{temp} \cdot r_{g(p),t},$$

burada , ilini içeren NUTS-2 alt bölgesini ifade etmektedir. Belirtilmeye değer bir istisna vardır: Kilis (2004) ve Tunceli (2009), parantez içinde belirtilen yıllarda ziyaret edilmemiş, ancak TÜİK bunların kendi alt bölgelerindeki diğer iller tarafından temsil edildiğini varsaymıştır. Bu varsayım il düzeyinde ayrıştırma için kabul edilebilir değildir ve bir ilin açıkça ziyaret edilmediği yıllarda, alt bölge düzeyindeki toplam sayı, eksik ilin nüfusunu bilinçli olarak dışarıda bırakmaktadır. $g(p)p$

3. İç ve Dış Veri Doğrulaması ve Uygulama

Yeniden yapılandırılan il kimlikleri, TÜİK tarafından doğrudan açıklanmak yerine algoritmik olarak geri kazanıldığından, bunların ampirik geçerliliğinin ortaya konması, sonraki analizler için gerekli bir ön koşuldur. Tanımlanan panelin istatistiki bir yapaylık değil, gerçek idari yargı alanlarını yansıttığını göstermek amacıyla, yöntemi; içsel bir yapısal yanlışılama testini, Sosyal Güvenlik Kurumu (SGK) kayıtları ve Nüfus Sayımı/ADNKS'ye karşı dışsal idari kıyaslamayla birleştiren iki aşamalı bir doğrulama sürecine tabi tutuyorum.⁶

Veri kümesindeki alt bölge sırası ve plaka numarası. Deterministik bir mekanizma yoluyla birbirinden ayrıştırılamayan altı il çiftinin bulunması, belirli il düzeyi tahminlerinin doğruluğu açısından olumsuz bir durumdur; ancak bu özellik, TÜİK'in mikroveri toplarken plaka numaralarını kullandığına dair kanıt sunan sağlam bir yanlışılama testi ortaya koymaktadır. Bunun yerine, illerin alt bölgeler içinde rastgele sıralandığı varsayılsaydı (bu durum plaka numarası argümanını geçersiz kılardı), belirsizliğin özellikle ve yalnızca ardışık plaka numaralarına sahip il çiftleri arasında ortaya çıkması için hiçbir neden olmazdı. "Gerçek bir sıra yok" sıfır hipotezi altında, aynı NUTS-2 alt bölgesi içindeki herhangi bir il çiftinin ayırt edilemez bir blok oluşturma olasılığı eşit olurdu; oysa döngünün tamamındaki tek altı belirsiz durum, tam olarak ardışık plaka numaralarına sahip altı il çiftidir. Dahası, bu tek seferlik bir örüntü de değildir: aynı altı çift ve yalnızca bu altısı, 2004'ten 2013'e kadar her yıl, on iki alt tur boyunca -toplamda örnekleme döngüsünün 120 tekrarında- tutarlı biçimde ayırt edilemez bir blok üretmektedir. Rastgele atama altında, çakışan illerin kimliğinin her döngüde tam olarak aynı altı çifti yeniden üretmek yerine döngüden döngüye değişmesi beklenirdi. Bu kusursuz ve kalıcı yoğunlaşma, tam olarak plaka sıralaması hipotezinin öngördüğü imzadır ve rastgele sıralamaya dayalı herhangi bir alternatif hipotezle bağdaştırılması esasen mümkün değildir.

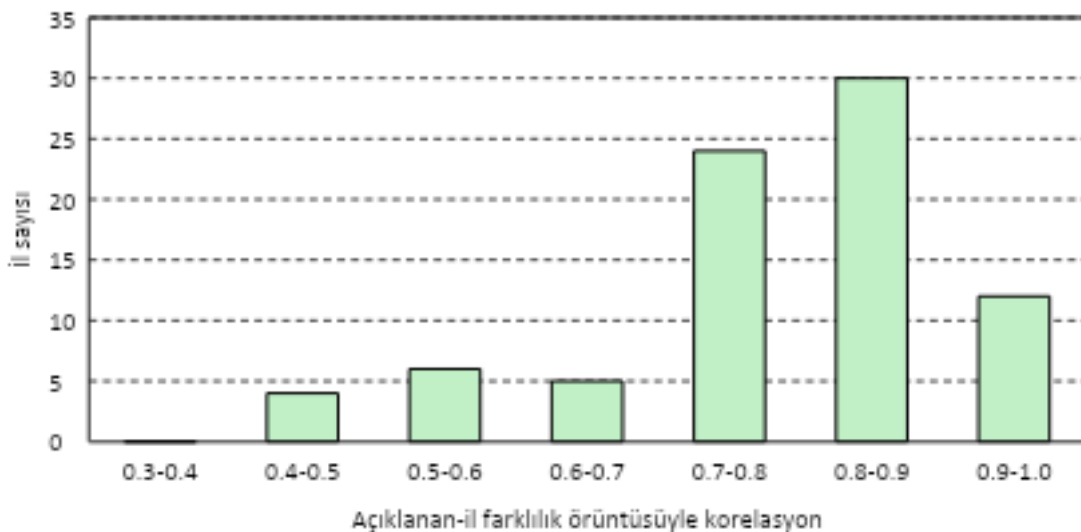
Örneklem büyüklüklerinin gerçek nüfus paylarına göre göreceli önemi. İçsel geçerliliği değerlendirmek amacıyla, her ilin 2004-2013 dönemine ait toplulaştırılmış örneklem payını, ADNKS tabanlı il nüfus toplamlarından elde edilen, aynı döneme ait gerçek yerleşik nüfus payıyla karşılaştırıyorum. Örneklem ve nüfus payları, 81 ilin tamamında yüksek düzeyde

⁶ Adrese Dayalı Nüfus Kayıt Sistemi (ADNKS), Türkiye'de yerleşiklerin kaydedilmesi için kullanılan resmi kayıt sistemidir.

