

DE MACHT VAN HET GETAL

Een statistische kritiek van de bevolkingsstatistieken in België, 1801-1976

- **Sven Vrielinck** -

*Il me reste une montagne de milliers de chiffres et
je me sens, au pied de la montagne, tout misérable*
(Jean Stengers, 1969)

De afgelopen twee eeuwen heeft de overheid zich in toenemende mate inspanningen getroost om de maatschappelijke realiteit in cijfers samen te vatten. Een groeiend aantal statistieken over uiteenlopende sectoren van de samenleving zagen het licht. Maar in hoeverre hebben deze inspanningen vruchten afgeworpen? Hoeveel geloof kunnen we hechten aan het cijfermateriaal dat het verleden ons heeft nagelaten? Kwalitatieve studies over de ontwikkeling van het statistische apparaat wekken de indruk dat het waarheidsgehalte is toegenomen. De vraag blijft of dit vermoeden klopt en hoe dit moderniseringsproces is verlopen. We onderzoeken langs kwantitatieve weg hoe de betrouwbaarheid is geëvolueerd. We richten ons op de meest elementaire statistieken over de bevolking vanuit een lange termijn perspectief. We maken gebruik van technieken die nog niet eerder zijn toegepast, waaronder de zogenaamde Wet van Benford.

I. Het feilbare getal

Het statistisch onderzoek in de sociale wetenschappen heeft de laatste decennia een hoge vlucht genomen¹. Door de voortschrijdende informatisering en de vooruitgang van de communicatietechnologie is het gebruik van kwantitatieve methoden sterk toegenomen. Elke vorser beschikt vandaag over een computer die voldoende krachtig is voor de meest ingewikkelde berekeningen. Steeds meer bronnen komen binnen handbereik van sociale wetenschappers. De Open Access beweging wint veld onder de dataproducenten, terwijl overheidsinstellingen grootschalige databanken met betrekking tot de samenleving en de burgers open stellen voor wetenschappelijk onderzoek². Deze ontwikkelingen heffen fysieke barrières op waarmee statistici in het verleden te kampen hadden. Een onoverzienbare massa aan kwantitatieve en kwantificeerbare gegevens vindt hierdoor een weg naar allerhande toepassingen.

Het stijgend gebruik van numerieke datacollecties doet de vraag naar de betrouwbaarheid van deze gegevens toenemen. Niettegenstaande de grote precisie van computerberekeningen blijft een degelijke bronnenkritiek noodzakelijk. Voorzichtigheid is geboden bij de interpretatie van data, wanneer deze worden aangewend voor toepassingen waarvoor ze oorspronkelijk niet

zijn opgesteld. De bronnenkritiek houdt lang geen gelijke tred meer met de groeiende productie van cijfermateriaal. Het aantal studies waarin de kwaliteit van de data wordt onderzocht, is verhoudingsgewijs gering.

Reeds in 1969 schreef de gezaghebbende historicus Jean Stengers over de groeiende kloof tussen dataproductie en -kritiek³. In een pessimistisch gestemde lezing wees hij op de vele valkuilen waarin gebruikers van statistieken die de tijd ons heeft nagelaten, konden vallen. Voor hem was het kritisch doorlichten van elke cijferreeks niet alleen noodzakelijk, maar ook praktisch onhaalbaar door de hoeveelheid arbeid die dit vergde. Door dit onvermogen stelde hij het gebruik van deze bronnen voor het sociaal-wetenschappelijk onderzoek in vraag. Ondanks dit scepticisme is kwantificering een integraal en onvervreemdbaar onderdeel geworden van de historische wetenschap. De meeste sociaal-historici in België zijn zich bewust gebleven van de valkuilen die Stengers beschreef. De schijnbare overmoed waarmee sommige volgelingen van de Franse *Annales*-school of de Amerikaanse *New Economic History* de kwantitatieve benadering propageren, is hier nauwelijks te vinden. Velen geven blijk van een gematigde aanpak en hanteren een mengsel van traditionele en moderne technieken. Sinds de jaren zestig werden onder impuls van leerlingen van de Gentse contemporanist Jan Dhondt en anderen

1. Mijn dank gaat uit naar Eric Vanhaute en Isabelle Devos voor het kritisch becommentariëren van de tekst, alsook naar Koen Aerts voor de inspirerende steun. 2. Als voorbeeld verwijzen we naar het programma Agora van Federaal Wetenschapsbeleid dat sinds 1999 federale ambtelijke databanken voor wetenschappelijk onderzoek ontsluit. 3. JEAN STENGERS, "L'histoire devant l'abondance statistique", in GINETTE KURGAN, PHILIPPE MOUREAUX (dir.), *La quantification de l'histoire*, Bruxelles, 1973, p. 7-38.

waardevolle bijdragen aan de statistische kritiek geleverd⁴. Desondanks blijft de scepsis over de kwaliteit van het cijfermateriaal latent aanwezig, bij gebrek aan technieken waarmee het kaf van het koren kan worden gescheiden.

Hier volgt een bijdrage om deze leemte te vullen. De berg cijfers die bij Stengers een gevoel van machteloosheid opwekte, is vandaag beter beheersbaar. Dankzij de automatisering staan ons middelen ter beschikking waarmee oude onderzoeksvragen kunnen worden beantwoord. Deze studie belicht het verloop van de datakwaliteit vanuit een lange termijn perspectief. Centraal staat de vraag hoe de geloofwaardigheid van het cijfermateriaal doorheen de 19^{de} en 20^{ste} eeuw is geëvolueerd. Door middel van een vergelijkende studie van een aantal elementaire statistieken willen we een beeld krijgen van de mate waarin de staat capabel was om de maatschappelijke realiteit te vatten. Dit onderzoek richt zich op de officiële statistieken van de bevolking die bewaard zijn gebleven.

Die bronnen bevatten de neerslag van wat gedurende een lange periode als de belangrijkste vorm van statistische activiteit van de overheid mag worden beschouwd. Nog steeds ontbreekt een omvattende en

datagefundeerde kritiek van deze voorname geschiedbronnen. We pogen een oordeel te vellen op grond van een kwantitatieve analyse van de bevolkingscijfers op het laagst beschikbare niveau, dat van de gemeenten. We maken gebruik van eenvoudige technieken die nog niet eerder op historische statistieken werden toegepast. Onze analyse steunt grotendeels op een vergelijking van de cijferfrequentie met de voorspelde verdeling volgens de zogenaamde Wet van Benford die verder wordt toegelicht. Als hypothese poneren we dat deze techniek in staat is om structurele tekortkomingen van de statistische informatieverwerking bloot te leggen. Zowel de voorwaarden waarin deze techniek toepasbaar is, als de drijvende krachten die verantwoordelijk zijn voor de vastgestelde afwijkingen worden onderzocht.

De statistieken die onder de loep worden genomen, zijn afkomstig uit de belangrijkste bronreeksen die ter beschikking staan. De algemene volkstellingen en de archiefreeks *Mouvement de la population* die het verloop van de bevolking weergeeft. Achtereenvolgens stellen we deze bronnen voor en bespreken een aantal eigenschappen die relevant zijn voor de externe bronnenkritiek, waaronder de kostprijs en de praktische organisatie van de tellingen (deel II). Nadien volgt een analyse

4. Een overzicht van de literatuur m.b.t. de historische kritiek van statistische gegevens wordt gegeven in CHRIS VANDENBROEKE, ERIC VANHAUTE, "Statistiek", in JAN ART (red.), *Hoe schrijf ik de geschiedenis van mijn gemeente? Deel 3b: hulpwetenschappen*, Gent, 1996, p. 69-73. Zie tevens de literatuurverwijzingen aangaande de bevolkingsstatistieken verder in deze bijdrage. Ook in andere vakgebieden, waaronder de landbouw en het gerecht, zijn kritische studies beschikbaar, zoals JEAN GADISSEUR, *Le produit physique de la Belgique, 1830-1913: présentation critique des données statistiques. Vol. 1a, Introduction générale: agriculture*, Bruxelles, 1990, p. 26-69; VIRGNIE DETIENNE, FRÉDÉRIC VESENTINI, "La sérialisation des chiffres de la poursuite pénale: formalisation, problèmes, perspectives et automatisation", in FRÉDÉRIC VESENTINI (dir.), *Les chiffres du crime en débat. Regards croisés sur la statistique pénale en Belgique (1830-2005)*, Louvain-la-Neuve, 2005, p. 217-233.

van het cijfermateriaal zelf (deel III). In dit deel confronteren we de numerieke uitkomsten met de resultaten van de hoger beschreven externe bronnenkritiek ter verificatie van de hypothese en ter verklaring van de vastgestelde evolutie van de datakwaliteit.

Voor de vergelijking van het cijfermateriaal met de verwachte verdeling volgens Benford zijn grote aantallen waarnemingseenheden vereist. Daarom beperkt deze studie zich tot de periode vóór 1977, toen het aantal gemeenten op Belgische bodem omvangrijk was, 2.777 in 1801, 2.359 in 1976⁵.

Deze studie zou niet mogelijk zijn geweest zonder LOKSTAT, de Historische Databank van Lokale Statistieken in België⁶. Die infrastructuur herbergt een omvangrijke collectie van historische data met betrekking tot zeer uiteenlopende domeinen die werden gedigitaliseerd in archieven en bibliotheken. Onze bevindingen steunen grotendeels op de analyse van de teluitkomsten, ruim twee miljoen, die door de databank ter beschikking werden gesteld.

II. De pijlers van de bevolkingsstatistiek in België

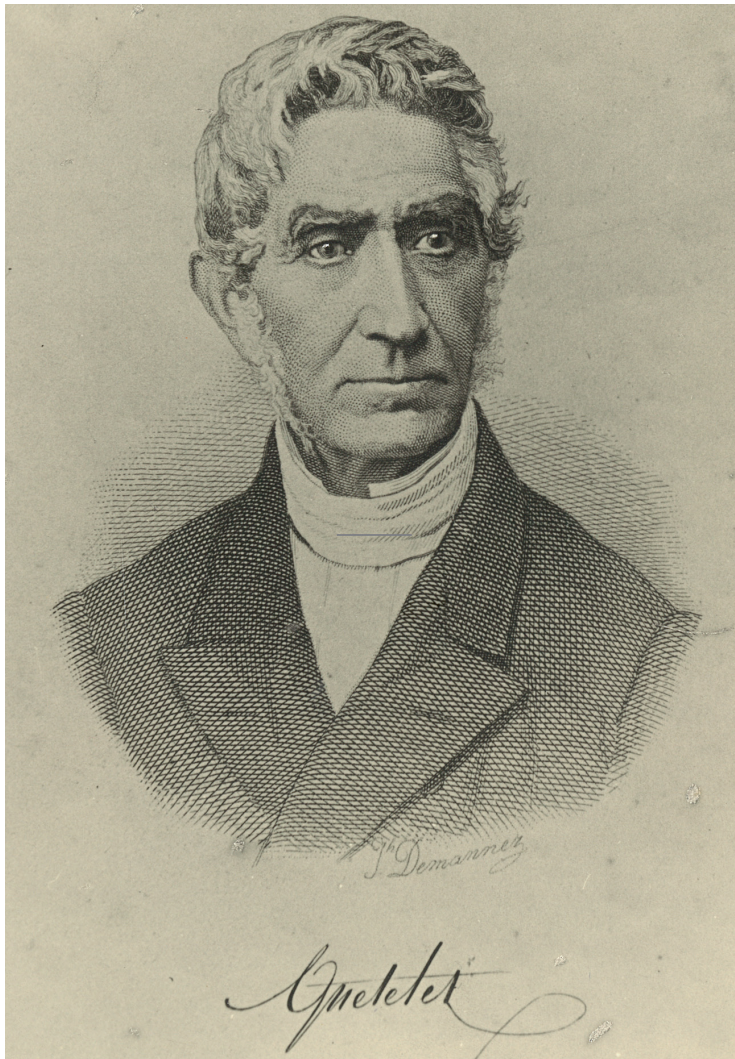
De algemene volkstellingen

Opzet en inhoud

In de beschouwde periode vonden 16 volkstellingen op het huidige grondgebied van België plaats⁷. Het waren grootschalige operaties die met een interval van 5 tot 17 jaar door de centrale overheid op touw werden gezet. Oorspronkelijk bleef de opzet louter van administratieve aard. De bevolkingscijfers werden vastgesteld voor diverse bestuurlijke verrichtingen, zoals de zetelverdeling in het parlement en de repartitie van het contingent dienstplichtigen over de gemeenten. Daarnaast dienden de tellingen voor de opmaak van de bevolkingsregisters waarin alle burgers moesten worden genoteerd. Gaandeweg breidde de vragenlijst zich uit en kreeg de census een meer wetenschappelijk karakter. Behalve de demografische kenmerken van de bevolking, werden in de beschouwde periode inlichtingen ingewonnen over de beleden religie, het taalgebruik, de pendel

5. Ingevolge samenvoegingen slonk het aantal gemeenten in 1977 van 2.359 naar 596.

6. De Historische Databank van Lokale Statistieken in België (LOKSTAT) werd ontwikkeld onder leiding van Eric Vanhaute en Sven Vrielinck aan de Vakgroep Geschiedenis van de Universiteit Gent, als onderdeel van het netwerk HISSTAT dat financiële steun kreeg van de Herculesstichting Vlaanderen, de Universiteit Gent en de Vrije Universiteit Brussel (*Interface Demography*). Bij het digitaliseren van de bronnen heeft LOKSTAT strenge kwaliteitsnormen toegepast. Alle data werden manueel ingetikt en systematisch op invoerfouten nagezien. Hierdoor bleef de foutenmarge klein en verwaarloosbaar voor dit onderzoek. 7. Alleen volkstellingen waarvan de uitkomsten zijn gekend voor het volledige grondgebied van het huidige België worden in beschouwing genomen. De voornaamste uitzonderingen op deze regel zijn de census van 1801 (zonder het departement Ourthe) en de tellingen die in de periode 1816-1910 doorgingen (zonder de zogenaamde Oostkantons). De telling van 1811 valt buiten beschouwing omdat deze slechts op een schatting van de bevolking berust. De census uit het jaar VIII wordt in deze tekst naar het voorbeeld van andere studies met het jaartal 1801 aangeduid.



*Adolphe Quetelet (1798-1874), astronoom, socioloog en grondlegger van de moderne statistiek. Als voorzitter van de "Commission centrale de Statistique" had hij de wetenschappelijk leiding over de volkstelling van 1846 in handen.
(Archief Universiteitsbibliotheek Gent)*

en de huisvesting. Tabel 1 bevat een overzicht van de thema's waaraan statistieken werden gewijd in de opeenvolgende tellingen⁸.

De gemeenten speelden een sleutelrol in de uitvoering van elke census. De lokale besturen stonden in voor de aanstelling van de telagenten die de gegevens op het terrein moesten inzamelen. Sinds 1830 werden voorgedrukte formulieren gebruikt die aan de

gezinshoofden werden uitgedeeld. Daarop moesten de gegevens worden genoteerd. De depoullering van de bulletins, d.w.z. het tellen van de individuele gegevens en het samenvatten van deze gegevens in tabellen, viel lange tijd eveneens onder de verantwoordelijkheid van de gemeenten. Pas vanaf 1920 werden de bulletins op het bevoegde ministerie gecentraliseerd en verwerkt. In dat jaar deden machines hun intrede bij het telwerk.

