Metadata

Min-Yang Lee

April 09, 2018

Contents

This has a purpose and a target audience. I'm not quite sure what that is yet. One of the audiences is "future me."

Dealer Databases

Tables: CFDETSyyyy; CFDERSyyyy

Location: Sole

Schema: CFDBS

Overview

The dealer data are transaction-level pricing at the level of the "market-category." These data are primarily generated through mandatory reporting by federally-permitted fish dealers. The federal reporting is supplemented with data from non-federally-permitted (state-only) fish dealers. Data are currently reported electronically in partnership with ACCSP through SAFIS.

- CFDETSyyyy contains "detailed species data" for 1994-2003
- CFDERSyyyy contains "detailed species data" for 2004-present

Current Collection Methods

These data are the result of mandatory federal dealer reporting at the "trip-level", supplemented by state-level, aggregated reporting. Federally permitted fish dealers that are required to report purchases of all fish to NMFS.

Changes to Collections Methods

- The number of species triggering these requirements have increased over time, which has implications for completeness (50 CFR 648.6). For example, mandatory dealer reporting for Monkfish, herring, and hagfish began in 1999, 2001, and 2007 respectively.
- Mandatory electronic reporting began in 2004. This improved quality of data, in particular, the collection of VTRSERNO,which improves matching to VTR data.

Tips 'n Tricks

- A dealer-veslog link can be made reasonbly well starting in 2005. To make this link, match the the CFDBS.VTRSERNO to VESLOG.SERIAL_NUM. Chances are that you care about Trip-level outcomes: be careful, because a vessel may have more than one SERIAL_NUM per TRIPID in the VESLOG tables.

General Caveats

- Dealers are only required to record one VTR serial per trip.
- Outlier prices are always possible. Filter these out carefully.
- The following species are sketchy:
  - Surfclam and Ocean Quahog dealer reports are contained in the SFCLAM schema (separate from CFDBS). It is unclear whether reports of SC and OQ in CFDETS and CFDERS are duplicates or not, particularly for landings of Maine Mahogany Clam [Walden].
  - Giant Bluefin Tuna dealer are supposed to be reported individually and should be in a different schema. Giant Bluefin Tuna in CFDERS are either misreporting or duplication [George Silva, NMFS HMS].

  - Herring stock assessments do not necessarily use Dealer data as the source for assessments [Jon Deroba, PDB]. This is because Maine DMR has collected herring data and comprises the population of catch. The dealer data does not match the ME DMR data for herring quantity landed.

- There are many species that have two NESPP3 codes (As of March,2018).
  - Cod (080, 081)
  - Monkfish (011, 012).
  - Winter flounder (119, 120)
  - Yellowtail flounder (122, 123)
  - Plaice (124, 125)
  - Haddock (147, 148)
  - White hake (153, 154). 155 is Red/White mixed
  - Pollock (269, 270)
- Some species were/are lumped together.
  - Tilefish, which starts as NK, but becomes Blueline and Golden.
  - Skates
- The same permit number does not necessarily link to the same vessel through all years, as vessels may be upgraded or replaces.
- The same permit number does not necessarily link to the same owner over time, as vessels (with accompanying permits) can be sold.
- Data derived from "state" reporting may not include all fields that are populated by "federally reported" dealer reports. This may affect the PORT, COUNTY, PORT2, PERMIT, HULLNUM, VTRSERNO, MONTH, and DAY fields.
- Many NESPP4 codes will not match well to VTR's SPPCODE. For example, VTR cod is all 0818 (unclassifed round). Almost all Cod will eventually be classified when sold; there is very little 0818 in dealer data.
- Ports are inconsistently encoded over time.
  - Some "port groups" were split into mutiple ports. (Hampton, Seabrook, and Hampton/Seabrook, NH is a good example).
  - Many records are entered only at the "state" or "county" level. This is particularly frequent for "older" records and non-federal reports that are received through SAFIS.
  - The names corresponding to the port codes may or may not match to Census "units." The 2 digit state code does not correspond to FIPS codes.

Sample Projects

- Construct prices for fish [Lee, Demarest].
- Commercial Landings and Revenues for the "Community Profiles." [Olson, Colburn]
- "The record" of commercial landings for use in stock assessment. [PopDy]
- Construct entity-level gross revenues from commercial fishing for Regulatory Flexibility Act Analyses [Lee].

Update Frequency and Completeness

- Nightly updates. Expect approximately 300 changes or additions to the current and previous year of data per day.
- Data is "complete" 6-9 months after the end of the calendar year; however, small changes are always occurring.

- This has consequences for reproducibility if you do not store a copy of the data.

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/12205
- NEFSC's Data Dictionary http://nova.nefsc.noaa.gov/datadict/
- Preceded by: "Weighout" (WODETSyy and WODETTyy)
- Succeeded by: n/a

Related Tables

- CFDERSyyyyAA tables - "perform some Area Allocation"
- CFDETSyyyyAA tables - "perform some Area Allocation"
- CFDETTyyyy contains "detailed trip data" for 1994-2003
- CFSUMTyyyy, CFSUMSyyyy - "summary tables" for 1994-2003
- CFLENyyyy -fish-level port sampling data.
- CFAGEyyyy - fish-level port sampling data.

Support Tables

- PORT
- VALID_PORTS
- GEAR
- SPECIES_ITIS_NE decodes into names, links to the species_itis system
- CFSPP - decodes NESPP3 and NESPP4 into names

Table 1: Unique fields

Column	Description
SPPLNDLB	
SPPVALUE	
UTILCD	Quality unknown
DISPOSITION_CODE	Quality unknown
REPORTED_QUANTITY	
UOM	
GRADE_CODE	
MARKET_CODE	
SPPLIVB	Certain NESPP4 codes (monkfish livers, cod milt) convert into zero "live pounds." This is done to prevent potential double counting during the stock assessment.