ilişkilidir ($r = 0,964$). Karşılaştırmayı, doğrudan ve deterministik biçimde tanımlanan 69 ille sınırladığımızda -bölünmesi nüfus paylarıyla eşleşecek şekilde tasarım gereği oluşturulan ve bu nedenle bağımsız bir kanıt niteliği taşıyamayan altı çift plakalı il hariç tutulduğunda-korelasyon, aksine, hafifçe daha da yüksektir ($r = 0,966$). Bu 69 il, herhangi bir orantılı atama mekanizması söz konusu olmaksızın, yalnızca plaka numarası sırasındaki konumlarından geri kazanıldığı için, bu yakın örtüşme mekanik bir yapaylık değildir: geri kazanılan il kimliklerinin gerçek nüfus yapısını yakından takip ettiğini göstermektedir. Bir ilin örneklem payı ile nüfus payı arasındaki ortalama mutlak fark yalnızca 0,29 yüzde puanıdır ve paneldeki en büyük farklar bile -TÜİK'in orantısal ulusal örneklemeden ziyade katmanlı NUTS-2/kentsel-kırsal kota tasarımını yansıtan İstanbul ve Ankara gibi birkaç çok büyük ilde yoğunlaşan farklar- bir yüzde puanının altında kalmaktadır. Belirsizlik olmadan tanımlanan küçük iller, nüfus paylarını özellikle yakından takip etmektedir: örneğin Burdur, örneklemin %0,30'unda yer alırken nüfusun %0,35'ini oluşturmaktadır. Türkiye'nin 81 ilinin tüm büyüklük yelpazesinde gözlenen bu sıkı, il bazında örtüşme, plakaya dayalı tanımlamanın gerçek il yapısını geri kazandırdığına dair bir başka kanıttır.

İstihdam yapısı ve iller arası heterojenlik. Toplulaştırma yanlılığının çarpıcı bir örneği, doğrudan yeniden yapılandırılan il göstergesinden ortaya çıkmaktadır. Birleştirilmiş TR41 alt bölgesi içinde, Bursa'nın tanımlanan istihdam profiline, ilin Türkiye'nin en büyük otomotiv üreticilerinden biri olan Tofaş'a ev sahipliği yapmasıyla tutarlı biçimde, motorlu taşıt ve römork imalatı (çalışanların %7,96'sı) ve tekstil imalatı (%9,80) egemendir. NUTS-2 ortaklarından biri olan Bilecik ise tamamen farklı bir profil sergilemektedir: yalnızca diğer metalik olmayan mineral ürünlerin imalatı, çalışanların %19,38'ini oluşturarak, ilin köklü seramik ve fayans sanayisini (Bien Seramik, Vitra Fayans vb.) yansıtan, tarım dışı en büyük tek kategoriye meydana getirmektedir. Diğer ilde her iki sektörde de buna benzer bir yoğunlaşma görülmemektedir. NUTS-2 raporlaması altında görünmez olan bu keskin farklılaşma, tam olarak yeniden yapılandırılmış il panelinin geri kazanmak üzere tasarlandığı türden bir bölge içi heterojenliktir.

Şekil 1. İl bazında sektörel farklılık örüntüsü



Kaynak: TÜİK HLFS 2007-2013 mikroversi, SGK 4a il düzeyinde ekonomik faaliyet istatistikleri ve yazarın kendi hesaplamaları. Notlar: Her il için, küçük hücre sayıları nedeniyle tarım ve madencilik hariç tutarak, yedi geniş sektörde HLFS (anket) istihdamının SGK (idari) istihdamına oranını hesaplıyorum. Ardından, bir referans farklılık örüntüsü elde etmek için bu sektör bazlı oranları, TÜİK'in doğrudan açıkladığı üç il (İstanbul, Ankara, İzmir) genelinde ortalamasını alıyorum. Grafik, 81 ilin tamamında, her bir ilin kendi 7-sektörlü oran profili ile bu referans örüntü arasındaki Pearson korelasyonunun dağılımını göstermektedir. Yüksek bir korelasyon, bir ilin HLFS-SGK farklılığının, doğrudan açıklanan ile aynı sektör bazlı yapıyı izlediğini göstermektedir.

Bu örüntünün Bursa örneğinin ötesinde ne ölçüde genellenebildiğini değerlendirmek amacıyla, özel şirketlerdeki çalışanlar arasında formal istihdamın sayısına ilişkin il düzeyi bilgiyi iki kaynaktan, HLFS ve Türkiye Sosyal Güvenlik Kurumu'ndan derliyorum. Bu işlem, 2007'den 2013'e kadar her yıl için yedi sektörde -imalat (ISIC rev.4, 10-33 kodları), enerji üretimi (35-39), inşaat (41-43), ticaret ve konaklama (45-47, 55-56), ulaştırma ve haberleşme (49-53, 58-63), finans, sigorta ve diğer iş faaliyetleri (64-82) ve diğer tüm hizmetler (84-99)- gerçekleştirilmektedir. 2004-2006 yıllarının mevcut olmadığını ve daha küçük illerde formal çalışan hücre sayılarının istikrarlı oranlar üretemeyecek kadar seyrek olduğu tarım ve madencilik sektörlerinin dışarıda bırakıldığını lütfen belirtmek isterim. Buradaki amaç, kaynaklar arasındaki rakamları karşılaştırmak değildir; zira HLFS istihdamı referans hafta boyunca ölçerken SGK bunu yıllık düzeyde ölçtüğünden, bu rakamlar doğrudan karşılaştırılabilir değildir ve SGK rakamları tanım gereği HLFS rakamlarından her zaman daha yüksektir. Buradaki fikir, yeni bulunan 79 ilin HLFS/SGK farklılık profili ile TÜİK'in doğrudan açıkladığı üç ilin (İstanbul, Ankara ve İzmir) ortalama farklılık profili arasındaki korelasyonu karşılaştırmaktır. İl tanımlaması güvenilir olmasaydı, yeniden yapılandırılmış bir ilin farklılık örüntüsünün açıklanan illerinkine benzemesi için hiçbir neden olmazdı. Bunun yerine, 81 ilin tamamında ortalama korelasyon 0,81'dir ve bazı iller (Gaziantep 0,95, Trabzon 0,93, Antalya 0,92) İstanbul (0,97), Ankara (0,94) ve İzmir (0,93) korelasyonlarına yakın değerler sergilemektedir. Yalnızca yeniden yapılandırılan 78 ile sınırlandırıldığında, ortalama korelasyon 0,78'dir (aralık: 0,44-0,98). Örnek olay incelememiz olan Bursa, 0,75 ile bu daha geniş örüntüyle uyumludur. En düşük korelasyonlar, tanımlama başarısızlığından ziyade örnekleme gürültüsüyle tutarlı biçimde, HLFS örneklem büyüklüğü en küçük olan iller (örneğin Tunceli, Osmaniye) arasında yoğunlaşmaktadır. Bu dağılım, hanehalkı anketi ile idari kayıt verisi arasındaki yapısal farktan kaynaklanan HLFS ve SGK arasındaki sektörel farklılığın, TÜİK'in tesadüfen doğrudan açıkladığı illerle sınırlı olmayan, Türkiye'nin illeri genelinde sistematik olarak paylaşılan bir özellik olduğunu göstermektedir.