8. Toelichting bij tabel 1. Telplichtige gemeenten (2) : exclusief gemeenten die in 1801 en 1806 buiten het huidige grondgebied van Frankrijk en België lagen; Superviserende statistische organen (3) : de telling van 1801 werd georganiseerd door het *Bureau administratif* van het ministerie van Binnenlandse Zaken te Parijs. Een half jaar na de beëindiging van de telling in de Belgische departementen kwam in de schoot van dit ministerie een *Bureau de Statistique* tot stand [MARCEL REINHARD, "La statistique de la population sous le Consulat et l'Empire. Le bureau de Statistique", in *Population*, (Paris), 5, 1950 (1), p. 108]. De telling van 1816 werd georganiseerd door het ministerie van Binnenlandse Zaken, afdeling militie; Depoullering (4) : voornaamste instanties belast met het tellen van de individuele gegevens en de redactie van de statistieken; Bijzondere controle-organen (5) : instellingen of administraties met als voornaamste opdracht het nazicht van de door de gemeenten meegedeelde staten en tabellen. Daarnaast behoorde de controle van de gemeenten eveneens tot het vaste takenpakket van de prefecten en onderprefecten, resp. gouverneurs en arrondissementscommissarissen. Thema's (6) : statistieken waarin de officiële uitkomsten van de telling zijn samengevat. Er wordt een onderscheid gemaakt tussen teluitkomsten die gekend zijn voor alle gemeenten uit België of uit minimaal drie provincies (symbool X) en bronreeksen waar dit niet het geval is (symbool 0); Duur (7) : periode van de teldatum tot de datum waarop de laatste uitkomsten in de vorm statistieken officieel werden bekrachtigd of gepubliceerd. De termijn vermeld in 1830 is een uitzondering op deze regel. Het slaat op de periode tot september 1830, toen de revolutie de werkzaamheden van de telling in de Belgische provincies onderbrak. De telresultaten waren alsdan nog niet bekrachtigd door het centrale bestuur. De duurtijd van de teloperaties in de jaren 1880-1970 zijn ramingen met een foutenmarge van maximaal één jaar; Bijzondere kredieten (8) : De vermelde bijzondere kredieten werden ter beschikking gesteld door het centrale gezag om de kosten van de volkstelling te dekken. De bedragen die deels werden gebruikt voor de financiering van andere tellingen die op dezelfde teldatum doorgingen, worden tussen haakjes weergegeven. Een klein deel van het vermelde budget voor de volkstelling 1910 werd ook gebruikt voor de handels- en nijverheidstelling die op dezelfde datum plaatsvond. De telagenten van deze census waren immers ook belast met de uitdeling en de ophaling van de formulieren voor de handels- en nijverheidstelling. Voor de rest verliepen de beide operaties gescheiden. De handels- en nijverheidstelling beschikte over een eigen budget (450.000 fr. voor de dienstjaren 1911 t.e.m. 1914); Bijzondere kredieten, per inwoner (8) : De gemiddelde kredieten per inwoner werden berekend aan de hand van de index van consumptieprijzen (zie noot 9). Voornaamste bronnen tabel 1 : SVEN VRIELINCK, *De territoriale indeling van België (1795-1963)*, Leuven, 2000, p. 2036-2102; MARCEL REINHARD, "La statistique...", p. 103-120; officiële uitkomsten van de volkstellingen uitgegeven door de ministeries van Binnenlandse Zaken (1846-1930) en Economische Zaken (1947-1970); ERIC VANHAUTE, "Bevolking, arbeid, inkomen", in *Inleiding tot de lokale geschiedenis van de 19de en de 20ste eeuw*, Gent, 2003, p. 115-150

	1801	1806	1816	1830	1846	1856	1866	1880	1890	1910	1920	1947	1961	1970
1 Territoriaal bereik	Frans Republiek	Koninkrijk Frankrijk	Belgische provincies	Koninkrijk Nederlanden	Koninkrijk België:									
2 Telplichtige gemeenten	44.309	± 40.000	2895	3844	2521	2531	2551	2583	2596	2617	2629	2638	2671	2670
3 Superviserend statistisch orgaan	Geen	Bureau de statistique (Bln, Ziklen)	Geen	Commissie voor de Statistiek; Bureau voor de statistiek	Centraal Statistiekcommissie (1841-1946), Hoger Raad voor de Statistiek (1946-); Bureau voor Algemene Statistiek (1846-1925); Centrale Dienst voor de Statistiek (1932-1946); Nationaal Instituut voor de statistiek (1946-)									
4 Depouillering van telformulieren	Gemeenten				Gemeenten				Nationaal bureau					
5 Bijzondere controleorganen	Geen	Geen	Geen	Provinciale commissies	Provinciale telbureaus				Provinciale bureaus					
6 Thema statistieken	Nationaal bureau													
Geslacht	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Burgerlijke staat	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Leeftijd					0	0	0	0	0	0	0	0	0	0
Beroep					0	0	0	0	0	0	0	0	0	0
Geboorteplaats					X	X	X	X	X	X	X	X	X	X
Verblijfplaats					X	X	X	X	X	X	X	X	X	X
Scholing					X	X	X	X	X	X	X	X	X	X
Taal					X	X	X	X	X	X	X	X	X	X
Religie			X	X										
Pendel														
Huisvesting				X	X	X	X	X	X	X	X	X	X	X
Wooncomfort				X	X	X	X	X	X	X	X	X	X	X
Gezinnen				X	X	X	X	X	X	X	X	X	X	X
Religieuze orden					X	X	X	X	X	X	X	X	X	X
7 Duur (jaar)	0,9	1,8	0,8	-0,7	3	4	4	4	3	4	6	6	8	5
8 Bijz. kredieten: (x 1.000 fr.)	Geen	Onbekend	Geen	Is bedrag onbekend	(612)	(200)	(740)	(900)	450	747	575	3.886	13.625	(60.000)
Fr. per inwoner (index 1914)				Onbekend	-0,15	-0,04	-0,15	-0,18	0,09	0,13	0,13	0,13	0,21	-0,24
9 Getijdtijge tellingen	Geen	Geen	Geen	Geen	Landb. Nijverh.	Landb.	Landb. Nijverh.	Landb. Nijverh.	Geen	Geen	Geen	Geen	Handel, Nijverh.	Handel, Nijverh.

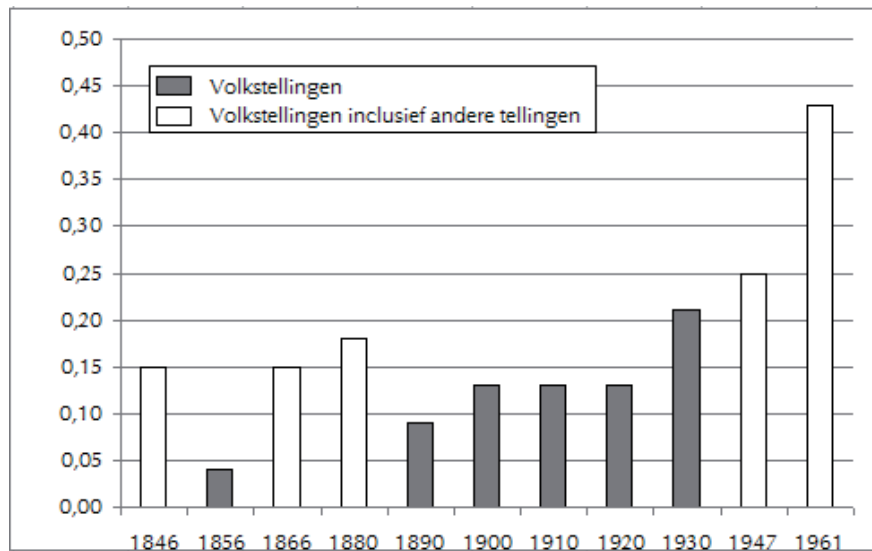
Tabel 1:

Overzicht van de algemene volkstellingen in België: organisatie, inhoud statistieken, duur en kostprijs (X: uitkomsten per gemeente; 0: uitslagen niet bekend gemaakt voor alle gemeenten)

Kostprijs

Volkstellingen die op basis van een individuele bevraging van de burgers tot stand komen, zijn grootschalige ondernemingen die een aanzienlijk budget vergen. Van 1830 af maakte het centraal gezag significante sommen vrij voor de vergoeding van de tellers,

het aanwerven van tijdelijk personeel die de telbureaus bemanden en het drukken van de telresultaten. De hoogte van de kredieten ging algemeen in stijgende lijn, ook wanneer we de bedragen per inwoner op een vergelijkbare basis naast elkaar plaatsen (figuur 1)⁹.



Figuur 1:
Gemiddelde kostprijs van de volkstellingen per inwoner (frank, geïndexeerd basis 1914)
(1846, 1866, 1880, 1947, 1961 inclusief tellingen van landbouw of nijverheid)¹⁰

De toekenning van de bedragen ging niet altijd zonder slag of stoot. In 1856 roerde het parlement zich sterk¹¹. Velen vonden een nieuwe census te duur en overbodig. Dat de telling er toch kwam, is te danken aan de

regering die sterk snoeide in de budgetten en het utilitair karakter van de operatie dik in de verf zette. Door de besparingsdrift was de census van 1856 de meest goedkope die in België heeft plaatsgevonden. Welke

9. De kostprijs werd berekend op basis van het indexcijfer van de consumptieprijsen zoals becijferd door het Leuvense *Institut des Sciences économiques* en de Ministeries van Arbeid en Economie voor de periode 1831-1975 (VALÉRY JANSSENS, *De Belgische frank*, Antwerpen, 1976, p. 433). Figuur 1 is louter bedoeld om de algemene evolutie weer te geven. 10. De budgetten zijn terug te vinden in de parlementaire stukken (1920, 1930, 1947, 1961) en in de publicaties waarin de officiële telresultaten werden bekend gemaakt (1846-1920). 11. NELE BRACKE, *Een monument voor het land : overheidsstatistiek in België, 1795-1870*, Gent, 2008, p. 334-335.

effecten de krappe budgetten hadden op de datakwaliteit, proberen we verder te achterhalen¹².

Volledigheidshalve dient vermeld dat de tellingen vóór 1830 zonder extra geldmiddelen tot stand kwamen¹³. De opzet van deze ondernemingen was een stuk minder ambitieus, zoals uit de lijst van de behandelde thema's valt af te leiden (tabel 1).

Bronkritiek

Wie wil weten welke census meer of minder vertrouwen verdient, komt in de literatuur onvoldoende aan zijn trekken. Hoewel de oude bevolkingscijfers in uiteenlopende publicaties nog steeds gretig worden gebruikt, ontbreekt een coherent oordeel over elke telling afzonderlijk. De schaarse studies die de bronnen systematisch doorlichten, beperken zich doorgaans tot een analyse van de richtlijnen en de organisatie van de verrichtingen¹⁴. Op basis van deze elementen

heerst de indruk dat de kwaliteit van de tellingen toeneemt naarmate de tijd vordert. De richtlijnen werden gedetailleerder, de omkadering nam toe en de overheid leerde uit eerdere ervaringen. De census van 1846 springt hierbij in het oog door de minutieuze voorbereiding en de geleverde inspanningen die hun tijd ver vooruit waren.

Maar er klinken ook andere geluiden. Sommige historici zien geen globale verbetering van de datakwaliteit¹⁵. De onzekerheid is een gevolg van de toegepaste methode die indirect is en onvolledig blijft zonder een grondige evaluatie van eigenlijke uitkomsten. Er bestaan weliswaar studies die de teluitslagen systematisch hebben geanalyseerd, maar deze hebben enkel betrekking op de eerste helft van de 19^{de} eeuw. Ze tonen aan dat de tellingen van 1816 en vooral het jaar VIII (1801) aan onderregistratie lijden¹⁶. De kwaliteit van de statistieken gaat in stijgende lijn in 1806, 1830 en 1846¹⁷. Die bevindingen steunen

12. Zie verder (deel III, hoofdstuk Toenemende divergentie). **13.** Hoewel niet gekend is of in 1806 extra middelen voor de uitvoering van de census werden uitgetrokken, gaan we ervan uit dat dit nauwelijks of niet het geval is geweest. De bescheiden wijze waarop deze telling is uitgevoerd, wijst in die richting (SVEN VRIELINCK, *De territoriale...*, p. 2053-2061). **14.** Voor een kritisch oordeel over de verschillende volkstellingen is men nog steeds aangewezen op oudere werken, waaronder "De algemene tellingen van de bevolking, de handel, de nijverheid en de landbouw sedert 1846 in België gehouden", in *Algemene volks-, nijverheids- en handelstelling op 31 december 1947. Deel I*, Brussel, 1949, p. 25-27; GUY GADEYNE, "De volkstelling van 1814 in West- en Oost-Vlaanderen", in *Belgisch Tijdschrift voor Nieuwste Geschiedenis*, 1981, p. 59-76; Id., "De maatregelen uitgevaardigd door de Franse hoofdsturen in West-Vlaanderen betreffende bevolkingsregisters en -tellingen (1784-1814)", in *Handelingen van het Genootschap voor Geschiedenis*, 112, 1975 (3-4), p. 231-357; ANNE-MARIE BOGAERT-DAMIN, LUC MARÉCHAL, *Bruxelles : développement de l'ensemble urbain 1846-1961: analyse historique et statistique des recensements*, Namur, 1978. Zie ook noot 54. **15.** De scepsis ten aanzien van het verloop van de datakwaliteit kwam tot uiting op de studiedag "Spiegels van de werkelijkheid" die op 21 november 2003 op het Algemeen Rijksarchief te Brussel doorging. **16.** SVEN VRIELINCK, "Boerenbedrog : de volkstelling van het jaar VIII in de Belgische departementen", in *Docendo discimus. Liber amicorum Romain Van Eenoo*, Gent, 1999, p. 97-115; R. OLBRECHTS, "De loop van de bevolking der Belgische provinciën van het begin der XIX^e eeuw tot aan de omwenteling van 1830", in *Statistisch bulletin*, 33, 1947 (2), p. 1528-1530.

17. SVEN VRIELINCK, "De territoriale...", p. 2098-2100.

op klassieke statistische methoden waarmee grote onregelmatigheden aan het licht kunnen worden gebracht (frequentie van uitbijters, sexratio's), maar die helaas niet volstaan om afwijkingen na 1846 te detecteren.

Over specifieke deelgebieden zijn meer kritieken voorhanden. Vooral het taalgebruik dat in het kader van de volkstellingen werd gemeten, was het voorwerp van menige kritische studie en commentaar¹⁸. Algemeen overheerst het oordeel dat deze cijfers hoogstens een trend weergeven en geen accurate weergave van de realiteit. Het taalgebruik was van meet af aan een complex en beladen thema van grote politieke betekenis. Reeds in 1846 rees protest tegen vermeende manipulatie¹⁹.

Naarmate de communautaire tegenstellingen in België toenamen, groeiden ook de klachten over valse verklaringen, intimidaties, vooraf ingevulde formulieren en andere vormen van beïnvloeding. Het taalluik werd pas in 1961 uit de vragenlijst geschrapt. Minder talrijk maar even belangrijk zijn de onderzoeken die de beroepsstatistieken onder de loep nemen²⁰. Eén van deze studies uitgevoerd door Gubin

en Van Neck, in samenwerking met Stengers, reveleerde grote tekortkomingen in 1846²¹. Aan de hand van een steekproef en vergelijking met de bevolkingsregisters legde dit werk fundamentele gebreken bloot die volgens de auteurs een donkere schaduw werpen op de beroepsstatistiek. Dit harde oordeel wordt niet door iedereen gedeeld. Recente bijdragen tonen aan dat beroepsgegevens mits een doordacht en kritisch gebruik tot waardevolle onderzoeksresultaten kunnen leiden²². De verschillende bevindingen mogen in elk geval niet tot andere domeinen worden uitgebreid. Het gaat om specifieke casussen die elk gekenmerkt worden door een sterke complexiteit die maar moeilijk in statistieken valt samen te vatten.

