

Snyte skatt hvis mulig? Hva fortalte Verdiundersøkelsen 2007?

Jostein Lillestøl
Jonas Andersson

SNF



SNF

SAMFUNNS- OG NÆRINGSLEVINGSFORSKNING AS

- er et selskap i NHH-miljøet med oppgave å initiere, organisere og utføre ekstern-finansiert forskning. Norges Handelshøyskole og Stiftelsen SNF er aksjonærer. Virksomheten drives med basis i egen stab og fagmiljøene ved NHH.

SNF er ett av Norges ledende forskningsmiljø innen anvendt økonomisk-administrativ forskning, og har gode samarbeidsrelasjoner til andre forskningsmiljøer i Norge og utlandet. SNF utfører forskning og forskningsbaserte utredninger for sentrale beslutningstakere i privat og offentlig sektor. Forskningen organiseres i programmer og prosjekter av langsiktig og mer kortsiktig karakter. Alle publikasjoner er offentlig tilgjengelig.

SNF

CENTER FOR APPLIED RESEARCH AT NHH

- is a company within the NHH group. Its objective is to initiate, organize and conduct externally financed research. The company shareholders are the Norwegian School of Economics (NHH) and the SNF Foundation. Research is carried out by SNF's own staff as well as faculty members at NHH.

SNF is one of Norway's leading research environment within applied economic administrative research. It has excellent working relations with other research environments in Norway as well as abroad. SNF conducts research and prepares research-based reports for major decision-makers both in the private and the public sector. Research is organized in programmes and projects on a long-term as well as a short-term basis. All our publications are publicly available.

Arbeidsnotat nr. 03/14

**Snyte skatt hvis mulig?
Hva fortalte Verdiundersøkelsen 2007?**

av

**Jostein Lillestøl
Jonas Andersson**

SNF prosjekt 6568/6569

“Norwegian Center for Taxation (NoCeT)”

NORSK SENTER FOR SKATTEFORSKNING

Dette arbeidsnotatet inngår i en serie publikasjoner fra Norsk senter for skatteforskning/Norwegian Center for Taxation (NoCeT). Senteret har til formål å fremme undervisning og bidra til forskning på skatt. Senteret er finansiert av NHH, Skattedirektoratet og Norges forskningsråd og er en del av en større satsing på skatteøkonomi i Norge.

**SAMFUNNS- OG NÆRINGSLIVSFORSKNING AS
BERGEN, FEBRUAR 2014
ISSN 1503-2140**

© Materialet er vernet etter åndsverkloven. Uten uttrykkelig samtykke er eksemplarframstilling som utskrift og annen kopiering bare tillatt når det er hjemlet i lov (kopiering til privat bruk, sitat o.l.) eller avtale med Kopinor (www.kopinor.no)
Utnyttelse i strid med lov eller avtale kan medføre erstatnings- og straffeansvar.

Snyte skatt hvis mulig? Hva fortalte Verdiundersøkelsen 2007?

Av Jostein Lillestøl og Jonas Andersson (NHH/SNF¹)

1. Innledning

Verdiundersøkelsen 2007 er Norges siste bidrag til World Values Survey (WVS), som ved passende mellomrom gjennomføres i mer enn femti land verden over med identiske spørsmål. En ny verdiundersøkelse er gjennomført, og data ventes tilgjengelig i 2014. Undersøkelsen omfatter følgende spørsmål om holdningen til skattesnyteri:

Hvor på denne skalaen vil du plassere deg selv når det gjelder å rettferdiggjøre å

Snyte på skatten dersom du har mulighet til det (V200)

1 2 3 4 5 6 7 8 9 10
 Kan aldri rettferdiggjøres Kan alltid rettferdiggjøres

Svarene fordelte seg prosentvis slik (n=1017):

Aldri = 1	2	3	4	5	6	7	8	9	Alltid =10
49.7	21.7	11.4	3.9	5.9	2.3	1.7	1.7	0.7	1.1

Tabell 1: Norge: Snyte på skatten

eller dikotomisert med tilsvarende prosenttall for en del andre land:

Land	n	Aldri = 1	Ikke si aldri >1
Norge	1017	49.7	50.3
Sverige	996	52.7	47.3
Tyskland	2032	59.0	41.0
Italia	1012	61.5	38.5
USA	1190	66.7	33.3

Tabell 2: Fem land: Snyte på skatten (dikotomisert)

Ved vurdering av om forskjeller er statistisk signifikante kan grovt regnes feilmarginer +/- 2 prosent til hver side.

Verdiundersøkelsen inneholder et betydelig antall variable som muligens samvarierer med V200, og som kan belyse potensialet for lovbrudd av typen skatteunndragelse. I undersøkelsen kommer spørsmålene i

¹ Dette arbeidet er utført som del av et prosjektet Mønster-gjenkjenning av skatteunndragere ved Senter for skatteforskning (NoCeT) med finansiell støtte av Skattedirektoratet.

naturlige grupper, der vi gir de mest relevante gruppene merkelappene: Oppfatning av egen situasjon, Personlighetstrekk (etter egen vurdering), Prioritering av miljø, Politisk ståsted, Sosioøkonomiske synspunkter, Tillit til offentlige organer og organisasjoner, Rettferdiggjøring av ulike handlinger. Variabelen «Snyte på skatten dersom du har mulighet til det» (V200) kommer i den siste gruppen.

Det foreligger flere studier av dataene fra WVS 2007 der holdningen til skatteunndragelse (V200) inngår, i første rekke av Torgler (2003, 2007, 2011). I disse studiene er forhold knyttet til Norge ikke belyst spesielt, og vår studie tar sikte på nettopp det, og med til dels andre (og mer avanserte) metoder enn det Torgler har brukt. Den norske delen av WVS materialet er imidlertid studert i noe detalj av Fugleberg (2013), men vårt perspektiv er utvidet, og gjør bruk av mer avanserte metoder. Hvorfor foretar vi analyser av WVS 2007 når nye data er rett rundt hjørnet? Vi mener at skatteunndragelse i WVS 2007 ikke er gjennomanalysert, og at dette bør gjøres, blant annet for å muliggjøre sammenligninger med tilsvarende analyser på de nye data.

Studien i denne rapport er eksplorativ, vi ønsker altså å studere kjennetegn blant de som er tilbøyelige til å unndra skatt, ikke å trekke konklusjoner om hvorfor de gjør det. Vi vil imidlertid her notere noen muligheter til konfirmativ analyse som vi også mener at det er av interesse å gjennomføre. Vi merker oss at V200 dreier seg om rettferdiggjøring av unndragelse som ikke er presist det samme som at en i en situasjon der muligheten forelå ville gjøre det selv. Det er imidlertid mulig å tenke seg at rettferdiggjøringsvariablene er indikatorer på en latent variabel "Tilbøyelig til unndragelse". Det melder seg da et spørsmål om vi skal betrakte denne som en eksogen eller endogen variabel. I det siste tilfellet er det naturlig å se på de observerte variable i de andre gruppene som indikatorer for tilsvarende eksogene latente variable. Dette åpner for en analyse av kovariansstrukturer med såkalte strukturligningsmodeller. Ut fra det siste perspektivet er det også mulig å foreta eksplorative analyser, der en tar for seg variablene gruppevis og samlet, og spør om hvilke variabelgrupper og hvilke variable innen gruppen er best i stand til å skille ut de som i visse situasjoner kan rettferdiggjøre skatteunndragelse, og kanskje dermed ville kunne gjøre det selv. Dette kan belyses ved flere ulike metoder, bl.a. ved logistisk regresjon, som i stor utstrekning er analysemetoden som Torgler bruker. Vi finner at en trebasert analyse av typen CART («Classification And Regression Trees» er mer velegnet da denne metoden kan håndtere mange forskjellige variabeltyper (kontinuerlige, ordinale etc) simultant. Det er også av interesse å foreta tilsvarende analyser på de deler av materialet som Torgler har brukt, og se om de konklusjoner som trekkes er de samme.

2. Aktuelle variable

Følgende variable med antatt interesse og mulig sammenheng er valgt:

Oppfatning av egen situasjon

V46 Hvor mye valgfrihet og selvbestemmelse mener du har når det gjelder hvordan livet ditt blir?

1	2	3	4	5	6	7	8	9	10
Overhodet ingen									En hel del

Det kan også være av interesse å studere sammenhengen mellom rettferdiggjøring av skatteunndragelse og følgende grupper av variable (Skatteunndragelse er i den siste gruppen)

Sosio-økonomiske synspunkter: (alle på skala fra 1 til 10)

V116 Det bør bli større likhet mellom folks inntekter (1) Vi trenger større inntektsforskjeller som belønning for personlig innsats (10)

V117 Privat eiendomsrett til næringsliv og industri bør økes (1) Offentlig eiendomsrett til næringsliv og industri bør økes (10)

V118 Det offentlige bør ta større ansvar for å sikre at alle har det de trenger (1) Folk bør ta større ansvar for å klare seg selv (10)

V119 Konkurransen er sunt, det stimulerer folk til å arbeide hardt og utvikle nye ideer (1) Konkurransen er skadelig, det frembringer det verste i folk (10)

V120 På lang sikt gir hardt arbeid som oftest suksess i livet (1) ... Hardt arbeid gir ikke nødvendigvis suksess. Det er mer et spørsmål om tilfeldigheter og forbindelser (10)

V121 Folk kan bare samle rikdom ved at det går ut over andre (1) Rikdom kan øke slik at det blir nok til alle (10)

Tillit til offentlige organer og organisasjoner

V136 Politiet

V137 Domstolene

V138 Regjeringen

V139 Politiske partier

V140 Stortinget

V141 Offentlige etater

(1 = svært stor, 2 = ganske stor, 3 = ikke spesielt stor, 4 = ikke i det hele tatt)

Variable i denne gruppen (unntatt V138 og V139) er studert av Fugleberg (2013)

Rettferdiggjøring: Variablene 198, 199, 201, 202, 203, 204, 205, 206, 207, 208, 209, eksempelvis

Hvor på denne skalaen vil du plassere deg selv når det gjelder å rettferdiggjøre å

V198 Å kreve offentlige goder som du ikke har rett til?

Kan aldri rettferdiggjøres (1)Kan alltid rettferdiggjøres (10)

Vi vil nedenfor gjøre dataanalyser for den norske delen av materialet, omfattende både variablene gruppevis og samlet. Det er også gjort tilsvarende analyser for noen andre land, og noen sammenligninger gjøres i begrenset omfang.

