

Migration Outliers

Robert Dinterman

2017-06-06

Detecting Outliers for Net Migration

The IRS data currently spans across 1992 to 2013 for a total of 68,984 FIPS-years observations for net-migration. There were substantial changes in data formatting and collecting across this time. As such, there are likely some data problems like clerical errors in reporting values or misclassifications in FIPS codes, etc. This is meant to be a quick and dirty way of detecting outliers in the data.

The process for detecting an outlier used here is to group the in-migration data for each county across the years (on average 22 years for each) and calculate the following:

$$|X_i - \text{median}(X_{-i})| > 7 * \text{sd}(X_{-i})$$

Values for which the above is true are flagged as potential outliers as this indicates the observation is approximately a seven sigma event (p-value of approximately 2.6e-12) and in expectation we should see one in 5,663,374 migration datasets if each county follows an independent normal distribution.

There are three values of each variable (returns, exemptions, and AGI) for a county: in-flows, out-flows, and net-flows (which is the in-flows less the out-flows). For outlier detection, the in-flow, out-flow, and net-flow are all evaluated for whether or not there may be an error in the data that we have parsed. I also calculate a column titled `flag_factor` which is the flagged migration value for the county-year divided by the mean of the migration for that county across the entire sample (excluding the flagged year). The higher the value, the more unlikely the event.

Exemptions (proxies population)

This is only meant as a quick and dirty way of detecting outliers, but there are a total of 30 flagged FIPS-year combinations.

Exemptions do not appear to be that big of an issue and none of them appear to be obvious mis-codings. There are 31 flagged observations, so I would be fine leaving these outliers in the dataset.

Table 1: Flagged Outliers for In

year	fips	exmpt_in	mean_in	sd_in	flag_factor
1992	02016	2,213	559	427	4
1992	26103	4,283	2,218	523	2
1992	51840	8,798	2,535	1,483	3
1993	28163	1,542	880	167	2
1994	13125	1,217	214	232	6
1994	42031	2,581	1,266	310	2
1994	42121	5,723	1,732	904	3
1995	13087	9,530	1,602	1,808	6
1995	34033	3,965	2,517	394	2
1996	01055	4,866	3,607	334	1
1996	13261	2,253	1,240	251	2
1998	16015	1,364	560	199	2

year	fips	exmpt_in	mean_in	sd_in	flag_factor
2005	22033	27,467	16,780	2,855	2
2005	22103	19,355	12,299	1,758	2
2005	28109	4,559	2,488	514	2
2007	21051	781	481	79	2
2011	18123	1,079	655	109	2
2011	20029	6,542	733	1,298	9
2011	20089	2,667	261	539	10
2011	20157	4,306	402	873	11
2011	20183	3,150	292	639	11
2011	21077	1,261	538	179	2
2011	30007	562	327	60	2
2011	31149	283	74	49	4
2011	36025	2,849	1,785	278	2
2011	36049	2,772	1,006	404	3
2011	42115	2,563	1,593	249	2
2011	47061	1,326	483	193	3
2011	47067	517	203	74	3
2013	48453	208,145	63,645	35,598	3

Table 2: Flagged Outliers for Out

year	fips	exmpt_out	mean_out	sd_out	flag_factor
1996	13087	10,489	1,648	1,986	6
2005	22103	17,593	9,618	1,984	2

Table 3: Flagged Outliers for Net

year	fips	exmpt_net	mean_net	sd_net	flag_factor
1992	51840	5,297	221	1,157	24
1994	42031	1,221	-14	302	-86
1994	42121	3,823	42	851	90
1998	16015	1,000	121	214	8
2005	22033	8,416	-2,118	2,552	-4
2011	18123	565	19	142	30
2011	20029	6,426	246	1,381	26
2011	21077	745	69	170	11
2011	31149	225	-4	54	-56
2011	36049	1,584	-5	374	-332
2011	47061	797	39	180	21
2011	47067	308	17	73	18
2013	48453	121,251	11,226	26,453	11

Adjusted Gross Income (AGI)