Teneinde een klaar beeld te geven van het vermogen die de overheid heeft ontwikkeld om de maatschappelijke realiteit door middel van statistiek te beschrijven, dienen in de eerste plaats variabelen te worden gekozen die slechts elementaire kennis vergen en hierdoor gemakkelijk vergelijkbaar zijn door de tijd. De officiële bevolkingscijfers naar burgerlijke staat voldoen aan dit criterium en worden verder geanalyseerd.

18. Een overzicht van de literatuur en de problematiek aangaande talentellingen geeft MACHTELD DE METSENAERE, "Talentellingen", in *Nieuwe Encyclopedie van de Vlaamse Beweging*, Tielt, 1998, dl. 3, p. 2950-2955. **19.** JULES HANNES, GER SCHMOOK, "De volkstelling van 1846 en een protestbrief van H. Conscience", in *Tijdschrift van de VUB*, 22, 1981, p. 41-61. **20.** Voorbeelden van kritische studies zijn GUIDO DE BRABANDER, *De regionaal-sectoriële verdeling van de economische activiteit in België (1846-1979) : een kritische studie van het bronnenmateriaal*, Leuven, 1984, p. 36-39; NELE BRACKE, *Bronnen voor de industriële geschiedenis : gids voor Oost-Vlaanderen (1750-1945)*, Gent, 2000, p. 203-204. **21.** ELIANE GUBIN, ANNE VAN NECK, "La répartition professionnelle de la population belge en 1846 : un piège statistique", in *Histoire et méthode IV*, (Acta Historica Bruxellensia), Bruxelles, 1981, p. 269-365. **22.** Zie bv. BART VAN DE PUTTE, ERIK BUYST (eds.), "Measuring social stratification in historical research : an overview of old problems and new methods", in *Belgisch Tijdschrift voor Nieuwste geschiedenis*, 40, 2010 (1-2), 293 p.

De “Mouvement de la Population”

Opzet en inhoud

De met de Franse benaming *Mouvement de la population* aangeduide archiefreeks staat bekend als de tweede pijler van de bevolkingsstatistiek in België²³. Terwijl volkstellingen momentopnamen geven, biedt deze bron een dynamisch overzicht van de natuurlijke en migratorische bevolkingsbeweging. De reeks bestaat uit ruim 700 registers met diverse tabellen aangaande migratie, geboorten, huwelijken en sterfte uit de periode 1841-1976²⁴. De gegevens werden gebruikt door de overheid voor de berekening van de bevolkingscijfers in de jaren vallend tussen de algemene tellingen. Daarnaast leverden de tabellen informatie voor wetenschappelijke publicaties op vlak van demografie, volksgezondheid, ruimtelijke mobiliteit en andere studiegebieden.

De registers werden bijgehouden door ambtenaren van het ministerie van Binnenlandse Zaken, later Economische Zaken, op basis van de inlichtingen die ze via de arrondissementscommissarissen van de gemeenten ontvingen. Elk gemeentebestuur in het land was verplicht om jaarlijks de gegevens te berekenen. Ze moesten op voorgedrukte formulieren worden ingevuld en naar Brussel gestuurd. Voor het vaststellen van de cijfers waren ze aangewezen op de registers van de Burgerlijke Stand en de

bevolking, het doodsoorzakenregister en het vreemdelingenregister. De selectie van de numerieke variabelen was in handen van de Centrale Statistiekcommissie. Van bij haar oprichting in 1841 nam zij de wetenschappelijke leiding van de *Mouvement* waar. De interesse van de commissieleden is ook in de vragenlijsten zichtbaar. Naast kennis over elementaire zaken toonden de leden een opvallende belangstelling in meerlingen, onwettige geboorten en honderdjarigen. Conform de tijdgeest ging vanaf het midden van de 19^{de} eeuw veel aandacht naar dodelijke ziekten en aandoeningen uit.

De *Mouvement* weerspiegelde sinds haar ontstaan het belang die de overheid aan de reeks hechtte. Vier perioden van uiteenlopende activiteitsgraad vallen te onderscheiden²⁵. Het eerste decennium kenmerkte zich door grote inspanningen. De variabelen werden systematisch opgetekend en elk jaar in een afzonderlijk boekdeel gepubliceerd. Nadien volgde een terugval. De periode van 1851 tot het begin van de jaren 1880 kende minder activiteit. Alleen de belangrijkste cijfers kwamen in de registers terecht. De indeling van de tabellen wijzigde herhaaldelijk. Veel van de registers uit die jaren verkeren, allicht niet toevallig, in slechte staat. Nadien vond de *Mouvement* haar tweede adem. In de jaren 1881-1885 liet de overheid opnieuw de voornaamste cijfers per gemeente publiceren. Na een herschikking en uitbreiding van

23. MARIJ PRENEEL, “Over levendgeboren onwettige meerlingen in Wortel, Luikse emigratie naar Argentinië en epidemische meningitis in Zedelgem : de Mouvement de la Population et de l’État Civil : een nauwelijks ontgonnen bron voor historische demografische statistiek in België”, in THIERRY EGGERICKX, JEAN-PAUL ANDERSON (dir.), *Histoire de la population de la Belgique et de ses territoires*, Louvain-la-Neuve, 2010, p. 73-81. 24. De registers worden bewaard op het Algemeen Rijksarchief te Brussel. Een plaatsingslijst is beschikbaar. 25. MARIJ PRENEEL, “Over levendgeborenen...”, p. 77-80.



Een ambtenaar bezoekt een woonwagenfamilie voor een volkstelling (Nederland, 1925). (Bron <http://www.egoproject.nl/woonwagenbeleid/>)

de tabellen in 1886 verliep de registratie gedurende meer dan 70 jaar ononderbroken. Zowel vormelijk als inhoudelijk vertonen de tabellen tot 1959 een grote stabiliteit. Nadien volgde een afslanking en heroriëntatie deels ingevolge de intrede van de computer.

Of deze periodisering op basis van uiterlijke kenmerken ook sporen heeft nagelaten op de kwaliteit van het cijfermateriaal zelf, onderzoeken we verderop aan de hand van een aantal basisvariabelen uit de *Mouvement*.

Bronnenkritiek

Niettegenstaande het gros van het cijfermateriaal uit de *Mouvement* nooit is uitgegeven en bestudeerd, zijn in de wetenschappelijke literatuur een aantal fundamentele kritieken aangaande de officiële statistieken van het bevolkingsverloop te vinden. Uit lokale studies blijkt dat de omvang van de migratie bijna steeds werd onderschat. Mensen die verhuisden lieten vaak na dit te melden aan het gemeentebestuur, hoewel ze daartoe sinds 1856 wettelijk verplicht waren²⁶. Vooral de uitwijking zou aan onderregistratie lijden in de registers van de bevolking die de gemeenten bijhielden. Naarmate verder wordt teruggegaan in de tijd, neemt het fenomeen grotere proporties aan. Uit een steekproef van bevolkingscijfers van 371 Waalse gemeenten

bleekt dat de registratie van de migraties in de loop van de tweede helft van de 19^{de} eeuw gevoelig verbeterde²⁷.

Wat betreft het natuurlijke bevolkingsverloop boezemen de statistieken meer vertrouwen in. Afgezien van de Franse Tijd zijn geen vermeldenswaardige onregelmatigheden in de Burgerlijke Stand vastgesteld²⁸. Dit positief beeld is niet van toepassing op het doods-oorzakenregister dat de gemeenten sinds 1851 moesten bijhouden. Door de gebrekkige notatie van de individuele doodsoorzaken ingevolge de beperkte medische kennis, inefficiënte diagnosetechnieken en taboes die op sociaal gevoelige ziekten rusten, dienen de statistieken die op basis van dit register werden opgesteld met een stevige korrel zout te worden genomen. Historici zien de betrouwbaarheid van de medische statistiek slechts geleidelijk toenemen²⁹. De kritische evaluatie van de getelde doodsoorzaken vormt het onderwerp van een afzonderlijke studie en laten we verder buiten beschouwing. Het betreft een complexe problematiek die door de specificiteit los staat van het cijfermateriaal die we verder in de tekst zullen analyseren.

Een probleem met grotere reikwijdte was de moeizame medewerking die sommige gemeenten verleenden. Het was voor de

26. RENÉ LEBOUTTE, *Du registre de population au registre national : deux siècles de pratique administrative en Belgique, 1791-1991*, Badia Fiesolana, 1998, p. 13-14, 25-26. 27. THIERRY EGGERICKX, "Les migrations internes en Wallonie et en Belgique de 1840 à 1939 : un essai de synthèse", in Id., JEAN-PAUL ANDERSON (dir.), *Histoire...*, p. 298. 28. De oudste registers van de Burgerlijke Stand gaven daarentegen blijk van lacunes. Een steekproef uitgevoerd in het departement Jemappes reveleerde aanzienlijke hiaten in de registratie van zowel de geboorten als de sterfte in de Franse Tijd (ROGER DARQUENNE, "La conscription dans le département de Jemappes", in *Annales du Cercle archéologique de Mons*, 67, 1968-70, p. 232-241). 29. ISABELLE DEVOS, *Allemaal beestjes : mortaliteit en morbiditeit in Vlaanderen, 18de-20ste eeuw*, Gent, 2006, p. 49-50; KAREL VELLE, "Statistiek en sociale politiek: de medische statistiek en het gezondheidsbeleid in België in de 19de eeuw", in *Belgische Tijdschrift voor Nieuwste Geschiedenis*, 16, 1985, p. 229-239.

centrale overheid niet evident om meer dan 2500 lokale besturen in de juiste pas te doen lopen. Elk jaar doken gemeenten op die de formulieren met het cijfermateriaal niet of laattijdig verstuurd. Rijksambtenaren hebben herhaaldelijk met het sturen van commissarissen op kosten van de nalatige gemeenten moeten dreigen. Deze stap volstond om de nodige stukken te bekomen. Helaas bestaat geen overzicht van de weerbarstige gemeenten die het bezoek van een commissaris kregen. Een willekeurige steekproef door ons uitgevoerd in de provincie Namen bracht zeven van deze gemeenten aan het licht gedurende de jaren 1879 en 1880³⁰.

Tot slot dient de vraag te worden gesteld in hoeverre de individuele gegevens uit de registers van de Burgerlijke Stand en de bevolking correct werden geteld en doorgegeven. Sommige microstudies hebben verschillen vastgesteld tussen de nominatieve staten en de officiële cijfers³¹. Hoewel historici in geval van twijfel doorgaans de voorkeur geven aan de nominatieve lijsten, bestaat *a priori* geen reden om de een of andere bron als onbetrouwbaar te bestempelen. Statistieken zijn wel gevoeliger voor bepaalde vormen van manipulatie. Onregelmatigheden in het cijfermateriaal kunnen een gevolg zijn van moedwillige tegenwerking van lokale

autoriteiten die zich wilden onttrekken aan militieverplichtingen en andere lasten.

Volgens de gouverneur van Henegouwen was de fraude in 1842 wijdverbreid. Burgemeesters zouden hebben toegegeven zich schuldig te maken aan deze praktijk. *“Ils allèguent pour motif de cette espèce de fraude, que leur commune serait victime de leur sincérité s'ils faisaient connaître le véritable chiffre de leur population”*³². De omvang, de geografische spreiding en het verloop van dit misbruik is onbekend. Indien we de resultaten van een kritische studie over de volkstelling van het jaar VIII mogen uitbreiden tot latere perioden, waren het vooral de kleine gemeenten die zich schuldig maakten aan het bedrog³³. Het centrale gezag heeft getracht paal en perk te stellen aan die onregelmatigheden. Gouverneurs, arrondissementscommissarissen en ambtenaren van het ministerie ondernamen inspecties bij de gemeenten.

Ondanks de onzekerheid die heerst over de ernst van deze gebreken, overweegt de indruk dat de registratie van de bevolkingsbeweging in de loop de 19^{de} en 20^{ste} eeuw aan precisie heeft gewonnen. De slagkracht van het staatsapparaat nam in het algemeen toe, door de uitbouw van de ambtenarij, de technische vooruitgang en het streven naar

30. RIJKSARCHIEF NAMEN, Archives de l'Administration provinciale de Namur concernant la santé publique, la population et l'état civil (1830-1950), 234, nr. 1425. **31.** Enkele voorbeelden : LEON DHAENE, *Boeren en burgers : sociale geschiedenis van het negentiende eeuwse platteland : Zingem 1795-1900*, Zingem, 1986, p. 117; ERIC VANHAUTE, *Heiboeren : bevolking, arbeid en inkomen in de 19de-eeuwse Kempen*, Brussel, 1992, p. 338. **32.** Brief van gouverneur Liedts aan de minister van Binnenlandse Zaken, 21 juli 1842 (gepubliceerd in *Bulletin de la Commission centrale de statistique*, I, 1843, p. 346). In dezelfde periode toonde Quetelet langs kwantitatieve weg aan dat het officiële bevolkingscijfer van het rijk vermoedelijk met één tiende was onderschat (ADOLPHE QUETELET, “Sur la répartition du contingent des communes dans les levées de la milice”, in *Idem...*, p. 346-364). **33.** SVEN VRIELINCK, “Boerenbedrog...”, p. 97-115.

rationalisering³⁴. Statistiek was een nuttig instrument voor de overheid om de greep op de samenleving te versterken. Steeds grotere budgetten werden vrijgemaakt voor het bekomen van goede data waarmee een efficiënt beleid kon worden gevoerd. Het blijft evenwel de vraag in hoeverre deze inspanningen vruchten hebben afgeworpen. We pogen tot een datagefundeerd oordeel te komen. We onderzoeken in welke mate de centrale overheid erin slaagde om de meest elementaire informatie adequaat te becijferen. Naast de resultaten van de volkstellingen selecteren we hiervoor een aantal basisstatistieken uit de *Mouvement*. Het aantal geboorten, overledenen, inwijkelingen en uitwijkelingen per geslacht, alsook de berekende bevolkingstotalen per gemeente.

III. Kwantitatieve analyse van de bevolkingscijfers

Het vinden van methoden voor het kritisch wegen van bevolkingstatistieken vormt een uitdaging voor historici. Het beschikbare instrumentarium uit Clio's laboratorium volstaat niet. Zoals Stengers aangaf, zijn de traditionele methoden die de authenticiteit van de individuele bron bepalen in de praktijk ondoenbaar door de overweldigende massa gegevens. Bovendien levert de klassieke aanpak vaak geen sluitend oordeel en uniforme vergelijkingsbasis op.

In deze studie wordt de kwaliteit van de statistieken gemeten aan de hand van de aanwezigheid van afwijkingen in het cijfermateriaal. De mate van afwijking of bias is detecteerbaar langs twee invalshoeken :

- Door het blootleggen van abnormale waarden (uitbijters). Numerieke verhoudingen in populaties beantwoorden vaak aan bepaalde wetmatigheden die statistisch uitdrukbaar zijn. De aanwezigheid van waarden die totaal niet passen in het vastgestelde patroon kan wijzen op onregelmatigheden in de informatievergaring. Deze methode levert alleen tastbare resultaten op bij significante afwijkingen, zoals die zich in de oudste volkstellingen voordeden.
- Door het vergelijken van de officiële uitkomsten met verwachte waarden. Er zijn geen geschiedbronnen voorhanden waarmee de oude datasets stelselmatig kunnen worden vergeleken. We moeten onze toevlucht nemen tot theoretische modellen als referentiekader. Onze keuze is gevallen op de zogenaamde Wet van Benford die in de exacte en economische wetenschappen wordt toegepast om numerieke datasets op validiteit te controleren. Achtereenvolgens beschrijven we die wet, stellen we de toepasbaarheid ervan voor ons onderzoek vast en toetsen we de bevolkingscijfers aan het model.