Kommentarer: Utvalget av variable kan gjøres ut fra ulike kriterier, om det er tale om å avdekke bakenforliggende forklarende variable slik en samfunns- eller adferdsforsker kan tenke seg eller foreta

klassifisering eller prediksjon, som godt kan skje på grunnlag av variable som ikke er noen bakenforliggende årsak, men kanskje samvarierer med Snyte-variabelen på grunn av en felles bakenforliggende årsak. I prinsippet kan man tenke seg at Snyte-tilbøyelighet ikke så lett innrømmes og at andre tilsynelatende mer ufarlige spørsmål svares mer sannferdig. Dette reiser to interessante spørsmål: Kan dette utnyttes på noe vis? Hva kan dette ha å si for dataanalysen og hvordan kan eventuelle skjevheter korrigeres?

3. Dataanalyser

3.1 Metodevalg

For å belyse sammenhengen mellom skattesnytevariabelen og de utvalgte variable er det mange aktuelle analysemetoder, fra enkel krysstabellering til avanserte multivariate metoder. En meget velegnet metode er statistisk tre, som håndterer ulike typer variable, manglende observasjoner, variabelutvelgelse og et forholdsvis enkelt og tolkbart resultat. Et statistisk tre lages ved en trinnvis algoritme der de gjenværende observasjonene på et trinn splittes i to grupper med henholdsvis høy og lav skår på rotvariabelen (her V200) ved hjelp av den variabelen som er mest egnet for splitting på dette trinn. Splittingen foregår så lenge som det fins variable som evner å skille i vesentlig grad. Dersom rotvariabelen er en numerisk variabel eller med minst ordinal skala kalles treet ofte et regresjonstre, mens dersom variabelen er kategorisk, typisk dikotom, kalles det et klassifikasjonstre.

Vår variabel V200 er ordinal, men gitt fordelingen av verdier gitt i tabellen ovenfor med halvparten har kodetall 1 og resten fordelt tynt over resten av skalaen, er det naturlig å dikotomisere, slik at kodetall 1 -> «Aldri snyter» og (2,3,4,5,6,7,8,9,10) -> «Mulig snyter». Vi kaller den nye variabelen Skatt, og har altså Skatt=1 -> «Aldri snyter» og Skatt=2 -> «Mulig snyter». Når ordene «Snyter», «Mulig snyter», «Aldri snyter», «Frikjent» brukes nedenfor så refererer det seg til spørsmål V200 som angår rettferdiggjøring av dette. Respondentene vil muligens oppfatte rettferdiggjøring generelt, og det vil ikke nødvendigvis bety at vedkommende selv ville snyte dersom anledningen bød seg. Dette kan imidlertid også slå den andre veien, en kunne snyte uten å ha en rettferdiggjøring.

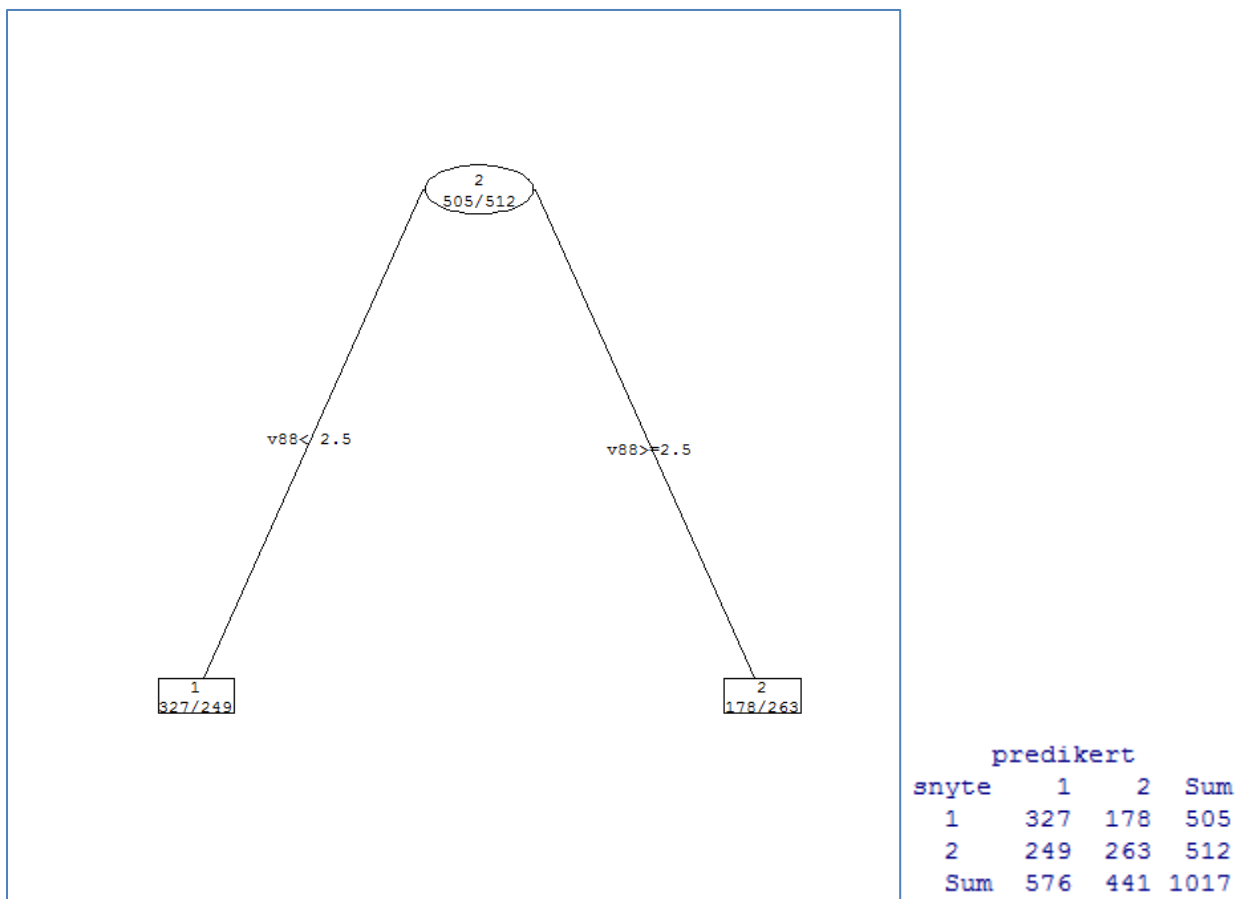
Vi har gjort omfattende analyser for Norge, for hver gruppe av variable og samlet. Sammenligninger med de fire andre landene ovenfor (Sverige, Tyskland, Italia, USA) er ønskelig, men blir vanskelig gjort ved at flere av landene har svært mange manglende observasjoner på enkelte av variablene, ikke nok observasjoner til å bygge et tre. I tilfeller der vi bare har roten av treet er det ikke opplagt om dette skyldes at det ikke er en sammenheng eller at det ikke er nok observasjoner til å avklare noe som helst.² Vi rapporterer likevel forskjeller, der vi har muligheten til det. Analysene må i hovedsak betraktes som eksplorative, med vekt på å forstå hvordan ulike spesifikasjoner manifesterer seg i ulike trær, og at disse ikke produserer urimeligheter. I den utstrekning trærne gir opphav til «gode karakteristikk» av mulige snytere må disse ikke tas altfor alvorlig.

² Det gjelder spesielt Italia, men andre aktuelle sør-Europeiske land var ikke annerledes.

I våre analyser har vi brukt programmet Rpart, som er nedlastbar som del av det fritt tilgjengelige R. Det er også gjort analyser med påbygningen rpartOrdinal, for å se om ordinaliteten kan utnyttes. Som input-parametere til Rpart kan en velge det minste antall observasjoner som kan splittes (minsplit). Velges denne stor får et enkelt tre med mindre detaljert klassifikasjon og kanskje ikke så treffsikker prediksjon. Velges denne liten kan en få et mer detaljert tre, som nok kan tilsynelatende være mer treffsikkert, men som på grunn av såkalt overtilpasning kanskje ikke gir like gode prediksjoner anvendt på nye data («out of sample»). Vi har i hovedsak brukt minsplit=200, men prøver for illustrasjonens skyld også verdiene 100 (mer detaljert) og 400 (mindre detaljert).³

3.2 Peronlighetstrekk

Det første treet er for personlighetstrekk-variablene V80-V89 og med minsplit=400 (Figur 1)



Figur 1: Norge V80-89 (minsplit=400)

Vi ser at toppnoden består av de n=1017 observasjonene, der 505 svarer til Skatt=1 og 512 til Skatt=2. Den beste første splittevariabel viste seg å være V88 («Å ta vare på naturen og verne om miljøet»), som hadde kodetall fra 1 til 6. Den beste splitteverdien (i henhold til default kriterier) viste seg å være 2.5, slik at kodetall (1,2) -> Predikert Skatt=1 (til venstre) og kodetall (3,4,5,6) -> Predikert Skatt=2 (til høyre). Her

