Estimating Persistence in Employee Business Expense Correspondence Examinations using Hidden Markov Models[1]

Julie Buckel, Anne Parker, Sarah Shipley, Christine Oehlert
US Internal Revenue Service

Abstract

We use Hidden Markov Models to study persistence of the compliance impact for tax examinations of Employee Business Expense (EBE). Using panels of yearly returns for taxpayers reporting EBE, we compare future filing behavior of those audited to those not audited, by fitting and comparing Hidden Markov Models for both groups. The Markov state space is EBE reporting compliance. The observation vectors are a function of reported line item amounts for a series of annual returns filed two years after the year of the tax audits, the baseline year. The functions used to create the observation vectors are proxies for compliance, and the unobserved Markov state space is true compliance. The observations have a probability distribution that is conditional upon unobserved compliance status. Our fitted models give some evidence that a no change audit may worsen compliance slightly.

Key Words: Hidden Markov Model, Tax Compliance

1. Introduction

The United States tax system relies in part on individuals and businesses self-reporting their tax liabilities. Given the voluntary nature of the system, there is a gap between the taxes that should be paid and those that are actually paid – the Tax Gap. Recent estimates of the underreporting tax gap are $387 Billion with $264 Billion attributed to Individual Income Tax.[2]

IRS Examination programs support voluntary compliance with tax laws by addressing compliance issues through audits – field examinations having multiple issues and conducted face-to-face with the taxpayer, and correspondence examinations focused on a single issue and conducted through written correspondence with the taxpayer.

Audits may have direct effects (immediate results from an examination – change/no change for example), and indirect effects (persistence – changes in the future filing behavior after an audit) on voluntary compliance. Given the tax gap and the cost of examinations, audit persistence has become increasingly important.

We focus on the question of whether a taxpayer changes his/her future filing behavior as a consequence of an audit, and whether any such changes are enduring or only fleeting.

---

[1] The views expressed in this presentation are those of the authors and do not necessarily reflect the views of the U.S. Internal Revenue Service or the Department of the Treasury.

[2] https://www.irs.gov/PUP/newsroom/tax%20gap%20estimates%20for%202008%20through%202010.pdf

1.1 Related Research

Recent work (DeBacker et al., 2015) on audit persistence using Tax Year 2006 - 2009 National Research Program (NRP)[3] examinations together with a randomized control group found that for individuals overall, an audit increased reportable income by an average of $1,000 with a persistence impact of at least six years. However, taxpayers with different income sources and claim eligibility (i.e. wages, self-employment income, refundable tax credits, and certain deductions), appear to respond differently to audits in both the increase in average reportable income and persistence effect. In addition, the persistence effect associated with audit type, correspondence or field, is larger for correspondence audits.

Using a similar approach, but using only operational administrative tax data for non-farm self-employed segment (Beer et al., 2016) found the indirect effects on an audit persist for at least three years but depend on the audit outcome – change with adjustment (positive impact) and no change (negative impact).

1.2 Problem Description

Our work builds on the existing literature by using a Hidden Markov Modeling framework, a stochastic modeling approach, to explore patterns of taxpayer filing behavior over time for Employee Business Expense Correspondence Exams (EBE). EBE is reported on a Form 1040 Schedule A line 21 under "Job Expenses and Certain Miscellaneous Deductions."

We chose this particular work stream for several reasons. First, the EBE audit program started around Tax Year 2003 and continues today so there is a reliable time series of data.[4] Second, EBE exams are straightforward single issue correspondence exams allowing us to test whether our modeling approach can be successfully applied to understanding audit persistence in this segment over the period TY2003 – TY2012. Table 1 provides an overview of EBE filings over the period. EBE has remained fairly stable over time with a slight downward trend.

We track the future EBE filing behavior[5] of three groups of taxpayers who reported EBE; those who were audited for EBE and had a positive tax change, those who were audited for EBE but did not have a positive tax change (no-change), and those who were not audited but whose reporting behavior suggested possible EBE non-compliance. We fit Hidden Markov Models to each group and compare the results.