3.3 Ampirik Uygulama: Kayıt Dışı İstihdamın Mekânsal Değişimi

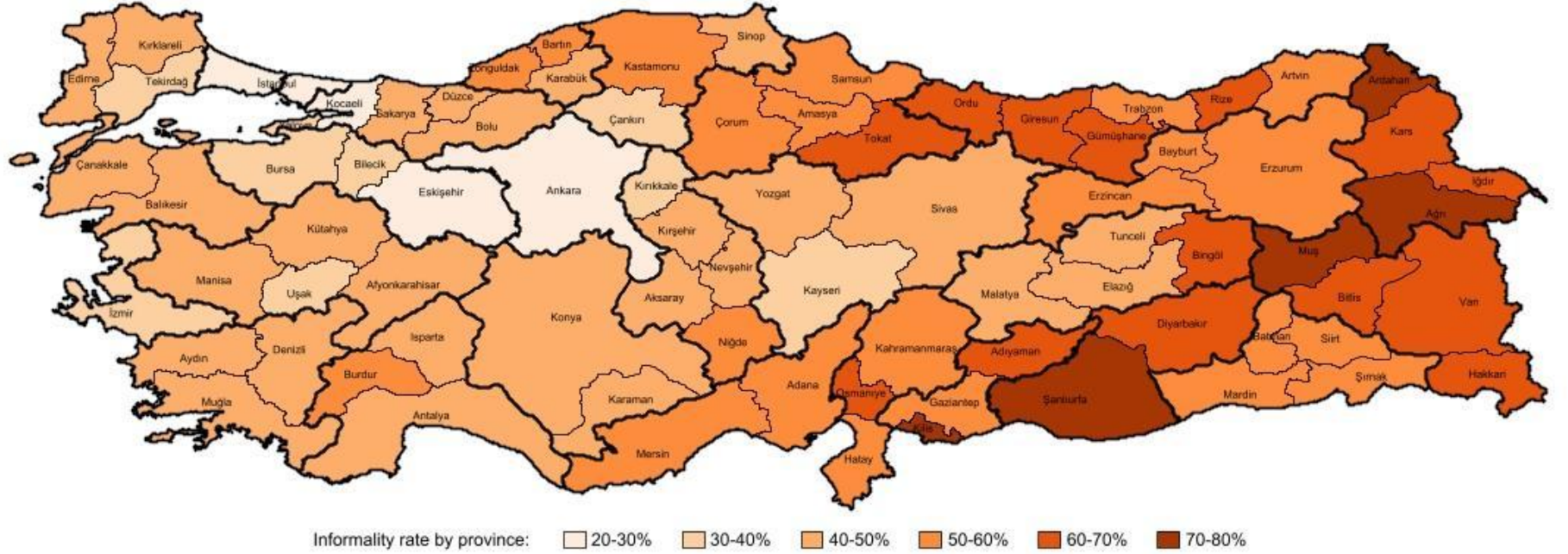
2000'li yıllardan ve 2010'ların başından ampirik literatür -örneğin Angel-Urdinola ve Tanabe (2012) ile Tansel ve Kan (2012)- kayıt dışı istihdama ilişkin temel analizleri kaba bölgesel toplulaştırmalarla sınırlı kaldığından, yüksek çözünürlüklü mekânsal mikroversiden yararlanabilirdi.⁷

⁷ İl düzeyinde mikroversinin faydası, tarihsel bir gereklilikten çok uzaktır. Yerinden edilmiş nüfusların yerel işgücü piyasasına katılımına ilişkin Del Carpio ve Wagner (2015) ile Tumen (2016) gibi, ya da yapısal ücret eksik bildirimi ve mekânsal ücret eğrilerine ilişkin Bağır, Küçükbayrak ve Torun (2021) ile Baltagi ve Başkaya (2022) gibi daha yakın tarihli çalışmalar, ulus-altı heterojenliğin Türk işgücü ekonomisi açısından merkezi öneme sahip olmaya devam ettiğini göstermekte, ancak bu çalışmalar daha güncel veriler kullanmaktadır.

Bu deęerlendirmeler, 26 NUTS-2 bölgesi yerine 81 illik HLFS'i kullanabilseydi, bulguları belki de belediye denetimi darboęazlarını ve yerelleşmiş ücret bozulmalarını ayırt eden, daha uygulanabilir ve hedefli kamu politikalarına dönüşebilirdi. TÜİK, kamuya açık mikroveri yayınlarında resmi il düzeyi kimliklerini gizlemeye devam ettięinden, burada sunulan yeniden yapılandırma yöntemi, geçmiş yerel politikaların deęerlendirilmesine imkân tanımak açısından hayati önemini korumakta ve en azından Türkiye'nin yakın tarihinde önemli bir dönemi kapsamaktadır.

Bu bölüm, alt bölge içi kayıt dışı istihdam oranı heterojenlięini göstermek amacıyla, HLFS'te 81 ili geri kazandıran tanımlama stratejisini uygulamaktadır; bu durum, NUTS-2 alt bölge sınırlarının il sınırlarının üzerine bindirildięi bir harita gösteren Şekil 2'de sunulmaktadır. Haritada görülebileceęi üzere, belirli NUTS-2 ortalamaları, yüksek düzeyde sanayileşmiş alanları komşu tarım illeriyle birleştirmekte, bu makalede geliştirilen il çözünürlüğü ise formal üretim koridorlarına bitişik, yüksek kayıt dışılıęa sahip yerelleşmiş cepleri ayırt etmektedir.

Şekil 2. İl bazında kayıt dışılık oranı, 2004-2013



Kaynak: TÜİK 2004-2013 Hanehalkı İşgücü Anketi mikroverisi ve yazarın kendi hesaplamaları. Notlar: Harita, 2004'ten 2013'e kadar olan veri kümelerinin toplulaştırılmasıyla elde edilen il düzeyindeki kayıt dışı istihdam oranını göstermektedir. NUTS-2 alt bölge sınırları (kalın çizgiler) bilgilendirme amacıyla gösterilmiştir.

Bu ayrıntılı haritalama, Türkiye'de işgücü kayıt dışılığı'nın bir dereceye kadar ile özgü olduğunu göstermekte ve hedefli belediye denetimi ile bölgesel politika değerlendirmesi için NUTS-3 çözünürlüğünün gerekliliğini vurgulamaktadır.