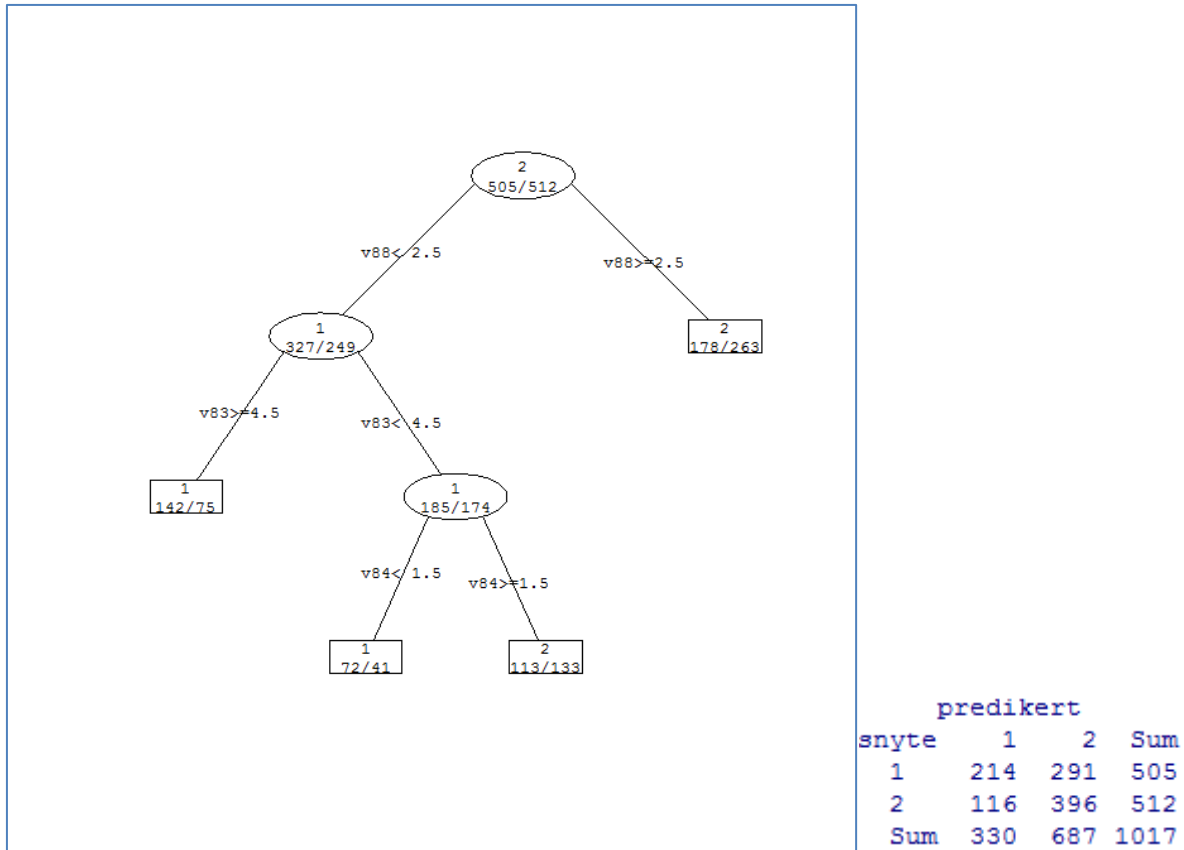
³ Samtidig innebærer det at antall som tillates i en endenode minbucket=minsplit/3

endte treet etter bare et trinn i hver sin endenode markert med henholdsvis 1 og 2 i henhold til prediksjonen. Tallene 327/249 i den venstre endenoden betyr at av de 576 som ble predikert til Skatt=1 tilhørte 327 virkelig Skatt=1, mens 249 i virkeligheten tilhørte Skatt=2. Tilsvarende betyr tallene 178/263 i den høyre endenoden betyr at av de 441 som ble predikert til Skatt=2 tilhørte 263 virkelig Skatt=2, mens 178 i virkeligheten tilhørte Skatt=1. Antallet korrekt klassifiserte blir dermed $327+263=590$, mens antall feilklassifiserte blir $249+178=427$ med sum lik 1017. I dette datasettet manglet V88 en observasjon, men likevel har vi fått fullstendig klassifikasjon. Det skyldes at algoritmen tillater bruk av substituttvariable, der man selv kan spesifisere i hvilken grad det skal tillates.⁴

Som grov oppsummering kan vi si at analysen så langt har gitt et visst grunnlag for følgende utsagn: «De som ikke avviser helt å snyte på skatten er mindre opptatt av å ta vare på naturen og verne om miljøet enn de som helt avviser å snyte på skatten».

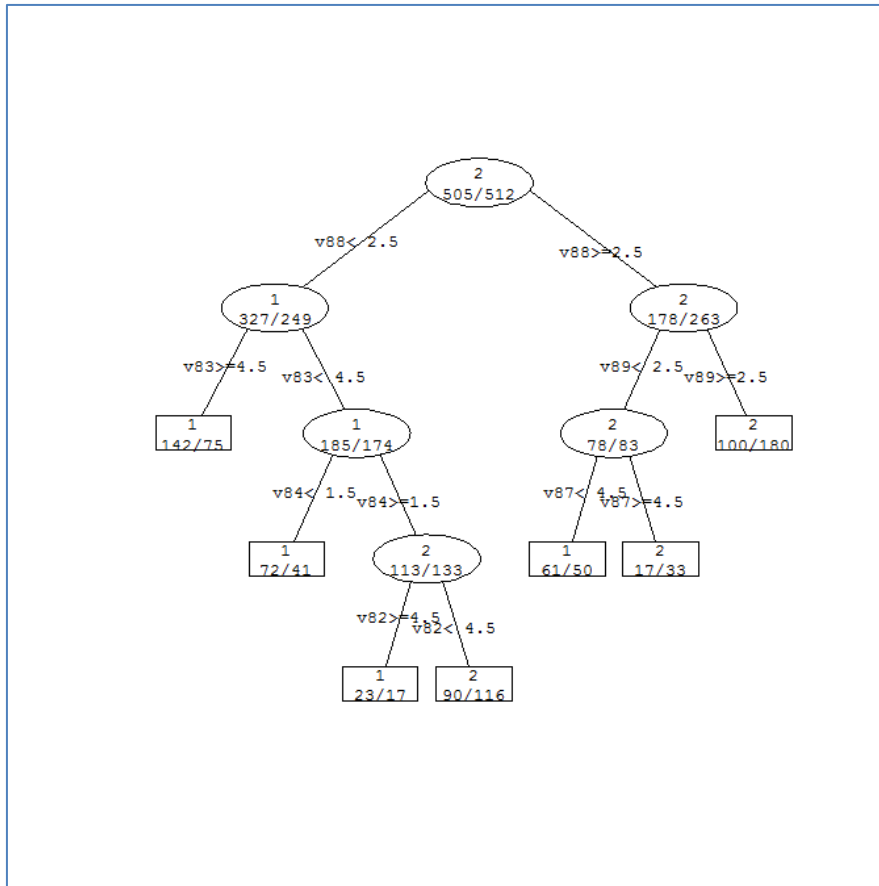
I Figur 2 har vi åpnet for flere splitt ved at de minste antall gjenværende observasjoner som tillater ny splitt er 200. Vi ser at $178+263=441$ observasjoner med høy verdi på V88 går rett til endenoden Predikert Skatt=2 som før, mens de $326+249=575$ med liten verdi på V88 blir splittet ytterligere, først mhp V83 («Å ha det gøy og skjemme bort seg selv») og deretter splittes de med liten verdi på denne variabelen mhp V84 («Å hjelpe menneskene rundt seg og sørge for at de har det bra»), der de med kodetall=1 («Veldig lik») frikjennes. Her blir antallet som predikeres korrekt lik $(142+72)+(133+263)=610$, mens antallet som predikeres feil lik $(75+41)+(112+178)=407$. Vi har altså predikert 20 flere korrekt i forhold til ovenfor. Grovt sagt har vi nå en viss dekning for utsagnet: «De som ikke avviser helt å snyte på skatten er mindre opptatt av å ta vare på naturen og verne om miljøet enn de som helt avviser å snyte på skatten, mens blant de som er opptatt av naturen og verne om miljøet finnes også noen, nemlig de som liker å ha det gøy og skjemme bort seg selv og samtidig ikke er opptatt av å hjelpe menneskene rundt seg og sørge for at de har det bra».

⁴ Vi har brukt maxcompete=4, maxsurrogate=4, usesurrogate=2.



Figur 2: Norge V80-89 (minsplit=200)

I Figur 3 har vi åpnet for ytterligere splitt (minsplit=100), og da dukker tre nye variable opp, nemlig V89, V87 og V82. Nå blir antallet som predikeres korrekt 627, mens antallet som predikeres feil 390. Vi har altså igjen økt antall korrekte, nå med 17. Det som skjer er at 61+50=111 av de 441 som gikk direkte til Predikert Skatt=2, nå blir «frikjent» ved kombinasjonen liten V89 og liten V87. Det er de som legger vekt på «å ta vare på tradisjoner og følge de skikker man har lært av familie og religion» samtidig som man legger vekt på «å alltid oppføre seg korrekt og unngå å gjøre ting som andre mener er galt». Dette gir vel ganske god mening! Vi ser også at 23+17=40 av de 112+133=245 miljøbevisste som fikk klassifikasjonen Skatt=2 også frikjennes. Det skjer for store verdier av variabelen V82 («Å leve i trygge omgivelser og unngå alt som kan være farlig»). Dette er imidlertid ikke like lett å tolke, og vi har trolig nådd grensen for hva som er forsvarlig å trekke ut av personlighetstrekkvariablene (om ikke allerede før).



Figur 3: Norge V80-89 (minsplit=100)

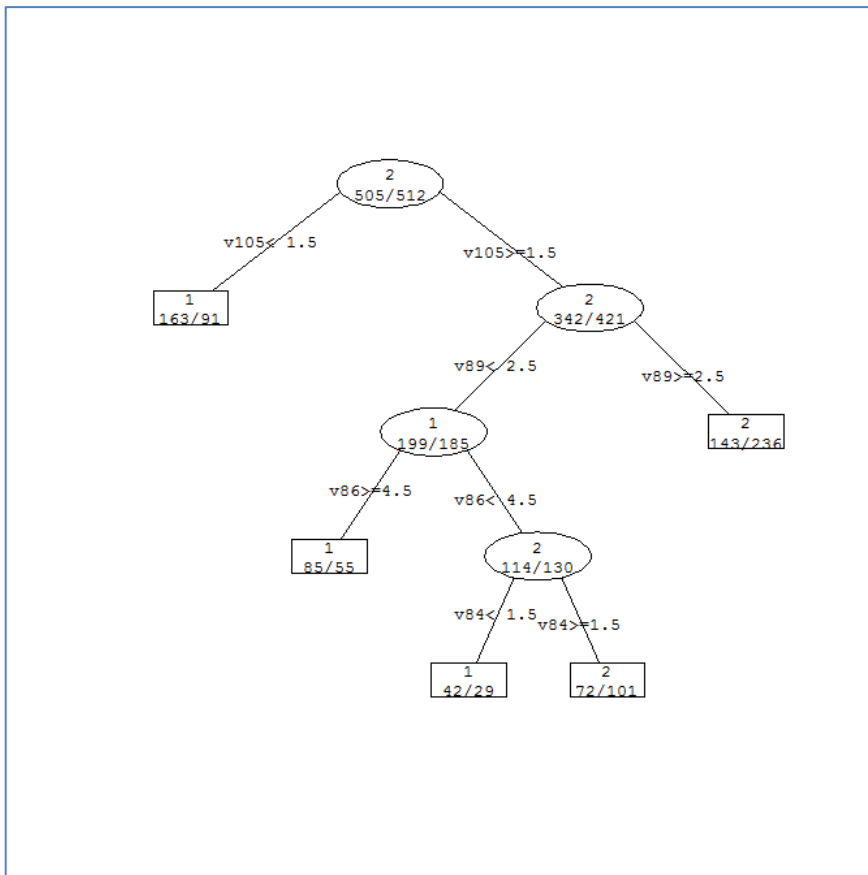
Den samme spesifikasjonen for de andre utvalgte landene er utført (i tilfellet minsplit=200) og de variable som opptrådte er oppsummert i tabellen