[3] https://www.irs.gov/pub/irs-soi/mazur.pdf
[4] All following references to years are Tax Years.
[5] We start our data vectors for the Hidden Markov Models with Tax Year 2005, to allow for lags between filing and auditing of tax returns. An audit can have no impact on future compliance until the taxpayer is aware of it. Correspondence audits have a fairly quick turnaround time, so by the time TY2005 returns must be filed, most audits of TY2003 returns will have been completed.

Table 1. Unreimbursed Employee Expenses filings for Tax Years 2003 – 2012

Tax Year	*1040 filing (count)*	*Schedule A Unreimbursed employee expenses (count)*	*Schedule A Unreimbursed employee expenses ($1,000)*	*Form 2106/2106EZ filings (count)*	*Form 2106/1206EZ filings ($1,000)*	*Percent of 1040 filings*
2003	130,423,626	14,896,433	63,210,079	6,813,407	44,791,299	11.42
2004	132,226,042	15,545,955	68,497,230	7,483,103	49,666,017	11.76
2005	134,372,678	15,920,218	75,824,189	7,825,703	56,639,758	11.85
2006	138,394,754	15,985,244	75,600,830	8,664,367	53,303,582	11.55
2007	142,978,806	16,479,370	82,105,794	8,966,892	58,925,639	11.52
2008	142,450,569	15,790,907	82,225,607	9,206,616	63,467,240	11.09
2009	140,494,127	14,942,268	75,607,218	8,704,483	57,855,103	10.64
2010	142,892,051	14,631,890	72,143,485	8,351,710	54,728,296	10.24
2011	145,370,240	14,730,817	76,857,890	8,709,898	58,552,419	10.13
2012	144,928,472	14,604,311	81,428,583	8,757,770	62,064,311	10.08

Source: IRS Statistics of Income, Individual Income Tax Returns Line Item Estimates, Publication 4801 (Rev. 10-2014)

2. Methodology

Prior research suggests that audit impact may persist six or more years, so to allow for as long a follow-up period as possible we selected study groups of taxpayers who claimed EBE for Tax Year 2003. We created three groups, those whose audited EBE deductions resulted in a positive tax change, those whose audited EBE deductions did not result in a positive tax change, and a random sample of those whose EBE reporting suggested possible non-compliance but were not audited. Comparing how these three groups behaved in the years following TY2003 would help us determine how an EBE audit impacts future EBE compliance, how long the impact persists, and how these impacts differ depending on the outcome of the audit.

Most of the taxpayers whose EBE claims for TY2003 were audited, were not audited again in the follow-up period.[6] As a result, we cannot know their EBE compliance with certainty. Instead we study their behavior by fitting Hidden Markov Models (HMM) to indirect indicators of compliance.

The HMM framework[7] posits a Markov process whose state membership cannot be observed directly, only inferred by the behavior of an emission variable whose probability distribution differs across the hidden Markov states. The states we are interested in are valid EBE claims (which we refer to as compliant) and invalid EBE (which we refer to as non-compliant). For our purposes only over-stated claims are considered non-compliant. A taxpayer who claims less EBE than the amount to which he may be entitled, or a taxpayer who claims no EBE at all, is by definition considered compliant.

[6] A small percentage of such taxpayers were re-audited, in some instances several times, until their EBE claim was found to be compliant. When these repeat audit taxpayers were analyzed separately, the number of repeat audits had an approximately geometric distribution. This small set of taxpayers whose EBE compliance was repeatedly and directly observed did have a sojourn time in the non-compliant state that is at least superficially consistent with the notion that EBE compliance operates like a two state Markov process.