İl verilerinin analizi ayrıca, tek bir NUTS-2 idari toplulaştırması altında eşleştirilmiş, ekonomik açıdan birbirine zıt illerin ayrıştırılmasının önemine de işaret etmektedir. TRC1 bölgesi içinde Gaziantep, geniş formal tedarik zincirlerine sahip, yüksek düzeyde sanayileşmiş bir üretim gücü olarak işlev görürken, komşu Adıyaman ağırlıklı olarak küçük ölçekli tarıma ve daha yüksek kayıt dışı işgücü bağımlılığına dayanmaktadır. Benzer şekilde, TRC2 bölge sınırını paylaşımlarına rağmen, Şanlıurfa'nın işgücü piyasası yüksek kayıt dışı istihdam oranlarıyla büyük ölçüde tarımsal üretime dayalıyken, Diyarbakır önemli ölçüde daha yüksek bir formal kamu sektörü istihdamı ve hizmet odaklı faaliyet yoğunluğuna sahiptir. Bir başka örnek olarak, TR32 kapsamında Muğla'nın ekonomisi temelde mevsimsel, hizmet odaklı kıyı turizmi tarafından yönlendirilirken, Denizli tamamen farklı bir formal istihdam yapısına sahip, iç kesimde yer alan bir tekstil ve ihracat üretim merkezini temsil etmektedir.

3.4 Örneklem büyüklüğü kısıtları ve güven aralıkları

TÜİK'in Hanehalkı İşgücü Anketi'nin (HLFS) örneklem tasarımı resmi olarak NUTS-2 düzeyinde temsil edicidir; bu nedenle daha az nüfuslu NUTS-3 illeri için daha küçük yıllık örneklem büyüklükleri () ortaya çıkmaktadır. Bu durum, örneklem büyüklüğünün kabul edilebilir güven aralıklarıyla çeşitli kesitlere hâlâ olanak tanıdığı Şanlıurfa gibi büyük illerde bir sorun teşkil etmemekte, ancak daha küçük illerde yapılan belirli analiz türleri için bir kısıt oluşturmaktadır. Bu belirsizliği azaltmaya yönelik öneri, 2004-2013 döneminin birden fazla yılına ait mikroveriyi toplulaştırmaktır. Bu on yıllık dönem, etkin örneklem büyüklüğünü () önemli ölçüde artırma, standart hataları azaltma ve ampirik politika analizine uygun, sağlam ve güvenilir il düzeyinde kayıt dışı istihdam tahminleri sağlama potansiyeli taşımaktadır. Bu hassasiyet sınırlarını belgelemek amacıyla, standart hatalar ve %95 güven aralıkları Ek C'de raporlanmaktadır.⁸ⁿ

Bununla birlikte, bir başka gerçek kısıt daha bulunmaktadır: Türkiye HLFS'inin kamuya açık mikroveri dosyaları, birincil örneklem birimi (PSU) tanımlayıcısı içermemektedir. TÜİK'in örneklem tasarımı, ilk aşamada yaklaşık 100 hanehalkından oluşan bloklar şeklindeki PSU'ların seçildiği iki aşamalı katmanlı küme tasarımıdır; bu PSU tanımlayıcısı dış araştırmacılara sunulmamakta ve blok düzeyi tanımlayıcılarına ilişkin herhangi bir anket belgesi kamuya açık değildir. Bunun bulunmaması nedeniyle, birincil örneklem birimini svysset tanımında hanehalkı tanımlayıcısı (famid) kullanarak yaklaşık olarak temsil ediyor, katmanları ise TÜİK'in kendi tasarımında kullanılan gerçek katmanlaşma değişkenleri olan alt bölge (NUTS-2) ve kentsel/kırsal statünün etkileşimi olarak tanımlıyorum. Bu seçim, hanehalkı içi korelasyonu yakalamakta, ancak aynı PSU içindeki birden fazla hanehalkının yayımlanan veride

⁸Bu çalışmada kullanılan mikroveri, Türkiye İstatistik Kurumu (TÜİK) tarafından standart bir Kamu Kullanımı Mikro Veri Lisansı (Dağıtımında Kısıtlama Olmayan Mikro Veri Kullanım Taahhütnamesi) kapsamında sağlanan Hanehalkı İşgücü Anketi'nden (HLFS) elde edilmektedir. Bu makalede önerilen analitik yöntem, 5429 sayılı Türkiye İstatistik Kanunu'nda belirtilen tüm veri gizliliği ve mahremiyet ilkelerine sıkı sıkıya bağlı kalmaktadır. Yeniden yapılandırılan coğrafi kimlikler (NUTS-3 düzeyi) yalnızca toplulaştırılmış il istatistikleri üretmek amacıyla derlenmektedir. Bu yöntem, hiçbir aşamada bireysel hanehalklarını veya katılımcıları tanımlamaya çalışmamakta, böylece bireysel veri gizliliğinin tam olarak korunması sağlanmaktadır.

birbirinden ayırt edilememesi nedeniyle, gerçek tasarımdaki blok içi (PSU düzeyi) korelasyonun tamamını yakalayamamaktadır.

Sonuç olarak, bu makale boyunca raporlanan standart hatalar, gerçek örnekleme varyansının muhafazakâr bir alt sınırı olarak yorumlanmalıdır: gözlemlenemeyen bir PSU'yu paylaşan komşu hanehalkları bağımsız çekimler olarak ele alındığından, bu değerler, TÜİK'in gerçek kümeleme yapısı altında elde edilecek standart hatalardan büyük olasılıkla daha dar olacaktır. Tahminlerin hassasiyetini abartmak yerine bu kısıtı açıkça bildiriyorum; bu, kamuya açık mikroverinin destekleyebileceği sınırları gizlemek yerine belgelemeye yönelik bu makalenin genel yaklaşımıyla tutarlıdır.

4. Sonuç

Bu makale, 2004-2013 dönemi boyunca TÜİK'in Hanehalkı İşgücü Anketi'ndeki il (NUTS-3) kimliklerini geri kazanmaya yönelik sistematik bir yöntem sunmaktadır. Ham mikroveride alt bölgelerin görüldüğü sıralı düzenden -resmi Türk araç plaka kodlama sistemiyle (Türkçede plaka) kesinlikle örtüşen bu düzenden- yararlanan önerilen algoritma, blok düzeyindeki gözlemleri karşılık gelen il idari birimlerine atamaktadır. Ayrıca, örnekleme tasarımı değişikliklerini ele almak ve demografik tutarlılığı sağlamak amacıyla, TÜİK'in yayımladığı ulusal genişletme toplamlarıyla tam uyum içinde kalırken kentsel-kırsal temsili koruyan il düzeyi anket ağırlıkları üreten bir post-stratifikasyon düzeltme matrisi geliştirilmiştir. Bu mekânsal geri kazanım sürecinin sağlamlığı, yeniden yapılandırılan işgücü piyasası göstergelerinin, çeşitli il büyüklükleri ve sektör sınıflandırmaları genelinde Sosyal Güvenlik Kurumu (SGK) idari kayıtlarıyla yüksek korelasyon gösterdiği iç ve dış doğrulama çalışmalarıyla ortaya konmaktadır.