Land	Variable	Topp-variabel
Norge	V88+V83+V84	Å ta vare på naturen og verne om miljøet
Sverige	V86+V84+V81+V85	Å ha opplevelser, ta sjanser og leve et spennende liv
Tyskland	V84+V82	Å hjelpe menneskene rundt seg og sørge for at de har det bra
Italia	Ingen splitt/Mangler	-
USA	V84+V81	Å hjelpe menneskene rundt seg og sørge for at de har det bra

Tabell 3: Fem land: Snytte og Personlighetstrekk

3.3 Prioritering av miljø

Vi fortsetter med Norge, og siden personlighetstrekket som hadde med miljø å gjøre kom ut på topp som første splitt, er det av interesse å se hvilken rolle de konkrete miljøprioritering-variablene V105, V106 og V107 spiller, alene eller sammen med personlighetstrekkene. Spesifiserer vi (V105, V106, V107) med minsplitt=200 kommer V105 ut alene («Gå ned i lønn dersom jeg er sikker på at pengene blir brukt til å forhindre forurensing av miljøet»). Det gir 583 riktig predikerte (mindre enn med V88 alene) og 423 feilpredikerte. Spesifiserer vi V80-V89 med tillegg av (V105,V106, V107) kommer V105 på topp som erstatning for V88 etterfulgt av en litt annen utvalgskombinasjon av V80-89 variable. Dette gir 627 riktig predikerte (mot 610 uten tillegget) og 380 feilpredikerte (mot 407 uten tillegget).



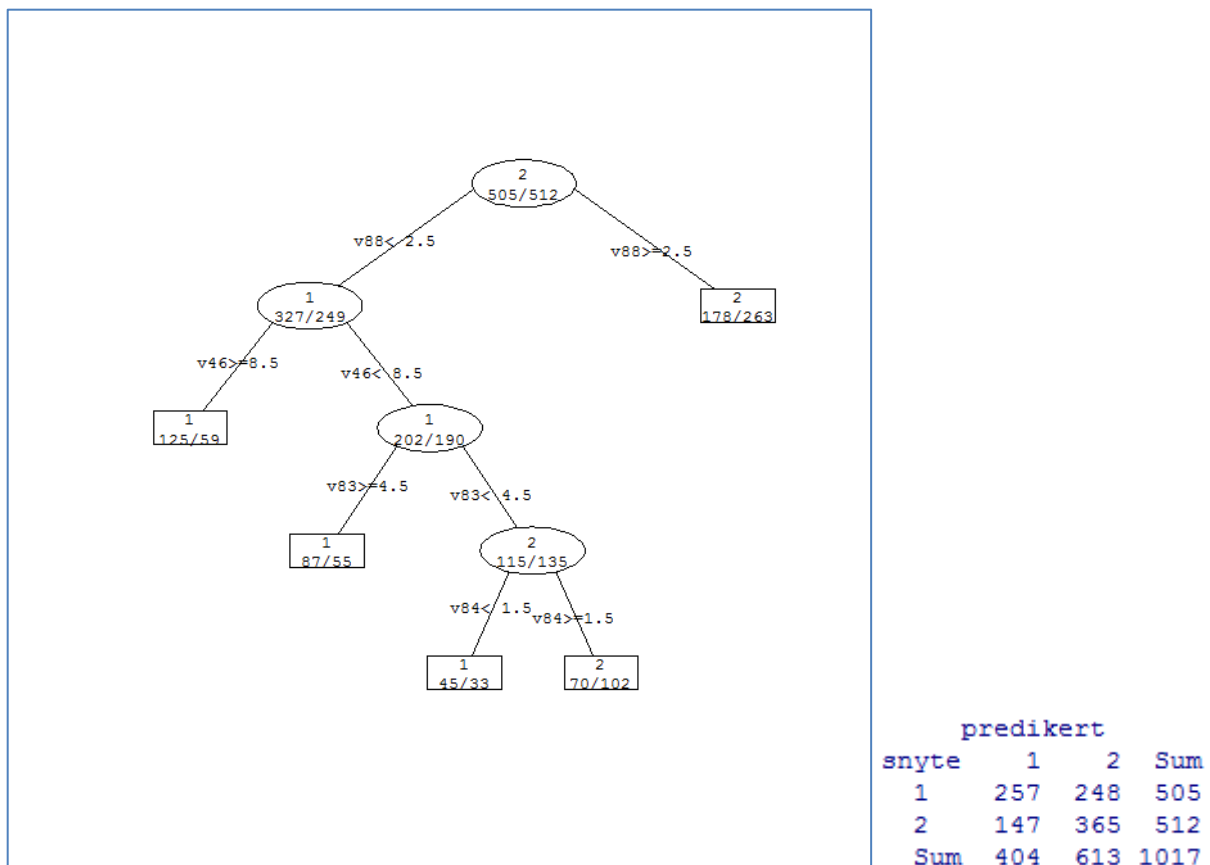
Figur 4: Norge: V80-89, V105-107 (minsplitt=200)

I den tilsvarende analyse for Sverige med de tre miljøvariablene (V105,V106, V107) kom isteden alene ut V106 («Gå med på skatteøkning dersom den ekstra skatten blir brukt til å forhindre forurensning av miljøet»), mens dersom vi spesifiserer V80-V89 med tillegg av (V105,V106, V107) kommer ingen av miljøvariablene med. For de øvrige tre landene gir disse miljøvariablene ingen splitt alene eller sammen med de andre. De må som ventet antas å være uten særlig verdi i skattemessig sammenheng.

3.4 Oppfatning av egen situasjon

Vi trekker nå inn variablene V46-47 som belyser oppfatningen av egen situasjon. Spesifiserer vi kun (V46, V47) kommer begge disse inn med V46 på topp («Hvor mye valgfrihet og selvbestemmelse mener du at du har nå det gjelder hvordan livet ditt blir»), men mønsteret er noe uryddig. Antall riktig og feilpredikerte er henholdsvis 595 og 422. Øker en minsplitt fra 200 til 300 kollapser treet til bare V46 der kodetall for mye likhet (9,10) -> SKATT=1 og kodetall for mindre likhet (1,2,3,4,5,6,7) -> SKATT=2. Antallet korrekt predikerte reduseres da til 561. For Sverige var bildet det samme, splitt først med V45 og deretter med V47. For de øvrige tre land oppnådde en ingen splitt.

Tar vi utgangspunkt i vår tidligere spesifikasjon (V80, V81, V82, V83, V84, V84, V86, V87, V88, V89) og legger til (V46, V47) får vi i Figur 5 et lignende tre som ovenfor, der V46 blir skutt inn på annet trinn, og der de miljøbevisste (lav V88) som føler stor grad av valgfrihet straks blir «frikjent», mens de som i mindre grad føler det, må gå gjennom et par kriterier til angående V83 og V84. Antall korrekte prediksjoner er nå 622, mens antall feilaktige er 395. Vi ser at det er omtrent på samme nivå som når V80-89 ble spesifisert sammen med V105-107 og V105 kom ut på topp som erstatning for V88.

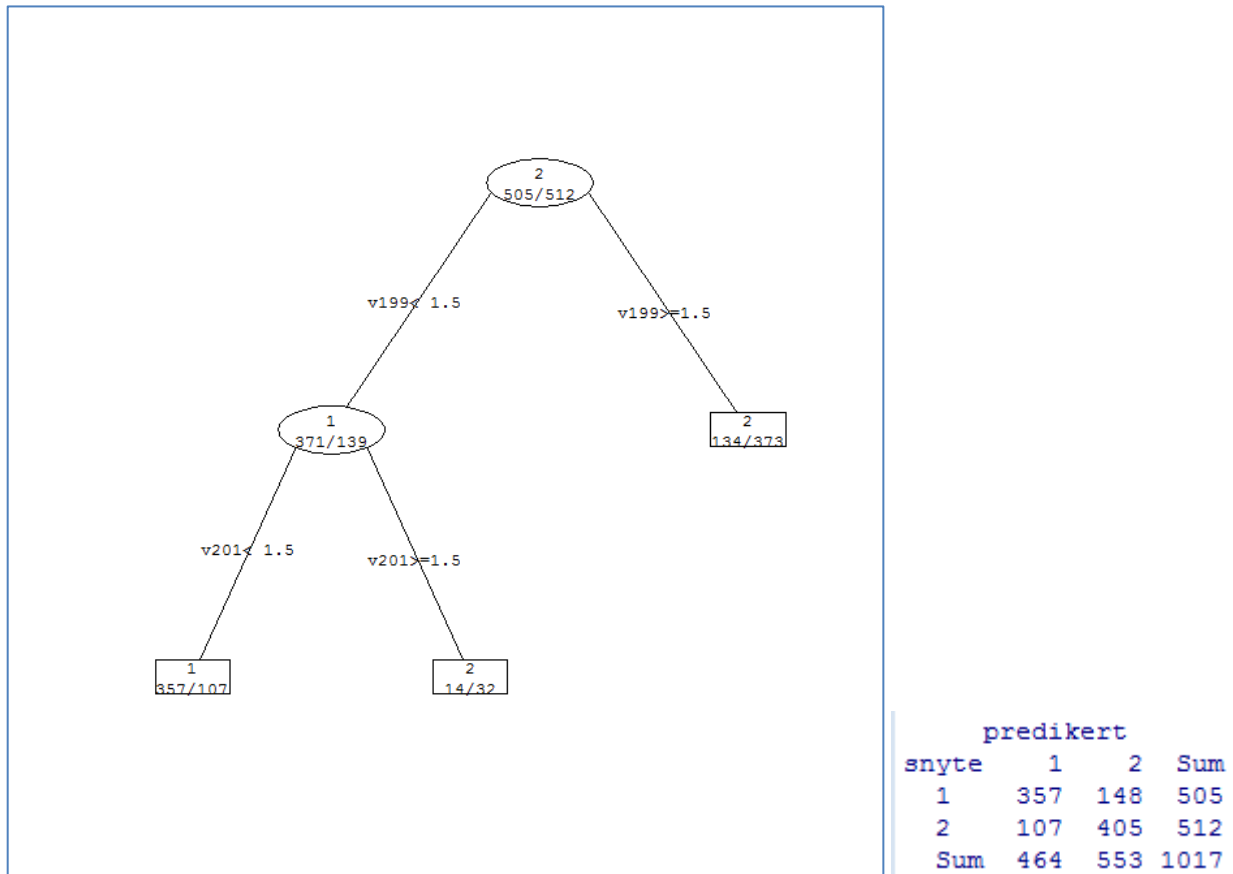


Figur 5: Norge: V80-89, V46-47 (minsplitt=200)

Spesifiserer vi alle tre gruppene (V80-89), (V105-107) og (V46-47) kommer den siste gruppen ikke med og resultatet blir det samme som i Figur 4.