34. We baseren ons op studies die de praktijk van de statistische informatieverwerving en de ontwikkeling van de administratieve statistiek vanuit institutioneel oogpunt hebben onderzocht, waaronder NELE BRACKE, *De economie gekend, de bevolking getemd : overheidsstatistiek in België, 1795-1870*, onuitgegeven proefschrift universiteit Gent, 2004, 2 dln.; Id., *Een monument...*, 467 p.; DIETER DE CNIJF, *Statistisch denken in de Belgische officiële statistiek (1895-1920)*, onuitgegeven verhandeling Universiteit Gent, 2006, 171 p.

De Wet van Benford

Principes

In 1938 publiceerde de Amerikaanse fysicus Frank Benford een wetenschappelijk artikel waarin hij, op basis van empirisch onderzoek, een wet beschreef die stelt dat de frequentie van de begincijfers van getallen uit bepaalde datasets logaritmisch is verdeeld en geen uniforme verdeling volgt zoals intuïtief wordt aangevoeld³⁵. Hij verzamelde duizenden getallen uit zeer uiteenlopende domeinen, variërend van rivierlengten, atoomgewichten en aantallen dagbladen, en toonde aan dat ze alle dezelfde grondregel volgden. Wanneer van elk getal uit de datasets telkens het eerste cijfer wordt genomen, ongeacht de grootte van het getal, treden steeds dezelfde verhoudingen op tussen de geselecteerde cijfers. Ongeveer 30 % begint met een 1, 18 % met een 2, enz³⁶. Een vergelijkbare regel, maar dan minder uitgesproken, is van toepassing op de cijfers op de volgende posities.

De theoretische onderbouw van dit fenomeen, door Benford de *Law of anomalous numbers* genoemd, heeft lang op zich laten wachten³⁷. Pinkham toonde in 1961 langs wiskundige weg aan dat de wet schaalinvariant is. Een mathematische verklaring werd pas in 1996

door Theodore Hill verschaft, die bewees dat bepaalde datasets waarvan de verdeling van de getallen het resultaat is van de combinatie van andere verdelingen, beantwoorden aan de regel die Benford eerder op het spoor was gekomen³⁸.

Opdat een dataset aan het vastgestelde fenomeen beantwoordt, moeten een aantal voorwaarden worden vervuld :

- De getallen mogen niet ontsproten zijn aan het menselijke brein of willekeurig zijn gekozen. Prijzen, postcodes, lottogetallen en bankrekeningnummers zijn voorbeelden van cijfercombinaties die los staan van de beschreven natuurlijke verdeling.
- De getallen in de dataset mogen geen ingebouwd minimum of maximum hebben. Zo zijn hoogtemetingen van het reliëf ongeschikt omdat de zwaartekracht het vormen van hoge bergen tegenwerkt.
- De getelde eenheden moeten hetzelfde attribuut beschrijven.
- De dataset moet voldoende groot zijn, bij voorkeur minstens 10 000 observaties³⁹.

Sinds die ontdekking in 1938 zijn verschijnselen die voldoen aan de wet in de meest uiteenlopende domeinen teruggevonden. Op empirische waarneming gebaseerde

35. FRANK BENFORD, "The law of anomalous numbers", in *Proceedings of the American Philosophical Society*, 78, 3.1938 (4), p. 551-572. De eerste die de wet beschreef was de Amerikaanse astronoom Newcombe in 1881, van wie de bijdrage onopgemerkt bleef. 36. Wiskundig wordt de Wet van Benford uitgedrukt met de formule $P = \frac{1}{a} \log(1 + \frac{1}{a})$ met "a" als eerste cijfer. In deze bijdrage beperken we ons tot het begincijfer van elk getal. Cijfers die niet vooraan staan, komen gemiddeld even vaak voor en zijn hierdoor minder bruikbaar om anomalieën in de LOKSTAT-dataset op te sporen. 37. Een historisch overzicht van het onderzoek aangaande de wet van Benford wordt beschreven in MARK NIGRINI, *Forensic analytics : methods and techniques for forensic accounting*, Hoboken (New Jersey), 2011, p. 89-97. 38. THEODORE HILL, "A Statistical derivation of the significant-digit law", in *Statistical Science*, 10, 1996 (4), p. 354-363. 39. PHILIP DRAKE, MARK NIGRINI, "Computer assisted analytical procedures using Benford's Law", in *Journal of Accounting Education*, 18, 2000, p. 132.

THE LAW OF ANOMALOUS NUMBERS

FRANK BENFORD

Physicist, Research Laboratory, General Electric Company,
Schenectady, New York

(Introduced by Irving Langmuir)

(Read April 22, 1937)

ABSTRACT

It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9. A compilation of some 20,000 first digits taken from widely divergent sources shows that there is a logarithmic distribution of first digits when the numbers are composed of four or more digits. An analysis of the numbers from different sources shows that the numbers taken from unrelated subjects, such as a group of newspaper items, show a much better agreement with a logarithmic distribution than do numbers from mathematical tabulations or other formal data. There is here the peculiar fact that numbers that individually are without relationship are, when considered in large groups, in good agreement with a distribution law—hence the name “Anomalous Numbers.”

A further analysis of the data shows a strong tendency for bodies of numerical data to fall into geometric series. If the series is made up of numbers containing three or more digits the first digits form a logarithmic series. If the numbers contain only single digits the geometric relation still holds but the simple logarithmic relation no longer applies.

Titelblad van het artikel uit 1938 waarin Frank Benford de wet beschreef die naar hem werd genoemd en die sindsdien ruime bekendheid heeft verworven. (www.gatsby.ucl.ac.uk)



*Frank Benford (1883-1948),
Amerikaans fysicus, gaf
bekendheid aan de wet die
naar hem werd genoemd
en die eerder door de
astronoom Newcomb was
ontdekt.*

(www.gatsby.ucl.ac.uk)

data uit de natuurwetenschappen, economie en andere disciplines blijken de regel te gehoorzamen⁴⁰. Een bijzondere toepassing van de wet is het opsporen van anomalieën in datasets. De econoom Hal Varian wees als eerste op de mogelijkheid om fraude op te sporen in tabellen met socio-economische gegevens⁴¹. Mensen die zelf getallen bedenken, zijn geneigd om de cijfers uniform te verdelen. Door de eerste cijfers van de getallen te vergelijken met de verwachte frequentieverdeling van Benford kunnen onregelmatigheden aan het licht komen. De techniek, uitgewerkt door Mark Nigrini en anderen, wordt in toenemende mate toegepast door auditeurs, belastingcontroleurs en economen voor de doorlichting van financiële en macro-economische data⁴².

Validatie van bevolkingsstatistieken

Verscheidene auteurs, waaronder Benford, hebben proefondervindelijk vastgesteld dat bevolkingscijfers aan de hoger beschreven wet van grote getallen beantwoorden. De

eerste cijfers kunnen worden aangewend om de accuratesse van inwonertallen van geografische eenheden te meten. Dat werd door de Japanner Sakai op grond van een vergelijking van statistieken van ontwikkelingslanden aangetoond⁴³.

In deze studie vergelijken we datareeksen afkomstig uit de volkstellingen en de *Mouvement* met de verdeling van Benford en op grond hiervan leiden we conclusies over het verloop van de datakwaliteit. Voor zover gekend, is deze toepassing uniek in de historische wetenschap⁴⁴.

Niettegenstaande de fascinerende eenvoud en de universaliteit van de *Law of anomalous numbers* moeten we ons bewust zijn van de grenzen ervan. Enerzijds zijn datasets die voldoen aan de wet niet noodzakelijk waarheidsgetrouw. De kans dat onregelmatigheden kunnen worden gedetecteerd, daalt naarmate in de productie van de data minder systematiek valt te bespeu-

40. De toepasbaarheid van de wet in uiteenlopende disciplines blijkt o.a. uit MALCOLM SAMBRIDGE, HRVOJE TKALČIĆ, ANDREW JACKSON, "Benford's law in the natural sciences", in *Geophysical Research Letters* (Washington), 37, 2010, p. 1-5. 41. HAL VARIAN, "Benford's law", in *The American Statistician*, 26, 1972, p. 65-66. 42. CHRISTINA GEYER, PATRICIA WILLIAMSON, "Detecting fraud in data sets using Benford's Law", in *Communication in statistics*, 33, 2004 (1), p. 229-246; CINDY DURTSCHI, WILLIAM HILLISON, CARL PACINI, "The effective use of Benford's law to assist in detecting fraud in accounting data", in *Journal of Forensic Accounting*, 5, 2004 (1), p. 17-34; AZIZ MOHAMMADI, *Fraude detecteren met behulp van de Wet van Benford: een simulatiestudie*, Amsterdam, 2008; GEORGE JUDGE, LAURA SCHECHTER, "Detecting problems in survey data using Benford's law", in *The journal of human resources*, 44, 2009 (1), p. 1-24. 43. HIROMICHI SAKAI, "Frequency distribution of first digit among population", in *Jinko mondai kenkyu - Journal of population problems*, 47, 1991, (3), p. 74-78. 44. De consultatie van de bibliografische databank *Historical Abstracts* en de citatie-index *Web of Science* bracht geen geschiedkundige werken aan het licht die gebruik maken van de Wet die Benford heeft beschreven. De Benford Online Bibliography (www.benfordonline.net) vermeldt 754 studies die wereldwijd met betrekking tot de wet zijn verschenen (situatie april 2013). Slechts een beperkt aantal publicaties uit de lijst richten zich op bevolkingsstatistieken. Relevant voor deze studie is het werk van F. Sandron die onderzocht in hoeverre bevolkingscijfers de verdeling van Benford volgen op grond van theoretische groeicijfers over 172 tijdperiodes (FRÉDÉRIC SANDRON, "Do populations conform to the law of anomalous numbers?", in *Population*, 57, 2002 (4), p. 753-761).

ren. Anderzijds wijst de Wet van Benford deviaties aan zonder een uitspraak te doen over de aard en de oorzaak van de afwijking. Fraude, manipulatie of structurele tekortkomingen zijn mogelijke oorzaken van distortie die niet met elkaar mogen worden verward. De wet is niet meer dan een hulpmiddel voor de vorser die zelf de uitkomsten moet interpreteren en elementen uit de externe bronkritiek bij zijn oordeel betrekken. De bevindingen van deze studie kunnen niet zonder meer worden uitgebreid tot andere bronnen of datareeksen.

Algemene conformiteitproef

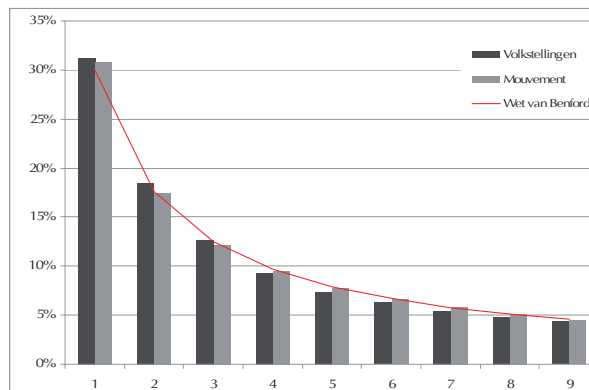
De dataset die we aan de vergelijkende test onderwerpen, omvat ruim 2,6 miljoen getallen die samen de periode 1801-1976 over-

spannen. Zoals hoger beschreven, selecteerden we volgende variabelen uit de bronnen zoals die officieel per gemeente werden geregistreerd :

- Bevolkingscijfers naar burgerlijke staat en geslacht per census (185.837 uitkomsten).
- Aantallen geboorten, sterften, inwijkingen en uitwijkingen per geslacht, plus het aantal huwelijken (2.466.740 getallen).

Deze variabelen zijn in digitale vorm beschikbaar voor de volledige periode met uitzondering van de volkstellingen 1920, 1930 en 1947 en de jaargangen 1851-1879 uit de *Mouvement*.

De mate van overeenkomst tussen de verdeling van de bevolkingscijfers en van Benford valt af te lezen uit volgende figuratieve voorstelling en bijhorende tabel.



	% Eerste cijfers									Totaal N	Conformiteit	
	1	2	3	4	5	6	7	8	9		χ^2	MAD
Wet Benford	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6	-	-	-
Volkstellingen	31,3	18,4	12,7	9,2	7,3	6,3	5,4	4,8	4,4	185 837	0,225	0,005
Mouvement	30,9	17,6	12,2	9,5	7,8	6,6	5,8	5,1	4,5	2 466 740	0,037	0,002

Figuur 2 - tabel 2 :

Frequenties van elk eerste cijfer per getal in de bevolkingsstatistieken en volgens Benford (bron LOKSTAT)

Om de mate van verwantschap te bepalen tussen de verdelingen, bedienen we ons van de chi-kwadraat (χ^2) *goodness-of-fit*toets die voor deze vorm van vergelijking wordt toegepast⁴⁵. De toets gaat na of waargenomen aantallen systematisch afwijken van verwachte (of gemiddelde) aantallen, en berekent het totaal van de gewogen kwadratische afwijkingen tussen deze aantallen. Hoe hoger de uitkomst van de chi-kwadraattoets, hoe sterker de afwijking met het theoretische model van Benford. Uit figuur 2 blijkt dat de samenhang tussen de drie reeksen bijzonder groot is. De volkstelling- en de *Mouvementsets* volgen de verdeling van Benford van dichtbij. De conformiteit is dicht tot aanvaardbaar volgens criteria die Drake en Nigrini op grond van de gemiddelde absolute afwijking (MAD) hanteren⁴⁶.

De onvoorspelbaarheid van kleine getallen

Vooraleer tot een analyse van de bronnen over te gaan, onderwerpen we de historische dataset aan een bijkomende proef. Indien afwijkingen het gevolg zijn van foutieve registraties, moeten vooral kleine aantallen vatbaar zijn geweest voor vertekening. Een onjuiste optekening van 5 in plaats van 6 inwoners zal vaker hebben voorgekomen dan 500 in plaats van 600 bewoners. In tabel 3 zijn de afwijkingen uitgesplitst naar grootte van het bevolkingsgetal. Doordat het aantal observaties tussen bepaalde groottegrenzen onvoldoende talrijk zijn voor een betrouwbare meting van de overeenkomst met de Benfordverdeling hebben we de categorieën cumulatief ingedeeld.

45. Gelet op de opzet van deze bijdrage en de beschikbare ruimte kunnen alleen de voornaamste uitkomsten van de statistische analyses worden weergegeven. De chi-kwadraat *goodness-of-fit*toets werd uitgevoerd op basis van de formule $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$, waarbij c het aantal categorieën is, f_o de geobserveerde en f_e de verwachte celfrequenties. Conform de standaard procedure werden de berekende chi-kwadraten vervolgens vergeleken met een zgn. cutoff value op basis waarvan wordt bepaald of er sprake is van conformiteit of niet. Alle chi-kwadraten berekend in deze bijdrage liggen lager dan de bijhorende cutoff-values (met vrijheidsgraden=8 en p-waarde=0,05). Dit betekent dat de nulhypothese, die stelt dat er sprake is van conformiteit met het model van Benford, op statistische gronden niet kan worden verworpen. We verwijzen naar de statistische handboeken voor meer toelichting in dit verband. 46. Aangezien de chi-kwadraattoets minder geschikt is voor grote datasets (meer dan 5.000 records), maken we daarnaast gebruik van de gemiddelde absolute afwijking (MAD) en plaatsen deze op de schaal die Drake en Nigrini hebben afgebakend om de mate van conformiteit met Benfords verdeling in te schatten (zie noot 49).