Using the same methodology as above, we can also detect any anomalies in the net-migration data that deal with AGI. This ends up being more of an issue than exemptions. There are some implausible values here, especially when considering that the AGI variable is measured in terms of \$1,000s. There are 226 flagged

observations, which is a bit too large and should be further investigated. Below are the flagged outliers:

Table 4: Flagged Outliers for In

year	fips	agi_in	mean_in	sd_in	flag_factor
1992	42039	87,090	41,976	11,710	2
1993	36117	125,708	67,573	14,828	2
1994	21137	-990,239	-31,330	214,201	32
1994	42031	40,230	19,699	5,405	2
1994	42121	105,096	28,213	17,540	4
1994	51630	-942,328	14,742	214,785	-64
1994	53011	-688,996	441,795	276,107	-2
1996	48137	14,011	2,459	2,668	6
1997	13301	-1,090	3,146	1,064	0
1997	13307	-9,341	717	2,282	-13
1998	38071	26,587	8,874	4,486	3
1998	46041	9,856	2,710	1,711	4
1998	46117	68,357	7,958	13,628	9
2000	36099	79,033	24,571	12,841	3
2000	47173	54,956	14,837	9,592	4
2000	53071	130,781	50,054	21,327	3
2001	54027	64,227	17,586	11,310	4
2002	26149	87,490	38,069	12,531	2
2003	48333	26,196	5,277	4,879	5
2004	20017	32,419	3,604	6,479	9
2005	28023	28,395	11,503	4,411	2
2005	29151	48,537	8,732	8,995	6
2005	47081	42,676	16,593	6,579	3
2005	54059	41,071	9,968	7,172	4
2006	31097	25,562	4,017	4,878	6
2006	41001	40,127	14,204	6,118	3
2006	48243	15,035	3,456	2,713	4
2006	48327	9,299	2,248	1,758	4
2006	48399	23,658	8,139	3,907	3
2006	48501	20,933	6,198	3,917	3
2006	51113	38,948	13,637	6,751	3
2007	21139	18,466	7,639	2,798	2
2007	56045	25,162	7,134	4,729	4
2009	18071	63,390	27,724	9,316	2
2009	30009	299,182	26,620	61,117	11
2009	46065	136,271	23,428	26,513	6
2010	56015	49,206	12,897	9,384	4
2011	17171	7,910	3,232	1,160	2
2011	19121	-16,210	14,979	7,980	-1
2011	20029	136,476	11,763	27,871	12
2011	21077	21,506	7,120	3,652	3
2011	31149	5,693	1,091	1,099	5
2011	36049	51,445	14,685	8,769	4
2011	47067	6,990	2,319	1,161	3
2011	56019	47,705	12,845	8,917	4
2012	05119	711,432	342,482	97,470	2
2012	13241	53,345	17,796	9,382	3
2012	16013	278,882	49,329	52,446	6
2012	20075	-20,255	545	4,696	-37

year	fips	agi_in	mean_in	sd_in	flag_factor
2012	28119	18,907	3,814	3,404	5
2012	29023	49,180	24,571	6,353	2
2012	30057	129,227	14,349	25,795	9
2012	31167	22,251	4,664	4,385	5
2012	36001	1,019,758	315,616	169,914	3
2012	40115	79,417	22,524	14,341	4
2012	46011	83,253	27,400	13,945	3
2012	46029	-19,634	20,178	9,753	-1
2012	48283	23,642	4,346	4,699	5
2012	54089	-13,647	4,408	4,137	-3
2012	55085	122,782	46,445	19,045	3
2012	56033	168,574	41,232	29,970	4
2013	08027	37,523	10,257	6,991	4
2013	26163	2,125,746	939,266	297,196	2
2013	48095	26,993	3,586	5,286	8
2013	48417	15,313	3,635	3,139	4
2013	53055	-106,733	34,999	35,841	-3