[7] We follow the approach described in (Rabiner, 1989)

Our notation for the HMM is as follows: EBE claims fall into one of k compliance categories, labeled $C_1, \ldots, C_k$ In the simplest example, $k = 2$ and the categories are $C_1 =$ "valid or understated EBE" and $C_2 =$ "overstated EBE". A given taxpayer's series of EBE claims generates a series of random variables $x(t)$, where each $x(t) = C_j$ for some $j = 1, \ldots, k$ and $0 <= t <= m$. Our model assumes these $x(t)$'s are a one-step Markov process with $\{C_1, \ldots, C_k\}$ as the state space. Elements of the transition matrix are denoted:

$$p_{ij} \equiv P[\,x(t+1) = C_j \mid x(t) = C_i] \text{ for all } j, i = 1 \ldots k$$

We call the emission function $y(t)$. Each element of the Markov state space, C_i, generates a conditional probability distribution F_i for the $y(t)$values. Making an EBE claim is of course voluntary, and the need to do so depends on circumstances that may change from year to year. As a result, some taxpayers claim EBE only intermittently. In contrast to the compliance of a non-zero EBE claim whose compliance cannot be directly observed, we can tell with certainty when no EBE claim was made and call this a form of compliance. We fit two HMMs that deal with intermittent claimants differently. For our first fit, we focus solely on the sub-population of continuous EBE claimants. For the second model fit we include intermittent EBE claimants and posit an additional "no-claim" Markov state.

HMMs are fit using an Expectation-Maximization (E-M) algorithm, in our case the one implemented in the R package *mhsmm* (O'Connell and Højsgaard, 2011, 2015). Since *mhsmm* performs an iterative fit, it requires several starting values. The first is an initial Markov transition matrix, $[p_{ij}]$ which we specify with each entry equal to $\frac{1}{k}$, where k is the number of Markov states. The package also requires initial values for the proportion of units initially occupying the Markov states, which we call $(p_1(0), p_2(0), p_3(0), \ldots, p_k(0))$. For our fits we specified equal proportions in all states. Finally, *mhsmm* requires specification of the starting distributions $F_1, \ldots, F_k$ of the emissions variable, conditional on each Markov state. We specified Normal distributions and provided initial location and spread parameters.

Prior to setting up the data for fitting models, we needed to address a missing data problem. EBE claim amounts are not generally available in IRS research databases for Tax Years prior to 2006. Using Statistics of Income (SOI) data, we built imputation models to fill gaps in our study subjects' EBE claim histories. SOI data is a large, multipurpose, complex weighted statistical sample of filed tax returns. It is the primary source of authoritative estimates of tax reporting.[8] We used these completed EBE histories both to determine who qualified to be in our non-audited control group, and when necessary to construct portions of the emissions data vectors. Just as we had two versions of our HMM framework, we also had two versions of our imputation methods. These are described in detail in the Data Set-up section.

The structure of the Form 1040 Schedule A Itemized Deductions imposes algebraic constraints on EBE reporting, which we exploited whenever possible. EBE is combined with other expenses into a preliminary total we refer to as "Gross Miscellaneous Deductions". This is reduced by 2% of Adjusted Gross Income (AGI), and the result is recorded as Net Miscellaneous Deductions.[9] Although EBE amounts are not available for

---

[8] https://www.irs.gov/pub/irs-soi/14indescofsample.pdf

[9] If the reduction yields a negative number, Net Miscellaneous Deductions is set to zero. It is thus possible for a taxpayer to record an EBE amount that, because of the 2% of AGI reduction has no

all Tax Years in our research databases, Net Miscellaneous Deductions and AGI are. The availability of these data fields allows us to cap our imputed EBE claims by Gross Miscellaneous Deductions, and to set to zero the effective EBE claims (imputed or known) for all taxpayers whose Net Miscellaneous Deductions amounts are zero.

The other components of Gross Miscellaneous Deductions include tax preparation expenses.[10] Taxpayers with multiple sources of substantial non-wage income tend to file added schedules, where they report details of their non-wage income. They also tend to incur higher tax preparation expenses, due to their higher income levels and return complexity. We took advantage of the association between added schedules and the non-EBE components of Gross Miscellaneous Deductions, to improve imputations and to create emissions functions that are correlated with compliance.

Because we are interested in the persistence of the impact of a single audit, we did not allow any observation vectors to extend beyond the year of a second audit, if one occurred within our timeframe. We also truncated the observation vector of any taxpayer who did not file a tax return for any year within our time frame. Although ceasing to file returns is one possible response to an audit, we do not include that option in our Markov state space. We regard filing a return, when required by law, to be a pre-condition to reporting accurate amounts (including EBE claims) on the return.