Getallen kleiner dan	Volkstellingen		Mouvement	
	N	χ^2	N	χ^2
10	8863	9176,4	902737	3690,8
25	28374	2818,7	1625305	629,8
50	51149	2451,6	2023964	201,2
75	67007	1328,4	2177804	52,7
100	80039	1311,2	2258038	21,1
250	127977	573	2392315	12,4
500	156582	321,2	2434581	4
750	167772	75,4	2445589	2,8
1000	173259	12,3	2451307	2,5
2500	182204	25,4	2461624	4,2
5000	184303	27,1	2464807	3,7
Totaal	185837	22,5	2466740	3,7

Tabel 3 :

Verdeling van de geregistreerde bevolking naar grootte en mate van afwijking (chi-kwadraat $\chi^2 \times 0,01$) met de verdeling van Benford (eerste cijfer) (bron : LOKSTAT)

Er is duidelijk sprake van een toenemende afwijking met het model van Benford naarmate de getallen kleiner zijn. Alleen de hoogste getallen volgen dit patroon niet. Deze algemene bevindingen sluiten aan bij het uitgangspunt. Ze zullen in overweging moeten worden genomen wanneer de historische statistieken in comparatief verband op betrouwbaarheid worden beoordeeld.

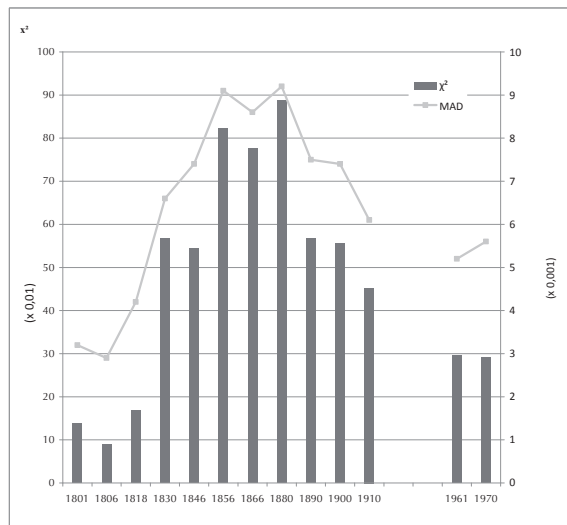
Analyse van de gemeentelijke uitslagen van de volkstellingen

Vergelijking met de frequentieverdeling van Benford

In welke mate sluiten de verschillende tellingen bij de verdeling van Benford aan en mogen vastgestelde afwijkingen aan de mindere betrouwbaarheid van

het cijfermateriaal worden toegeschreven? Het antwoord op deze vraag is mogelijk door middel van een vergelijkende *goodness-of-fit* test per bron waarbij de resultaten met de gekende elementen uit deel II worden geconfronteerd. We base-

ren ons op de cijfers aangaande burgerlijke staat en geslacht die een geschikte basis vormen voor een vergelijking van de tellingen doorheen de tijd. Deze gegevens zijn beschikbaar voor 13 volkstellingen⁴⁷.



Volkstellingen	1801	1806	1818	1830	1846	1856	1866	1880	1890	1900	1910	1961	1970
N	9459	16516	15744	5480	15123	15179	15304	15496	15571	15701	15789	15968	14527
Variabelen ⁴⁸	4	6	6	6	6	6	6	6	6	6	6	6	6
Gemeenten	2365	2761	2624	914	2521	2530	2551	2583	2596	2617	2629	2663	2379
MAD (x 0,001)	32	29	42	66	74	91	86	92	75	74	61	52	56
χ^2 (x 0,01)	13,8	8,8	16,8	56,7	54,5	82,3	77,7	88,7	56,7	55,5	45,3	26,6	29,2

Figuur 3 - tabel 4 :

bevolking naar burgerlijke staat en geslacht geregistreerd in de algemene tellingen: conformiteit met de frequentieverdeling van Benford (chi-kwadraattoets χ^2 en gemiddelde absolute afwijking MAD) (eerste cijfer) (bron : LOKSTAT)

47. Ter vervanging van de census in 1816, waarvan de uitkomsten grotendeels zijn verloren gegaan, maken we gebruik van bevolkingsstatistieken uit 1818 die in de meeste provincies teruggaan op eerstgenoemde telling (SVEN VRIELINCK, *De territoriale...*, p. 2068-2077). De uitkomsten van de census in 1830 zijn slechte ten dele teruggevonden (914 gemeenten of 37 % van het totaal). Verder bevatten de telresultaten uit 1920, 1930 en 1947 geen bevolkingsverdeling naar burgerlijke staat op het niveau van de gemeenten, waardoor de gegevens uit deze jaren niet voorkomen in onze analyse. De gemeentelijke uitkomsten zijn in 1930 en 1947 te reconstrueren met behulp van de kladformulieren die op het Algemeen Rijksarchief te Brussel worden bewaard, maar de middelen ontbraken om in het kader van dit onderzoek van deze mogelijkheid gebruik te maken. 48. Het aantal personen dat uit de echt is gescheiden, werd van 1880 af in de tellingen opgenomen en laten we omwille van de vergelijkbaarheid buiten beschouwing. In 1801 werden gehuwden en ooit gehuwden in één categorie gegroepeerd.

Behoudens lichte verschillen zien we dezelfde trend voor de verschillende afwijkingsmaten (χ^2 en MAD) in figuur 3. De juxtapositie van de tellingen laat

ons toe om drie perioden te onderscheiden waarin deze maten *grosso modo* dezelfde orde van grootte bereiken.

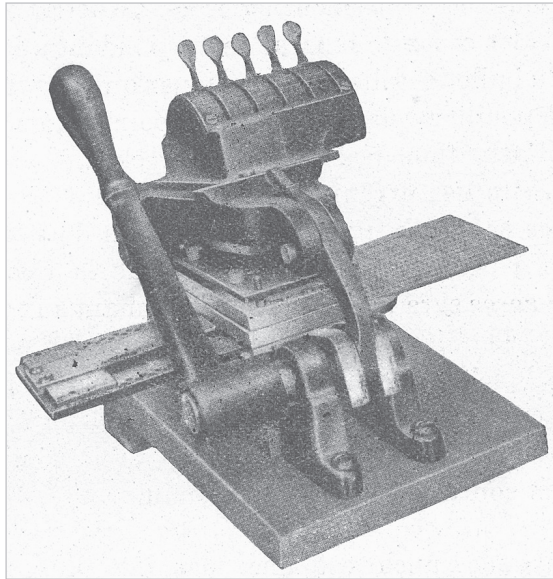
Tellingen	Vastgestelde afwijking	Score conformiteit ⁴⁹
1801-1818	Geen onregelmatigheden gedetecteerd	1 (groot)
1830-1910	Aanhoudende significante deviatie - tellingen 1830-46, 1890-1910: bescheiden afwijking - tellingen 1856-80: sterke afwijking	2 (acceptabel) 3 (zwak)
1961-1970	Terugval tot een niveau dat als overeenstemmend met de Wet van Benford mag worden beschouwd.	1 (groot)

a) Op zoek naar de verklaring

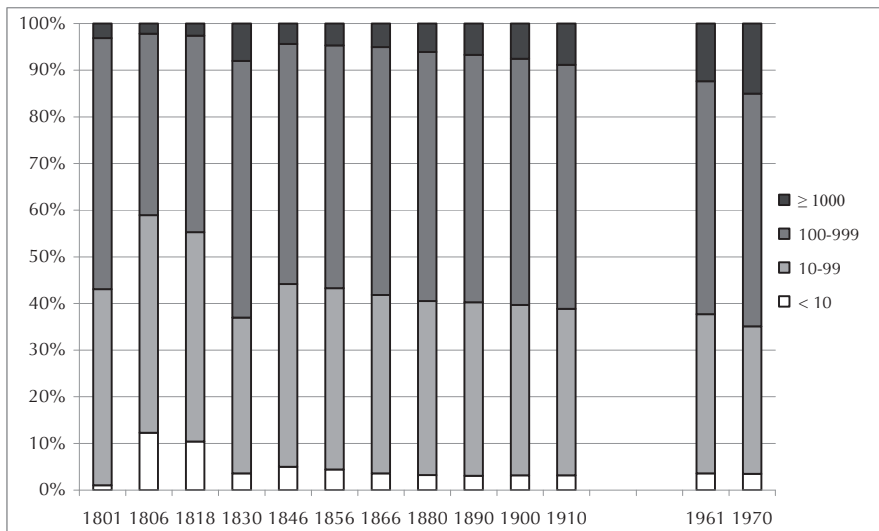
Op het eerste gezicht mag de grote conformiteit van de volkstellingen uit de jaren 1801-1818 aan het model van Benford verrassend heten. Eerder onderzoek heeft immers uitgewezen dat deze tellingen allesbehalve accuraat zijn⁵⁰. De toegepaste methode is ontoereikend om bias in de oudste datasets te detecteren. De redenen hiervoor staan verder beschreven. Om de betrouwbaarheid van de vroegste tellingen te testen, kunnen we terugvallen op spreidingsmaten⁵¹.

Verskillende elementen kunnen een rol hebben gespeeld in het algemene patroon. Een mogelijke verklaring voor de vastgestelde ontwikkeling, schuilt in de grootte van de geregistreerde bevolkingsaantallen. Zoals we hebben aangetoond, neemt de volgzaamheid aan de regel van Benford af naarmate de getallen in grootte slinken. Er moet worden nagegaan of de uiteenlopende scores die de tellingen halen, mag worden toegeschreven aan het aandeel van lage bevolkingsaantallen in elke census.

⁴⁹. De conformiteitscores werden toegekend op grond van de criteria die Drake en Nigrini hanteren (0-0,004 : *close conformity*; 0,004-0,008 : *acceptable conformity*; 0,008-0,012 : *marginally acceptable conformity*; >0,012 : *nonconformity*). Deze classificatie steunt op de proefondervindelijke ervaring van beide auteurs (PHILIP DRAKE, MARK NIGRINI, "Computer assisted...", p. 133). ⁵⁰. SVEN VRIELINCK, *De territoriale...*, p. 2098-2100. ⁵¹. Zie verder (Deel III, hoofdstuk De detectie van uitbijters).



*Toestel voor het perforeren van ponskaarten gebruikt bij de census van 1920. Deze volkstelling was de eerste in België die gecentraliseerd en machinaal werd verwerkt. (Uit Armand Julien, *Principes de statistique théorique*, Bruxelles, 1921, p. 268)*



Figuur 4 :
Bevolking naar burgerlijke staat en geslacht geregistreerd per gemeente in de algemene tellingen :
verdeling naar grootte (bron : LOKSTAT)

Het verloop van de bevolkingsgrootte is geen factor die de onderlinge verschillen in conformiteitscores van de tellingen kan verklaren, zoals uit figuur 4 blijkt. De bronnen met het grootste aandeel lage getallen (1806, 1818) zijn deze die het dichtst aansluiten bij de verdeling van Benford. Dit staat in tegenspraak met de stelling die de oorzaak van het verloop in figuur 3 aan de ongelijke aanwezigheid van kleine inwonertallen toeschrijft. Mogelijk heeft de toename van hoge bevolkingsaantallen in de loop van de 20^{ste} eeuw de algemene deviatiegraad doen dalen, maar het effect hiervan mag men niet overschatten.

De sleutel tot de verklaring dient gezocht in de analyse van de breuklijnen die uit de kwantitatieve analyse te voorschijn komen. Door de veranderingen die zich

toen hebben voorgedaan te bekijken, kunnen we de factoren die van fundamenteel belang zijn voor de vastgestelde evolutie isoleren. De grootste sprongen in de berekende afwijkingsgraad deden zich voor in 1830, 1856 en 1890. Vooral de laatste verschuiving is van beslissend belang voor een goed begrip van het mechanisme dat ten grondslag ligt aan de waargenomen fluctuaties.

b) Het ontstaan van de discrepantie (1830)

Met de census van 1830 kwam een einde aan de volkstellingen die een sterke gelijkens met Benfords verdeling tonen. Tot de Eerste Wereldoorlog zou de afwijking het laagste niveau niet meer bereiken. Twee fundamentele veranderingen zijn in 1830 aanwijsbaar die de plotse discrepantie kunnen verklaren.

De census van 1830 was de eerste die onder toezicht stond van statistische organen die stappen hebben gezet om de betrouwbaarheid te verhogen. De Commissie en het Bureau van de Statistiek waren betrokken bij de voorbereiding en de uitvoering van de operatie. Provinciale commissies zagen het licht die belast waren met het verifiëren van de telresultaten. In tegenstelling tot de voorgaande periode waar het aan statistische organen ontbrak die de massa cijfermateriaal nakeken, verschijnen in 1830 instanties die potentieel in staat waren om de censusresultaten van vele gemeenten aan te passen⁵². Mogelijk hebben de rijksambtenaren een deel van de geregistreerde inwonertallen bijgewerkt in een poging om de (vermeende) werkelijkheid te benaderen. Dat heeft de 'natuurlijke orde van de cijfercombinaties' verstoord. Of dit in 1830 effectief gebeurd is, valt evenwel te betwijfelen. Door de Revolutie die in september 1830 uitbrak, hebben de controleorganen vermoedelijk nauwelijks gefunctioneerd⁵³.

Figuur 5 :
Voorbeeld van standaard telformulieren gebruikt in 1806 (boven) en 1830 (onder) (bron LOKSTAT).

123

Commune de Heuvelmuntster
Rapportement Général de la population au
Premier Janvier 1806.

Naam	Plaats	Ammer Muntster	Ammer Muntster	Wijf	Man	Totaal	Opmerkingen
74	70	10	10	2	3	239	

L'Altre de la Commune de Heuvelmuntster
certifie le recensement ci-dessus vérifié le 1^{er} Janvier 1806.
J. Kiers

Gemeente Waareghem MODEL B.

STAAT van Bevolking op den 1^{sten} Januarij 1830, der Provincie West-Vlaanderen, naar den verschillende ouderdom der ingezetenen.