3.5 Rettferdiggjøring av handlinger

La oss se på variabelgruppen V198-208 som har med rettferdiggjøring av handlinger, der vår hovedvariabel V200 er med. Den eneste variabelen som splitter er V199 («Å la være å betale på kollektive transportmidler»), der kodetall 1 -> SKATT=1 og Kodetall>1 -> SKATT=2. Antall riktig predikerte er nå 744 og antall feilpredikerte er 273, altså en god del bedre enn tidligere spesifikasjoner.



Figur 6: Norge: V198-208 unntatt V200 (minsplit=100)

Utvider en treet (sett minsplit=100) kommer også V201 med («At noen tar imot bestiktelser på jobben»), slik at det bare er de som svarer «Kan aldri rettferdiggjøres» på begge spørsmål som «frikjennes». Antall riktige prediksjoner blir nå 762, mens antall feilprediksjoner blir 255, det beste oppnådd hittil.

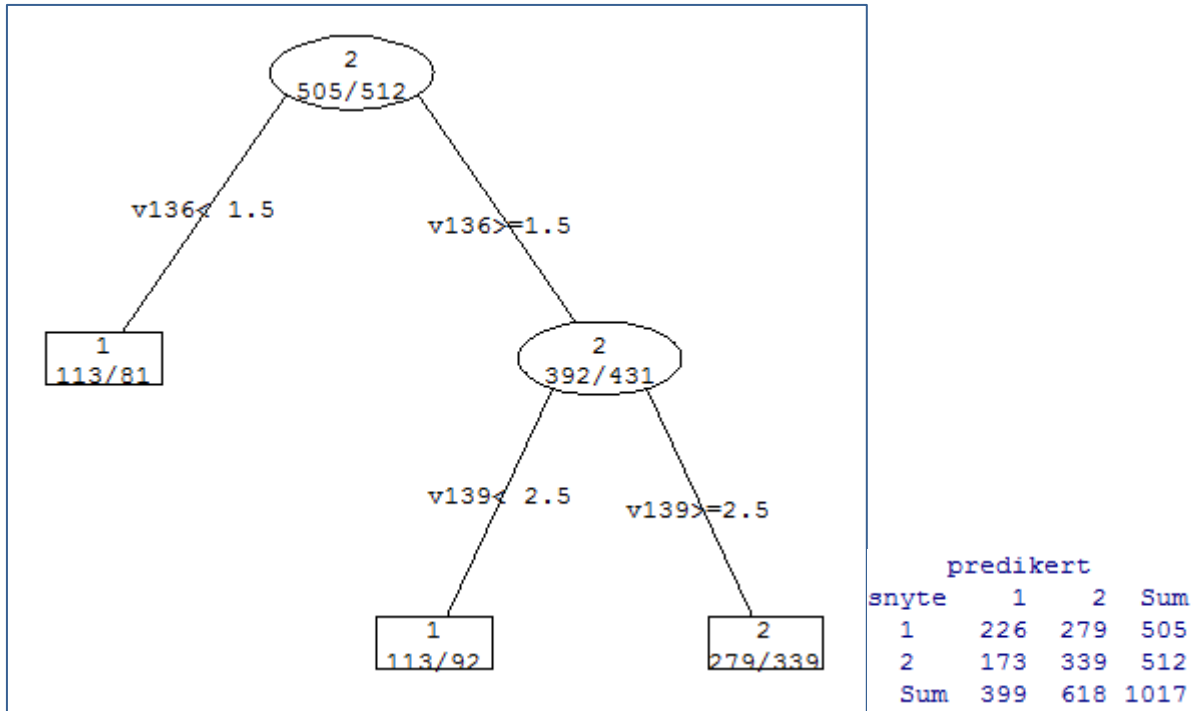
En sammenligning av de fem landene ga følgende (minsplit=200) nokså ensartede bilde:

Land	Variable	Topp-variabel
Norge	V199	Å la være å betale på kollektive transportmidler
Sverige	V199+V201+V203	Å la være å betale på kollektive transportmidler
Tyskland	V199	Å la være å betale på kollektive transportmidler
Italia	V199	Å la være å betale på kollektive transportmidler
USA	V201+V199	At noen tar imot bestiktelser på jobben sin

Tabell 4: Fem land: Snyte og Rettferdiggjøring av handlinger

3.6 Tillit til offentlige organer

Vi har hittil holdt variabelgruppen om tillit til offentlige organer og organisasjoner utenfor, dvs. politiet (V136), domstolene (V137), regjeringen (V138), politiske partier (V139), Stortinget (V140) og offentlige etater (V141). Vi vil her også se på denne gruppen og samtidig gjøre noen sammenligninger og betraktninger rundt de rapporterte analyser i Fugleberg (2013) som er på ikke dikotomiserte data. Vår dikotomiseringen for denne gruppen av variable er: 1 = svært stor tillit, (2=ganske stor, 3=ikke spesielt stor, 4=ikke tillit i det hele tatt) -> 2. Treet for minsplit=100 er gitt i Figur 7.



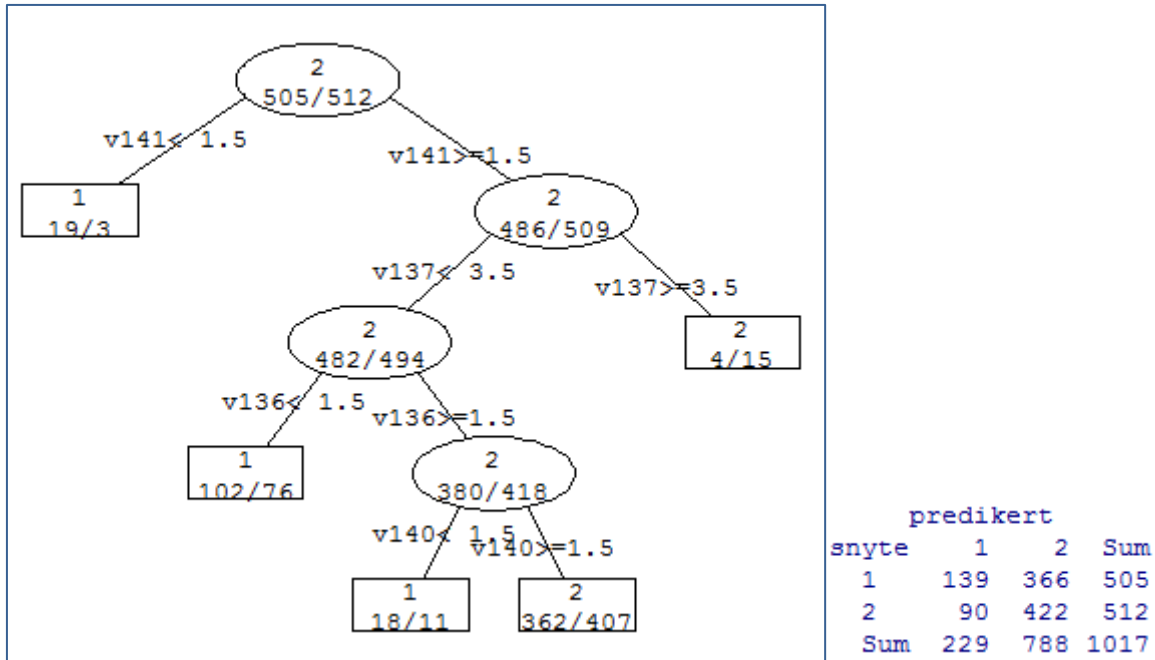
Figur 7: Norge: V136-141 (minsplit=100)

Vi ser at V136 (graden av tillit til politiet) er den av disse variablene som splitter best, og slik at i første omgang bare de med svært stor tillit klassifiseres som «ikke snytere». De øvrige splittes så med V139 (grad av tillit til de politiske partier), og slik at både de med svært stor tillit og ganske stor tillit klassifiseres som «ikke snytere». Vi ser at blant de 279 + 339 = 618 som fortsatt er «under mistanke» er det likevel en stor andel $279/618=45\%$ som sier at snyting aldri kan rettferdiggjøres. Andelen korrekte klassifikasjoner er $(226+339)/1017=55.5\%$, altså ikke mye å skryte av.

For Sverige kom V136 (politiet) ut alene, mens Tyskland hadde som Norge V136 (politiet) og deretter V139 (de politiske partier). De to andre landene ga ingen splitt.

Fugleberg (2013) studerer fire av variablene i denne gruppen (med unntak av V138, V139) hver for seg og uten å dikotomisere. Konklusjonen hans er at jo mer tillit folk har til Stortinget, forvaltningen, rettsvesen og politi, jo høyere skattemoral har de. Denne konklusjonen er det nok et visst belegg for, og står ikke i motstrid med ovenstående. Kjører vi analysen med hans fire variable (V136, V137, V140, V141) får vi imidlertid kun en splitt, den første i Figur 7 ovenfor. Andelen korrente klassifikasjoner er da 53.4%, ikke så mye bedre enn å knipse kron/mynt.

