

Kjetil Telle

Effekter av Klifs tilsyn

Resultater fra produktkontrollen 2007-2010

Rapporter I denne serien publiseres analyser og kommenterte statistiske resultater fra ulike undersøkelser. Undersøkelser inkluderer både utvalgsundersøkelser, tellinger og registerbaserte undersøkelser.

© Statistisk sentralbyrå, juni 2011 Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Standardtegn i tabeller	Symbol
ISBN 978-82-537-8109-9 (trykt)	Tall kan ikke forekomme	.
ISBN 978-82-537-8110-5 (elektronisk)	Oppgave mangler	..
ISSN 0806-2056	Oppgave mangler foreløpig	...
Emne: 01.06	Tall kan ikke offentliggjøres	:
Trykk: Statistisk sentralbyrå	Null	-
	Mindre enn 0,5 av den brukte enheten	0
	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

Forord

Denne rapporten oppsummerer resultatene fra kontroller som Klima- og forurensningsdirektoratet (Klif) har utført med importører av faste, bearbejdede produkter i 2007-2010, og den er en videreføring av en tidligere oppsummering for årene 2007 og 2008. Rapporten er utarbeidet av Kjetil Telle, SSB, på oppdrag fra Klif.

Kjetil Telle (SSB) har hatt 2 møter i Klifs lokaler (01.12.10 og 11.03.11) der følgende personer fra Klif har deltatt: Ragnhild Orvik, Einar Knutsen, Line Telje Høydal og Inger Marie Haaland. På disse møtene er forhold omkring denne rapporten drøftet. I forkant av det andre møtet utarbeidet Kjetil Telle et foreløpig notat, og dette notatet ble diskutert på møtet. Innspill fra dette møtet er forsøkt innarbeidet i foreliggende versjon.

Sammendrag

Klima- og forurensningsdirektoratet (Klif) driver tilsyn med importører av faste, bearbejdede produkter. Siden 2007 er dette tilsynet utført på en måte som innebærer at tallene er representative for hele den aktuelle populasjonen av importører, og tilsynsmetoden gjør det sågar mulig å teste effekter av Klifs tilsynsvirksomhet.

Til sammen bestod tilsynspopulasjonen av 1801 importører i 2007/8, 1766 i 2009 og 1625 i 2010. Av disse kontrollerte Klif 202 i 2007/2008, 165 i 2009 og 167 i 2010 (totalt 534).

I analysene benyttes flere resultatindikatorer, herunder om Klif har avdekket avvik i kontrollen. Resultatene tyder på dårligere overholdelse av regelverket over tid; andelen importører med avvik var 33 prosent i 2008, 40 prosent i 2009 og 44 prosent i 2010. Resultatene kan også tyde på store forskjeller i overholdelsen i ulike undergrupper av importører, og avviket er også svært høyt i noen grupper av importører med lav kontroll sannsynlighet.

Importørene er trukket tilfeldig med hensyn til tilsyn og tilsynsform, noe som gjør det mulig å teste effekten av Klifs tilsyn. Resultatene viser at tilsynet har en individualpreventiv effekt: Å kontrollere en importør reduserer importørens sannsynlighet for å ha avvik neste år med omlag 40 prosent ved brevkontroll og omlag 70 prosent ved besøkskontroll. Resultatene viser også at grundigere tilsyn (besøkskontroll) avdekker flere avvik enn mindre grundig tilsyn (brevkontroll). Vi finner imidlertid ingen tegn til at varsel om høyere kontrollfrekvens (allmennpreventiv effekt) påvirker overholdelsen. På bakgrunn av disse resultatene drøfter vi aktuelle omprioriteringer innenfor tilsynsvirksomheten.

Abstract

The Norwegian Climate and Pollution Agency (Klif) is responsible for enforcing environmental regulations of firms that import manufactured substances. Since 2007 the enforcement activities are conducted in a way securing representative information of the overall population of firms. Indeed, the method enables us to test the effects of Klifs enforcement activities.

The population comprised 1801 importing firms in 2007/8, 1766 in 2009 and 1625 in 2010. Among these, Klif inspected 202 in 2007/2008, 165 in 2009 and 167 in 2010 (total 534).

Firm behavior and effects of enforcement activities are measured in several ways in the analyses, including whether Klif detected violations in the inspection. The results indicate less compliance over time; the percent of importing firms with violation was 33 in 2008, 40 in 2009 and 44 in 2010. The results also indicate substantial variation in compliance across subgroups of firms, and violations are very high in some groups of firms where the inspection probability is low.

Enforcement activities are randomly assigned to firms, which enables us to test the effects of Klif's activities. The results show a specific deterrent effect: Inspecting a firm in the previous year reduces the firm's propensity to be in violation in the preceding year by about 40 percent in on-site inspections and about 70 percent in self-reported inspection. The results also show that more thorough inspections (on-site) detect more violations than less thorough inspections (self-reported). However, we find no sign that letters notifying the firm of higher inspection frequency (general deterrence) raises compliance. Based on these results, we discuss relevant reallocations of enforcement activities.

Innhold

Forord.....	3
Sammendrag.....	4
Abstract.....	5
1. Innledning	7
2. Bakgrunn, data og frafall	7
3. Resultatindikatorer	8
4. Representative gjennomsnitt	9
5. Testing av hypoteser.....	12
6. Skisse til metode for systematisk utvelgelse av importører med antatt dårlig score.....	22
7. Konklusjoner og oppsummerende anbefalinger	22
Referanser.....	24
Vedlegg A: Fremgangsmåte og forutsetninger for de statistiske analysene.....	25
Figurregister	26
Tabellregister.....	27

1. Innledning

I denne rapporten oppsummeres tall fra Klima- og forurensningsdirektoratets (Klifs) tilsynsvirksomhet overfor importører av faste, bearbejdede produkter i 2007 – 2010. Dette tilsynet utføres på en måte som innebærer at tallene vil være representative for hele den aktuelle populasjonen. Tilsynsmetoden muliggjør også testing av forskjeller i resultatindikatorer på tvers av grupper og år, og forskjeller som følge av ulike måter å drive tilsyn på. Denne tilsynsmetoden gjør det mulig for Klif å stadig lære hvordan ulike tilsynsmetoder påvirker kontrollobjektene overholdelse av regelverket. Dette læringsbaserte tilsynet startet opp i 2007, og tallene som rapporteres her gjelder for de første tre rundene av kontroller utført med denne metoden. Den første runden var i 2007/8, den andre i 2009 og den tredje i 2010.

2. Bakgrunn, data og frafall

Populasjonen er definert gjennom lister over importører som Klif har mottatt fra Tollvesenet. Populasjonen omfatter importører av de tre gruppene (1) forbrukerprodukter, (2) EE-produkter og (3) byggevarer som importerte varer fra Asia, og der vekten per år av de importerte varene ikke var ubetydelig. Til sammen var denne populasjonen på 1801 importører i 2007/8, 1766 importører i 2009 og 1625 importører i 2010.¹ De fleste importørene var med i populasjonen i alle de tre rundene, mens noen bare var med i en eller to av rundene. Til sammen består disse tre årspopulasjonene av 1975 ulike importører. De fleste importørene tilhørte gruppen forbrukerprodukter (1298), mens gruppen innenfor byggevarer var minst (138), se tabell 1. Hver importør var i gjennomsnitt med i populasjonen i 2,6 av de tre rundene. Totalt gir dette 5192 år/importør-observasjoner.

Det ble trukket tilfeldige utvalg på 249, 250 og 244 (totalt 743) importører til å motta kontroll i hhv. 2007/2008², 2009 og 2010. Antall og andel importører trukket ut til kontroll varierte over gruppene (strata). Det ble trukket flest importører for kontroll i forbrukerprodukter (430), men sannsynligheten for å bli trukket ut til kontroll var klart høyest i byggevarer (141). I tillegg til oversampling i byggevarer ift. forbrukerprodukter, var det også oversampling fra ulike under-strata innenfor hver av de tre gruppene. Hver gruppe ble for det første delt i forskjellige antall undergrupper. For det andre ble hver undergruppe delt i (normalt) tre nivåer, der nivå 1 ble oversamlet ift. nivå 2 (som igjen ble oversamlet ift. nivå 3). Mens det var kunnskap om hva slags varer som ble importert som lå til grunn for dannelsen av undergrupper (og grupper), så var det vurderinger av risiko som lå til grunn for plasseringen på nivå. Av de totalt 5192 år/importør-observasjonene i datasettet, var det 1261 år/importører som lå i nivå 1 og 2592 i nivå 3, mens det ble trukket 421 år/importører til kontroll i nivå 1 og bare 105 i nivå 3. Dette medførte at det var betydelig oversampling fra noen strata.

Når importører var trukket ut til kontroll, ble det igjen ved tilfeldig trekking bestemt om importøren skulle motta brev- eller besøkskontroll. Fordi det er mer ressurskrevende å utføre besøkskontroller, ble det trukket langt flere brevkontroller enn besøkskontroller. Av totalt antall kontroller utgjorde besøkskontrollene snaut 20 prosent.

¹ Noen få importører er flyttet fra en gruppe til en annen mellom to år. I analysene tegnes slike importører den gruppen de tilhørte i det aktuelle året, mens når jeg omtaler totaltall for alle år samlet tegnes disse gruppen de tilhørte i 2008. I en tidligere rapport, der resultater fra tilsynet i 2007/8 er analysert, inngår det 4 flere importører enn her. Disse er utelatt her fordi vårt oppdaterte datasett ikke inneholder informasjon om gruppetilhørighet

² I 2007 ble EE-produkter (2) og byggevarer (3) kontrollert, mens forbrukerprodukter (1) ble kontrollert i 2008.

Antall importører som ble kontrollert var 202 i 2007/2008, 165 i 2009 og 167 i 2010 (totalt 534). Det var altså 209 importører som var trukket ut til kontroll, men som likevel ikke ble kontrollert. Dette utgjør et frafall på opp mot 30 prosent, og frafallet var større i 2009 og 2010 enn i 2008. Importørene har naturligvis ikke anledning til å unndra seg kontroll, så seleksjonsproblemene som normalt oppstår i utvalgsundersøkelser fordi noen ikke ønsker å delta skal ikke forekomme her. Den vanligste årsaken til frafall er at importøren ikke lenger driver forretningsvirksomhet eller at virksomheten ikke lenger importerer aktuelle produkter, men på brevkontrollene er det også noen som rett og slett ikke har svart. Frafallsprosenten varierer en del over de tre gruppene, med 25 prosent i forbrukerprodukter, 38 prosent i EE-produkter og 24 prosent i byggevarer.

I 2008 ble det trukket veldig få importører til kontroll i den store undergruppen *forbrukerprodukter ellers*, og på grunn av frafall ble det ikke gjennomført en eneste kontroll i to store strata (162 og 163). Også for noen andre, men meget små undergrupper, ble det gjennomført meget få kontroller (213, 222, 143, 153). Dette frafallet vil kunne medføre enkelte skjevheter i noen typer estimer, for eksempel endringer over tid. Det burde imidlertid ha liten betydning for testingen av hypotesene under. Analysene i dette notatet bygger således på et datasett med 4265 år/importør-observasjoner.