Table 2: Primary Source fields - These fields are firsthand data.

Column	Description
YEAR	This may not be the same as the year in which fish was caught.
MONTH	This may not be the same as the month in which fish was caught. May be zero.
DAY	This may not be the same as the day in which fish was caught. May be zero.
DEALNUM	Dealer Identification number

Column	Description
NESPP4 [3]	There are 4 species codes (NESPP3, NESPP4, WHSPP, SPECIES_ITIS).
WHSPP	
SPECIES_ITIS	
STATE_DNUM	

Table 3: Secondary Source Fields. These fields might be more accurate somewhere else.

Column	Description
PORT	Concatenation of state, port, county
COUNTY	Data dictionary claims this is a string, but it is a 2 digit code.
NEGEAR	
NEGEAR2	
NEMAREA	
AREA	
HARVEST_AREA	
DEPTHCD	
SUBTRIP	
TONCLASS [TONCL1,TONCL2]	
FZONE	
PERMIT	
HULLNUM	
VTRSERNO	
SPRATIO	
FIPS_STATE	
FIPS_PLACE	
FIPS_COUNTY	
CF_LICENSE	
NEGEAR_VTR	

Table 4: QAQC columns. Quality Control or Auditing fields.

Column	Description
LINK	
DOCN	
EFFIND	
SOURCE	
DERSOURCE	
PARTNER_ID	
DEALER_RPT_ID	
DOE	
LANDING_SEQ	

VESLOG Databases

This is very incomplete.

Tables: VESLOGyyyyT; VESLOGyyyyG; VESLOGyyyyS;

Location: Sole

Schema: VTR

Overview

The veslog data contains everything collected through the Vessel Trip Report System. These data are primarily generated through mandatory reporting by federally-permitted fish dealers.

- 1994 to present

Current Collection Methods

These data are the result of mandatory federal vessel reporting. Federally permitted vessels are required to submit one VTR report per "gear-mesh-area" fished.

Changes to Collections Methods

- I dunno. But the VTR form has changed slightly over time.
- Electronic VTRs start in 2011.

Tips 'n Tricks. very incomplete.

- A dealer-veslog link can be made reasonbly well starting in 2005. To make this link, match the the CFDBS.VTRSERNO to VESLOG.SERIAL_NUM. Chances are that you care about Trip-level outcomes: be careful, because a vessel may have more than one SERIAL_NUM per TRIPID in the VESLOG tables.

General Caveats. very incomplete.

- Electronic VTRs have very long serial numbers. Some software doesn't like this (Excel, stata) – you might want to do this:
  - select to_char(g.serial_num) from veslog2014g g
- Some of the older numbers (from 1994-1995) are non-numeric.
- All quanties are "Hail Weights," which are the operator's best estimate of catch.
- SPPCODES will not match well to dealer's NESPP4 codes. For example, VTR cod is all 0818 (unclassifed round). Almost all Cod will eventually be classified when sold; there is very little 0818 in dealer data.
- The following species are sketchy:
  - Surfclam and Ocean Quahog dealer reports are contained in the SFCLAM schema (separate from CFDBS). It is unclear whether reports of SC and OQ in CFDETS and CFDERS are duplicates or not, particularly for landings of Maine Mahogany Clam [Walden].

  - Giant Bluefin Tuna dealer are supposed to be reported individually and should be in a different schema. Giant Bluefin Tuna in CFDERS are either misreporting or duplication [George Silva, NMFS HMS].
- The same permit number does not necessarily link to the same vessel through all years, as vessels may be upgraded or replaces.
- The same permit number does not necessarily link to the same owner over time, as vessels (with accompanying permits) can be sold.
- PORTLND1 and PORTLND2 are inconsistently encoded over time.
  - The names corresponding to the port codes may or may not match to Census "units." The 2 digit state code does not correspond to FIPS codes.
- CAREA
- 1994-1995 are kind of sketchy

Sample Projects very incomplete.

Update Frequency and Completeness very incomplete.

- Nightly updates. Expect approximately 300 changes or additions to the current and previous year of data per day.
- Data is "complete" 6-9 months after the end of the calendar year; however, small changes are always occurring.
- This has consequences for reproducibility if you do not store a copy of the data.

Other Metadata sources

- INPORT. https://inport.nmfs.noaa.gov/inport/item/1423
- NEFSC's Data Dictionary http://nova.nefsc.noaa.gov/datadict/
- Preceded by: "none"
- Succeeded by: n/a

Related Tables very incomplete.

- CATCH, IMAGES, DOCUMENT - these are the "GARFO" version of VTR.

Support Tables very incomplete.

- VLSPP decodes SPPCODES into names
- SPECIES_ITIS_NE decodes into names, links to the species_itis system

Table 5: Unique fields

Column	Location	Description
QTYKEPT	A	Placeholder
QTYLAND	A	Placeholder

Table 6: Primary Source fields - These fields are firsthand data.

Column	Location	Description
TRIPID	A	Placeholder
GEARID	A	Placeholder

Table 7: Secondary Source Fields. These fields might be more accurate somewhere else.

Column	Location	Description
IDK	A	Placeholder
IDK	A	Placeholder

Table 8: QAQC columns. Quality Control or Auditing fields.

Column	Location	Description
IDK	A	Placeholder
IDK	A	Placeholder