Kanıtlanmış doğruluğuna rağmen, tanımlama çerçevesi hem algoritmik stratejiye hem de TÜİK'in altta yatan örnekleme tasarımına içkin birkaç yapısal kısıt altında işlemektedir. Metodolojik olarak, 81 ilin çoğunluğu sıralı taramalar yoluyla deterministik biçimde geri kazanılırken, 12 daha küçük il yerel demografik dağılımlara dayalı olasılıksal atama gerektirmektedir. Anket düzeyinde, örnekleme çerçevesi belirli küçük illeri -en dikkat çekici olarak Tunceli ve Kilis'i- bazı yıllarda veri toplama sürecinin dışında bırakmıştır; bu da mekânsal kapsamın her dalgada neredeyse ama tam olarak değil, kapsayıcı olduğu anlamına gelmektedir. Ayrıca, daha düşük nüfuslu illerde, belirli kırsal veya kentsel katmanlar daha küçük örneklem büyüklükleri sergilemekte, bu da yüksek düzeyde ayrıştırılmış alt grup tahminleri etrafındaki güven aralıklarını doğal olarak genişletmektedir.

Bu teknik kısıtlar, yeniden yapılandırılmış veri kümesinin sağladığı önemli analitik kazanımlarla telafi edilmektedir. Bu yüksek çözünürlüklü işgücü anketi setinin değerine bir örnek, resmi Türk mikroverisi kullanılarak il düzeyinde kayıt dışılık oranlarının ilk kez hesaplanmasıyla sunulmakta ve standart NUTS-2 raporlaması altında görünmez olan çarpıcı bölge içi farklılıkları ortaya çıkarmaktadır. Nihayetinde, bu yöntem 2004-2013 dönemine ait il göstergelerini kurtarmakla birlikte, sonraki anket yeniden yapılandırmasının illeri ne belirlenebilir bir ayırım ne de bir mantık taşıyan ortak bloklar halinde gruplandırması nedeniyle, 2014 sonrası HLFS dalgalarına doğrudan uygulanamamaktadır. Yine de, bu makalenin temel mesajı, sürekli bir adli veri yeniden yapılandırmasını savunmak değil, TÜİK'in NUTS-3 mekânsal kimliklerini resmi olarak yayımlamasının gerekliliğini vurgulamak; araştırmacılara ve karar alıcılara, Türkiye'de etkili, yerel öncülüğünde ve kanıta dayalı politika tasarımı için gerekli olan idari hassasiyeti sunmaktır.

Kaynakça

- Angel-Urdinola, D. F., & Tanabe, K. (2012). Micro-determinants of informal employment in the Middle East and North Africa region (Social Protection and Labor Discussion Paper No. 1201). World Bank.
- Aydemir, A., & Kırdar, M. G. (2017). Low wage returns to schooling in a developing country: Evidence from a major policy reform in Turkey. *Oxford Bulletin of Economics and Statistics*, 79(6), 1046–1086.
- Bağır, Y. K., Küçükbayrak, M., & Torun, H. (2021). Declining labor market informality in Turkey: Unregistered employment and wage underreporting (CBRT Working Paper Series No. 21/19). Central Bank of the Republic of Turkey.
- Baltagi, B. H., & Başkaya, Y. S. (2022). Spatial wage curves for formal and informal workers in Turkey. *Journal of Spatial Econometrics*, 3(1), Article 1–24.
- Del Carpio, X. V., & Wagner, M. (2015). The impact of Syrian refugees on the Turkish labor market (World Bank Policy Research Working Paper No. 7402). World Bank.
- Dinur, I., & Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (pp. 202–210). Association for Computing Machinery.
- Enevoldsen, N. (2023, February 6). Filling gaps in India's official labor force survey: Constructing an urban panel (EGC Perspectives). Yale University.
- Garfinkel, S., Abowd, J. M., & Martindale, C. (2019). Understanding database reconstruction attacks on public data. *Communications of the ACM*, 62(3), 46–53.
- Helderop, E., Nelson, J. R., & Grubestic, T. H. (2023). "Unmasking" masked address data: A medoid geocoding solution. *MethodsX*, 10, Article 102090.
- Pinedo Caro, L. (2020). Identifying Syrian refugees in Turkish microdata. *IZA Journal of Development and Migration*, 11(1), 1–21.
- Pinedo Caro, L. (2021). The impact of COVID-19 on the Turkish labour market during 2020–2021: Results from a nowcasting model using Google mobility data (ILO Research Brief Series). International Labour Organization.
- Tansel, A., & Kan, E. Ö. (2012). The formal/informal employment earnings gap: Evidence from Turkey (IZA Discussion Paper No. 6556). Institute of Labor Economics (IZA).
- Tumen, S. (2016). The economic impact of Syrian refugees on host countries: Quasi-experimental evidence from Turkey. *American Economic Review*, 106(5), 456–460.
- Tunalı, İ. (2023). Unofficial user's guide to the Household Labor Force Survey, Turkey. *Ekonomi-tek*, 12(2).

Vilhuber, L., Abowd, J. M., Lewis, E., Goldschlag, N., Ashmead, R., Kifer, D., Leclerc, P., Rodríguez, R. A., Adams, T., Darais, D., Dey, S., Garfinkel, S. L., Moore, S., & Tadros, R. N. (2023). A simulated reconstruction and reidentification attack on the 2010 U.S. Census (CES Working Paper 23-63R). U.S. Census Bureau.

Ek A: HLFS Veri Döngüleri ve Örneklem Örüntüleri

Aşağıda, mikroveride "alt bölge" (NUTS2) değişkeninin görüldüğü sıra, tahmin edilen il ve tahminin arkasındaki mantığın açıklaması olarak plaka numaralandırması birlikte gösterilmektedir. Yer kısıtları nedeniyle yalnızca 2004 yılı kanıt niteliğinde sunulmakta olup, diğer tüm yıllara ait tablolar talep üzerine mevcuttur. Sonraki yıllarda yalnızca kapsamda iyileşme görülmekte, önemli bir değişiklik gözlenmemekte ve bu nedenle tanımlama stratejisinin kalitesi etkilenmemektedir.