3. Resultatindikatorer

Som resultatindikatorer benyttes grad av overholdelse av internkontrollforskriften (*ik*), fagscore (*fs*) og et (uvektet) gjennomsnitt av disse (*gs*). Scoren på *ik* og *fs* bygger på vurderinger gjort av Klif etter hver enkelt kontroll. I tillegg har Klif gjort en totalvurdering av funnene i kontrollen (*ts*). Denne totalscoren (*ts*) trenger ikke å være sammenfallende med *gs*, for eksempel fordi *ts* kan legge mye vekt på alvorlige avvik. *ik* kan anta verdiene 1, 3 eller 5, og i hovedsak gjelder det samme for *fs*, mens *ts* i tillegg også kan ha verdiene 2 og 4.³ Resultatoppnåelsen er bedre jo høyere tallet er. Noe grovt kan vi si at tallet 1 indikerer avvik, 3 anmerkning og 5 at det ser bra ut. Vi har derfor også laget ytterligere to dikotome resultatindikatorer; *avv* er satt til 1 hvis *ik* eller *fs* er 1 (ellers 0), og *anm* er satt til 1 dersom *ik* eller *fs* er 3 eller lavere (ellers 0). I tillegg har Klif fra 2008 skilt ut om det er avdekket alvorlige avvik (*aa*) hos importøren (altså ikke bare om det er avdekket avvik).

Det er grunn til å tro at disse resultatindikatorerne fanger mange interessante sider ved importørens overholdelse av regelverket. Men de fanger naturligvis ikke alle interessante sider. Klif bør hele tiden vurdere om settet av resultatindikatorer bør endres eller utvides. Et par mulige utvidelser, som kan gi mer utfyllende informasjon, inkluderer antall avvik, antall alvorlige avvik og type frafall.

Hver importør er trukket tilfeldig innenfor sitt stratum, og i utgangspunktet er derfor utvalget i hvert stratum representativt for hele populasjonen i stratumet. Vi kan da bruke standard metoder for å fremskaffe representativ informasjon om ulike delpopulasjoner, eller for hele populasjonen. I det følgende beregner vi først noen representative gjennomsnitt og tester om resultatoppnåelsen er ulik over tid og i ulike delpopulasjoner. Deretter tester vi ulike hypoteser knyttet til effekter av Klifs tilsyn.

³ I data antar *ik* kun verdiene 1, 3 og 5. Der er imidlertid eksempler på at *fs* også antar heltallene 2 og 4, samt desimaltall, noe som skyldes at *fs* i noen tilfeller (multiimportører) er en sammenveining av fagscore innenfor ulike fagområder. Også *ts* er i noen tilfeller oppgitt med desimaler.

Tabell 3.1. Oversikt over populasjonen og kontrollene

	Totalt	Forbruker (1)	EE-produkter (2)	Byggevarer (3)
Antall i populasjonen (N)	1 975	1 295	539	138
Fordelt på år				
2008	1 801			
2009	1 766			
2010	1 625			
Fordelt på nivå				
1	665			
2	365			
3	945			
Antall trukket ut til kontroll (n)	743	430	172	141
Fordelt på år				
2008	249			
2009	250			
2010	244			
Fordelt på nivå				
1	421			
2	217			
3	105			
n/N	0,38	0,33	0,32	1,02
Antall trukket til besøk (n_{be})	147	99	29	19
Antall trukket til brev (n_{br})	596	331	143	122
n_{be}/n	0,20	0,30	0,20	0,16
Antall kontroller gjennomført (n_g) ...	534	321	106	107
Antall besøkskontroller gjennomført	114	77	24	13
Antall brevkontroller mottatt	420	244	82	94
Frafall kontroller (n_f)	209	109	66	34
n_f/n	0,28	0,25	0,38	0,24

4. Representative gjennomsnitt⁴

I dette avsnittet presenteres representative gjennomsnitt for resultatindikatorerne, både for hele populasjonen over tid og for delpopulasjoner (se tabell 4.1). I teksten vil jeg konsentrere oppmerksomheten om resultatindikatorerne *ts* og *avv*, men resultatene for de andre indikatorerne er også oppgitt i flere av tabellene.

Gjennomsnittet av resultatindikatoren *ts* viser ingen endring fra 2008 (3,2) til 2009 (3,2), men deretter en forverring i 2010 (2,6). For mer alvorlige forhold (*avv*) er det en forverring over tid; andelen importører med avvik var 33 prosent i 2008, 40 prosent i 2009 og 44 prosent i 2010. Imidlertid er ingen av disse forskjellene mellom årene statistisk signifikante. Forverringen over tid i avvikene kan gjenspeile at det faktisk er mer avvik nå enn tidligere, men den kan også gjenspeile at Klif ved sine kontroller avdekker en større andel av de faktiske avvikene nå enn de første årene.⁵ Det siste ville særlig kunne være aktuelt dersom skjemaene som benyttes ved kontrollene er endret for å bedre fange opp avvikene. Det ser på den annen side ut til å ha vært en reduksjon i importører med alvorlige avvik.

Figur 4.1 gjengir denne utviklingen over tid totalt (som i tabell 4.1), og for hver av gruppene. Vi ser at det har vært et betydelig fall for byggevarer, mens forbrukerprodukter har steget noe.

Over alle årene samlet er resultatindikatoren *ts* 2,8, 3,2 og 3,3 for hhv. Forbrukerprodukter (1), EE-produkter (2) og byggevarer (3), mens *avv* er hhv. 42 prosent, 39 prosent og 34 prosent. Men heller ikke disse forskjellene er statistisk signifikante.

⁴ Se vedlegget for nærmere redegjørelse for forutsetningene bak beregningene i dette og de etterfølgende avsnittene.

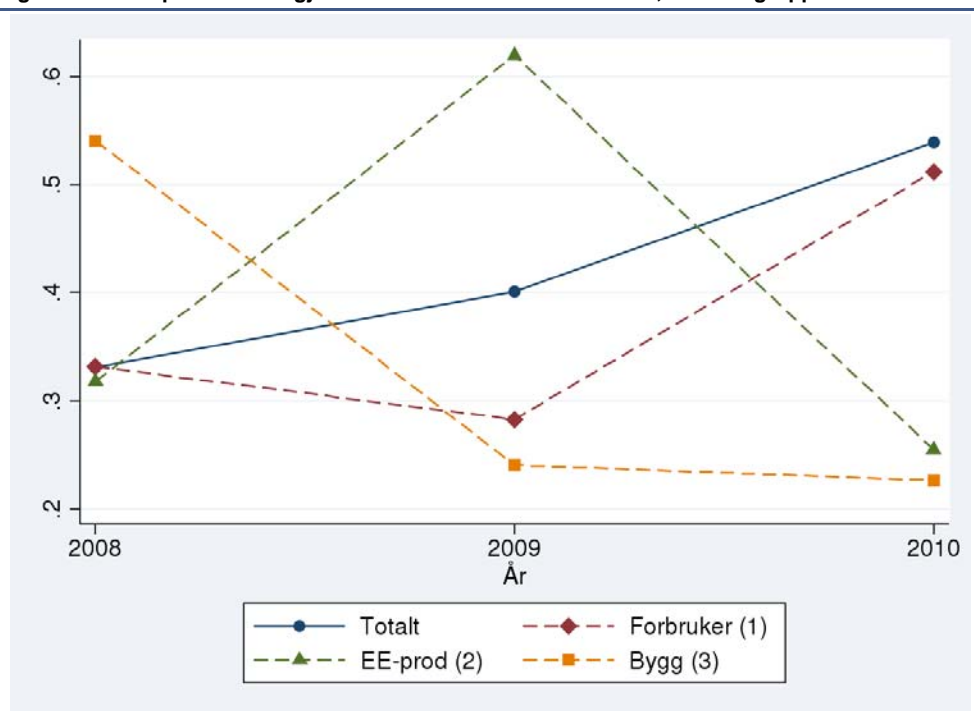
⁵ Merk at ved en metode som ikke trakk tilsynsobjekter tilfeldig, ville forverringen også kunne skyldes at Klif var blitt flinkere til å velge ut importører som hadde mer avvik til tilsyn (og dermed ble flere avvik avdekket). Den metoden som benyttes her, med tilfeldig trekking av importører innenfor klart definerte strata (og der alle importører har strengt positiv tilsynssannsynlighet), sikrer at vi får representative gjennomsnitt og således er ikke slike læringseffekter hos Klif en feilkilde for oss.

Tabell 4.1. Representative gjennomsnitt av resultatindikatorer over tid og gruppe

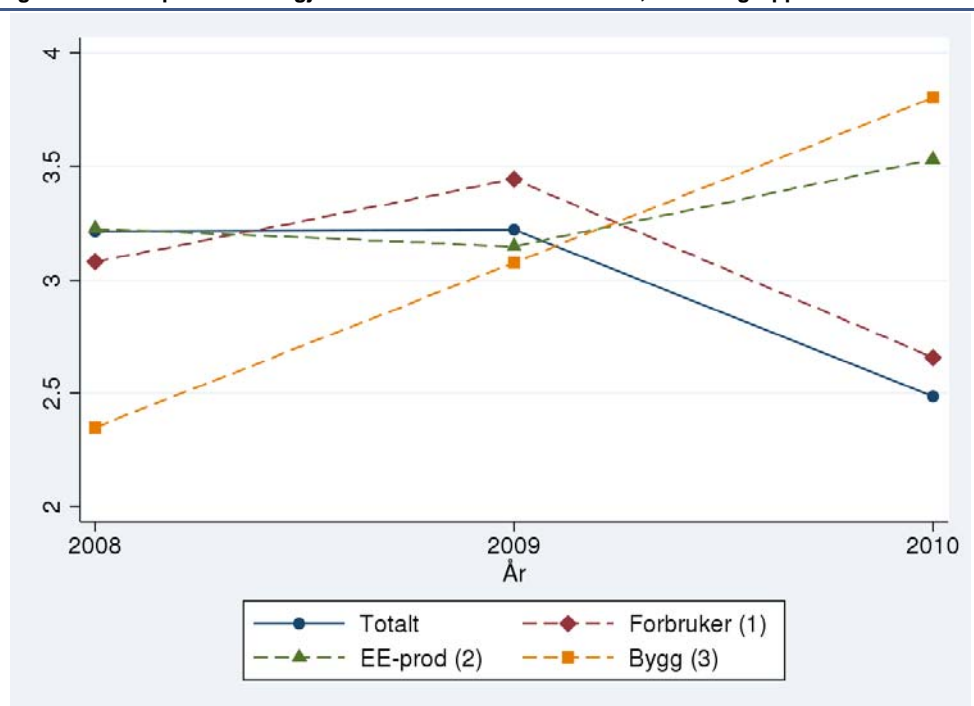
	ts	gs	ik	fs	avv	aa	anm
År 2008	3,2	3,4	3,0	3,9	0,33	0,03	0,80
År 2009	3,2	3,3	3,4	2,6	0,40	0,01	0,61
År 2010	2,6	3,1	3,0	3,2	0,44	0,00	0,74
Gruppe 1 – forbrukerprodukter	2,8	3,1	3,0	2,9	0,42	0,01	0,64
Gruppe 2 – EE-produkter	3,2	3,3	3,4	3,3	0,39	0,01	0,81
Gruppe 3 – Byggevarer	3,3	3,3	3,1	4,1	0,34	0,00	0,65

Resultatene bygger på data fra de 534 kontrollene som er utført i årene 2008-2010. Det er tatt hensyn til stratifisering og oversampling ved utregning av gjennomsnitt (og standardfeil benyttet i testene referert til i teksten. Som utdypet i vedlegget, er det grunn til å tro at disse standardfeilene er for store slik at nullhypotesen forkastes for sjelden).