OUDERDOM.	INWONERS VAN HET							
	MANNELIJK GESLACHT.				VROUWELIJK GESLACHT.			
	Opgelinden.	Colenden.	Widwenen.	Totaal.	Opgelinden.	Colenden.	Widwenen.	Totaal.
Bemeden het jaar van 1 tot 2	22	.	.	22	23	.	.	23
3 - 4	34	.	.	34	35	.	.	35
5 - 6	46	.	.	46	47	.	.	47
7 - 8	143	.	.	143	144	.	.	144
9 - 10	16	.	.	16	17	.	.	17
11 - 12	146	.	.	146	147	.	.	147
13 - 14	145	.	.	145	146	.	.	146
15 - 16	132	.	.	132	133	.	.	133
17 - 18	113	.	.	113	114	.	.	114
19 - 20	166	.	.	166	167	.	.	167
21 - 22	132	.	.	132	133	.	.	133
23 - 24	340	14	.	354	340	20	1	361
25 - 26	182	21	2	205	186	25	1	212
27 - 28	41	132	3	235	166	136	5	307
29 - 30	31	146	3	180	62	131	5	248
31 - 32	51	113	3	167	63	144	14	261
33 - 34	25	44	7	76	77	160	7	244
35 - 36	10	61	6	77	23	96	16	145
37 - 38	2	37	2	41	7	25	23	55
39 - 40	0	35	15	50	8	39	14	61
41 - 42	4	33	5	42	11	23	18	52
43 - 44	5	23	2	30	5	17	15	37
45 - 46	1	16	7	24	2	9	8	20
47 - 48	1	13	1	15	1	2	2	5
49 - 50	1	8	1	10	1	2	2	5
51 - 52	1	7	4	12	1	5	4	10
53 - 54	1	2	5	8	1	2	1	4
55 - 56	1	3	3	7	1	4	4	9
57 - 58	1	1	2	4	1	4	4	9
59 - 60	1	1	1	3	1	1	1	3
61 - 62	1	1	1	3	1	1	1	3
63 - 64	1	1	1	3	1	1	1	3
65 - 66	1	1	1	3	1	1	1	3
67 - 68	1	1	1	3	1	1	1	3
69 - 70	1	1	1	3	1	1	1	3
71 - 72	1	1	1	3	1	1	1	3
73 - 74	1	1	1	3	1	1	1	3
75 - 76	1	1	1	3	1	1	1	3
77 - 78	1	1	1	3	1	1	1	3
79 - 80	1	1	1	3	1	1	1	3
81 - 82	1	1	1	3	1	1	1	3
83 - 84	1	1	1	3	1	1	1	3
85 - 86	1	1	1	3	1	1	1	3
87 - 88	1	1	1	3	1	1	1	3
89 - 90	1	1	1	3	1	1	1	3
91 - 92	1	1	1	3	1	1	1	3
93 - 94	1	1	1	3	1	1	1	3
95 - 96	1	1	1	3	1	1	1	3
97 - 98	1	1	1	3	1	1	1	3
99 - 100	1	1	1	3	1	1	1	3
Totaal	2176	177	98	3451	2677	270	211	3158

52. Navorsingen in de departementale, provinciale en nationale archieven van vóór 1830 leverden geen aanwijzingen op van enige vorm van systematische controle van de censusgegevens door het rijksbestuur. Behoudens uitzonderingen werden de getallen die de gemeenten meedeelden kritiekloos overgenomen in de nationale registers. Zoals uit tabel 1 blijkt, was de census van 1806 de enige die door een statistisch bureau in de schoot van het ministerie van Binnenlandse Zaken werd gesuperviseerd. Het gering aantal ambtenaren die op het bureau werkten, kunnen onmogelijk de staten van meer dan 44.000 telplichtige gemeenten hebben nagezien. 53. De oprichting van de controlecommissies vond plaats kort vóór de septemberrevolutie die de normale werking van deze organen heeft gehinderd (oprichting resp. 28 juli in Limburg en 30 augustus 1830 in Antwerpen) (RIJSARCHIEF ANTWERPEN, *Provinciaal Archief. Reeks J*, nr. 158 B; Rijksarchief Maastricht, *Provinciaal bestuur, 1814-1913*, nr. 310, p. 4473).

De tweede oorzaak voor de breuk met het verleden in 1830 ligt wellicht in de toegenomen complexiteit van de berekeningen. Vanaf dat jaar werden de gegevens over de burgerlijke staat gekoppeld aan de leeftijdsverdeling van de bevolking. Waar in voorgaande tellingen (1801, 1806 en 1818) de totale bevolking over slechts 6 tot 8 categorieën werd verdeeld, voorzagen de telformulieren in 1830 niet minder dan 384 in te vullen cellen. Als totalen onderaan het formulier waren de gegevens over de burgerlijke staat gevoeliger voor fouten die in de vele rijen daarboven konden sluipen. Het doen kloppen van de totalen op het formulier vergde meer inspanningen (of creativiteit) van de tellers. Het uitzicht van de telformulieren onderging geen fundamentele veranderingen meer vóór de Eerste Wereldoorlog. De koppeling leeftijd – burgerlijke staat bleef tijdens de verdere periode gehandhaafd in de officiële statistieken die de uitslagen van de tellingen weergeven.

c) Toenemende divergentie (1856)

De gelijkens met het model van Benford was het zwakst in het derde kwart van de 19^{de} eeuw tot en met 1880. Tussen 1846 en 1856 voltrok zich een grote verwijdering. Zoals werd aangegeven, contrasteerden de tellingen uit deze jaren sterk wat de ingezette middelen betreft. Voor de census van 1846 werden kosten noch moeite

gespaard⁵⁴. De onderneming was groots in opzet, voorbereiding en ondersteuning. Niettegenstaande op hetzelfde moment tellingen van de landbouw en de nijverheid doorgingen, beschikten de organisatoren over een ruim werkingsbudget. De uitwerking stond model voor andere tellingen in het buitenland. De Nederlandse historicus Paul Klep noemt de census “de meest uitgebreide en kwalitatief allerbeste van alle Europese tellingen in de 19de eeuw”⁵⁵.

Daartegen leek de census van 1856 op een flauw afkooksel. Methodologisch werd nauwelijks afgeweken van de procedures tien jaar eerder, maar de omvang van het project bleef minimaal. Het parlement was niet bereid om een nieuwe grootschalige operatie te financieren en had de doelstellingen teruggeschoefd. De census werd gerealiseerd met een aanzienlijk kortere vragenlijst en met het laagste budget ooit voor een census in het onafhankelijke België (figuur 1). De provinciale bureaus die belast waren het depouilleren van de individuele bulletins kampten met een personeelstekort en een strakke timing die de kosten moest drukken⁵⁶. In Henegouwen slaagde het bureau er zelfs niet in om de klus te klaren. Een deel van bulletins werd naar Limburg overgebracht om daar te worden geteld. Noodgedwongen zag men af van de verwerking van bepaalde inlichtingen op de telformulieren.

54. De teloperatie in 1846 wordt beschreven in NELE BRACKE, “L’élément statistique par excellence : op zoek naar het perfecte bevolkingscijfer: de Belgische statistieken tussen 1841 en 1870”, in THIERRY EGGERICKX, JEAN-PAUL SANDERSON (dir.), *Histoire...*, p. 312-319; RAYMOND GOEMAERE, “De volkstelling van 15.10.1846 te Kortrijk”, in *Handelingen van de Koninklijke Geschied- en Oudheidkundige Kring van Kortrijk. Nieuwe reeks*, 60, 1994, p. 205-255.

55. PAUL KLEP, *Politieke strubbelingen rond de volkstelling 1859*, Voorburg, 2007. 56. Brief van de West-Vlaamse gouverneur aan de Minister van Binnenlandse zaken, 22.8.1865; verslagen van de Centrale Statistiekcommissie (*Bulletin de la Commission centrale de Statistique*, 11, 1869, p. 55, 108).

De repercussies van deze problemen op de geloofwaardigheid van het cijfermateriaal, werd nog niet eerder vastgesteld. Onze analyse doet op zijn minst het vermoeden rijzen dat de haast, het geldtekort en de daaruit voortvloeiende moeilijkheden tot de grote discrepantie met het model van Benford hebben geleid. De aanwijzingen zijn sterk, maar niet sluitend. We moeten wachten tot 1890 om het harde bewijs te kunnen leveren dat ernstige afwijkingen in het patroon van Benford het gevolg zijn van onjuiste data.

Na 1856 bleven de onderzochte statistieken de voorspelde verdeling volgens Benford nauwelijks gehoorzamen. De waarden uit 1866 en 1880 weken in dezelfde orde van grootte af als die uit 1856 (figuur 3). De sterke deviatie was vermoedelijk een gevolg van het gebrekkige toezicht op de telverrichtingen in de gemeenten. Zoals het volgende blijkt, heeft de centrale overheid pas van 1890 af meer controle op de gemeenten gaan uitoefenen, wat de betrouwbaarheid van de teluitkomsten heeft doen toenemen. De laksheid van de regering ten aanzien van de gemeenten vóór 1890 ontging sommige tijdgenoten niet. Bijzonder scherp was de kritiek ten aanzien van de census van 1866 in *'t Jaer 30*, een Brugse oppositiekrant die het populisme niet schuwde⁵⁷. “Deze statistiek heeft het land vele duizende frankskes gekost en ‘t schoonste van geheel die kostelyke historie is, dat er geen een cifer van dit colossal werk juiste en is : de gemeentebedienden sloegen er overal een slag in, gelyk men gemeenlyk zegt, en alzoo kwam een werk tot stand, dat een leugen is van aen de kop tot aen de steert.

Wy hebben gemeentesecretarissen daerover dingen hooren vertellen, die u zouden doen bersten van lachen of gy wilt of niet”.

d) De kentering (1890)

Tussen 1880 en 1890 tekent zich opnieuw een sterke verschuiving af in de beschreven conformiteitproef (figuur 3). Na decennia lang het patroon van Benford in geringe mate te hebben gevolgd, sloot het cijfermateriaal in 1890 weer dichter bij het wiskundig model aan. Deze wending moet worden toegeschreven aan een toename van de betrouwbaarheid van de data ingevolge meer controle op de gemeenten. Het is de enige verklaring die na onderzoek uit de lijst van mogelijke oorzaken overblijft. Alle andere elementen met betrekking tot de organisatie en de uitvoering van de tellingen blijken (zo goed als) identiek te zijn in 1880 en 1890 en mogen bijgevolg als verklaringsfactor worden geschrapt. De geobserveerde verschuivingen in de cijferverdeling kunnen evenmin worden toegeschreven aan maatschappelijke ontwikkelingen die zich in de gemeenten hebben voorgedaan. De periode tussen de beide tellingen was voor deze processen te kort om de berekende cijferfrequenties op significante wijze scheef te trekken.

Het verstrengde toezicht op de gemeenten tijdens de teloperatie in 1890 was een beslissing van de minister van Binnenlandse Zaken Burlet. In juli 1891, toen de verwerking van de censusdata nog in volle gang was, riep de minister een bureau in het leven dat de bevolkingsstatistieken van de gemeenten op

57. *'Jaer 30 of politike wegwyzer voor trefelyke lieden*, 19.2.1870, nr. 294, p. 1, kol. C.

onregelmatigheden moest nazien. Een budget van 61.000 fr. werd voor deze opdracht uitgetrokken. De beslissing getuigt van weinig vertrouwen in de bestaande voorzieningen die een correcte toepassing van de richtlijnen moesten verzekeren. Vermoedelijk handelde Burlet op aandringen van het parlement. Voorafgaandelijk aan de census had de Gentse volksvertegenwoordiger Lippens om bijzondere maatregelen gevraagd die de juistheid van de teluitkomsten moesten verzekeren⁵⁸. Volgens het parlementslid werkten nieuwe financieringsmechanismen van de gemeenten fraude in de hand omdat de verdeling van bepaalde budgetten aan het officiële inwonertal werd gekoppeld. De minister wilde verder een herhaling voorkomen van fouten die zich bij de vorige teloperatie in 1880 hadden voorgedaan en die naar aanleiding van de vastlegging van de taalgrens in 1889 tot moeilijkheden hadden geleid⁵⁹.

De ijver die de ambtenaren van het controlebureau in 1891 aan de dag legden, was ongezien. Ruim 35000 numerieke tabellen werden systematisch op gebreken nagekeken. Hierbij kwamen talrijke fouten en vergissingen aan het licht. 4.829 statistieken (13 %) werden voor rectificatie terug naar de gemeenten gezonden. In totaal stuurde het bureau 3.530 brieven en 440 rappels naar lokale besturen⁶⁰. Deze ervaring indachtig, heeft de regering soortgelijke controleorganen bij de daaropvolgende tellingen ingesteld. Vanaf 1920 stond het ministerie ook in voor het depouilleren van de nominatieve bulletins.

De studie van de ommekeer die in 1890 plaatsgreep, levert het bewijs dat de afwijkingen van het Benfordpatroon te wijten zijn aan fouten en andere onregelmatigheden die bij de verwerking van de bevolkingsgegevens optraden. Het is een antwoord op de onderzoeksvraag die we bij aanvang van de analyse van de volkstellingen hebben gesteld.

De detectie van uitbijters

Uit de voorgaande analyse blijkt dat Benford een bruikbare maatstaf biedt om de betrouwbaarheid van onze historische dataset aan af te meten. Maar deze techniek heeft zijn grenzen. Zoals Nigrini en andere fraudeonderzoekers opmerken, garandeert een sterke conformiteit met de wet niet noodzakelijk een hoge betrouwbaarheid van de bron. Indien de onderzochte data afkomstig zijn van talrijke instanties die onafhankelijk van elkaar werken zonder inmenging (of manipulatie) van hogeraf en zonder dat er veel systematiek bij de productie van de data komt kijken, gaat de bewuste *Law of anomalous numbers* niet op. Deze voorwaarden zijn van toepassing op de tellingen van vóór 1830. De telformulieren uit die periode waren bijzonder eenvoudig en werden nauwelijks gecontroleerd door de rijksoverheden in Den Haag en Parijs.

Om een globale indruk te krijgen van de datakwaliteit gaan we de vroegste tellingen onderzoeken op de aanwezigheid van waarden die sterk afwijken van het algemeen patroon. Bevolkingscijfers die ver van de histo-

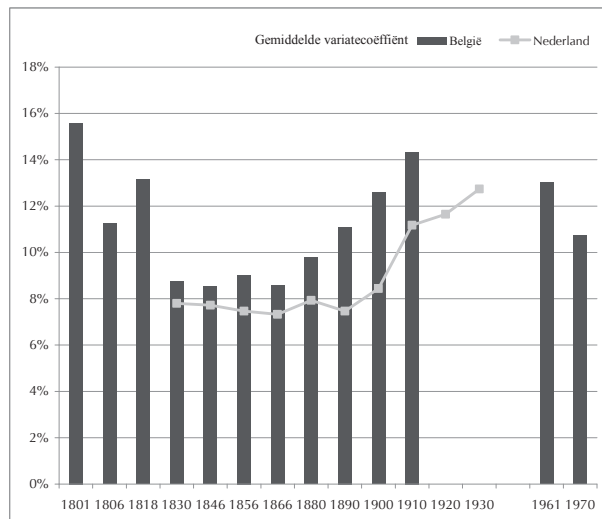
58. *Chambre des Représentants. Annales parlementaires*, zitting van 17 mei 1890, p. 1445.

59. De talenstatistiek uit 1880 bevatte verschillende blunders die moesten worden rechtgezet door het K.B. van 3 mei 1889, dat de Vlaamse gemeenten met het oog op het taalgebruik in strafzaken afbakende (*Recueil des Lois*, 76, 1889, p. 204-222).

60. *Population de la Belgique. Recensement général du 31 décembre 1890*, Bruxelles, 1893, p. VI.

rische realiteit staan, liggen vaak ook ver verwijderd van het rekenkundig gemiddelde. Een manier om 'abnormale' waarden op het spoor te komen, bestaat uit het vergelijken van de numerieke verhoudingen tussen bevolkingsgroepen in de verschillende gemeenten.

Aan de hand van het onderscheid dat tussen celibatairen en gehuwden (incl. ooit gehuwden) in de bronnen wordt gemaakt, berekenen we de relatieve spreidingsmaten⁶¹. We betrekken de resultaten van volkstellingen uit Nederland bij de analyse als basis voor vergelijking.



Figuur 6 – Tabel 5 :
Bevolkingsverdeling naar burgerlijke staat in de gemeenten met meer dan 100 inwoners : procentueel aandeel van de categorieën celibatairen en gehuwden (incl. ooit gehuwden) per geslacht en per gemeente in het bevolkingstotaal. Spreiding per census⁶².