Tablo 1. Örneklem veri örüntüsü, plaka sırası ve tanımlama potansiyeli

Döngüdeki sıra	Alt bölge (NUTS2)	Plaka	İlişkili il	Belirsiz konum mu?	Birleşik il (ardışık plaka numaraları nedeniyle belirsizse)
1	TR-62 Adana	1	Adana	Hayır	
2	TRC-1 Gaziantep	2	Adıyaman	Hayır	
3	TR-33 Manisa	3	Afyonkarahisar	Hayır	
4	TRA-2 Ağrı	4	Ağrı	Hayır	
5	TR-83 Samsun	5	Amasya	Hayır	
6	TR-51 Ankara	6	Ankara	Hayır	
7	TR-61 Antalya	7	Antalya	Hayır	
8	TR-90 Trabzon	8	Artvin	Hayır	
9	TR-32 Aydın	9	Aydın	Hayır	
10	TR-22 Balıkesir	10	Balıkesir	Hayır	
11	TR-41 Bursa	11	Bilecik	Hayır	
12	TRB-1 Malatya	12	Bingöl	Hayır	
13	TRB-2 Van	13	Bitlis	Hayır	
14	TR-42 Kocaeli	14	Bolu	Hayır	
15	TR-61 Antalya	15	Burdur	Hayır	
16	TR-41 Bursa	16	Bursa	Hayır	
17	TR-22 Balıkesir	17	Çanakkale	Hayır	
18	TR-82 Kastamonu	18	Çankırı	Hayır	
19	TR-83 Samsun	19	Çorum	Hayır	
20	TR-32 Aydın	20	Denizli	Hayır	
21	TRC-2 Şanlıurfa	21	Diyarbakır	Hayır	
22	TR-21 Tekirdağ	22	Edirne	Hayır	
23	TRB-1 Malatya	23	Elazığ	Hayır	
24	TRA-1 Erzurum	24	Erzincan	Evet	Erzurum (25)
25	TR-41 Bursa	26	Eskişehir	Hayır	
26	TRC-1 Gaziantep	27	Gaziantep	Hayır	
27	TR-90 Trabzon	28	Giresun	Evet	Gümüşhane (29)
28	TRB-2 Van	30	Hakkari	Hayır	
29	TR-63 Hatay	31	Hatay	Hayır	
30	TR-61 Antalya	32	Isparta	Hayır	
31	TR-62 Adana	33	Mersin	Hayır	
32	TR-10 İstanbul	34	İstanbul	Hayır	
33	TR-31 İzmir	35	İzmir	Hayır	

34	TRA-2 Ağrı	36	Kars	Hayır	
35	TR-82 Kastamonu	37	Kastamonu	Hayır	
36	TR-72 Kayseri	38	Kayseri	Hayır	
37	TR-21 Tekirdağ	39	Kırklareli	Hayır	
38	TR-71 Kırıkkale	40	Kırşehir	Hayır	
39	TR-42 Kocaeli	41	Kocaeli	Hayır	
40	TR-52 Konya	42	Konya	Hayır	
41	TR-33 Manisa	43	Kütahya	Hayır	
42	TRB-1 Malatya	44	Malatya	Hayır	
43	TR-33 Manisa	45	Manisa	Hayır	
44	TR-63 Hatay	46	Kahramanmaraş	Hayır	
45	TRC-3 Mardin	47	Mardin	Hayır	
46	TR-32 Aydın	48	Muğla	Hayır	
47	TRB-2 Van	49	Muş	Hayır	
48	TR-71 Kırıkkale	50	Nevşehir	Evet	Niğde (51)
49	TR-90 Trabzon	52	Ordu	Evet	Rize (53)
50	TR-42 Kocaeli	54	Sakarya	Hayır	
51	TR-83 Samsun	55	Samsun	Hayır	
52	TRC-3 Mardin	56	Siirt	Hayır	
53	TR-82 Kastamonu	57	Sinop	Hayır	
54	TR-72 Kayseri	58	Sivas	Hayır	
55	TR-21 Tekirdağ	59	Tekirdağ	Hayır	
56	TR-83 Samsun	60	Tokat	Hayır	
57	TR-90 Trabzon	61	Trabzon	Hayır	
58	TRC-2 Şanlıurfa	63	Şanlıurfa	Hayır	
59	TR-33 Manisa	64	Uşak	Hayır	
60	TRB-2 Van	65	Van	Hayır	
61	TR-72 Kayseri	66	Yozgat	Hayır	
62	TR-81 Zonguldak	67	Zonguldak	Hayır	
63	TR-71 Kırıkkale	68	Aksaray	Hayır	
64	TR-52 Konya	70	Karaman	Hayır	
65	TR-71 Kırıkkale	71	Kırıkkale	Hayır	
66	TRC-3 Mardin	72	Batman	Evet	Şırnak (73)
67	TR-81 Zonguldak	74	Bartın	Hayır	
68	TRA-2 Ağrı	75	Ardahan	Evet	Iğdır (76)
69	TR-42 Kocaeli	77	Yalova	Hayır	
70	TR-81 Zonguldak	78	Karabük	Hayır	
71	TR-63 Hatay	80	Osmaniye	Hayır	
72	TR-42 Kocaeli	81	Düzce	Hayır	

Kaynak: TÜİK'ten HLFS 2004 mikroversi, İçişleri Bakanlığı'ndan Türkiye plaka numaraları ve yazarın kendi hesaplamaları. **Notlar:** Tablo, alt bölge değişkeninin değiştiği anlara dayalı olarak 2004 HLFS veri gruplarının sırasını, Türkiye'nin plaka numaralarının artan sırasını ve bu plaka numarasıyla ilişkili ilin adını birlikte göstermektedir. Bu tablo, 2004 yılındaki on iki alt tur döngüsünden ilkinin karşılık gelmektedir; bu belirli döngüde bulunmayan iller (örneğin Kilis, Bayburt, Tunceli) aynı yıl içindeki diğer döngülerde görünebilir - tüm yıla ait kapsam için bkz. Ek B. Birleşik iller, ilgili ilin nüfus payına eşdeğer bir gözlem payının artan sırada atandığı orantılı bir yöntemle ayrıştırılmaktadır.

Ek B: Veri erişilebilirlik matrisi

Tablo B.1 Zamansal kapsam, mevcut il ve kırsal/kentsel katmanların varlığı.