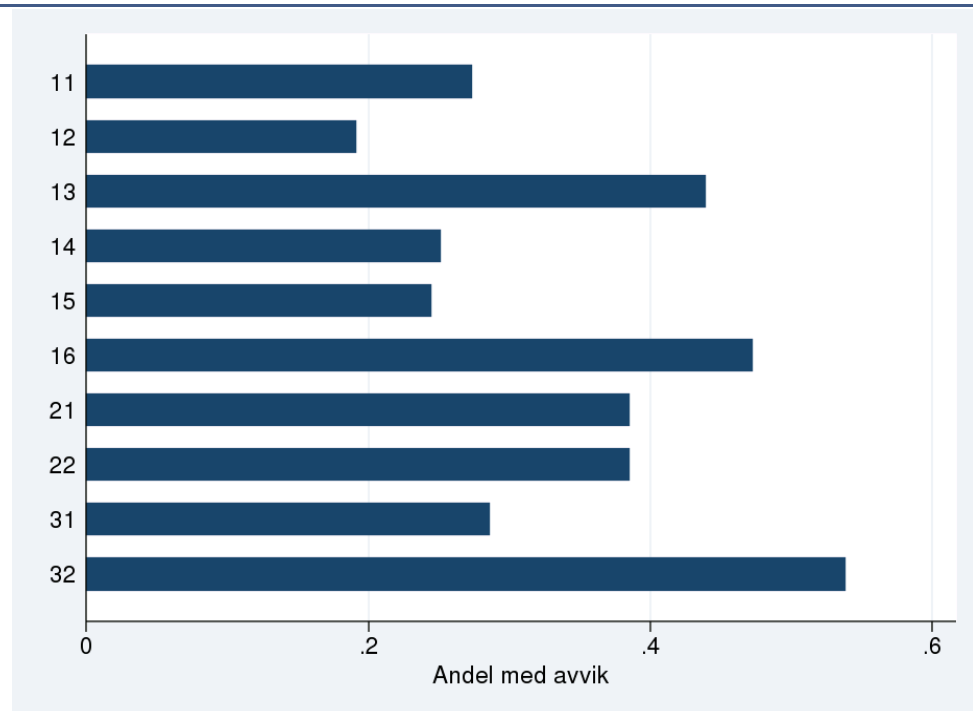
Figur 4.1a. Representative gjennomsnitt for avv over tid. Totalt, for hver gruppe



Figur 4.1b. Representative gjennomsnitt for ts over tid. Totalt, for hver gruppe

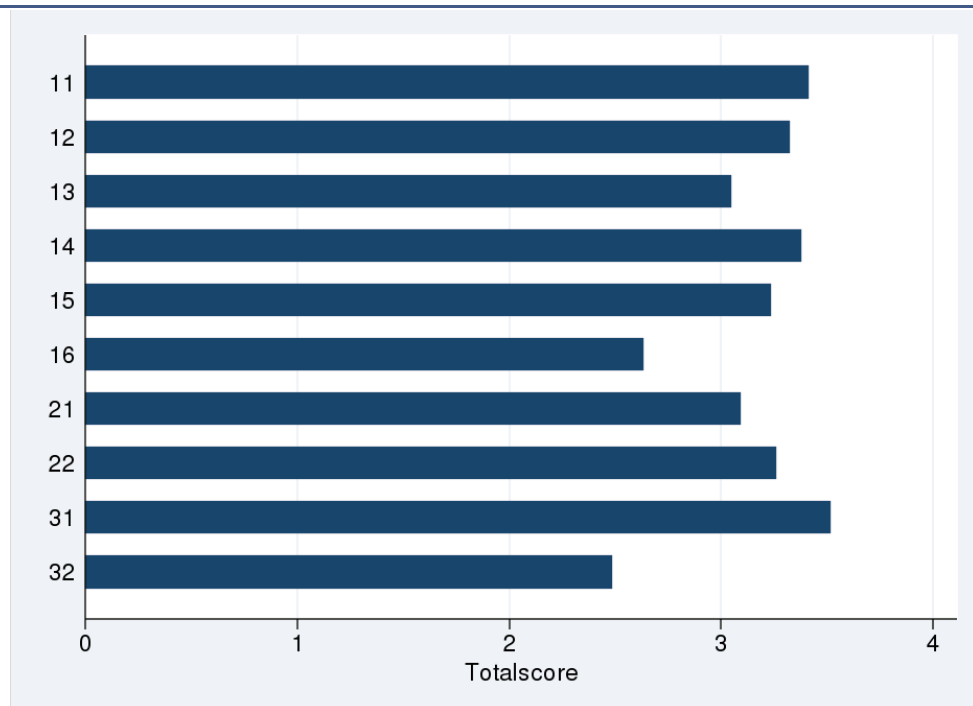


Figur 4.2a. Representative gjennomsnitt for avv innen hver undergruppe (alle år samlet)



Note: 11=forbrukerprodukter forlag, 12=forbrukerprodukter leker- og småbarnsprodukter, 13=forbrukerprodukter tekstiler, 14=forbrukerprodukter smykker/bijouteri, 15=forbrukerprodukter multi-importører (ting-og-tang), 16=forbrukerprodukter ellers, 21=EE-produkter, 22=EE-produkter ikke prioritert, 31=byggevarer, 32=byggevarer ikke prioritert.

Figur 4.2b. Representative gjennomsnitt for ts innen hver undergruppe (alle år samlet)



Note: 11=forbrukerprodukter forlag, 12=forbrukerprodukter leker- og småbarnsprodukter, 13=forbrukerprodukter tekstiler, 14=forbrukerprodukter smykker/bijouteri, 15=forbrukerprodukter multi-importører (ting-og-tang), 16=forbrukerprodukter ellers, 21=EE-produkter, 22=EE-produkter ikke prioritert, 31=byggevarer, 32=byggevarer ikke prioritert.

De to undergruppene med best resultat på *ts* er 31 og 14 (3,5 og 3,4), mens de to med dårligst er 32 (2,5) og 16 (2,6). Det finnes imidlertid ingen undergruppe som skiller seg signifikant fra en annen for *ts*, og heller ikke forskjellen mellom de to beste undergruppene (31 og 14) og de to dårligste (32 og 16) er signifikant. Heller ikke for *avv* er det mange statistisk signifikante forskjeller, men det er nesten dobbelt så stor andel med minst et avvik i noen undergrupper enn i andre (se figur

4.2). Særlig er forskjellen stor mellom undergruppene 12 og 15 (hhv. 19 og 24 prosent med avvik) på den ene siden og undergruppene 16 og 32 (47 prosent og 54 prosent) på den andre.

Det er mye variasjon i dataene vi har sett på her, og trendene i figurene kan derfor skyldes tilfeldigheter. I tillegg skal vi være varsomme med å knytte disse trendene til Klifs tilsynsvirksomhet (selv om tilsynet har vært mest omfattende for byggevarer, der fallet i avvik også har vært størst) da det er mange andre forhold enn Klifs tilsyn som kan ha påvirket overholdelsen av regelverket. Hvorvidt Klifs tilsynsvirksomhet har hatt en effekt på importørens overholdelse av regelverket undersøker vi i neste avsnitt.

5. Testing av hypoteser

Besøkskontroll vs. brevkontroll

En kontroll vil sjelden avdekke alle faktiske avvik, men noen kontrolltyper kan være bedre egnet til å avdekke avvik enn andre. Vi kan eksempelvis tenke oss at flere avvik avdekkes ved en besøkskontroll enn ved en brevkontroll. Da kontrolltype er trukket tilfeldig, er det ingen grunn til at faktiske avvik skal være (systematisk) forskjellig hos importører som mottar brev og besøk. Om måten Klif vurderer resultatene på ikke påvirkes av kontrolltype, så vil observerte forskjeller i resultatindikatorene mellom disse to gruppene avspeile kontrolltypens evne til å avdekke avvik.

Totalt sett avdekkes det i gjennomsnitt minst ett avvik (*avv*) i 30 prosent av brevkontrollene og i 54 prosent av besøkskontrollene, eller man kan si at vi forventer å avdekke 0,30 importører med avvik per brev og 0,54 per besøk. Denne forskjellen er statistisk signifikant. Vi forventer altså å avdekke 0,23 flere importører med avvik ved besøk enn ved brev. Det er særlig på *ik* at scoren er dårligere ved besøk enn ved brev. Dette kan altså tyde på at besøk er best egnet til å avdekke ureglementerte forhold knyttet til *ik*, mens for *fs* er ikke forskjellen så stor mellom brev og besøk. Alvorlige avvik er kun avdekket i besøk. Forskjellene mellom besøk og brev ser ut til å være noe større innen forbrukerprodukter enn for de to andre gruppene.

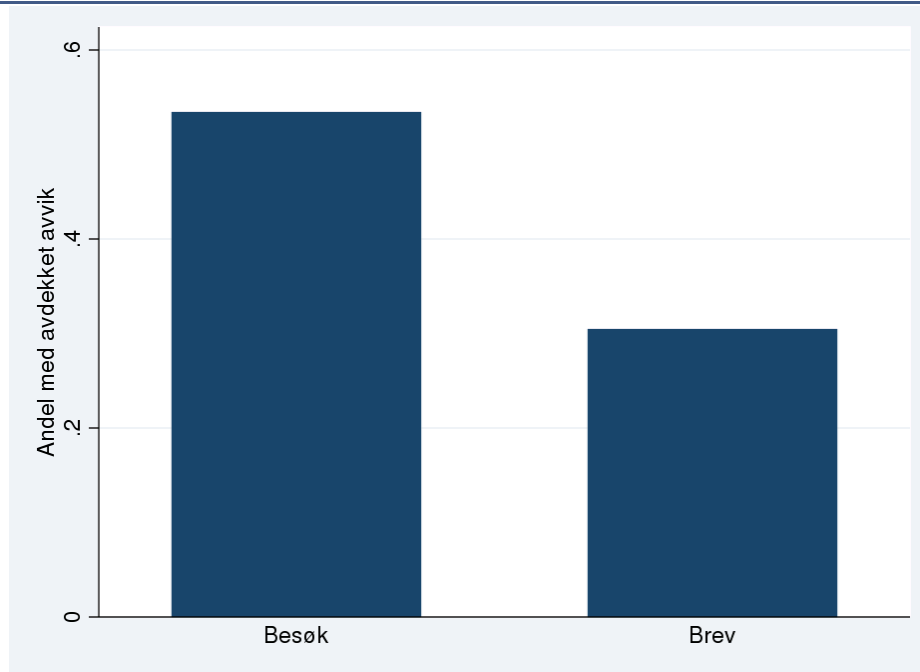
Disse resultatene er vel som vi skulle forvente, og de har iallfall to mulige forklaringer. Den ene er at hvordan Klif setter *ts*, *ik* og *fs* påvirkes av kontrolltype.⁶ Den andre er at brevkontrollene ikke har en like stor evne til å avdekke ureglementerte forhold som besøkskontrollene. Isolert sett kan det siste tale for at Klif utfører flere besøkskontroller på bekostning av brevkontroller. Men besøkskontroller krever imidlertid mye mer ressurser. Klif vil kunne avdekke flere importører med avvik ved å bytte ut én besøkskontroll (dét forventes å gi 0,54 færre importører med avdekkede avvik) med to brevkontroller (dét forventes å gi 0,60 flere importører med avdekkede avvik). Fordi et besøk krever langt mer ressurser enn to brev, vil totalt antall importører med avdekkede avvik dermed gå opp.