Volkstellingen		1801	1806	1818	1830	1846	1856	1866	1880	1890	1900	1910	1961	1970
Gemeenten	A	2239	2677	2581	913	2514	2525	2546	2580	2593	2613	2625	2653	2370
Gehuwde mannen	G	19,6	17,9	17,8	16,6	17,5	17,6	18,4	18,3	18,3	19	19,9	27	27
	S	3,4	2,1	2,5	1,7	1,6	1,7	1,7	2	2,2	2,7	3,1	2,7	2,1
	V	17,4	12,2	13,9	10,3	9,3	9,9	9,3	10,9	12,4	14	15,5	10	8,1
Gehuwde vrouwen	G	20,6	19,3	19,4	18,3	18,9	18,9	19,5	19,4	19,6	20,4	21,4	30	30,7
	S	3,4	2,3	2,8	1,9	1,7	1,9	1,8	2,1	2,5	3	3,6	3,6	3
	V	16,5	12,1	14,6	10,1	9,5	9,8	9,4	11,1	12,7	14,7	16,6	12,1	9,7
Ongehuwde mannen	G	30,3	32,4	31,5	33,2	32,9	33,3	32,7	32,7	32,6	31,9	30,8	22,8	22,6
	S	4,2	3,3	3,5	2,5	2,6	2,8	2,6	2,7	3,1	3,4	3,9	3,5	2,9
	V	14	10,2	11,3	7,7	8	8,4	7,8	8,2	9,5	10,8	12,7	15,4	13,1
Ongehuwde vrouwen	G	29,5	30,3	31,2	32	30,6	30,3	29,5	29,6	29,5	28,7	27,9	20,2	19,7
	S	4,2	3,2	4	2,2	2,2	2,5	2,3	2,6	2,9	3,1	3,6	3,3	2,7
	V	14,4	10,5	12,9	7	7,3	8,1	7,8	8,8	9,7	11	12,8	16,5	14
Gemiddeld	V	15,6	11,2	13,2	8,8	8,6	9	8,6	9,8	11,1	12,6	14,4	13,5	11,2

A= absolute aantal; G = procentueel aandeel; S= standaarddeviatie; V = variatiecoëfficiënt

⁶¹. Omwille van de vergelijkbaarheid van de verschillende tellingen baseren we ons op de bevolkingsindeling die in 1801 werd toegepast. ⁶². De variatiecoëfficiënten in Nederland zijn berekend op basis van de algemene volkstellingen in 1839, 1849, 1859, 1869, 1879, 1889, 1899, 1909, 1920 en 1930 (teljaren bij benadering weergegeven in de grafiek). Niet alle uitkomsten zijn in 1879, 1889 en 1899 gekend (resp. 857, 845 en 202 getelde gemeenten) (bron: www.volkstelling.nl).

Afgezien van de begin- en eindperiode zien we de numerieke verhoudingen tussen de beschouwde bevolkingsgroepen in toenemende mate divergeren. Dit proces is zowel in België als in Nederland merkbaar. De divergentie was een gevolg van de aangroeiende migratie, het verdwijnen van de streekgebondenheid en de afbrokkeling van traditionele maatschappelijke structuren. De drie vroegste tellingen steken schril af bij die ontwikkeling. De spreidingsmaten, staan haaks op wat verwacht mag worden. Geen enkel historisch proces kan deze hoge waarden verklaren. De grote spreiding wijst op de aanwezigheid van data die ver afwijken van de realiteit. Dat de census van 1801 het slechtst scoort, wekt geen verwondering. Eerder onderzoek toonde aan dat deze telling frauduleus was door de grootschalige ontduiking van inwoners op het platteland ingevolge de ongunstige politieke situatie die in de Belgische departementen heerste⁶³.

Analyse van de data betreffende het bevolkingsverloop

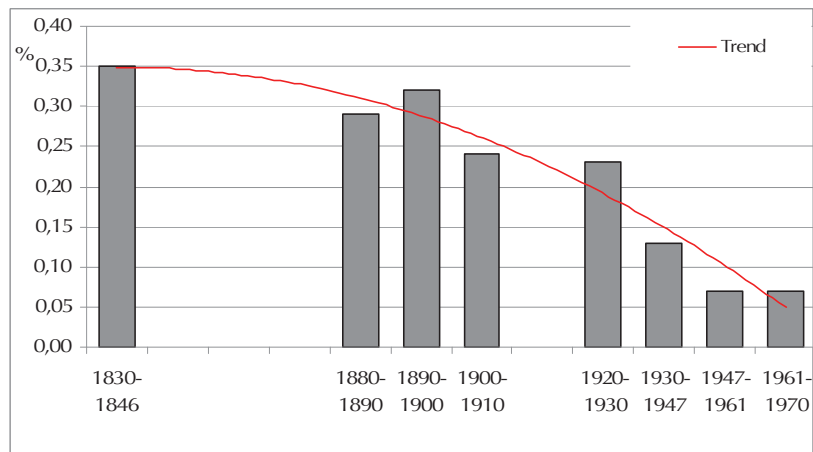
Vergelijking met de volkstellingen

Een voor de hand liggende manier om zicht te krijgen op de precisie van de registratie van de bevolkingsbeweging, is een vergelijking van de inwonertallen die op grond van de *Mouvement* werden berekend en deze die door middel van de klassieke tellingen werden vastgesteld. De omvang van de bevolking werd jaarlijks in de *Mouvement* becijfert aan de hand van de tabellen met

gegevens over geboorten, overlijden, in- en uitwijking afkomstig van de gemeenten. In theorie moeten de inwonertallen uit de beide bronreeksen in de teljaren identiek zijn, maar in de praktijk was dit natuurlijk zelden het geval. De bevolkingscijfers bevonden zich aan het eind van lange informatieketens waarvan slechts één schakel volstond om vertekening op te leveren. Men is het erover eens de afwijkingen grotendeels toe te schrijven aan de gebrekkige registratie van de migratie⁶⁴.

Reeds in de 19^{de} eeuw werden de getelde en de berekende bevolkingscijfers tegenover elkaar geplaatst teneinde het nut van de algemene tellingen aan te tonen. Nog nooit werden de getallen op het niveau van alle Belgische gemeenten systematisch vergeleken. Behoudens regionale studies baseerde men zich steeds op nationale en provinciale totalen die een onvolledig beeld verschaffen. Achter de totalen schuilen immers overschotten en tekorten die door het optellen van de gemeentelijke cijfers grotendeels worden afgevlakt. In figuur 7 geven we de gemiddelde afwijking weer tussen de beide bronnen op basis van gemeentelijke data.

63. SVEN VRIELINCK, "Boerenbedrog...", p. 97-115. 64. Zie hoger (Deel II, *Mouvement*, hoofdstuk bronkritiek).



Figuur 7 :

Jaarlijkse gemiddelde absolute afwijking tussen de berekende en de getelde bevolking in alle Belgische gemeenten. Procentueel aandeel van de totale bevolking (bron : LOKSTAT)⁶⁵.

Hoewel de onderzochte dataset niet alle jaren bestrijkt, is duidelijk sprake van een afname van de deviaties in de loop van de onderzochte periode. Voor elke afwijking die zich na de Tweede Wereldoorlog voerde, waren er gemiddeld vier in de beginperiode. De grootste verbetering deed zich na de eeuwwisseling en in het Interbellum voor. De resultaten van de analyse wijzen onmiskenbaar op een toename van de betrouwbaarheid van de datareeksen uit de *Mouvement*. Deze bevinding vergt verder onderzoek. We weten immers niet in hoeverre fouten in de volkstellingen doorwegen in het becijferde resultaat. Evenmin komen we te weten wat de rol is van de verschillende variabelen in de evolutie. Hier-

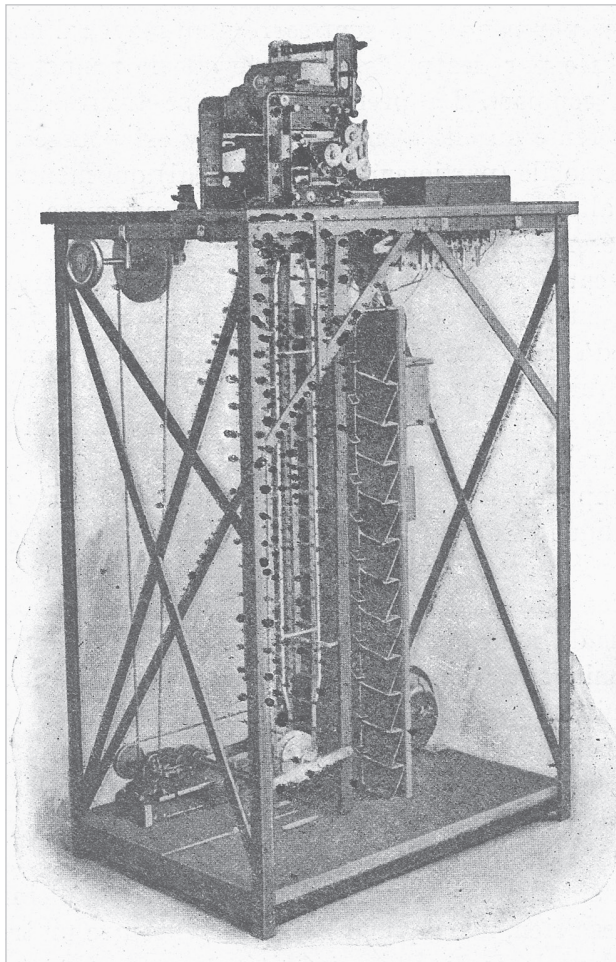
voor doen we een beroep op de Wet van Benford.

Vergelijking met de frequentieverdeling van Benford

- Conformiteit van de basisvariabelen

Welke vorm van registratie uit de *Mouvement* is gevoeliger voor vertekening? Verdienen de officiële statistieken van de burgerlijke stand meer vertrouwen dan deze die teruggaan op de bevolkingsregisters zoals algemeen wordt aangenomen? De *Law of anomalous numbers* is een geschikt referentiekader om de geselecteerde basisvariabelen uit onze dataset in comparatief verband aan af te meten. We passen dezelfde *goodness-of-fittoets* toe als hoger beschreven⁶⁶.

65. De gemiddelde afwijking werd bekomen door het totaal van de absolute verschillen per gemeente in de censusjaren (1846, 1890, 1900, 1910, 1930, 1947, 1961, 1970) te percenteren ten opzichte van de totale wettelijke bevolking zoals vastgesteld door de census in het betreffend jaar. Deze methode is onmogelijk voor de jaren 1846-1880 (wegens onvolledige datareeksen) en 1910-1920 (wegens oorlogsomstandigheden). 66. Zie hoger (Deel III, hoofdstuk Algemene conformiteitsproef)..



Machine voor het sorteren van ponskaarten waarop individuele teluitkomsten waren overgebracht. Deze toestellen, ontwikkeld door de Amerikaan Hollerith, werden ingezet bij de volkstelling van 1920. (Uit Armand Julien, Principes de statistique théorique, Bruxelles, 1921, p. 270)

	N	χ^2	MAD	Score conformiteit ²²
Geboorten	558812	0,041	0,002	1 (hoog)
Sterfte	557838	0,239	0,004	2 (acceptabel)
Huwelijken	277256	0,142	0,003	1 (hoog)
Inwijking	535145	0,323	0,006	2 (acceptabel)
Uitwijking	537689	0,276	0,005	2 (acceptabel)
Totaal	2466740	0,037	0,002	1 (hoog)

Tabel 6 :

Officiële statistieken van het bevolkingsverloop in alle Belgische gemeenten 1841-1850, 1880-1976 : conformiteit met de frequentieverdeling van Benford (chi-kwadraattoets χ^2 en gemiddelde absolute afwijking MAD) (eerste cijfer) (bron : LOKSTAT).

Alle onderzochte variabelen volgen het patroon van Benford, maar niet in even sterke mate. De migraties vertonen de grootste afwijking, zoals uit boven staande tekst verwacht kon worden. Het onvermogen van de overheid om de migratiestromen accuraat op te tekenen wordt hier tot op zekere hoogte zichtbaar. De hoger beschreven veronderstelling dat vooral de registratie van de mensen die de gemeenten verlieten mank liep, wordt daarentegen niet bevestigd. De statistiek van inwijkingen scoort immers niet beter dan die van de uitwijking.

Opmerkelijk is de relatief slechte score van de sterfte in bovenstaande tabel. In tegenstelling tot de geboorten en de huwelijken, valt dit afgeleide product van de Burgerlijke Stand buiten de hoogste categorie van conformiteit die Drake en Nigrini hebben afgebakend. Dit betekent dat er significante deviaties zijn zonder evenwel het algemeen patroon te verstoren. Het is raden naar de reden van dit onderscheid. Misschien is de afwijking

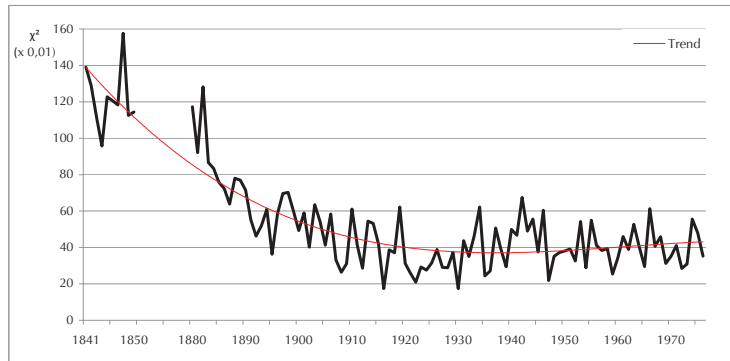
te wijten aan een gebrekkige registratie van personen die buiten de woonplaats stierven. Volgens de wet moest elk overlijden worden opgetekend in de registers waar de overledene wettelijk verbleef. De communicatie tussen de gemeenten verliep in de 19^{de} eeuw stroef, wat tot vertekening in de statistieken kan hebben geleid. Het valt evenwel te betwijfelen of deze zwakte voldoende gewicht had om het algemeen patroon te beïnvloeden.

- Conformiteit doorheen de tijd

Een verdere uitsplitsing van de onderzochte variabelen op jaarbasis geeft ons inzicht in de evolutie van de datakwaliteit en de onderliggende oorzaken. Hiermee naderen we de limieten van de toegepaste methode. Door het lager aantal observaties per jaar (ca. 4000 tot 5000) treden grotere afwijkingen op in het Benfordpatroon. Dit fenomeen is eigen aan de wet⁶⁷. De *goodness-of-fit*toets levert hierdoor tijdreeksen op met fluctuaties, waaruit niettemin een duidelijke trend valt waar te nemen⁶⁸

67. Zie hoger (deel III, Wet van Benford, principes). 68. De dataset van huwelijken heeft door de beperkte omvang een te smalle basis voor een jaarlijkse vergelijking met de verdeling van Benford en wordt om die reden verder buiten beschouwing gelaten.

De statistiek van de geboorten

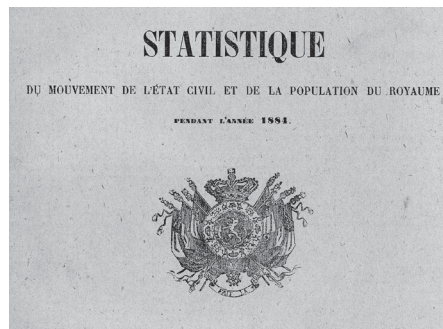


Figuur 8 :

Aantal geboorten geregistreerd per jaar en per geslacht in de Belgische gemeenten, 1841-1976 (uitgezonderd 1851-1879) : conformiteit met de frequentieverdeling van Benford (chi-kwadraattoets χ^2) (eerste cijfer) (bron : LOKSTAT).

In tegenstelling tot wat de algemene gelijkvormigheidstest (tabel 6) laat vermoeden, maakt de geboortestatistiek in die periode een sterke ontwikkeling door. Op de schaal van Drake en Nigrini evolueerde het cijfermateriaal van non-conformiteit naar grote gelijkheid met de Wet van Benford. De verschuiving voltrok zich grotendeels in de laatste decennia van de 19^{de} eeuw, waar na de consolidatie volgde. Vooral in de jaren 1881-1886 was de afgelegde afstand aanzienlijk. De sterke deviatie die zich in de vroegste decennia voordeed, beïnvloedt de score van de volledige datareeks (tabel 6) nauwelijks, wegens het beperkt aantal jaren waarvoor deze oudste gegevens zijn gekend.