Table 5: Flagged Outliers for Out

year	fips	agi_out	mean_out	sd_out	flag_factor
1992	26069	61,445	21,133	9,394	3
1992	42031	72,815	23,507	11,841	3
1994	21001	-996,416	-38,987	213,850	26
1994	48027	-682,191	367,810	252,627	-2
1995	26085	25,348	8,168	4,321	3
1995	40067	12,066	4,629	1,957	3
1997	13197	-6,752	3,433	2,562	-2
1997	39119	109,467	44,803	16,008	2
1997	49017	34,867	5,305	6,686	7
1998	19063	21,072	7,789	3,255	3
1999	20021	32,052	13,888	4,808	2
1999	30035	25,760	7,771	4,306	3
1999	30059	5,136	1,610	887	3
1999	31129	14,705	3,679	2,551	4
1999	38069	10,746	3,712	1,693	3
1999	42083	66,471	23,080	10,147	3
2000	39135	65,710	29,588	9,292	2
2000	46111	6,862	1,968	1,209	3
2001	29117	25,054	10,422	3,813	2
2002	21065	17,263	6,490	2,832	3
2003	35011	3,711	1,454	582	3
2004	18175	67,251	17,213	11,807	4
2005	22071	1,819,128	441,026	311,877	4
2005	22087	486,666	51,047	97,620	10
2005	28045	108,434	37,197	18,235	3
2005	39067	52,310	11,071	9,393	5
2005	48253	40,626	14,214	6,706	3
2006	27133	27,102	7,760	4,512	3
2006	46065	71,484	20,930	12,575	3
2006	48211	20,827	4,988	4,091	4

year	fips	agi_out	mean_out	sd_out	flag_factor
2007	12123	42,449	12,198	7,208	3
2007	20097	12,555	3,245	2,162	4
2008	13141	14,191	4,932	2,377	3
2008	38031	13,598	3,290	2,428	4
2008	48293	52,654	16,626	9,088	3
2009	46121	120,372	8,819	24,929	14
2010	19167	70,594	20,375	11,983	3
2010	50009	12,432	5,051	1,951	2
2011	13099	15,165	6,018	2,267	3
2011	13257	52,157	16,743	8,370	3
2011	18061	128,607	30,301	22,920	4
2011	20127	67,684	7,172	13,548	9
2011	20201	367,087	20,001	77,526	18
2011	22031	260,604	28,796	52,048	9
2011	30057	-41,459	5,842	12,231	-7
2011	33007	67,732	23,181	10,894	3
2011	42131	51,840	22,895	7,744	2
2011	46005	54,750	14,823	9,438	4
2011	46067	15,814	5,294	2,726	3
2011	48283	-29,599	2,726	7,935	-11
2011	54059	82,607	15,689	15,093	5
2011	54085	14,877	5,278	2,470	3
2012	01091	44,900	13,396	7,333	3
2012	05069	350,028	65,694	63,847	5
2012	08037	567,174	110,996	106,820	5
2012	08039	538,839	52,857	109,090	10
2012	12073	896,335	292,246	148,896	3
2012	17121	54,445	23,991	7,756	2
2012	17191	18,258	8,237	2,647	2
2012	22037	40,418	15,220	6,526	3
2012	27023	-38,278	8,075	10,710	-5
2012	27039	47,353	18,043	7,543	3
2012	28027	39,752	15,491	5,926	3
2012	30075	8,041	1,395	1,508	6
2012	30091	13,241	3,247	2,465	4
2012	31037	29,031	9,359	5,054	3
2012	37061	89,431	27,529	15,923	3
2012	40021	102,989	28,683	17,787	4
2012	46051	15,084	5,851	2,433	3
2012	46117	30,887	5,422	5,884	6
2012	47021	150,789	41,027	25,803	4
2012	48017	-10,672	4,587	3,710	-2
2012	48065	25,990	7,711	4,345	3
2012	48311	28,657	2,313	6,015	12
2012	48327	15,161	2,333	2,941	6
2012	56009	58,897	16,741	10,966	4
2013	06023	-127,418	67,345	46,392	-2
2013	20099	36,748	14,234	5,367	3
2013	20109	-3,187	2,256	1,346	-1
2013	23013	96,575	33,929	16,531	3
2013	45063	546,961	235,558	81,515	2
2013	48417	14,318	3,360	2,631	4