Et viktigere spørsmål er imidlertid om Klif kunne oppnådd omtrent den samme overholdelsen av regelverket med langt flere brevkontroller og færre besøkskontroller.⁷ Ved en slik avveining kan man ikke bare legge vekt på at to brev avdekker flere importører med avvik enn ett besøk, men også på at effekten på framtidig importøratferd kan være annerledes for de to kontrolltypene. I det følgende ser vi på om avvikene i etterfølgende år påvirkes av om det tidligere er blitt utført besøks- eller brevkontroll.

⁶ For å utelukke muligheten for at forskjeller mellom scoren til kontrolltypene skyldes ulik evaluering fra Klif side, bør personen i Klif som setter score på skjemaene ikke kjenne til om skjemaet som vurderes stammer fra brev eller besøk (blind).

⁷ Det kan naturligvis også hende at besøkene, men ikke brevene, evner å avdekke de mest alvorlige avvikene, men dette har vi (foreløpig) ikke data til å undersøke. I praksis vil vel likevel avdekking av avvik være et viktig skritt på veien til å avdekke de mest alvorlige forholdene.

Figur 5.1. Importører med avdekkede avvik i besøkskontroll og brevkontroll



Tabell 5.1. Resultater fordelt på kontrolltype

Indikator	Kontrolltype	Totalt	Forbrukerprodukter	EE-produkter	Byggevarer
avv	Brev	0,30	0,27	0,33	0,35
	Besøk	0,52	0,53	0,50	0,46
	Differanse	-0,22*	-0,26*	-0,17	-0,11
ts	Brev	3,3	3,3	3,4	3,3
	Besøk	3,0	3,0	2,8	3,2
	Differanse	0,36*	0,35+	0,61*	0,04
ik	Brev	3,5	3,7	3,4	3,2
	Besøk	2,8	2,7	2,8	3,0
	Differanse	0,78*	1,00*	0,69+	0,17
fs	Brev	3,8	3,9	3,4	3,9
	Besøk	3,4	3,5	3,3	3,5
	Differanse	0,36*	0,41*	0,18	0,41

Resultatene bygger på data fra de 534 kontrollene som er utført i årene 2008-2010. * (+) indikerer signifikans på 5 (10) prosent nivå.

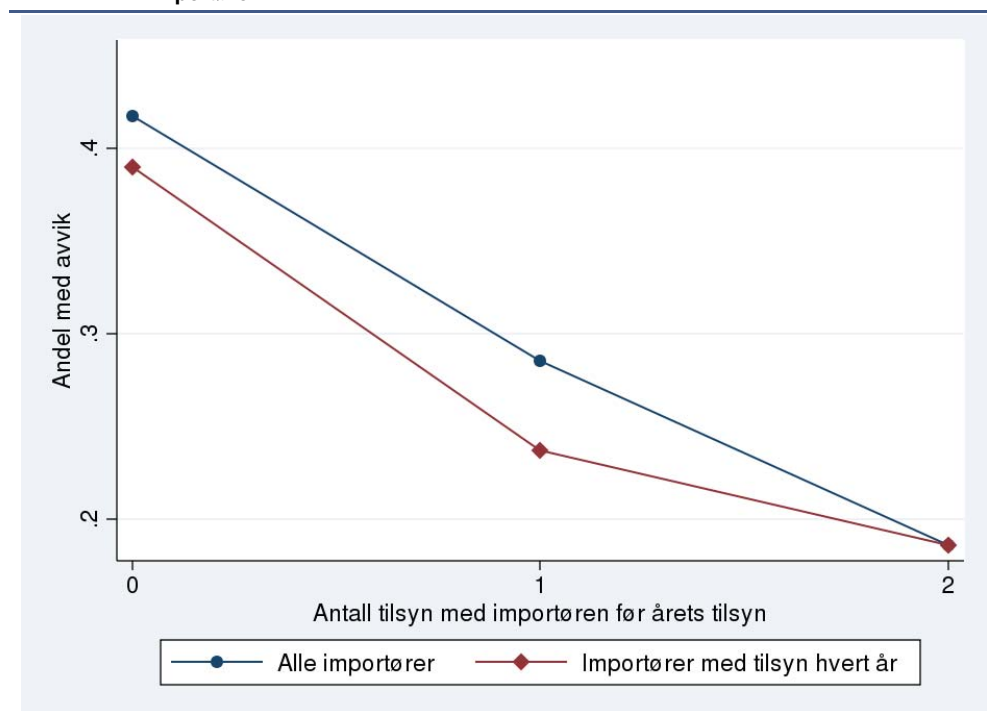
Effekt av tilsyn på framtidige avvik

Påvirkes overholdelsen i framtidige år av årets kontroll? Er en slik framtidig effekt avhengig av hva slags kontrolltype bedriften ble utsatt for i år (og i fjor)?

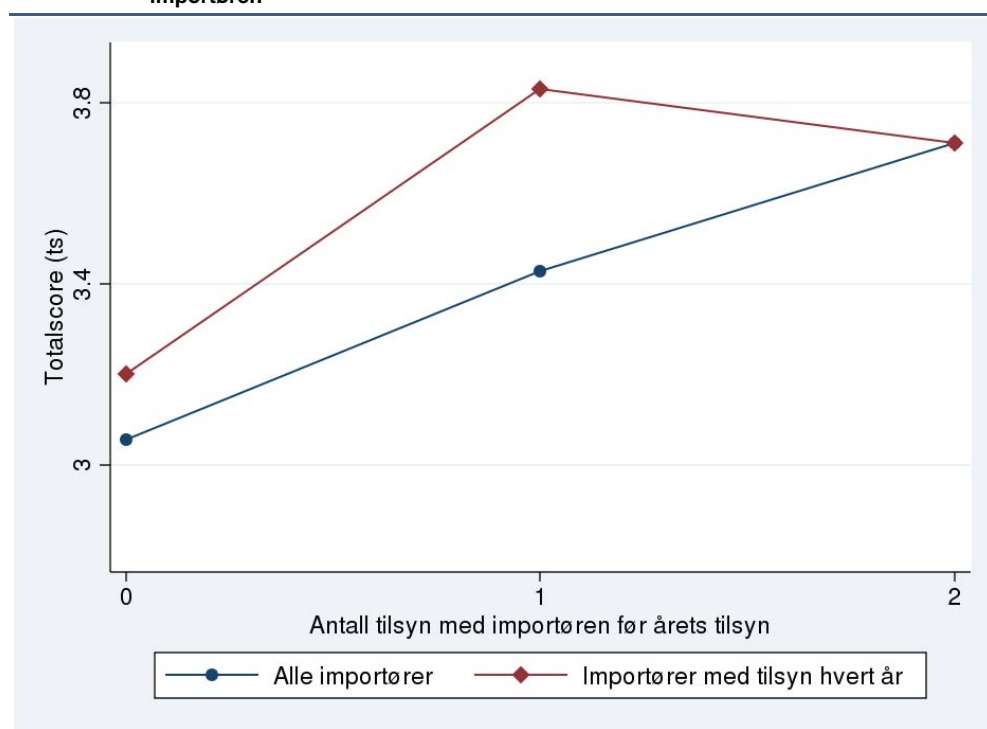
I en enkel teori der importørene kun overholder regelverket i den grad det lønner seg for dem (sannsynlighet for å bli tatt og straff hvis de blir tatt), så vil ikke en kontroll i år påvirke framtidig overholdelse av regelverket (de vil bare ta straffen og fortsette som før). I en mer realistisk verden kan vi imidlertid tenke oss mange grunner til at en kontroll skal påvirke framtidig overholdelse av regelverket. En kontroll kan ha en opplæringseffekt der importøren lærer hvordan reglene er og derfor lettere kan overholde dem. En kontroll kan også påvirke importørens kunnskaper om sannsynligheten for å bli tatt hvis han har avvik, samt hva slags kostnader som vil påløpe ved uoverenstemmelser med Klif. Ofte vil Klif også øke oppmerksomheten og kontrollfrekvensen (og dermed straffemulighetene) for importører med lav overholdelse eller liten samarbeidsvillighet. Det virker rimelig at alle disse forholdene vil variere med tilsynsformen importøren møter.

Figur 5.2a viser hvordan de avdekkede avvikene faller med antall ganger importøren tidligere er blitt kontrollert av Klif. Figuren ser på to ulike grupper av importører; i) importører som er blitt kontrollert i hvert år (dvs. 2 ganger tidligere), og ii) alle importører som er kontrollert minst en gang. Den andre gruppen (ii) blir identisk med den første (i) for importører som er kontrollert to ganger tidligere.

Figur 5.2a. Utviklingen i avvikene (avv) som funksjon av antall tidligere kontroller av importøren



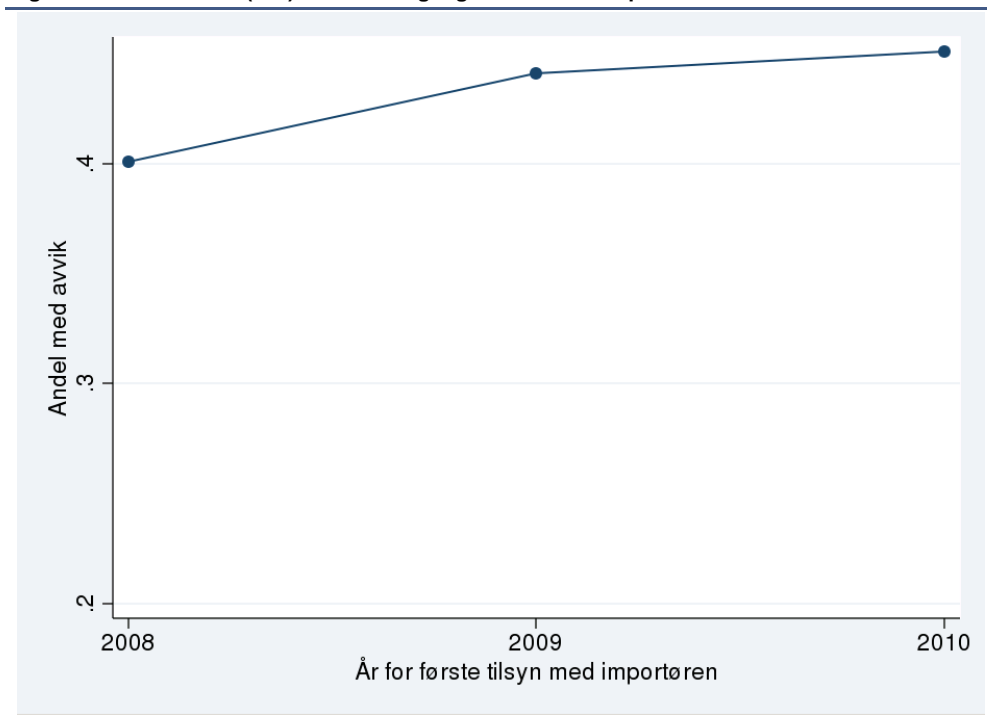
Figur 5.2b. Utviklingen i totalscoren (ts) som funksjon av antall tidligere kontroller av importøren