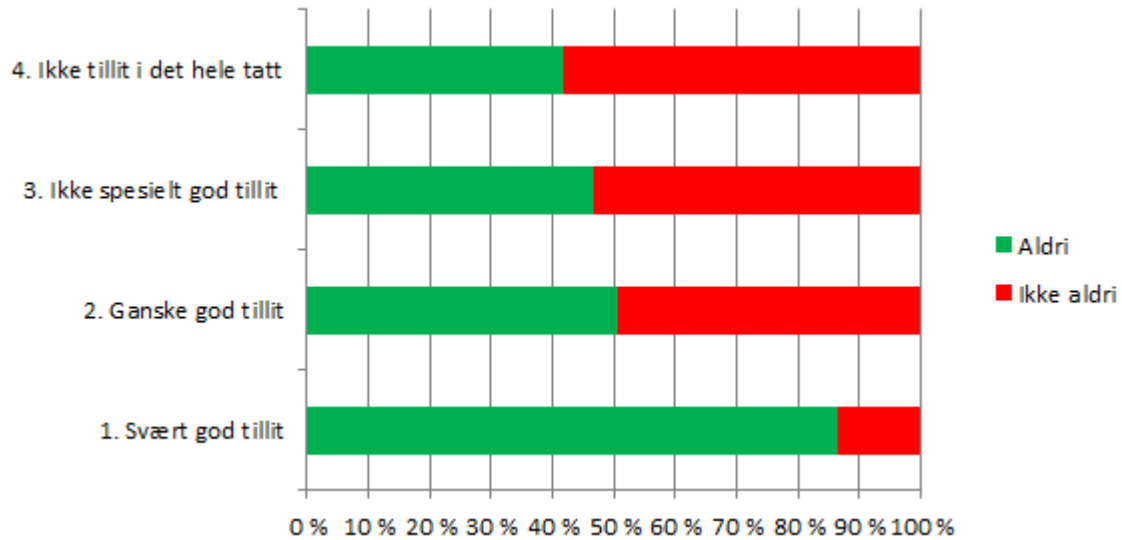
Svarene på spørsmålene i denne gruppen gir nok uttrykk for en generell grunnholdning, der hvert spørsmål er en indikator på denne. At ikke flere variable kommer med skyldes selvfølgelig at de andre variablene er sterkt korrelerte med de to som er med og dermed ikke forbedrer separasjonen. For å kunne sammenligne med Fugleberg sine resultater kan det likevel ha en viss interesse å prøve en utvidet splitt ved å sette minsplit=50. Da får vi et mer komplisert bilde:



Figur 8: Norge: V136-141 (minsplit=50)

Her er første splitt isteden med V141 (grad av tillit til offentlige etater) som frikjenner et svært lite antall. De øvrige splittes så på V137 (grad av tillit til domstolene), der et lite antall med ingen tro på disse blir mistenkeliggjort. De som fortsatt ikke er mistenkeliggjort splittes så med V136 (graden av tillit til politiet), der de med svært stor tillit frikjennes. Blant de uten svært god tillit til politiet frikjennes de med svært stor og ganske stor tillit til Stortinget. Andelen korrekte klassifiseringer $(139+422)/1017=55\%$, altså bare litt bedre enn å knipse kron/mynt. Vi ser at vi ikke har oppnådd noe med omsyn til andelen korrekte klassifiseringer. Når en tar omsyn til muligheten for overtilpasning, er det ikke tilrådelig å legge vekt på mulige fortolkninger av de utvidede analysene med sikte på klassifikasjoner av nye data med ukjent gruppetilhørighet.

Merk at den måten resultatene til Fugleberg (2013) er rapportert på hans Figur 1 kan gi inntrykk av en klar sammenheng. Her er for hver av de fire spørsmålsvariablene vist andelen som på spørsmålet om rettferdiggjøring av unndragelse svarer Aldri/Ikke aldri, for hver av de fire svarkategoriene. Dette er gjenskap for variabelen V141 (offentlige etater) som var mest iøynefallende.



Figur 9: Tillit til offentlige etater (V141).

Dette gir inntrykk av vi kan få ganske god separasjon for de to yttergruppene, spesielt vil de fleste av de som har svært god tillitt til den offentlige administrasjon korrekt bli satt utenfor mistanke. For de to midtkategoriene er fordelingen omtrent 50-50. Imidlertid spiller det her en avgjørende rolle hvor mange som er i de fire svarkategoriene. Dette var

v141					
evasion	1	2	3	4	Sum
1	19	303	172	10	504
2	3	297	196	14	510
Sum	22	600	368	24	1014

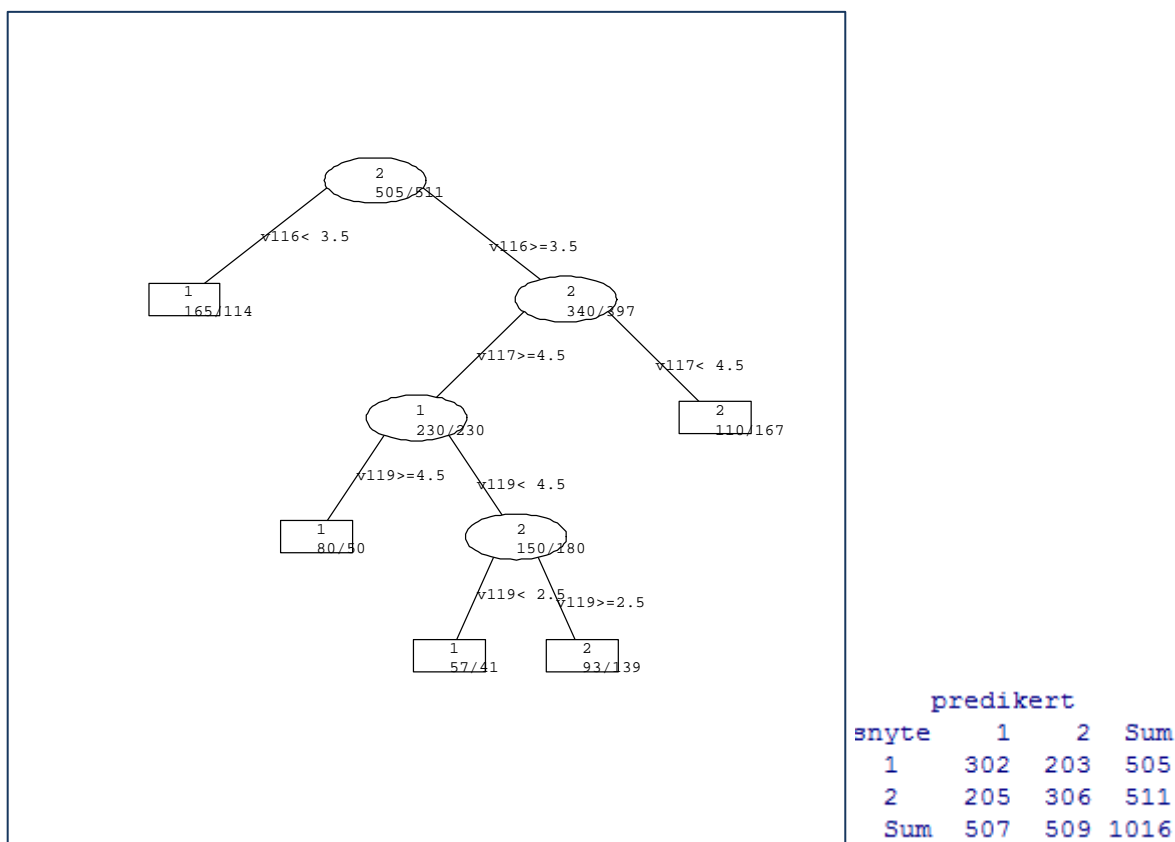
Tabell 5: Norge: Snyte og Tillit til offentlige etater (V141)

Herav ser vi at bare 4.5% av respondentene er i de to ytterkategoriene og hele 95.5% i de to midtkategoriene. Figuren gir derfor et falskt bilde av hva det er mulig å oppnå ved en klassifiseringsregel på dette grunnlag. Dette har vi allerede konkludert med ovenfor.

3.7 Sosio-økonomiske synspunkter

Vi ser så på gruppen av sosio-økonomiske variable (V116-V121), der det viser seg at det er variablene som angår økonomiske aspekter som i første rekke trer frem. I Figur 10 (minsplit=200) er første splitt mhp V116 (holdningen til inntektslikhet versus inntektsulikhet som incitament), der som ventet de mest egalitære frikjennes først. De gjenværende splittes etter V117 (favorisering av privat versus offentlig eiendomsrett til næringsliv og industri), der de første «mistenkeliggjøres». De som fortsatt ikke er klassifisert, splittes to ganger etter V119 (konkurransen sunt versus skadelig som incitament), og slik at det er personer på de to ytterpunktene på skalaen som frikjennes, uten at vi skal prøve å fortolke det

her. Dersom en setter minsplit=400 forsvinner V119, men isteden gjentas V116 som da reflekterer samme type ikke-monotonitet. Med minsplit=50 kommer V121 inn på annet nivå i treet som da frikjenner et mindre antall som mener at rikdommen kan øke, slik at det blir nok til alle, men her har en nok igjen gått for langt i retning av overtilpasning.



Figur 10: Norge: V116-121 (minsplit=200).

Sammenligningen mellom de fem landene ga som resultat:

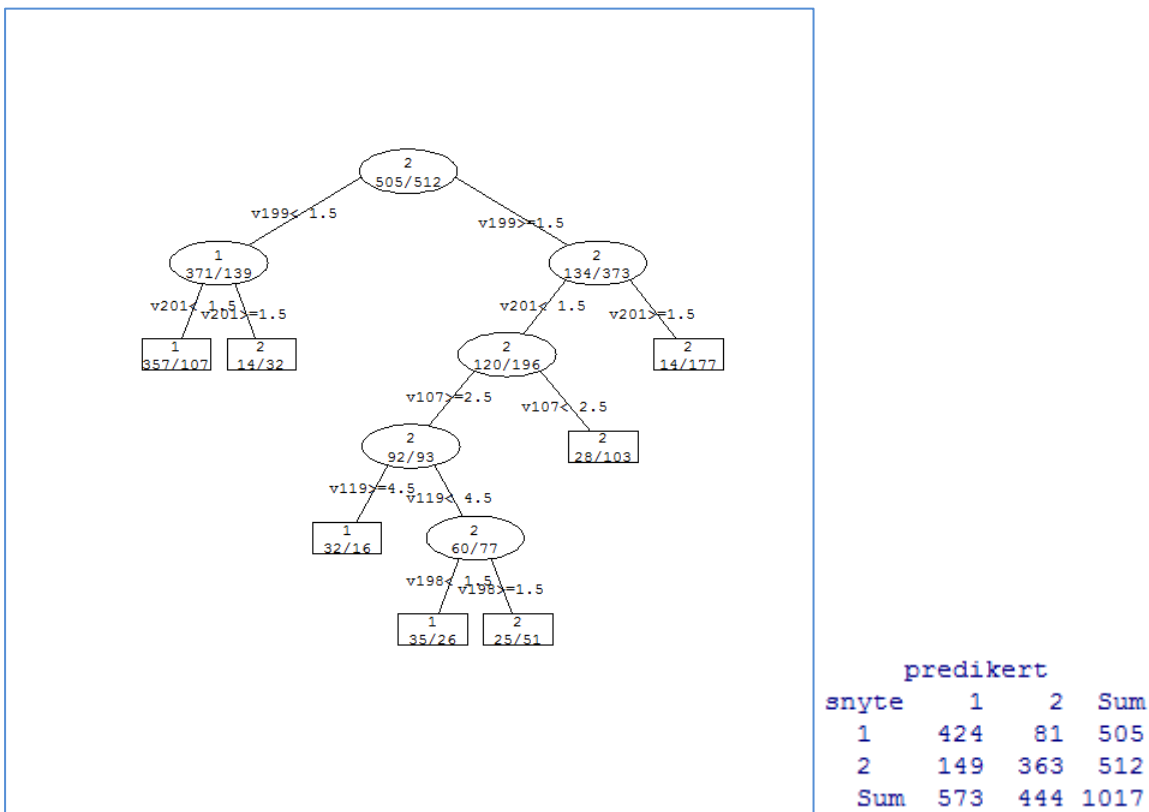
Land	Variable	Topp-variabel: Holdning til
Norge	V116+V117+V119	Inntektsforskjeller som belønning eller ikke
Sverige	V117+V121	Eiendomsrett: Privat versus offentlig eiendom
Tyskland	V119+V121+V118	Konkurransen som incitament eller ikke
Italia	Ingensplitt/Mangler	-
USA	V119+V118+V117	Konkurransen som incitament eller ikke