year	fips	agi_out	mean_out	sd_out	flag_factor
2013	51630	231,087	65,099	40,869	4

Table 6: Flagged Outliers for Net

year	fips	agi_net	mean_net	sd_net	flag_factor
1992	26069	-41,520	2,826	10,283	-15
1992	42031	-57,562	-3,808	13,652	15
1992	42039	61,231	-1,672	16,455	-37
1992	51840	69,163	1,870	16,260	37
1993	36117	74,300	-2,408	17,808	-31
1993	48491	-869,481	191,325	249,426	-5
1994	12003	26,419	4,333	5,711	6
1994	21001	1,002,284	46,283	213,527	22
1994	21137	-998,079	-43,157	213,299	23
1994	42121	80,308	728	18,236	110
1994	48027	963,156	30,011	211,857	32
1994	51630	-990,114	-50,357	212,281	20
1994	53011	-862,049	99,837	219,072	-9
1995	26085	-16,043	2,415	4,585	-7
1995	40067	-8,985	-779	2,083	12
1996	48137	12,822	409	2,907	31
1997	13197	10,873	843	2,341	13
1997	13301	-5,311	-277	1,334	19
1997	13307	-10,664	-473	2,330	23
1997	39119	-71,316	-5,684	15,395	13
1997	49017	-30,733	-1,193	6,633	26
1997	51660	-32,709	-6,387	7,031	5
1998	16015	23,501	3,414	4,639	7
1998	19063	-16,180	-2,430	3,284	7
1998	31177	-29,072	555	7,169	-52
1998	38071	16,467	-477	4,084	-35
1998	39057	113,205	-9,475	31,938	-12
1998	46041	6,999	-325	1,788	-22
1998	46117	61,520	2,535	13,932	24
1998	48255	27,400	811	6,048	34
1998	51141	-9,629	601	2,501	-16
1999	01015	-57,941	-8,956	12,838	6
1999	12107	-26,827	6,143	8,552	-4
1999	20021	-19,961	-1,634	4,734	12
1999	27035	-67,495	11,481	18,827	-6
1999	30035	-20,794	-1,878	4,339	11
1999	30059	-3,914	26	1,017	-151
1999	31129	-12,476	-1,293	2,600	10
1999	35023	-10,652	-1,451	2,393	7
1999	38069	-8,576	-842	1,878	10
1999	42083	-50,395	-6,143	10,087	8
1999	44001	-90,585	-5,381	21,370	17
2000	29061	-6,735	85	1,821	-79
2000	36099	53,121	138	12,382	386
2000	39135	-39,059	-2,438	9,025	16
2000	46111	-5,873	-463	1,299	13