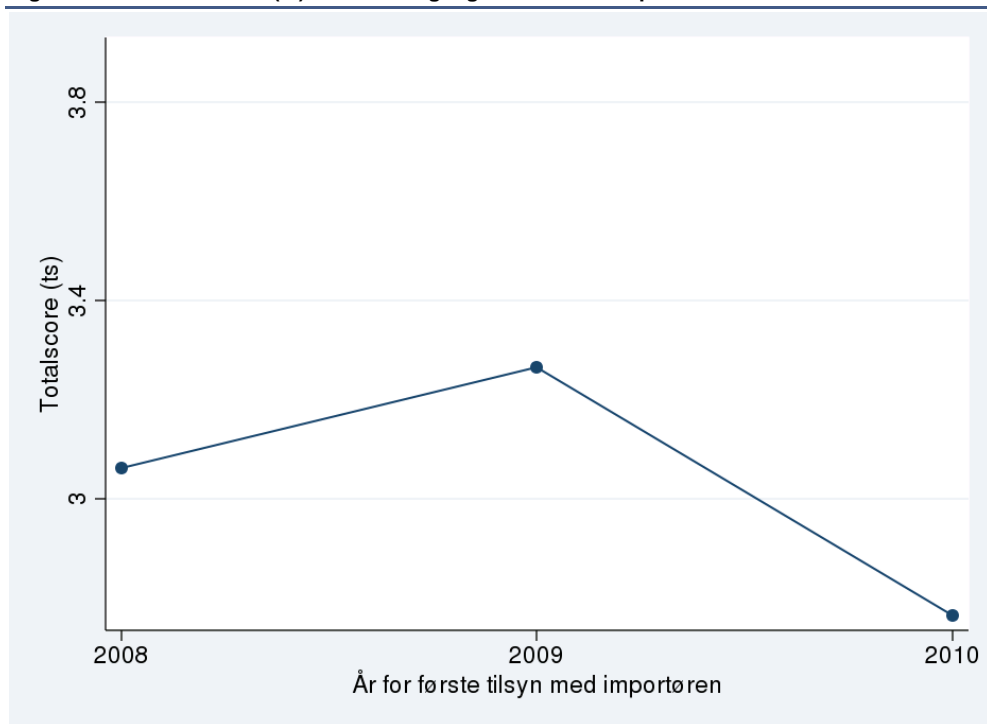
Rundt 40 prosent av importørene hadde avvik da de ble kontrollert første gang, mens i andre gangs kontroll er andelen med avvik falt til godt under 30 prosent. Avvikene hos de importørene (59) som ble kontrollert i alle tre år falt ytterligere, til snaut 20 prosent. Avvikene faller altså noe mer etter den første kontrollen enn etter den andre. I den tredje kontrollen er altså andelen med avvik halvert. Denne fallende trenden kunne tenkes å henge sammen med andre forhold enn Klifs tilsynsvirksomhet, for eksempel en generelt økende oppmerksomhet fra forbrukerne som importørene tar hensyn til eller at de utenlandske produsentene forbedrer sine rutiner. Figur 5a viser imidlertid at noen slik generell forbedring ikke har forekommet: Importører som ble kontrollert for første gang i 2008 hadde noe mindre

avvik enn importører som ble kontrollert for første gang i 2010.⁸ Sammen tyder således figur 5.2 og 5.3 på at Klifs tilsyn har en betydelig og gunstig effekt på de kontrollerte importørens overholdelse av regelverket. I det følgende tester vi dette formelt, ved å benytte importører som ikke tidligere er kontrollert som en kontrollgruppe for å påse at forbedringen i de kontrollerte importørens opptreden ikke skyldes forhold som ikke har noe med Klifs tilsyn å gjøre.

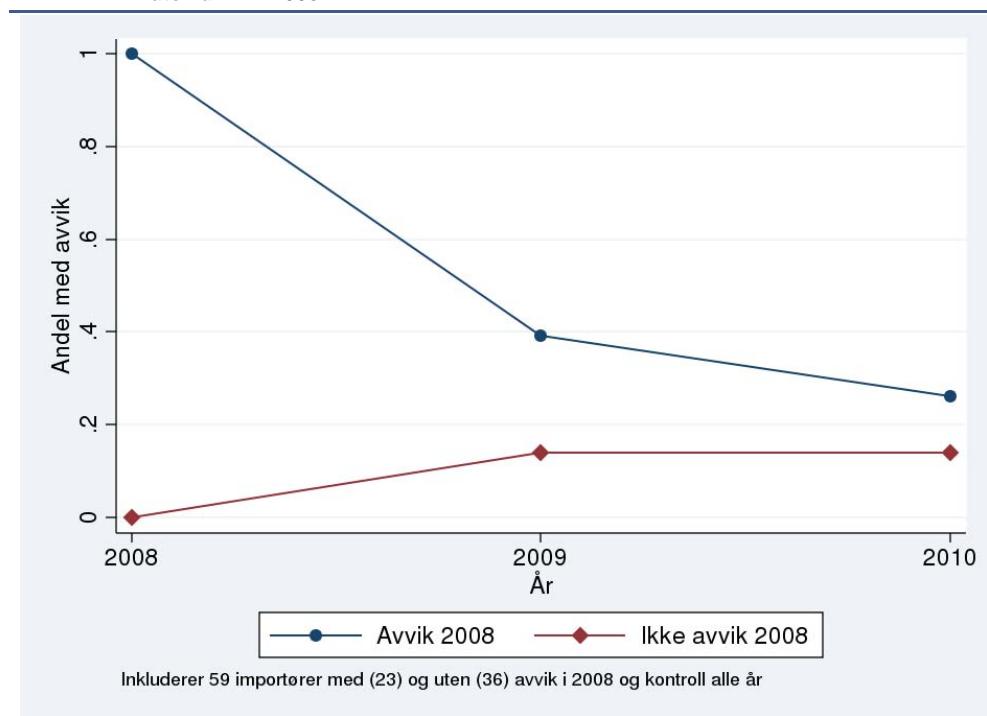
Figur 5.3a. Avvikene (avv) blant førstegangskontrollerte importører over tid



Figur 5.3b. Totalscoren (ts) blant førstegangskontrollerte importører over tid



⁸ Dersom Klifs tilsyn hadde en sterk allmennpreventiv effekt, så skulle vi vel forvente at avvikene falt over tid. På den annen side ville de avdekkede avvikene stige over tid dersom Klif ble bedre til å avdekke avvik gjennom sine kontroller. Andre forhold som ikke har noe som helst med Klifs tilsyn kan naturligvis også ha store effekter på utviklingen i antall avvik over tid.

Figur 5.4. Andelen importører med avvik (avv) i 2009 og 2010 fordelt på importører med og uten avvik i 2008

Vi undersøker nå om importørene som ble kontrollert i 2008 og 2009 hadde færre avvik i det påfølgende året (dvs. hhv. 2009 og 2010) enn de som ikke ble kontrollert i 2008 eller 2009. Alle tallene ses derfor i forhold til en kontrollgruppe av importører som ikke ble kontrollert året før.

Av de som ble kontrollert i året før, hadde 25 prosent avvik (avv), mens 42 prosent av dem som ikke ble kontrollert året før hadde avvik. Forskjellen er signifikant. Å ha blitt kontrollert året før reduserer altså sannsynligheten for avvik med 17 prosentpoeng. Eller med andre ord, Klif kan forvente at en kontroll i år reduserer sannsynligheten for at importøren har avvik neste år med 17 prosentpoeng. Dette innebærer at Klifs tilsyn reduserer importørens sannsynlighet for avvik det påfølgende året med 40 prosent (0,17/0,42).

Denne effekten varierer imidlertid med tilsynsformen. Blant de som ble kontrollert året før (25 prosent med avvik i påfølgende år, jf. over), var avviksprosenten 29 for dem som var blitt kontrollert med brev og 13 for dem som var blitt kontrollert med besøk. Et besøk reduserer importørens sannsynlighet for avvik det påfølgende året med hele 69 prosent. Forskjellen mellom besøk og brev medfører at Klif kan forvente at en besøkskontroll i år reduserer sannsynligheten for at importøren har avvik neste år med 16 prosentpoeng mer enn en brevkontroll i år.

Tabell 5.2. Effekt på avviket (avv) og totalscoren (ts) i 2009 og 2010 av angitt tilsynstype året før (dvs. hhv. 2008 og 2009)

Kontrolltype året før	n	Gjennomsnitt avv	Effekt avv	Gjennomsnitt ts	Effekt ts
Brev eller besøk	189	0,25	-0,17*	3,6	0,53*
Brev	143	0,29	-0,13*	3,4	0,37*
Besøk	46	0,13	-0,29*	4,1	1,09*

Kontrollgruppen består av 143 importører som ikke ble kontrollert året før. Disse importørene hadde en avv på 0,42 og ts på 3,0. * (+) indikerer signifikans på 5 (10) prosent nivå.

Det kan også tenkes mer langsiktige effekter. Og det kan tenkes at noen kombinasjoner av kontroller, som en besøkskontroll som året etter følges opp med en brevkontroll, er bedre enn andre kombinasjoner. Det er foreløpig litt begrenset tilgang på data for å undersøke dette, men vi ser her på noen ulike tilfeller i tabell 5.3. Alle tallene refererer seg til en referansegruppe som verken ble kontrollert i 2008 eller i 2009.

Som antydnet i figurene 5.2 og 5.3, viser tabellen at tilsyn reduserer framtidige avvik. Avvikene reduseres mest hos importørene som ble kontrollert både i 2008 og i 2009 (rundt 30 prosentpoeng), men reduksjonen er også stor blant importører som bare var kontrollert en gang (dvs. enten i 2008 eller i 2009). Som vi skulle forvente, faller avvikene som følge av den første kontrollen noe mer enn det ytterligere fallet som følge av to kontroller. Det er altså en positiv, men avtakende, effekt av antall kontroller på framtidig overholdelse.⁹ Isolert sett taler dette for at det er bedre å heller kontrollere mange importører sjelden enn å kontrollere få importører ofte.

Den gunstige effekten på framtidige avvik synes klart størst i kombinasjoner som inneholder besøk. Andelen med avvik hos importører som hadde besøk i ett av årene 2008 og 2009 er også lav (12 prosent). De importørene som fikk brev i 2008 eller 2009 har høyere avvik (29 prosent). Det er så å si ingen avvik hos importører som har fått en besøkskontroll som ble fulgt opp med en brevkontroll året etter, mens andelen med avvik er høyere for importører som hadde en brevkontroll som ble fulgt opp med en brevkontroll (25 prosent).

Fra tabell 5.4 ser vi at den absolutte effektene av tilsynet synes å være størst for forbrukerprodukter. Men vi skal legge merke til at avviket for kontrollgruppen også er høyest i forbrukerprodukter, slik at den relative effekten av tilsynet (målt som prosentvis reduksjon i andelen importører med avvik) er stor i alle gruppene. Mens kontroll i 2008 og 2009 reduserer andelen importører med avvik med 60 (0,29/0,48) prosent innenfor forbrukerprodukter, er reduksjonen 42 og 48 prosent for hhv. EE-produkter og byggevarer. Dersom Klifs mål var å redusere totalt antall avvik, så tilsier disse resultatene at man burde øke kontrollfrekvensen noe innenfor Forbrukerprodukter.