[illegible]

Kayseri	F	F	F	F	F	F	F	F	F	F
Kilis	M	R	R	R	R	F	F	F	F	F
Kocaeli	F	F	F	F	F	F	F	F	F	F
Konya	F	F	F	F	F	F	F	F	F	F
Kütahya	F	F	F	F	F	F	F	F	F	F
Kırklareli	F	F	F	F	F	F	F	F	F	F
Kırıkkale	F	F	F	F	F	F	F	F	F	F
Kırşehir	F	F	F	F	F	F	F	F	F	F
Malatya	F	F	F	F	F	F	F	F	F	F
Manisa	F	F	F	F	F	F	F	F	F	F
Mardin	F	F	F	F	F	F	F	F	F	F
Mersin	F	F	F	F	F	F	F	F	F	F
Muğla	F	F	F	F	F	F	F	F	F	F
Muş	F	F	F	F	F	F	F	F	F	F
Nevşehir	F	F	F	F	F	F	F	F	F	F
Niğde	F	F	F	F	F	F	F	F	F	F
Ordu	F	F	F	F	F	F	F	F	F	F
Osmaniye	F	F	F	F	F	F	F	F	F	F
Rize	F	F	F	F	F	F	F	F	F	F
Sakarya	F	F	F	F	F	F	F	F	F	F
Samsun	F	F	F	F	F	F	F	F	F	F
Siirt	F	F	F	F	F	F	F	F	F	F
Sinop	F	F	F	F	F	F	F	F	F	F
Sivas	U	F	F	F	F	F	F	F	F	F
Tekirdağ	F	F	F	F	F	F	F	F	F	F
Tokat	F	F	F	F	F	F	F	F	F	F
Trabzon	F	F	F	F	F	F	F	F	F	F
Tunceli	R	F	U	U	U	M	U	F	F	F
Uşak	F	F	F	F	F	F	F	F	F	F
Van	F	F	F	F	F	F	F	F	F	F
Yalova	U	F	F	F	F	F	F	F	F	F
Yozgat	F	F	F	F	F	F	F	F	F	F
Zonguldak	F	F	F	F	F	F	F	F	F	F
Çanakkale	U	F	F	F	F	F	F	F	F	F
Çankırı	F	F	F	F	F	F	F	F	F	F
Çorum	F	F	F	F	F	F	F	F	F	F
İstanbul	F	F	F	F	F	F	F	F	F	F
İzmir	F	F	F	F	F	F	F	F	F	F
Şanlıurfa	F	F	F	F	F	F	F	F	F	F
Şırnak	F	F	F	F	F	F	F	F	F	F

Notlar: Kapsam: 785/810 hücre TAM (%97,0). İller alfabetik sırayla listelenmiştir. 12 çift il (Erzincan/Erzurum, Giresun/Gümüşhane, Nevşehir/Niğde, Ordu/Rize, Batman/Şırnak, Ardahan/Iğdır), kimliklerinin her zaman doğrudan gözlemlenmek yerine tahmin edilmesi nedeniyle her yıl işaretlenmiştir.

F	TAM (kentsel + kırsal mevcut)
U	YALNIZCA KENTSEL (kırsal hücre boş)
R	YALNIZCA KIRSAL (kentsel hücre boş)

M	EKSİK (hiç gözlem yok)
	Çift il

Ek C: Güven Aralıkları ve Hassasiyet Ölçütleri

Tablo C.1. İl düzeyinde kayıt dışı istihdam oranı, Türkiye, toplulaştırılmış 2004-2013

Plaka numarası	İl	Kayıt dışılık oranı	Std. hata	%95 GA (alt)	%95 GA (üst)
1	Adana	0.515	0.0031	0.509	0.521
2	Adıyaman	0.649	0.0075	0.634	0.664
3	Afyonkarahisar	0.481	0.0048	0.471	0.490
4	Ağrı	0.748	0.0044	0.739	0.756
5	Amasya	0.547	0.0070	0.533	0.561
6	Ankara	0.225	0.0017	0.222	0.228
7	Antalya	0.417	0.0027	0.412	0.423
8	Artvin	0.516	0.0104	0.495	0.536
9	Aydın	0.496	0.0038	0.488	0.503
10	Balıkesir	0.472	0.0029	0.466	0.477
11	Bilecik	0.317	0.0087	0.300	0.334
12	Bingöl	0.613	0.0114	0.590	0.635
13	Bitlis	0.636	0.0089	0.618	0.653
14	Bolu	0.426	0.0089	0.409	0.443
15	Burdur	0.505	0.0102	0.485	0.525
16	Bursa	0.331	0.0023	0.326	0.335
17	Çanakkale	0.443	0.0054	0.433	0.454
18	Çankırı	0.387	0.0080	0.371	0.403
19	Çorum	0.554	0.0050	0.544	0.563
20	Denizli	0.404	0.0041	0.396	0.412
21	Diyarbakır	0.613	0.0046	0.604	0.622
22	Edirne	0.479	0.0070	0.465	0.493
23	Elazığ	0.480	0.0065	0.468	0.493
24	Erzincan	0.501	0.0078	0.486	0.516
25	Erzurum	0.566	0.0037	0.559	0.574
26	Eskişehir	0.267	0.0045	0.258	0.276
27	Gaziantep	0.527	0.0036	0.520	0.534
28	Giresun	0.605	0.0055	0.595	0.616
29	Gümüşhane	0.628	0.0102	0.608	0.648
30	Hakkari	0.618	0.0143	0.590	0.646
31	Hatay	0.576	0.0040	0.568	0.583
32	Isparta	0.410	0.0063	0.397	0.422
33	Mersin	0.550	0.0037	0.543	0.558
34	İstanbul	0.258	0.0012	0.256	0.260
35	İzmir	0.333	0.0019	0.330	0.337
36	Kars	0.653	0.0059	0.641	0.664
37	Kastamonu	0.531	0.0049	0.521	0.540
38	Kayseri	0.345	0.0042	0.337	0.354
39	Kırklareli	0.415	0.0058	0.403	0.426
40	Kırşehir	0.417	0.0100	0.397	0.436
41	Kocaeli	0.260	0.0032	0.253	0.266
42	Konya	0.478	0.0025	0.473	0.483