year	fips	agi_net	mean_net	sd_net	flag_factor
2000	47173	44,678	4,595	9,042	10
2000	48417	7,737	275	1,843	28
2000	51185	-25,498	-3,205	5,783	8
2000	53047	-38,610	3,704	10,052	-10
2000	53071	72,867	9,670	15,133	8
2001	16003	-8,008	520	2,212	-15
2001	29117	-15,885	-1,430	3,771	11
2001	54027	53,974	6,682	10,903	8
2002	21065	-9,483	-319	2,241	30
2002	21203	8,097	905	1,741	9
2002	26149	50,516	-88	12,434	-575
2003	19187	-27,040	-7,029	5,270	4
2003	23009	94,906	16,349	19,581	6
2003	35011	-2,696	-208	636	13
2003	39151	-138,651	-30,192	27,449	5
2003	48333	23,497	2,184	4,959	11
2003	48365	-30,987	-860	7,667	36
2004	18175	-50,337	-1,101	11,287	46
2004	20017	30,373	1,397	6,527	22
2005	17021	-32,008	-3,822	6,667	8
2005	22071	-1,652,772	-86,348	366,313	19
2005	22075	-60,356	-5,627	14,033	11
2005	22087	-482,997	-20,763	104,464	23
2005	22105	56,690	11,463	10,872	5
2005	28023	14,181	-260	3,693	-55
2005	28045	-60,223	6,720	16,129	-9
2005	28047	-223,725	-8,030	50,850	28
2005	29151	40,970	1,778	8,841	23
2005	39067	-43,573	-1,865	9,446	23
2005	47081	26,995	3,864	5,547	7
2005	48253	-26,999	-1,274	5,927	21
2006	13031	60,798	5,213	13,427	12
2006	19099	-39,507	-4,567	8,997	9
2006	27133	-20,745	-2,138	4,272	10
2006	31097	20,780	507	4,584	41
2006	31121	-6,047	-1,057	1,293	6
2006	41001	27,317	2,032	6,204	13
2006	42059	-38,381	-2,944	8,411	13
2006	48211	-16,435	-1,554	3,649	11
2006	48243	12,473	1,073	2,704	12
2006	48399	15,057	230	3,632	65
2006	48501	13,607	-2,237	3,822	-6
2006	51113	28,861	3,490	6,039	8
2007	12123	-31,555	-231	7,748	137
2007	20097	-11,110	-1,162	2,286	10
2007	21139	9,320	981	2,090	10
2007	30093	-40,299	-3,659	9,303	11
2007	48143	48,228	4,268	10,145	11
2007	55097	27,837	-2,492	7,449	-11
2007	56045	17,625	690	4,425	26
2008	13141	-10,433	-492	2,463	21
2008	30027	6,723	659	1,579	10

year	fips	agi_net	mean_net	sd_net	flag_factor
2008	38031	-11,375	-839	2,471	14
2008	48293	-34,188	-974	7,709	35
2008	55029	-19,769	10,227	8,045	-2
2009	18071	38,108	55	9,416	698
2009	30009	289,434	14,808	61,739	20
2009	46065	118,642	2,499	28,612	47
2009	46121	-117,279	-5,957	24,871	20
2009	50001	-14,933	3,283	4,688	-5
2010	19167	-54,620	-5,232	11,431	10
2010	50009	-7,454	299	1,989	-25
2010	55003	7,951	-3	2,067	-2,302
2010	56015	35,048	2,462	7,463	14
2011	13099	-8,510	-1,222	1,844	7
2011	13257	-34,457	-1,939	7,545	18
2011	16055	-100,148	47,239	38,486	-2
2011	17139	-16,136	-886	4,017	18
2011	17171	5,268	-181	1,365	-29
2011	18061	-95,980	-231	21,761	415
2011	18115	-7,599	-78	1,905	97
2011	18123	10,943	-4	2,892	-2,482
2011	19121	-34,870	1,586	8,395	-22
2011	20029	134,682	5,084	28,962	26
2011	20127	-63,192	-2,784	13,509	23
2011	20201	-364,440	-17,229	77,553	21
2011	21077	14,220	1,032	3,129	14
2011	21191	-10,146	170	2,570	-60
2011	22031	-230,472	-9,400	49,406	25
2011	27075	10,013	1,393	2,222	7
2011	31149	4,769	119	1,173	40
2011	33007	-37,575	-331	9,255	114
2011	36049	29,234	327	6,649	89
2011	40045	-5,494	-458	1,272	12
2011	40121	-20,550	722	5,622	-28
2011	42131	-15,203	-1,694	3,601	9
2011	45037	-22,533	1,254	5,988	-18
2011	46005	-41,177	-4,133	8,805	10
2011	46067	-9,572	-1,188	2,049	8
2011	47061	10,893	1,077	2,649	10
2011	47063	63,194	6,139	14,840	10
2011	47067	3,437	278	801	12
2011	47115	-10,715	1,310	3,129	-8
2011	47185	25,870	4,210	5,646	6
2011	48283	40,078	1,621	9,432	25
2011	48313	12,521	1,957	2,817	6
2011	54059	-70,282	-5,721	16,209	12
2011	54085	-8,933	180	2,219	-50
2011	54105	5,004	430	1,121	12
2011	56019	39,027	5,146	8,152	8
2012	01009	-29,760	6,586	9,544	-5
2012	01043	46,792	8,977	9,538	5
2012	01091	-31,008	-3,197	6,490	10
2012	05069	-311,398	-29,641	63,117	11