Tabell 5.3. Effekt på avvik (avv) og totalscore (ts) i 2010 av ulike kombinasjoner av tilsyn i 2008 og 2009

Kontrollsekvens	n	Gjennomsnitt avv	Effekt avv	Gjennomsnitt ts	Effekt ts
Kontroll 2008 og 2009	59	0,19	-0,26*	3,7	0,95*
Kontroll 2008 eller 2009	116	0,28	-0,18*	3,4	0,61*
Kontroll 2008, ikke kontroll 2009	24	0,29	-0,16	3,0	0,24
Ikke kontroll 2008, kontroll 2009	33	0,42	-0,03	3,0	0,27
Brev 2008 og brev 2009	36	0,25	-0,20+	3,5	0,76*
Brev 2008 eller brev 2009	95	0,29	-0,16+	3,3	0,55*
Brev 2008, ikke kontroll 2009	18	0,33	-0,12	3,0	0,24
Ikke kontroll 2008, brev 2009	21	0,52	0,07	2,6	-0,15
Brev 2008 og besøk 2009	6	0,17	-0,28	4,0	1,24+
Besøk 2008 og brev 2009	14	0,00	-0,45*	4,1	1,38*
Besøk 2008 og besøk 2009	3	0,00	-0,45	4,3	1,57+
Besøk 2008 eller besøk 2009	41	0,12	-0,33*	3,9	1,09*
Besøk 2008, ikke kontroll 2009	0				
Ikke kontroll 2008, besøk 2009	12	0,25	-0,20	3,8	0,99*

Kontrollgruppen består av 51 importører som verken ble kontrollert i 2008 eller i 2009. Disse importørene hadde avv på 0,45 og ts på 2,8. * (+) indikerer signifikans på 5 (10) prosent nivå.

Tabell 5.4. Effekt på avvik (avv) og totalscore (ts) i 2010 tilsyn i 2008 og 2009 på gruppenivå.

Kontrollsekvens	Kontrollgruppen			Behandlingsgruppen			Effekt i poeng	
	n	avv	ts	n	avv	ts	avv	ts
Kontroll 2008 og 2009								
Forbruker (1)	33	0,48	2,7	36	0,19	3,6	-0,29*	0,94*
EE-produkter (2)	14	0,43	2,7	8	0,25	3,6	-0,18	0,91
Byggevarer (3)	4	0,25	3,8	15	0,13	4,0	-0,12	0,25
Kontroll 2008 eller 2009								
Forbruker (1)	33	0,48	2,7	81	0,29	3,3	-0,19+	0,64*
EE-produkter (2)	14	0,43	2,7	19	0,26	3,3	-0,17	0,55
Byggevarer (3)	4	0,25	3,8	16	0,19	3,8	-0,06	0,06

Kontrollgruppen består av importører innen samme gruppe (dvs. 1, 2 eller 3) som verken ble kontrollert i 2008 eller i 2009. * (+) indikerer signifikans på 5 (10) prosent nivå.

⁹ Det kan også være interessant å se om effekten av en kontroll i år t avtar med antall år etter kontrollen. Vi ser få tegn til en slik avtakende effekt i våre data, men antall observasjoner blir svært lite når vi bryter opp på denne måten, og nærmere undersøkelser av slike mulige avtakende effekter må derfor vente til vi har data for flere år.

Selv om besøkskontroller øker framtidig overholdelse mer enn brevkontroller, så er det likevel et spørsmål om Klif bør ha flere brevkontroller på bekostning av besøk. Besøk krever mer ressurser, men fører også til større overholdelse av regelverket i ettertid. Hvordan vil en omprioritering fra besøk til brev forventes å påvirke avvikene året etter?

- 1 besøkskontroll mindre medfører at importørens sannsynlighet for avvik neste år øker fra 0,13 til 0,42, dvs med 29 prosentpoeng.
- 1 brevkontroll mer medfører at importørens sannsynlighet for avvik neste år reduseres fra 0,42 til 0,29, dvs med 13 prosentpoeng.

Én besøkskontroll reduserer altså sannsynligheten for at importøren har avvik året etter med 29 prosentpoeng, noe som om lag tilsvarende reduksjonen man oppnår med to brevkontroller (26). Da én besøkskontroll er mer ressurskrevende enn to brevkontroller, kan dette tyde på at Klif vil kunne øke overholdelsen som følge av en slik individualpreventiv effekt ved å øke antall brevkontroller på bekostning av besøkskontroller. Noen momenter kan likevel tale mot en slik omprioritering. Det kan tenkes at den allmennpreventive effekten av Klifs tilsyn avhenger av et visst omfang av besøkskontroller. Det kan også tenkes at den individualpreventive effekten vil bli redusert dersom sannsynligheten for besøkskontroll faller – besøkskontrollen kan fungere som et "ris bak speilet" som medfører mer oppriktig rapportering i brevkontrollene. Det er videre mulig at avvikene som avdekkes ved besøkskontroller er mer alvorlige enn de som avdekkes ved brev (jf. over der vi fant at alvorlige avvik kun var avdekket i besøkskontroller), noe som også kunne tale mot å benytte flere brevkontroller på bekostning av besøk.

Klif kan likevel vurdere å redusere antall besøk noe, men kanskje slik at besøk benyttes mer som oppfølging av brevkontroller med svak score. For å være sikker på å lavere sannsynlighet for besøk ikke reduserer kvaliteten på brevkontrollene, kan Klif vurdere å gjennomføre flere besøk (ved tilfeldig trekking) med noen års mellomrom (for eksempel hvert 5. år).

Skaper frafallet skjevheter?

Til tross for at det ikke skulle være anledning til det, kunne vi tenke oss at importører med avvik rett og slett lar være å besvare Klifs brevkontroller. Klif responderer i noen grad på slik atferd ved å øke besøkssannsynligheten for slike importører. Vi kan således undersøke om importører som hadde frafall i et år, hadde mer avvik (eller lavere *ts*) enn andre importører i det påfølgende året. Vi har sett på dette forholdet i tabell 5.5. Der fremkommer det at de 31 importørene med frafall i et år og kontrollresultat i det påfølgende året har signifikant høyere avvik enn importørene som ble kontrollert året før. Men de har om lag like høyt avvik som importørene som ble kontrollert for første gang.¹⁰ Dette kan tyde på at frafallet i liten grad foregår slik at importører med særlig mye avvik har en større tendens til å unndra seg tilsyn enn andre importører.

Frafallet er likevel bekymringsfullt. Dersom Klif ikke følger opp importører som ikke besvarer brevkontroller, vil det kunne spre seg et inntrykk av at man kan unndra seg Klifs tilsyn og dermed la være å følge regelverket. Det er mulig at dette vil kunne undergrave Klifs legitimitet blant mer seriøse aktører. Klif kan derfor vurdere å styrke dagens ordning med tettere oppfølging av importører som ikke besvarer Klifs henvendelser. Det kan videre vurderes å varsle bruk av tvangsmulkt allerede i følgebrevet til brevkontrollen, og det vil være svært interessant å teste om importører som får slikt varsel sammen med brevkontrollen har mindre frafall enn importører som ikke får slikt varsel (trekk tilfeldig blant importørene som får brevkontroll: halvparten får varsel og halvparten får det ikke).

¹⁰ En enkel regresjon med kontroll for tilsynstype viser at disse resultatene i liten grad påvirkes av at importører med frafall i år har en noe høyere sannsynlighet enn andre importører for å ha mottatt besøkskontroll (i stedet for brevkontroll) i det påfølgende året.

Tabell 5.5. Forskjell i avviket (avv) og totalscoren (ts) i 2009 og 2010 for importører med frafall året før (dvs. hhv. i 2008 og 2009)

De med frafall året før sammenliknet med dem med	n	Gjennomsnitt avv	Forskjell avv	Gjennomsnitt ts	Effekt ts
Brev eller besøk året før	189	0,25	0,20*	3,6	0,57*
Ingen kontroll året før	112	0,41	0,04	3,1	0,05

Det var 31 importører som hadde frafall året før (og altså ikke frafall i etterfølgende år). Disse importørene hadde avv lik 0,45 og ts lik 3,0 (i etterfølgende år). * (+) indikerer signifikans på 5 (10) prosent nivå.

Varselbrev vs. ikke varselbrev

I 2008 (kun forbrukerprodukter og byggevarer) og i 2009 sendte Klif ut brev med varsel om mulig kontroll til et tilfeldig utvalg av importører i forkant av kontrollene hvert år (brevet ble også sendt til importører som viste seg å ikke bli trukket ut til kontroll). I teorien kan man tenke seg at et varsel om kontroll øker importørens subjektive forventning om kontrollsannsynlighet (allmennpreventiv effekt), og derigjennom blir importørene mer påpasselige med at de oppfyller kravene. Dette vil i så fall gjenspeile seg i bedre regelkompetanse og færre avvik, altså bedre score på våre resultatindikatorer.

Tabell 5.6 viser forskjeller i resultatindikatorerne for gruppen av importører som mottok og som ikke mottok varselbrev. Vi ser at det *ikke* er tegn til at varselbrevet forbedrer resultatene i den påfølgende kontrollen.¹¹

Det er ikke hva vi hadde forventet (siden varselbrev skulle øke den subjektive sannsynligheten for kontroll), og kan forstås på iallfall to måter. For det første kan det forklares dersom personen i Klif som satte scoren visste om importøren hadde mottatt varselbrev eller ikke og at personen satte en strengere score overfor importører som hadde mottatt varselbrev (med informasjonsmaterieell). Det fremstår imidlertid ikke som veldig sannsynlig at denne personen skulle opptre slik, og særlig ikke fordi det ikke vil være lett å holde oversikt over hvilke importører som hadde mottatt slikt varselbrev.

For det andre kan dette forstås i lys av annen forskning som finner liknende resultater. Det foreligger noe forskning som tilsier at varselbrev som påpeker at kontrollsannsynligheten vil øke, bidrar til bedre overholdelse (Slemrod mfl. 2001). Men denne studien finner også at overholdelsen faller for en undergruppe. En ny studie tyder likeledes på at en positiv virkning av varselbrev bare forekommer dersom kontrollobjektene ikke hadde god kunnskap om den faktiske kontrollhyppigheten (Alm et al. 2009). Resultatene til Alm mfl. (2009) og for en undergruppe i Slemrod mfl. (2001) kan forstås slik at kontrollobjektene som varsles om høyere kontrollsannsynlighet setter seg mer inn i den faktiske kontrollsannsynligheten, regelverket og straffereaksjonene.¹² Dette kan medføre at importørens subjektive oppfatning av kontroll- og straffe-sannsynligheten og straffeomfanget faller, noe som kan lede til lavere overholdelse. Dette kan vi imidlertid ikke vite uten grundigere undersøkelser av importørenes oppfatninger før og etter varselbrevet. Klif kan likevel vurdere om det kan være noe med måten tilsynsvirkomheten drives på som kan bidra til at slike brev ikke har en forventet allmennpreventiv effekt.

På denne bakgrunn synes konklusjonen å være at det ikke kan forventes en allmennpreventiv effekt av slike varselbrev, og det er således neppe noen grunn til å sende ut slike brev.

¹¹ Klif mottok en del telefoner fra bekymrede importører som respons på disse varselbrevene, og Klif kan da ha beroliget importørene med at tilsynsregimet ikke var vesentlig strengere enn tidligere. Dette kan ha avvæpnet brevene og det kan således ha bidratt til at vi ikke finner noen effekt.

¹² Klif la et "faktaark" sammen med varselbrevet.