Tabell 6 Fem land: Snyte og Sosio- økonomiske synspunkter

Dette bildet med sin toppvariabel og sine sekundærvariable stemmer bra med rådende stereotyper om befolkningen i de ulike landene.

3.8 Samlet spesifikasjon

Til slutt ser vi på de seks variabelgruppene V46-47, V80-89, V105-107, V116-V121, V136-141, V198-208 samlet. Det viser seg igjen at V199 er den eneste variabelen som splitter (med minsplitt=200). Treet er da som i Figur 6, med siste forgrening skåret bort. Det innebærer at denne variabelen er såpass sterkt samvarierende med SKATT (dikotomisert V200) at ingen av de andre har noe vesentlig å tilføre. Åpner en for utvidelse (med minsplitt=100) får vi treet i Figur 11.



Figur 11: Norge: V46-47, V80-89, V105-107, V116-121, V136-141, V198-208 (minsplitt=100).

Her ser vi at V201 igjen kommer inn etter V199, og videre V107 («Myndighetene bør redusere forurensningen av miljøet, men det må ikke koste meg noe»), V119 (Holdning til konkurranse som incitament), V198 («Å kreve offentlige goder du ikke har rett til»). Vi merker oss spesielt at ingen variable fra gruppen Tillit til offentlige organer er kommet med, og som er grunnlaget for Fuglebergs artikkel om hva som former skattemoralen i Norge. Her blir 787 predikert korrekt, mens 230 blir feilpredikert mot bare ca. 450 dersom en kun bruker variablene fra gruppen om tillit til de offentlige

organene. Dette er det beste resultat oppnådd, men det er her en betydelig fare for at vi overtilpasser. De fortolkninger som byr seg fram i Figur 11 behøver altså ikke ha rot i virkeligheten!

Alle foretatte spesifikasjoner ovenfor, gruppevis og samlet, er også gjennomført for Norge med tillegg av V114, som gikk på politisk ståsted fra venstreside mot høyreside. Variabelen viste seg ikke noe sted, slik at trærne ble identiske med de allerede presenterte. Dette var for så vidt et interessant funn, men som gjerne kan studeres noe nærmere, f.eks om de forekommer i de nest beste trær eller se på forklaringen av denne variabel med de som vi har brukt ovenfor.

Her er sammenstillingen av splittingen i nivårekkefølge for de fem landene, der også variabelen om politisk ståsted (V114) tatt med. Tabellen viser minsplitt=100 og minsplitt=200 markert med fete typer.

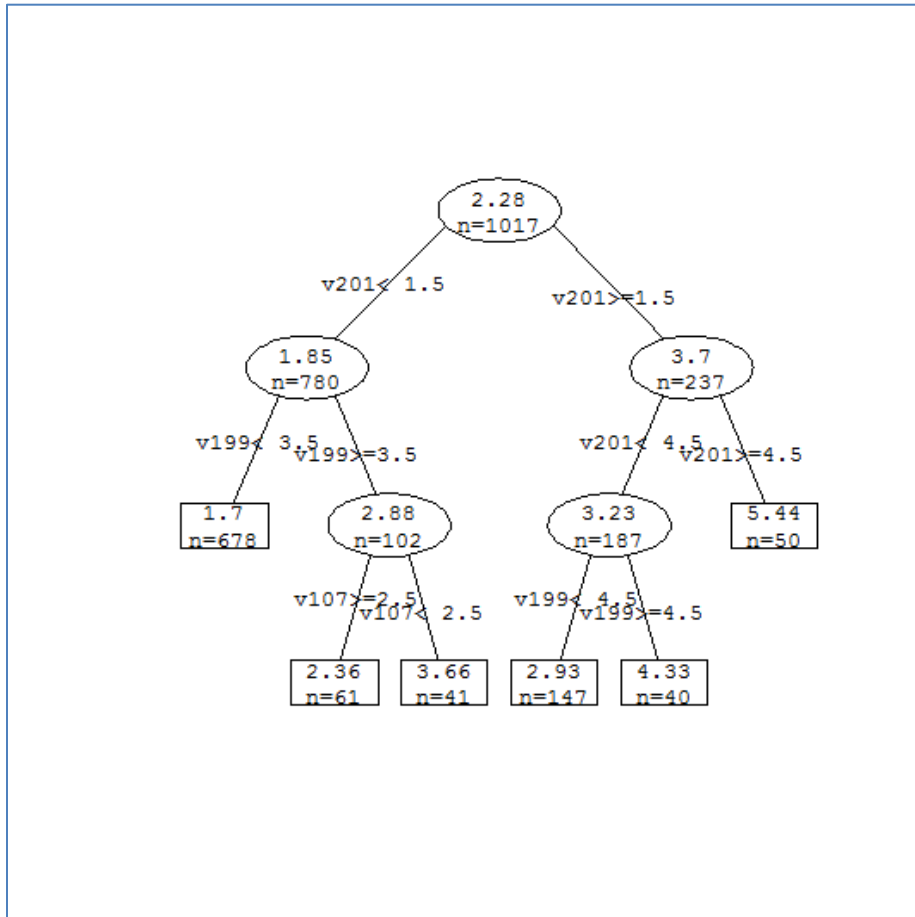
Land	Variable	Topp-variabel:
Norge	V199+V201+V107+V119+V198	Å la være å betale på kollektiv transport
Sverige	V199+V201+V46+V203	Å la være å betale på kollektivtransport
Tyskland	V199+V201+V203	Å la være å betale på kollektivtransport
Italia	V199+V198+V114	Å la være å betale på kollektivtransport
USA	V201+V199+V205+V46+V107	At noen tar imot bestillinger på jobben

Tabell 7: Fem land: Snyte og Alle variable

Hovedinntrykket ellers er store likheter mellom landene og god logisk konsistens, der V199 den dominerende. Forskjellene vil kunne fremstå som enda mindre, dersom en også hadde listet de nest beste trærne for hvert land. Husk at rapporteringen av et enkelt tre kan skjule at det er en portefølje av trær som er nesten like gode. Tabellen viser noen få spesielle trekk som fortjener en kommentar: Variabelen V114 (om politisk ståsted) dukket opp for Italia for minsplitt=100, der venstresiden «frikjennes». Variabelen V46 (om valgfrihet og selvbestemmelse) er kommet med for USA, der de som føler de har dette i stor grad frikjennes (et riktignok lite antall). Variabelen V107 er kommet med for både Norge og USA, men for Norge frikjennes de som er mest enig i at «Myndighetene bør redusere forurensningen av miljøet, men det må ikke koste meg noe», mens for USA er de som er mest uenig som frikjennes, riktignok i et lite antall.

Vi har også foretatt en del tilsvarende analyser som ovenfor, der vi ikke har dikotomisert V200. Dette kan gjøres i rpart slik at det istedenfor et klassifikasjonstre er et regresjonstre eller med tillegget rpartOrdinal, som gir et klassifikasjonstre, men tar omsyn til at kriterievariabelen (V200) er ordinal. Selv om vår V200 er ordinal og ikke intervallskalert spiller det antagelig liten rolle hva vi bruker, men sammenligninger kan være teoretisk interessante.