year	fips	agi_net	mean_net	sd_net	flag_factor
2012	05119	188,144	-32,575	52,808	-6
2012	08037	-371,533	3,845	88,682	-97
2012	08039	-466,221	-4,360	104,049	107
2012	12073	-571,503	-43,849	119,727	13
2012	13241	37,430	7,054	7,654	5
2012	13263	-5,776	104	1,561	-56
2012	16013	241,504	20,588	50,102	12
2012	17017	-14,465	-2,896	2,979	5
2012	17121	-27,915	-3,993	6,000	7
2012	17191	-10,667	-1,244	2,383	9
2012	18127	-42,430	11,944	14,426	-4
2012	19017	13,783	667	3,292	21
2012	20075	-22,198	-1,550	4,685	14
2012	22037	-20,940	698	5,350	-30
2012	24017	-152,517	6,501	39,622	-23
2012	27023	48,595	166	10,981	293
2012	27039	-26,305	15	6,582	-1,808
2012	28027	-27,518	-5,887	5,337	5
2012	28119	13,854	-496	3,254	-28
2012	29023	20,585	1,175	5,127	18
2012	30057	113,696	8,507	26,762	13
2012	30075	-5,804	-246	1,343	24
2012	30091	-7,691	-923	1,689	8
2012	31037	-17,363	-3,073	3,672	6
2012	31167	11,827	231	2,768	51
2012	33017	-85,504	2,568	22,194	-33
2012	36001	556,689	-27,752	131,683	-20
2012	37061	-45,276	-693	10,461	65
2012	40021	-63,402	-702	14,213	90
2012	40047	118,681	-1,806	27,797	-66
2012	40115	46,846	2,618	10,937	18
2012	40127	-17,514	-138	3,949	126
2012	46011	41,022	-3,500	11,412	-12
2012	46029	-50,491	-1,416	11,570	36
2012	46051	-9,560	-1,069	2,078	9
2012	47021	-90,438	3,321	21,552	-27
2012	48017	16,513	-703	3,985	-23
2012	48065	-18,422	-1,113	4,123	17
2012	48099	-92,898	-11,278	21,466	8
2012	48223	-16,224	1,770	4,716	-9
2012	48285	14,115	1,742	3,088	8
2012	48311	-26,140	-1,394	5,614	19
2012	48327	-13,355	-85	3,445	158
2012	51740	48,506	-12,197	16,109	-4
2012	54081	48,924	3,336	11,357	15
2012	54089	-22,130	-939	4,800	24
2012	55085	79,203	13,877	16,359	6
2012	56009	-33,023	-884	7,736	37
2012	56033	123,100	14,243	24,623	9
2013	05085	-38,674	12,059	13,242	-3
2013	06023	215,126	6,561	47,441	33
2013	08027	30,827	5,940	6,199	5

year	fips	agi_net	mean_net	sd_net	flag_factor
2013	17113	-117,100	-11,975	27,196	10
2013	20099	-22,251	-3,098	4,475	7
2013	20109	5,565	-161	1,439	-35
2013	23013	-48,214	5,230	13,401	-9
2013	26163	574,077	-480,722	258,147	-1
2013	27067	-30,011	-3,197	7,027	9
2013	31185	-15,419	-2,326	3,236	7
2013	34017	-991,716	-312,169	174,790	3
2013	38025	8,186	283	2,024	29
2013	45063	-166,324	33,436	50,765	-5
2013	46053	5,932	42	1,512	142
2013	48055	-37,971	2,984	10,040	-13
2013	48095	24,298	1,132	5,233	21
2013	53007	62,168	15,686	12,268	4
2013	53055	-136,595	16,851	37,699	-8