Tabell 5.6. Resultater med og uten varselbrev

Indikator	Varselbrev	Totalt	Forbrukerprodukter	EE-produkter	Byggevarer
avv	Ja	0,41	0,38	0,40	0,47
	Nei	0,33	0,30	0,39	0,35
	Differanse	0,08	0,08	0,01	0,12
ts	Ja	3,1	3,1	3,3	3,1
	Nei	3,4	3,5	3,5	3,1
	Differanse	-0,28+	-0,38+	-0,18	-0,01

Differansen er indikert som statistisk signifikant (*) dersom p-verdien av testen er under 0,05 (og med + dersom p-verdien er under 0,1). Utvalget er her kun grupper som mottok varselbrev, dvs. alle i 2009 og kun forbrukerprodukter og EE-produkter i 2008 (dvs. 327 kontroller).

Treffsikkerheten til det risikobaserte tilsynet

Klif driver såkalt risikobasert tilsyn, noe som betyr at importørene med størst skadepotensial og høyest sannsynlighet for avvik skal kontrolleres hyppigst. Det er således ingen entydig sammenheng mellom kontrollhyppigheten og avvikssannsynligheten. Hvis vi likevel antar at det er en positiv sammenheng mellom kontrollhyppigheten og score på våre resultatindikatorer, så kan vi få en indikasjon på hvor godt Klif treffer importørene med høy risiko når Klif velger ut hvilke importører som skal ha hyppig kontroll. Eller mer presist: er det slik at de importørene Klif har kontrollert med en høyere sannsynlighet også viser seg å score dårligere på indikatorene?

Vi kan undersøke dette på flere måter. For det første har Klif delt importørene inn i tre nivåer, der importører i nivå 1 har høyere prioritet enn importører i nivå 2, som igjen har høyere prioritet enn importører i nivå 3. Vi kunne altså forvente at det ble avdekket mer avvik i nivå 1 enn i nivå 2 (enn i nivå 3). Tabell 5.7 viser forskjellene i resultatindikatorerne for importører i de tre nivåene.

Andelen med avvik er lavere i nivå 1 enn i nivå 2 og 3, og totalscoren er også bedre i nivå 1 enn i nivå 2 og 3. Da importørene i nivå 1 har hatt høyere sannsynlighet for kontroll, betyr dette at Klif har brukt mer ressurser på å kontrollere importører som har vist seg å ha bedre resultatoppnåelse. I den grad målet til Klif har vært å plassere bedrifter med antatt dårligst resultatoppnåelse under nivå 1 (og de med bedre resultatoppnåelse under nivå 2 og 3), har man mislykkes: Ved å trekke importører helt tilfeldig (i stedet for å oversample nivå 1), ville Klif avdekket flere importører med avvik.¹³

Tabell 5.7. Representative gjennomsnitt av resultatindikatorer over nivå

	Ts	gs	ik	fs	avv	aa	anm
Nivå 1	3,5	3,6	3,6	3,8	,29	0,02	,72
Nivå 2	2,8	3,0	3,1	3,1	,47	0,01	,71
Nivå 3	2,8	3,1	3,0	2,5	,43	0,00	,68

Resultatene bygger på data fra de 534 kontrollene som er utført i årene 2008-2010. Det er tatt hensyn til stratifisering og oversampling ved utregning av gjennomsnitt (og standardfeil benyttet i testene referert til i teksten). Som utdypet i vedlegget, er det grunn til å tro at disse standardfeilene er for store slik at nullhypotesen forkastes for sjelden).

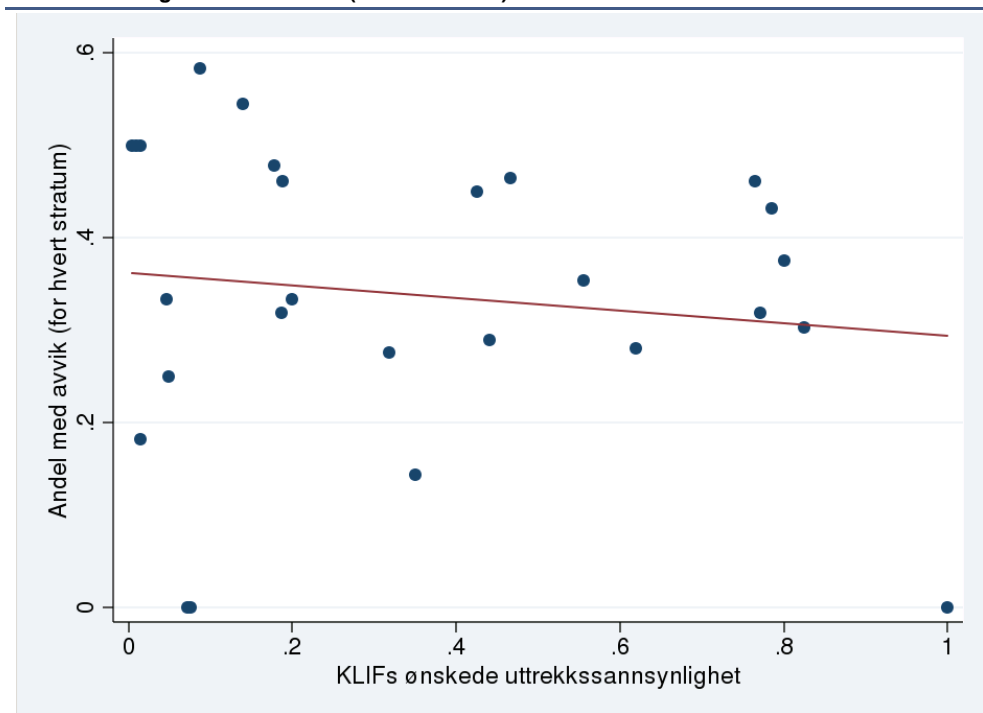
For det andre kan vi benytte oss av sannsynligheten for å bli trukket ut som indikator for hvor viktig Klif synes det er å kontrollere en bedrift. Dette er altså et relatert og alternativt mål for bruk av nivå 1, 2 og 3 over. Dette er illustrert i figur 5.5, der uttrekkssannsynligheten for hvert stratum (definert som gruppe, undergruppe og nivå, noe som gir 27 strata) er angitt på x-aksen (fra 0 til 1), og gjennomsnittet av avv (innen hvert stratum) på y-aksen. Dersom målet med å kontrollere noen importører hyppigere enn andre er å rette tilsynet mot importørene som Klif tror har mange avvik, så skal vi forvente at observert andel med avvik skal være høyere jo høyere uttrekkssannsynligheten er. Dette innebærer at

¹³ Dersom importørene forutser hvilken kontrollhyppighet de har, så kan men tenke seg at vi kunne observere akkurat det samme: Importører med høy kontrollsannsynlighet vil da redusere sine avvik fordi de forutser at de ikke vil slippe unna med avvik, mens importører med lav kontrollsannsynlighet ikke vil redusere sine avvik tilsvarende. Vi vil da kunne komme i en situasjon hvor avvikstilbøyeligheten er lik for importører med ulik kontrollfrekvens. Det re imidlertid lite som skulle tilsi at importørene har kunnet fremskaffe slik informasjon i vårt tilfelle.

punktene skal ligge langs en stigende linje i figuren. Som vi ser, er det ikke noen tegn til dette (regresjonsanalyser med vektorer viser det samme). Vi kan oppsummere med at de importørene Klif velger å kontrollere hyppig ikke synes å ha dårligere score på resultatindikatorene *avv* og *ts* enn de importørene Klif velger å kontrollere sjelden.¹⁴ Det ser sågar ut til at avvikene er aller flest i en gruppe med importører som kontrolleres meget sjelden.

Resultatindikatorene fanger i stor grad opp avvik, men i mindre (eller liten/ingen) grad opp konsekvensene ved et avvik. Begge deler inngår i Klifs vurdering av risiko (sannsynligheten for avvik multiplisert med konsekvensene ved et avvik), og begge elementene bør fanges opp av resultatindikatorer som benyttes til å evaluere hvorvidt Klifs risikobaserte tilsyn faktisk treffer importørene med størst risiko. Er det mulig å finne operasjonaliserbare resultatindikatorer som kan fange opp begge disse elementene? Kunne *ts* delt på importørens størrelse (for eksempel et eller annet mål på omsetning) vært en interessant slik resultatindikator? Eller kunne Klif gjort en vurdering av alle importørenes skadepotensial (på en skala fra 1 (lavt) til 10 (høyt)?), og så kunne en resultatindikator være *ts* multiplisert med dette skadepotensialmålet? I forkant av trekking til kontroll kan Klif vurdere å tallfeste hvor sannsynlig Klif tror det er at importøren har avvik hvor stort skadepotensial (konsekvens) mener Klif at importøren har. Disse vurderingene gjøres antakeligvis også i dag, men da mer implisitt. Eksempelvis legges importører i nivå 1 dersom de har stort skadepotensial. Men også importører med svake resultater i tidligere kontroller, eller som ikke har respondert på tidligere brevkontroller, legges i nivå 1. For å kunne evaluere hvor godt Klif lykkes i arbeidet med å gjøre tilsynet (mer risikobasert, kan Klif søke å tallfeste i) og ii).

Figur 5.5. Sammenhengen mellom ønsket kontrollhyppighet (vannrett akse) i hvert stratum og andel med avvik (loddrett aksene)



¹⁴ Alvorlige avvik (*aa*) avdekkes mest på nivå 1, og i den grad alvorlige avvik også gjenspeiler skadepotensial, så kan vel dette tolkes som et tegn på at Klifs tilsyn er risikobasert. Det foreligger imidlertid foreløpig for få observasjoner av alvorlige avvik til at verdien av ytterligere kvantitative analyser er begrenset.

6. Skisse til metode for systematisk utvelgelse av importører med antatt dårlig score

Kan noen av variablene vi har tilgang til forutsi hvilke importører – for gitt risiko – som har dårlig overholdelse? Kanskje har små importører dårligere overholdelse enn større importører, eller avhenger overholdelsen av opprinnelsesland eller omsetning?

Basert på importørene som ble kontrollert i 2008-2010 kan vi ved en regresjonsanalyse finne ut hvilke kjennetegn ved disse som predikerer dårlig score, som gruppe eller størrelse. Det er vel grunn til å tro at andre variable (som opprinnelsesland, landsdel, kommune, selskapsform, antall ansatte, omsetning, antall opprinnelsesland, tidligere overholdelse, tidligere kontroller, osv.) vil være vel så viktig for å predikere overholdelse. Når Klif utvikler operasjonaliserbare resultatindikatorer for risiko, kan metoden benyttes til å predikere hvilke importører som har høyest risiko. Dermed vil det være mulig å undersøke hvordan Klif kan systematisere utvelgelsen av importører som skal ha høy kontroll sannsynlighet for å forbedre det risikobaserte tilsynet.

7. Konklusjoner og oppsummerende anbefalinger

Tilsynet har en individualpreventiv effekt

Å kontrollere en importør reduserer importørens sannsynlighet for å ha avvik neste år med 40 (brevkontroll) til 70 (besøkskontroll) prosent. Grundigere tilsyn (besøkskontroll) avdekker flere avvik og har en gunstigere effekt på framtidige avvik enn mer overflatisk tilsyn (brevkontroll).¹⁵

Mer brev, mindre besøk?