Nader onderzoek van het cijfermateriaal wijst uit dat de relatieve aanwezigheid van kleine getallen, die zoals aangetoond de neiging hebben het patroon van Benford te verstoren, nauwelijks invloed heeft uitgeoefend op



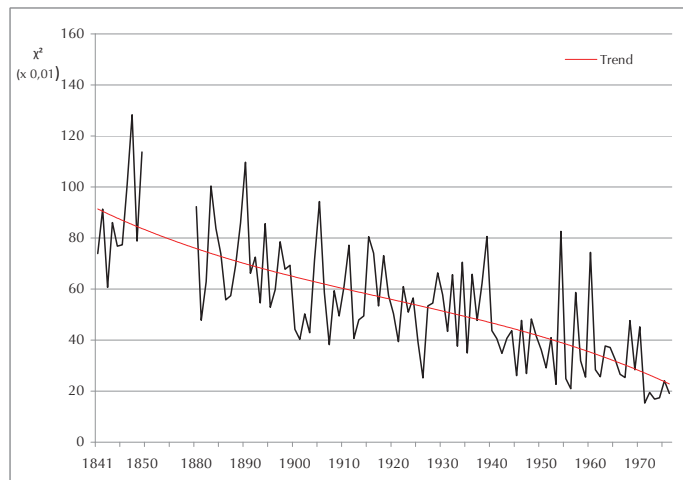
Figuur 9 : Titelblad van de Mouvement in het Staatsblad van 1885

de vastgestelde evolutie. Een reden voor het verloop moet daarentegen gezocht worden in de verbetering van de registratie. De spectaculaire verschuiving die zich bij aanvang van de jaren 1880 aftekende, is in dit verband veelzeggend. Op enkele jaren tijd bewogen de geboortecijfers zich sterk naar de Wet van Benford. Precies in die

periode situeren zich de inspanningen van de rijksoverheid om de Mouvement nieuw leven in te blazen. De toegenomen aandacht voor de reeks blijkt uit de publicatie van de gemeentelijke geboortecijfers in het *Staatsblad* (1883-1886). Door het officieel bekrachtigen en bekendmaken van de lokale statistieken werd paal en perk gesteld aan laksheid en misbruiken bij het invullen van

de telformulieren. De betrouwbaarheid van de kwantitatieve microgegevens steeg, wat zich in bovenstaande grafiek door een grotere conformiteit met de regel van Benford vertaalt. In het algemeen mag de weergegeven trend als een graadmeter worden aanzien van de efficiëntie van het overheidsapparaat in het verkrijgen van de meeste elementaire statistische informatie van de gemeenten.

De statistiek van de sterfte

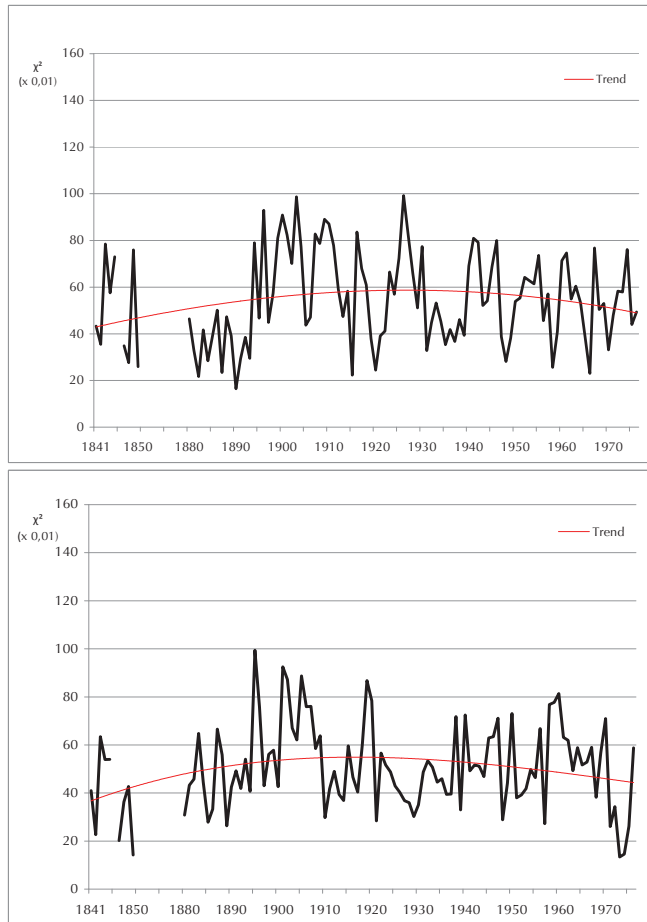


Figuur 10 :

Aantal sterften geregistreerd per jaar en per geslacht in de Belgische gemeenten, 1841-1976 (uitgezonderd 1851-1879) : conformiteit met de frequentieverdeling van Benford (chi-kwadraattoets χ^2) (eerste cijfer)
(bron : LOKSTAT).

De sterftcijfers die de gemeenten jaarlijks aan de hogere overheid meedeelden, volgden tussen 1841 en 1975 in groeiende mate de verdeling van Benford. Het was een geleidelijk proces dat alleen op lange termijn zichtbaar is. In tegenstelling tot de geboorten zijn geen versnellingen of vertragingen in de trend waar te nemen. De factoren die de

ontwikkeling sturen, zijn hierdoor moeilijker lokaliseerbaar. Op grond van het verband tussen de betrouwbaarheid van de data en de conformiteit aan de Wet van Benford dat hoger is aangetoond, mogen we concluderen dat de registratie van het aantal overledenen in de loop van de beschouwde periode aan accuratesse heeft gewonnen.

De statistiek van de migratie**Figuren 11-12 :**

Aantal inwijkelingen (boven) en uitwijkelingen (onder) geregistreerd in de Belgische gemeenten, 1841-1976 (uitgezonderd 1846, 1851-1879) : conformiteit met de frequentieverdeling van Benford (chi-kwadraattoets χ^2) (eerste cijfer) (bron : LOKSTAT)

De algemene trend van de jaarlijkse schommelingen loopt in beide grafieken opvallend gelijk. In tegenstelling tot de geboorte- en sterftecijfers starten de migratiestatistieken niet met extreme afwijkingen van Benford's patroon.

Afgezien van uitschieters, blijft de divergentie opvallend laag alvorens naar het einde van de 19^{de} eeuw licht toe te nemen. Zeer geleidelijk keert het tij in de loop van het Interbellum, niettegenstaande grote fluctuaties per jaar.

Zo ondubbeltzinnig de conformiteitscore evolueerde bij de statistieken van de geboorten en sterfte, zo mysterieus lijken de migratiestatistieken zich te gedragen. Hoe de relatieve sterke gelijkenis met de Wet van Benford uit de beginperiode verklaren wanneer alle bronnen het erover eens zijn dat de migratiegegevens toen sterk afwaken van de realiteit? De reden hiervoor dient te worden gezocht in de toegepaste methode. De conformiteitanalyse kan alleen onregelmatigheden aan het licht brengen wanneer de vertekening massaal en systematisch plaatsgreep. Deze voorwaarde was in de jaren 1840 afwezig om dezelfde redenen als die we bij de vroegste volkstellingen hebben onderscheiden⁶⁹. Enerzijds gaven de voorgedrukte telformulieren waarop de gemeentebesturen het aantal in- en uitwijkingen moesten inschrijven, geen aanleiding tot verwarring door de grote eenvoud van de vraagstelling. Terwijl de cijfers over de geboorte en de sterfte moesten worden geventileerd over verschillende subrubrieken, was dit niet het geval voor de migratie⁷⁰. Hierdoor daalde de kans op structurele fouten die bij het uitsplitsen en optellen van de gegevens konden ontstaan. Anderzijds oefenden de vertegenwoordigers van het centrale gezag geen systematische controle uit op de migratiegegevens die ze van de gemeenten ontvingen. Het ontbrak hen in de jaren 1840 aan de middelen om grote aantallen getallen na te kijken en desgevallend aan te passen. Er waren met andere woorden geen krachten aanwezig die

de ‘natuurlijke orde’ van het cijfermateriaal konden verstoren.

De factoren die het gebruik van de Wet van Benford als leugendetector beletten, verdwenen geleidelijk in de bestudeerde periode. De voorgedrukte migratietabellen die de gemeenten jaarlijks moesten invullen, werden uitgebreider en complexer. De staat ging zich ook meer moeien met de registratie van de migratie. Vanaf de tweede helft van 19^{de} eeuw moest ze het hoofd bieden aan uitdijende migratiestromen ingevolge de verstedelijking, de intercontinentale emigratie en de groeiende mobiliteit. De rijksoverheid nam initiatieven om de nauwkeurigheid van de bevolkingsregisters te verbeteren door het uitvaardigen van allerhande richtlijnen en besluiten (o.a. 1846, 1856, 1866)⁷¹. Om het telwerk te vergemakkelijken, moest elke gemeente van 1901 af een hulpregister van in- en uittredingen bijhouden. Ook de controle op de correcte uitvoering van deze verplichtingen werd geleidelijk aangescherpt. Deze maatregelen hebben ongetwijfeld de statistieken in gunstige zin beïnvloed, maar het precieze effect ervan valt niet te achterhalen via de uitgevoerde *goodness-of-fittest*. Van structurele afwijkingen of systematische manipulatie valt geen spoor te bekennen. Ondanks alle moeilijkheden die gepaard gingen met de registratie van de talrijke volksbewegingen, bleven de beschouwde migratiestatistieken relatief goed scoren op vlak van conformiteit met Benford's wet. Op de schaal van Drake en Nigrini schommelden

69. Zie hoger (hoofdstuk De detectie van uitbijters). 70. Voorbeelden van telformulieren zijn afgedrukt in *Bulletin de la Commission centrale de Statistique*, 2, 1845, p. 192-203 en *Idem*, 11, 1869, p. 178-197. 71. Een overzicht van de wetgeving aangaande bevolkingsregisters is te vinden in RENÉ LEBOUTTE, *Du registre...*; MICHEL POULAIN, “Le registre de population belge”, in THIERRY EGGERICKX, JEAN-PAUL SANDERSON (dir.), *Histoire...*, p. 83-116.

ze tussen categorie 1 (grote gelijkenis) en 2 (aanvaardbare verschillen). Deze bevinding doet geen uitspraak over gebreken die zich alleen op lokaal en individueel niveau afspeelden. Wanneer de vertekening in de verschillende gemeenten onvoldoende overeenkomst vertoonde om de cijfercombinaties scheef te trekken, blijft de bias onopgemerkt met de hier toegepaste methode.

IV. Besluit

Nu we de voornaamste bronreeksen hebben doorgelicht, komen we tot een eindbalans. Deze studie beoogt de vorming van een datagefundeerd oordeel over de effectiviteit van de statistische informatieverwerking vanuit een lange termijn perspectief. We poogden de betrouwbaarheid te bepalen van de meest elementaire bevolkingsstatistieken die door het overheidsapparaat werden opgesteld. Daarom vergeleken we grote aantallen bevolkingscijfers van de gemeenten met de frequentieverdeling van Benford. Deze studie toont via een aantal specifieke casussen aan dat de toegepaste methode bruikbaar is om de beschouwde datasets kritisch te evalueren. De gekozen invalshoek is geschikt om systematische anomalieën op supracommunaal niveau bloot te leggen, maar houdt geen appreciatie in van de individuele cijfers of de grootte van elke deviatie. Om tot een evenwichtige antwoord op de centrale onderzoeksvraag te komen, moeten de resultaten van de kwantitatieve en kwalitatieve onderzoeken samen op de weegschaal van de historische kritiek worden gelegd.

In grote lijnen bevestigt deze studie de vermoedens over het verloop van de datakwaliteit. De rijksoverheid kreeg tijdens de beschouwde periode meer greep op de lokale besturen en op de werkelijkheid, tenminste wat de verwerving van elementaire bevolkingsgegevens betreft. Achter de positieve evolutie die op lange termijn te bespeuren valt, schuilt een proces van vallen en opstaan. In de eerste helft van de 19^{de} eeuw boekte de overheid merkelijke vooruitgang in de organisatie van de volkstellingen, met als voorlopig hoogtepunt de census van 1846. Daarop volgde de inzinking. De onderzochte variabelen uit de klassieke tellingen van 1856 tot en met 1880 vertonen sterke afwijkingen met de verwachte theoretische waarden, wat op ernstige problemen met de informatieverwerking duidt. Dezelfde moeilijkheden zijn aantoonbaar in de statistieken van de natuurlijke bevolkingsbeweging van bij de aanvang van de reeks in 1841. Het voorlaatste decennium van de 19^{de} eeuw tekent zich in onze analyses af als een scharnierperiode. De significante vertekening van de datareeksen in de voorgaande jaren werd toen teruggedrongen en een proces van betere registratie kwam op gang. Vooral de cijfers van de geboorten en van de burgerlijke staat naderden op relatief korte tijd weer een aanvaardbare afwijkingmarge. De onregelmatigheden namen eveneens af in de sterftestatistieken, zij het op een meer geleidelijke manier. Onze analyse bracht niet aan het licht in hoeverre de migratiestatistieken die ontwikkeling volgden. Door de beperkingen van de toegepaste evaluatietechnieken kunnen de gegevens over de in- en uitwijking in de gemeenten

niet op een vergelijkbare basis worden beoordeeld.

Door de grote sprongen die zich voordeden in het geschetste proces aan een nader onderzoek te onderwerpen, kunnen factoren worden geïsoleerd die ten grondslag liggen aan de vastgestelde bias in de datasets. Van beslissende betekenis was de rol die het centrale gezag speelde. Naarmate het rijksgesetz middelen investeerde in de betrouwbaarheid van de statistieken, zien we de deviaties en de abnormale waarden in het onderzochte cijfermateriaal slinken. Belangrijke koerswijzigingen van het centraal beleid ten aanzien van de volkstellingen, met name de dalende interesse in 1856 en de prioriteit voor fraudebestrijding in 1890, tekenen zich af in het verloop van de kwaliteit van de bevolkingsgegevens. Maar niet alle inspanningen van de overheid wierpen vruchten af. Een drijvende kracht die de accuratesse van de data lijkt te hebben beïnvloed, was de structuur van de statistische vragenlijst die aan de gemeenten

werd voorgelegd. Systematische vertekening van de data werd wellicht in de hand gewerkt door de complexiteit van de tabellen waarop de originele gegevens moesten worden genoteerd. Dit verschijnsel biedt een verklaring voor de bias die door ons vanaf 1830 in de onderzochte dataset van de volkstellingen werd vastgesteld.

Tot slot moet benadrukt dat deze studie geen afbreuk doet aan de waarde van de bevolkingsstatistieken. Door de schijnwerpers op de tekortkomingen te zetten, mag de rijkdom en de hoge relevantie van deze bronnen voor het wetenschappelijk onderzoek niet worden onderschat. De toegepaste cijferanalyse staat niet toe om individuele getallen op betrouwbaarheid te beoordelen. Hoogstens geven we aan dat een bevolkingsgetal uit een bepaalde bron meer kans op vertekening geeft dan in een andere bron. Het blijft aan de vorser om te bepalen welke graad van nauwkeurigheid voor zijn of haar onderzoek vereist is.

SVEN VRIELINCK (°1964) is wetenschappelijk medewerker bij de Vakgroep Geschiedenis van de Universiteit Gent. Zijn onderzoek richt zich op historische informatiesystemen, administratieve statistiek en historische geografie. Van zijn hand verscheen het naslagwerk De territoriale indeling van België (1795-1963) bij de Universitaire Pers te Leuven. Als opdrachthouder bij Federaal Wetenschapsbeleid was hij lange tijd verantwoordelijk voor beleidsondersteunende onderzoeksprogramma's in het domein van de sociale wetenschappen. Hij coördineerde de publicatierexen Samenleving en Toekomst, Wetenschap en Maatschappij en Sociale Cohesie.