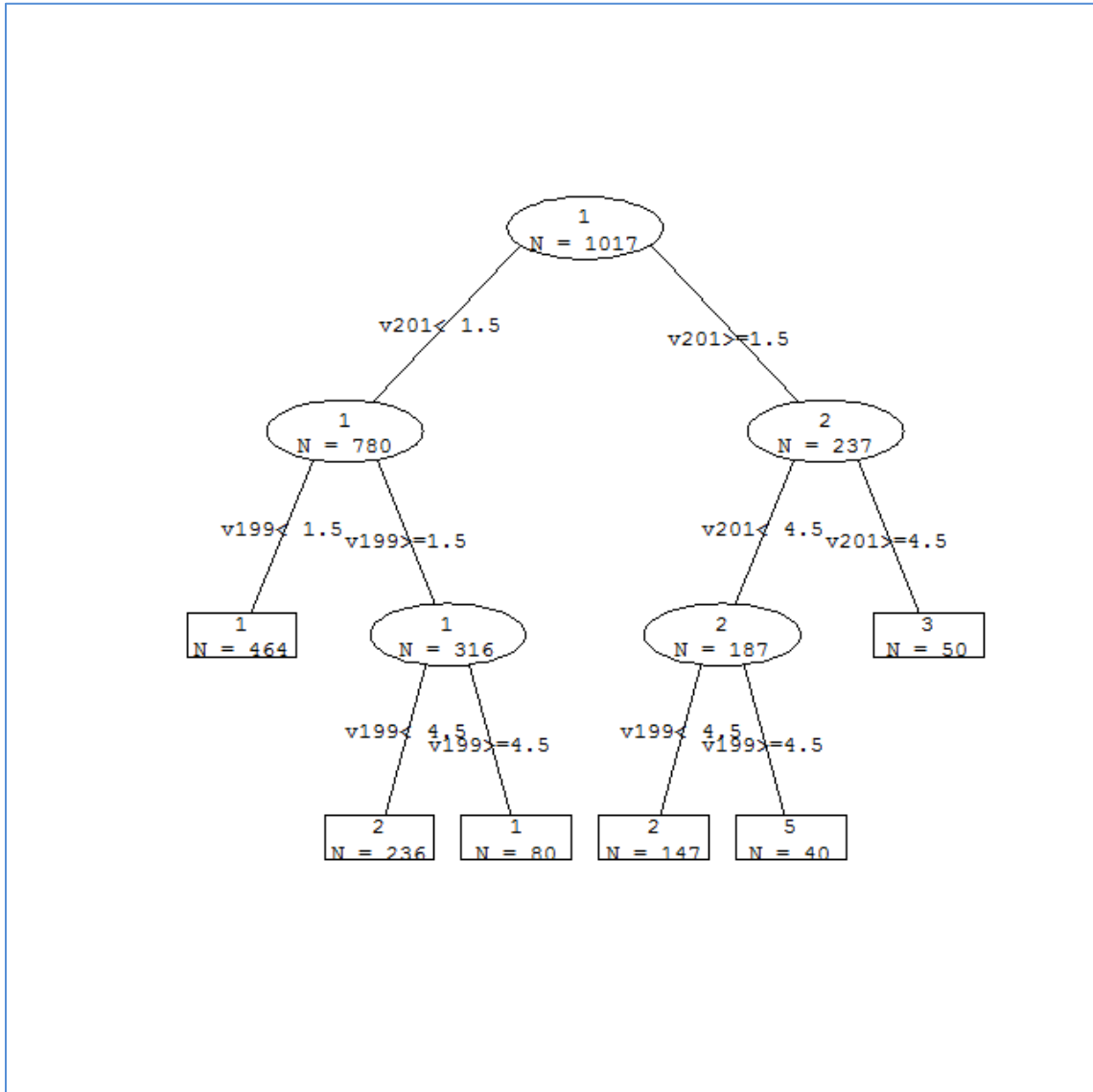
Hovedinntrykket med regresjonstre med rpart og ordinalt klassifikasjonstre med rpartOrdinal er at de samme variable kommer inn på de øverste nivåene, men noen ganger i motsatt rekkefølge, samt at noen variable lenger nede kan være byttet ut. Eksempelvis spesifikasjonen V80-89 (i tilfellene minsplitt=100 og 200) at V83 → (V88, V82), mens vi for den dikotomiserte variabelen ovenfor hadde V88 → V83 → V84. For spesifikasjonen med alle fem grupper variable ble resultatet for rpart og rpartOrdinal (i tilfellet minsplitt=200) at V201 → V199, mens det var kun V199 som splittet i klassifikasjonstreet. I tilfellet minsplitt=100 ble resultatene som i Figur 12 og Figur 13 med V201 på topp.



Figur 12: Regresjonstre Norge: V46-47, V80-89, V105-107, V136-141, V198-208 (minsplitt=100).

I regresjonstreet i Figur 12 er desimaltallet i noden gjennomsnittlig skårverdi på V200 for observasjonene som er igjen i noden, mens i Figur 13 er heltallet i noden klassifikasjonen i en av de ti ordinale kategoriene (etter nummereringen fra 1 til 10). Vi ser at de tre variablene V201, V199, V107 her også er med ovenfor i begge trær og at de betingede klassifikasjonsreglene som følger derav gir god formell logisk mening. Vi ser at det ordinale treet kun inneholder rettferdiggjøringsvariable, mens regresjonstreet også har med miljøprioriteringvariabelen V107. Klassifikasjonsreglene kan imidlertid

ofte lyde nokså sære, som i Figur 12, med en sekvens av utsagn som angår der bestiktelser på jobb, «snyte på buss og miljøprioritering.



Figur 13: Ordinalt tre Norge: V46-47, V80-89, V105-107, V136-141, V198-208 (minsplit=100).

Det interessante spørsmålet for oss er om man har tapt noe i prediktiv forstand ved å dikotomisere. La oss først liste opp antallene i hver predikert kategori:

Prediksjon	1	2	3	4	5	6	7	8	9	10	Sum
Antall	544	383	50	0	40	0	0	0	0	0	1017

Tabell 8: Norge: Ordinal Snyteklassifikasjon

Vi ser at metoden har endt opp med bare fire kategorier, og ikke «våget» å plassere noen i de øverste kategoriene. Det er ikke rart, når vi hadde svært få slike blant observasjonene selv (men aner vi her en slags regresjon mot gjennomsnittet). Den ordinale metoden har en åpenbar fordel fremfor dikotomisering ved at den gir anledning til å prioritere eventuelle undersøkelser fra toppen. Her blir altså de i kategori 5 prioritert fremfor de i kategori 3, som i sin tur kommer foran de i kategori 2.

Her følger en tabell som viser de observerte antall i de ti kategoriene sammen med hvor mange som blir klassifisert i hver av kategoriene:

Pred\Obs	1	2	3	4	5	6	7	8	9	10	Sum
1	385	55	40	13	32	6	3	4	1	5	544
2	116	156	59	17	11	8	7	6	2	1	383
3	4	2	8	6	7	5	5	5	4	4	50
5	0	8	9	4	10	4	2	2	0	1	40
Sum	505	221	116	40	60	23	17	17	7	11	1017

Tabell 9: Norge: Ordinal Snyteklassifikasjon vs. Observervert

Vi ser blant annet at noen observerte i toppen av skalaen blir klassifisert til bunnen, men i lite antall. For å sammenligne prediksjonsevnen med vår tidligere dikotomiserte analyse kan vi dikotomisere denne tabellen og får:

Predikert\Observervert	Snyte=1	Snyte>1	Sum
Snyte = 1	385	159	544
Snyte>1	120	353	473
Sum	505	512	1017

Tabell 10: Norge: Ordinal Snyteklassifikasjon vs. Observervert (dikotomisert)

Her blir andelen korrekt predikerte $(385+353)/1017=73\%$, mens andelen feilpredikerte blir 27%. I den tilsvarende dikotomiserte analysen var andelen korrekt predikerte $(424+363)/1017=77\%$ (se Figur 11). Det kan altså se ut en ordinal løsning som gir anledning til å prioritere fra toppen krever en pris.

4. Oppsummering

Vi har i denne artikkel analysert holdninger til skatteunndragelse i Norge. Hensikten med dette har vært å

- a. Få innsikt i variable som samvarierer, dvs. rent eksplorativ
- b. Få innsikt i prediksjonsmuligheter

Gjennom bruk av data fra verdiundersøkelsen i 2007 har vi med hjelp av statistiske trær avdekket variabler som ser ut til å samvariere med holdning til skatteunndragelse og i tillegg studert hvordan godt dette kan bidra med prediksjon av holdning.

Vår analyse indikerer at det er sterkest samvariasjon mellom holdninger til skatteunndragelse og rettferdiggjøring av andre handlinger, slik som å la være å betale for kollektivtransport. I tillegg markerer seg holdninger til natur og miljø i noen grad som en mulighet for prediksjon av holdninger til skatteunndragelse.

4 Litteratur

Fugleberg, O. (2013) Hva former skattemoral i Norge? Analysenytt. Skattedirektoratet.

Torgler, B. (2003) Tax Morale: Theory and Empirical Analysis of Tax Compliance. Dissertation.

Torgler, B. (2003) To evade taxes or not to evade: that is the question, *The Journal of Socio-Economics*, Elsevier, vol. 32(3), 283-302.

Alm, J. & Torgler, B. (2004) Culture differences and tax morale in the United States and in Europe, *CREMA Working Paper Series 2004-14*, Center for Research in Economics, Management and the Arts (CREMA).

Torgler, B. (2007) *Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*. Edward Elgar Publishing Inc. (bok basert på egne artikler fra perioden 2003-2007)

Alm, J. & Torgler, B. (2011) Do ethics matter? Tax compliance and morality, *Journal of Business Ethics*, Springer, vol. 101(4), 635-651.

Torgler, B. (2011) Tax morale and compliance: review of evidence and case studies for Europe, *Policy Research Working Paper Series 5922*, The World Bank.

Rpart: Recursive partitioning and regression trees.

<http://cran.rproject.org/web/packages/rpart/index.html>

RpartOrdinal: Ordinal classification tree functions

<http://cran.open-source-solution.org/web/packages/rpartOrdinal/index.html>

**PUBLICATIONS FROM THE NORWEGIAN CENTER FOR TAXATION (NoCeT)
2012 -**

- Jostein Lillestøl
Jonas Andersson
Snyte skatt hvis mulig?
Hva fortalte Verdiundersøkelsen 2007?
SNF Working Paper No 03/14
- Armando J. G. Pires
The business model of The British Virgin Islands and Panama
SNF Working Paper No. 31/13
- Daniel Jäck
The business commitment of a Norwegian company in Germany
SNF Working Paper No 17/13
- Chris Edson
The Capital Constraining Effects of the Norwegian Wealth Tax
SNF Working Paper No 41/12
- Ida Selseth
Norges selskapsbeskatning: Globaliseringsutfordringer og EUs forslag om CCCTB
SNF Arbeidsnotat Nr. 37/12
- Jonas Andersson
Jostein Lillestøl
Bård Støve
Kjennetegnsanalyser av skattytere som unndrar skatt ved å skjule formuer og inntekter i utlandet
SNF Report No 10/12
- Rasmus Møller
Erlend Andreas Nordal
FoU og overskuddsflytting i Norge
SNF Working Paper No 36/12

Denne rapporten er basert på data fra Verdiundersøkelsen 2007, som er foreløpig Norges siste bidrag til World Values Survey (WVS). Undersøkelsen omfatter følgende spørsmål om holdningen til skattesnyteri: «Hvor på denne skalaen (1 til 10) vil du plassere deg selv når det gjelder å rettferdiggjøre å snyte på skatten dersom du har mulighet til det?». I rapporten studeres dette i sammenheng med en rekke grupper av variable: Oppfatning av egen situasjon, Personlighetstrekk, Prioritering av miljø, Politisk ståsted, Sosio-økonomiske synspunkter, Tillit til offentlige organer og organisasjoner, Rettferdiggjøring av andre tvilsomme handlinger. Ved hjelp av såkalte klassifikasjonstrær (CART) avdekkes hvilke variable som er best egnet som prediktorer innen hver gruppe og samlet sett.

SNF



Samfunns- og næringslivsforskning AS

Centre for Applied Research at NHH

Helleveien 30
NO-5045 Bergen
Norway

P +47 55 95 95 00

E snf@snf.no

W snf.no

Trykk: Allkopi Bergen