Selv om besøk både avdekker flere avvik og bedrer framtidig overholdelse mer enn brev, så er besøk såpass mye mer ressurskrevende at Klif likevel kan vurdere å redusere bruken av besøk noe, og i stedet øke bruken av brev. Det er også et behov for å følge overholdelsen i de delene av populasjonen som nå har lav kontroll sannsynlighet tettere, noe brevkontroll er egnet for. Det kan også vurderes om besøk i hovedsak kun skal benyttes som reaksjon på dårlig score på brevkontroll (og frafall og tips), og da gjerne i kombinasjon med andre reaksjoner som for eksempel varsel om tvangsmulkt (gjerner med tilfeldig trekking for å kunne teste effektene av ulike reaksjoner).

Frafallet kan være bekymringsfullt

En ikke ubetydelig andel av importørene påstår at de ikke lenger er importører av de aktuelle produktene, eller de lar rett og slett være å svare på Klifs henvendelser.¹⁶ Dersom Klif ikke følger opp slike importører tettere, vil det kunne spre seg et inntrykk av at man kan unndra seg Klifs tilsyn og dermed la være å følge regelverket. Dette vil kunne undergrave Klifs legitimitet blant mer seriøse aktører. Da overholdelsen blant importører med frafall i fjor, men uten frafall i år, ikke skiller seg fra overholdelsen til importører som ikke ble kontrollert året før, kan Klif særlig ta for seg importører som lar være å svare i gjentatte kontroller.

¹⁵ Selv om disse resultatene er oppløftende, bør Klif likevel reflektere over følgende. Dersom importørene tror at Klif lar dem rette opp avvik etter at avviket er oppdaget, og kun straffer importører som ikke foretar slik oppretting, så vil ikke importørene ha incentiver til å (sette seg inn i og) overholde regelverket før Klif har avdekket avvik. Under slike forhold vil vi også lett observere sterke individualpreventive effekter, men det skyldes altså svak (eller ingen) allmennpreventiv effekt av Klifs tilsyn.

¹⁶ Et underordnet praktisk poeng er at Klif bør "vaske" populasjonen de får fra Tollmyndighetene før tilsynsobjektene trekkes (og ikke etterpå slik det ser ut til å ha vært gjort i 2009 og 2010).

Uklart om tilsynet er risikobasert

Klif ønsker å drive risikobasert tilsyn. Her veier Klif tilsynsobjektene sannsynlighet for avvik med konsekvensene av avvik. Det ser ikke ut som Klifs tilsyn er hyppigere overfor objekter med mye avvik – Klif ser altså ut til å bruke like mye ressurser på å kontrollere importører som har lite avvik som importører som har mye avvik. Det er sågar mest avvik i en gruppe importører som kontrolleres veldig sjelden. I prinsippet kan dette forsvares dersom konsekvensene av avvik er mer alvorlig for de importørene som Klif kontrollerer hyppig (men der Klif altså ikke finner mye avvik). Klif besitter imidlertid ikke data som tilsier at dette er tilfelle, selv om man kanskje kan hevde at slik informasjon implisitt ligger til grunn for plassering av importører i ulike undergrupper og nivåer. Klif kan gjøre denne implisitte informasjonen eksplisitt slik at det blir mulig å undersøke hvorvidt det risikobaserte tilsynet faktisk er risikobasert. Dette vil også gi Klif informasjon som vil gjøre det mulig å gjøre det risikobaserte tilsynet enda mer risikobasert.

Andre anbefalinger

I det videre arbeidet kan det legges vekt på å sette opp datasystemer som forenkler trekking av kontrollobjekter og kontrolltyper, samt som bidrar til å sikre korrekt og fullstendig registrering av endringer og registreringer. Det kan også satses noe mer på å kartlegge deler av populasjonen som så langt har hatt lav kontrollsannsynlighet. Klif kan vurdere å klargjøre ytterligere og skriftlig hva som er kriteriene for at det gis hhv. 1, 3 og 5 på *ik* og *fs*, samt kriteriene for fastsettelsen av *ts* og alvorlig avvik (*aa*). Det kan også vurderes i) å registrere antall (alvorlige) avvik ved hver kontroll (ville for eksempel gjøre det mulig å regne på hvor mange flere avvik en besøkskontroll avdekker sammenliknet med en brevkontroll), samt ii) å tallfeste importørenes risiko før uttrekk, gjerne ved skjønnsmessig fastsettelse av skadepotensial og avvikssannsynlighet.

Referanser

Alm, J., B. Jackson og M. McKee 2009. Getting the word out: Enforcement information dissemination and compliance behavior. *Journal of Public Economics* 93, 392-402.

Dickens, W. 1990. Error components in grouped data: Is it ever worth weighting? *Review of Economics and Statistics* 72 (2), 328-333.

DuMouchel, W. og G. Duncan 1983. Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association* 78 (383), 535-543.

Greene, W. 2000. *Econometric analysis*. Fourth Ed, Prentice Hall.

Korn, E. og B. Graubard 1999. *Analysis of Health Surveys*. John Wiley & Sons.

Lee, J. Y. og G. Solon 2011. The Fragility of Estimated Effects of Unilateral Divorce Laws on Divorce Rates. NBER 16773.

Pfreferrmann, D. 1996. The use of sampling weights for survey data analysis. *Statistical Methods in Medical Reserach* 5, 239-61.

Skinner, C. J. 2003. Introduction to Part B, Chapter 6 i: R. Chambers and C. Skinner (red.), *Analysis of Survey Data*. John Wiley & Sons, England.

Slemrod, J., M. Blumenthal og C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics* 79, 455-483.

Vedlegg A: Fremgangsmåte og forutsetninger for de statistiske analysene

Tilfeldig trekking fra hvert stratum medfører at gjennomsnittet innen hvert stratum er representativt for hele stratumpopulasjonen. Gjennomsnitt over flere strata er vektet med populasjonen i hvert stratum og er således også representative for de respektive stratapopulasjonene.¹⁷ De estimerte standardfeilene tar hensyn til stratifisering og er korrigert for trekking uten tilbakelegging fra endelig sampel, men det er ikke tatt hensyn til frafall. Jeg viser for øvrig til statistikkprogrammet Stata 10 (kommandogruppen `svy`), som er benyttet til å gjennomføre beregningene. For strata med kun en observasjon, har jeg benyttet option `scaled`.

Det har vært en debatt i faglitteraturen om når survey-vekter skal benyttes og når de ikke skal benyttes (Skinner 2003, s. 83). Det er enighet om at vekter skal benyttes når vi ønsker å estimere populasjonsstørrelser som representative gjennomsnitt (og det har jeg gjort her). Men når vi skal estimere effekter på importørens atferd av ulike tiltak, er det i og for seg ikke populasjonsparametre men atferdseffekter vi er ute etter, og det er for slike tilfeller det har pågått en debatt om hvorvidt det er korrekt å vekte. Det påpekes at dersom modellen er rimelig velspesifisert, så vil ikke samplingen introdusere skjevhet i de uvektede estimatene. Poenget er at dersom effektene er ulike i ulike strata (heterogene effekter), så vil vektene tillegge effektene fra noen strata større vekt i beregningen av den samlede effekten. I vårt tilfelle er punktestimatene oftest svært like i vektete og uvektede regresjoner, noe som taler for å benytte de enklere uvektede modellene (DuMouchel og Duncan 1983).

I tillegg har den vektete estimatoren normalt større varians enn den uvektede, noe som også taler mot å vekte (Pfeffermann 1996). I vårt tilfelle er det i tillegg meget stor variasjon i vektene, noe som gjør dette problemet enda større (Skinner 2003, Greene 2000, s. 522, Lee og Solon 2011, Dickens 1990). I tråd med denne litteraturen (Korn og Graubard 1999, sect. 4.6) har jeg av disse grunnene ikke brukt vekter når jeg har testet effekter av tilsynet på importørens overholdelse.

Testene det refereres til i notatet er t-tester (tosidige), og med mindre annet er angitt omtales en forskjell som signifikant dersom p-verdien er mindre enn 0.05. De relativt små utvalgene og det forhold at resultatindikatorne bare antar bestemte verdier kan tilsi at resultatene fra andre typer tester (for eksempel ikke-parametriske) ville kunne være noe annerledes og mer pålitelige. Når jeg har testet atferdshypoteser, har jeg derfor i noen tilfeller også utført slike ikke-parametriske tester (Wilcoxon-Mann-Whitney), og de ser ut til å gi liknende resultater som de som rapporteres i dette notatet. Vi skal også legge merke til at en gitt forskjell i større grad vil være signifikant mellom store grupper som følger av flere observasjoner.

Det kan være verdt å nevne et forhold ved testen av forskjeller mellom besøkskontroller og brevkontroller. I enkelte strata innenfor forbrukerprodukter i 2008 ble det bare trukket brevkontroll eller bare besøkskontroll. I tillegg er trekkingen av brev vs. besøk gjort på gruppenivå (ikke innenfor strata) i 2008 for byggevarer og EE-produkter, noe som i prinsippet vil kunne påvirke effektestimaterne dersom effektene er heterogene over strata.

¹⁷ Men i konstruksjonen av vektene har jeg ikke forsøkt å ta hensyn til frafall, noe som bl.a. innebærer at frafallet forutsettes å ikke være systematisk knyttet til overholdelsen. Resultatene i tabell 5.5 kan tyde på at denne forutsetningen ikke er urimelig.

Figurregister

4.1a. Representative gjennomsnitt for <i>avv</i> over tid. Totalt, for hver gruppe	10
4.1b. Representative gjennomsnitt for <i>ts</i> over tid. Totalt, for hver gruppe	10
4.2a. Representative gjennomsnitt for <i>avv</i> innen hver undergruppe (alle år samlet).....	11
4.2b. Representative gjennomsnitt for <i>ts</i> innen hver undergruppe (alle år samlet)	11
5.1. Importører med avdekkede avvik i besøkskontroll og brevkontroll	13
5.2a. Utviklingen i avvikene (<i>avv</i>) som funksjon av antall tidligere kontroller av importøren	14
5.2b. Utviklingen i totalscoren (<i>ts</i>) som funksjon av antall tidligere kontroller av importøren	14
5.3a. Avvikene (<i>avv</i>) blant førstegangskontrollerte importører over tid	15
5.3b. Totalscoren (<i>ts</i>) blant førstegangskontrollerte importører over tid	15
5.4. Andelen importører med avvik (<i>avv</i>) i 2009 og 2010 fordelt på importører med og uten avvik i 2008	16
5.5. Sammenhengen mellom ønsket kontrollhyppighet (vannrett akse) i hvert stratum og andel med avvik (loddrett aksene).....	21

Tabellregister

3.1. Oversikt over populasjonen og kontrollene	9
4.1. Representative gjennomsnitt av resultatindikatorer over tid og gruppe	10
5.1. Resultater fordelt på kontrolltype.....	13
5.2. Effekt på avviket (<i>avv</i>) og totalscoren (<i>ts</i>) i 2009 og 2010 av angitt tilsynstype året før (dvs. hhv. 2008 og 2009)	16
5.3. Effekt på avvik (<i>avv</i>) og totalscore (<i>ts</i>) i 2010 av ulike kombinasjoner av tilsyn i 2008 og 2009	17
5.4. Effekt på avvik (<i>avv</i>) og totalscore (<i>ts</i>) i 2010 tilsyn i 2008 og 2009 på gruppenivå. ...	17
5.5. Forskjell i avviket (<i>avv</i>) og totalscoren (<i>ts</i>) i 2009 og 2010 for importører med frafall året før (dvs. hhv. i 2008 og 2009)	19
5.6. Resultater med og uten varselbrev.....	20
5.7. Representative gjennomsnitt av resultatindikatorer over nivå	20