Plaka numarası	İl	Kayıt dışılık oranı	Std. hata	%95 GA (alt)	%95 GA (üst)
43	Kütahya	0.444	0.0045	0.435	0.453
44	Malatya	0.492	0.0047	0.483	0.501
45	Manisa	0.465	0.0039	0.458	0.473
46	Kahramanmaraş	0.507	0.0048	0.498	0.516
47	Mardin	0.570	0.0065	0.557	0.582
48	Muğla	0.452	0.0044	0.443	0.460
49	Muş	0.764	0.0065	0.751	0.776
50	Nevşehir	0.455	0.0083	0.439	0.471
51	Niğde	0.537	0.0072	0.523	0.551
52	Ordu	0.618	0.0038	0.610	0.625
53	Rize	0.601	0.0082	0.585	0.617
54	Sakarya	0.450	0.0044	0.442	0.459
55	Samsun	0.561	0.0036	0.554	0.568
56	Siirt	0.594	0.0149	0.565	0.623
57	Sinop	0.494	0.0064	0.482	0.507
58	Sivas	0.440	0.0069	0.427	0.454
59	Tekirdağ	0.344	0.0036	0.337	0.351
60	Tokat	0.654	0.0050	0.644	0.663
61	Trabzon	0.562	0.0037	0.555	0.569
62	Tunceli	0.448	0.0235	0.402	0.494
63	Şanlıurfa	0.737	0.0041	0.729	0.745
64	Uşak	0.381	0.0059	0.370	0.393
65	Van	0.696	0.0049	0.686	0.706
66	Yozgat	0.496	0.0070	0.482	0.510
67	Zonguldak	0.591	0.0039	0.584	0.599
68	Aksaray	0.489	0.0067	0.476	0.502
69	Bayburt	0.535	0.0162	0.503	0.566
70	Karaman	0.458	0.0070	0.444	0.471
71	Kırıkkale	0.341	0.0073	0.327	0.356
72	Batman	0.577	0.0083	0.561	0.594
73	Şırnak	0.595	0.0102	0.575	0.615
74	Bartın	0.575	0.0076	0.560	0.590
75	Ardahan	0.737	0.0175	0.701	0.770
76	Iğdır	0.699	0.0138	0.671	0.725
77	Yalova	0.339	0.0105	0.319	0.360
78	Karabük	0.407	0.0083	0.391	0.424
79	Kilis	0.742	0.0174	0.707	0.775
80	Osmaniye	0.604	0.0066	0.591	0.617
81	Düzce	0.409	0.0069	0.396	0.423

Kaynaklar: TÜİK HLFS 2004-2013 mikroverisi ve yazarın kendi hesaplamaları. **Notlar:** Tablo, yeniden yapılandırılan il kimliği, yeniden ağırlıklandırılmış anket ağırlığı ve alt bölge (NUTS2) ile kentsel/kırsal statünün etkileşimi olarak tanımlanan katmanlar kullanılarak, on anket yılının tamamında (2004-2013) toplulaştırılmış tahminleri göstermektedir. Gerçek PSU tanımlayıcısı kamuya açık mikroveride bulunmadığından, birincil örnekleme birimi hanehalkı tanımlayıcısıyla yaklaşık olarak temsil edilmektedir (bu kısıtın ayrıntılı tartışması için ana metne bakınız). 81 ilin tamamı için nokta tahminleri, doğrusallaştırılmış standart hatalar ve Wald %95 güven aralıkları raporlanmaktadır.

Ek D: Örneklem ile nüfus payı karşılaştırması, il düzeyinde bir karşılaştırma

Tablo D.1 Her ilin örneklemdeki payı ve toplam nüfustaki payı

İl	Örneklem %	Nüfus %	İl	Örneklem %	Nüfus %
Adana	2.73%	2.82%	Kahramanmaraş	1.19%	1.43%
Adıyaman	0.76%	0.82%	Karabük	0.38%	0.31%
Afyonkarahisar	1.19%	0.98%	Karaman	0.50%	0.32%
Ağrı	1.33%	0.75%	Kars	0.66%	0.43%
Aksaray	0.70%	0.52%	Kastamonu	0.95%	0.50%
Amasya	0.54%	0.46%	Kayseri	1.31%	1.66%
Ankara	5.32%	6.39%	Kırıkkale	0.59%	0.40%
Antalya	2.09%	2.64%	Kırklareli	0.54%	0.46%
Ardahan	0.21%	0.15%	Kırşehir	0.52%	0.31%
Artvin	0.27%	0.23%	Kilis	0.14%	0.17%
Aydın	1.26%	1.35%	Kocaeli	2.38%	2.07%
Balıkesir	2.27%	1.57%	Konya	4.55%	2.79%
Bartın	0.35%	0.26%	Kütahya	1.32%	0.81%
Batman	0.72%	0.68%	Malatya	1.05%	1.04%
Bayburt	0.18%	0.11%	Manisa	1.98%	1.84%
Bilecik	0.25%	0.28%	Mardin	1.19%	1.04%
Bingöl	0.44%	0.35%	Mersin	2.01%	2.26%
Bitlis	0.65%	0.46%	Muğla	0.86%	1.10%
Bolu	0.40%	0.38%	Muş	0.82%	0.57%
Burdur	0.30%	0.35%	Nevşehir	0.54%	0.39%
Bursa	3.36%	3.48%	Niğde	0.61%	0.47%
Çanakkale	0.93%	0.67%	Ordu	1.20%	1.02%
Çankırı	0.52%	0.26%	Osmaniye	0.68%	0.65%
Çorum	1.00%	0.75%	Rize	0.54%	0.45%
Denizli	1.05%	1.27%	Sakarya	1.06%	1.18%
Diyarbakır	2.06%	2.08%	Samsun	1.68%	1.71%
Düzce	0.46%	0.46%	Siirt	0.49%	0.41%
Edirne	0.60%	0.55%	Sinop	0.55%	0.28%
Elazığ	0.75%	0.76%	Sivas	0.66%	0.89%
Erzincan	0.56%	0.31%	Şanlıurfa	2.32%	2.23%
Erzurum	1.99%	1.09%	Şırnak	0.61%	0.59%

İl	Örneklem %	Nüfus %	İl	Örneklem %	Nüfus %
Eskişehir	0.84%	1.04%	Tekirdağ	1.11%	1.07%
Gaziantep	2.18%	2.25%	Tokat	0.94%	0.88%
Giresun	0.70%	0.59%	Trabzon	1.23%	1.06%
Gümüşhane	0.20%	0.19%	Tunceli	0.09%	0.12%
Hakkari	0.36%	0.36%	Uşak	0.58%	0.46%
Hatay	1.70%	1.96%	Van	1.47%	1.38%
İğdır	0.33%	0.25%	Yalova	0.25%	0.27%
İsparta	0.55%	0.59%	Yozgat	0.71%	0.68%
İstanbul	11.41%	17.72%	Zonguldak	1.12%	0.85%
İzmir	5.11%	5.28%			

Kaynak: TÜİK HLFS mikroverisi 2004-2013, ADNKS'den 2007-2013 il nüfusu ve yazarın kendi hesaplamaları. Notlar: Tablo, illerin 2007'den 2013'e kadar Türkiye'nin toplam nüfususundaki ve 2004-2013 toplulaştırılmış örneklem büyüklüğündeki paylarını göstermektedir.