To create the data[11] for running the *mhsmm* package we needed to determine what variable associated with non-compliance to use as our emission function. At first glance, the criteria used to select returns for audit might seem a good candidate, because they are *de facto* indicators of likely non-compliance. We decided against using audit selection criteria, because by design all the cases in the study groups met these criteria in the TY2003 baseline.

What we sought were variables associated with compliance, preferably unconditionally, but where necessary, conditionally upon meeting audit selection criteria. These variables determine what functions we use to populate our data vectors, and how we specify the emissions distributions for the HMM fits. The analysis results we used to make these determinations are described in the Data Set-up section. We specified Normally distributed emissions for the Markov states, with starting parameters estimated via descriptive analysis of known change and no-change audit cases. The final Normal parameters for each Markov state's emissions are one component of the fitted HMM model that *mhsmm* provides as output. These are included in our Results section.

We used different strategies to compensate for inflation impacts on the dollar values reported on a return. The imputation models are all specific to a single Tax Year, and hence require no adjustment. The emissions for our second fit were based on TY2012

effect on his tax. We focus in our analysis only on effective EBE claims, that is, those where the Net Miscellaneous Deductions amount is positive.

[10] None of these other components of Gross Miscellaneous Deductions were available for this research.

[11] In contrast to other HMM fitting packages, *mhsmm* allows the user to stack short vectors for a large number of subjects into one long vector, and to use that as the data input. Once the data are stacked in this way, some pairs of sequential entries in the long data vector correspond to a change in subjects, rather than representing sequential emissions from the same subject. The user also provides *mhsmm* with a list of integers that tells the package how many entries in the long data vector sequence are associated with each subject.

constant dollars, using a Bureau of Labor Statistics inflation calculator.[12] The emissions for our first fit used a ratio of EBE claims for two successive years where the ratio is in log scale. These were not adjusted for inflation.

Our strategy for compensating for any selection bias that results from using operational IRS audit cases for our treatment groups was to use a non-audited control group that met basic audit selection criteria in TY2003, the baseline year. Our purpose in fitting HMMs was to see if this type of model could help gauge the impact of a particular IRS audit program on future compliance. Audit selection methods for the EBE audit program have been similar from year to year, so conclusions based on TY2003 study groups can help gauge the ongoing impact of the program on the portion of the EBE claimant population commonly subject to EBE operational correspondence audits.

3. Data Set-up

We used several IRS data sources in our analysis. We have already mentioned NRP and SOI data. In addition, we used IRS research databases of filed Form 1040 returns and associated schedules, and operational audit results for the correspondence EBE audits project.

3.1 Imputing Missing Employee Business Expenses

Our first HMM models apply only to taxpayers who continuously claim EBE, and the Markov states are compliant and non-compliant. Our first set of imputation models was estimated to use with our first set of HMM fits. SOI data includes panels, so we exploited the EBE claimants subset of that data to estimate a multiyear imputation model, where known EBE claims for TY2006–2008 were available as predictors of missing EBE claims for TY2003–2005. Net Miscellaneous Deductions (NMD) is available for all tax years, so our models impute non-zero EBE only when NMD is non-zero. We performed no out-of-sample validation on this first set of imputation models.

We created another imputation model for our second HMM fits. The framework for this set of fits included "no EBE claim" as one of the states in the Markov space. Our imputation model correspondingly was designed to allow "no EBE claim" as one of its predicted values. The imputation proceeds in two stages. First, a logistic regression predicts whether there is a claim. Conditional upon having predicted there is a claim a second stage linear regression predicts the size of the claim. The data set for the stage two model included all EBE claimants, rather than all *predicted* claimants, which made it unwise to use conventional model fit summaries to assess how our imputation methods work. To address this problem, we reserved a 15% holdout sample to carry out final performance evaluations of our second HMM fit imputation models. We evaluated both the first stage logistic regression and the composite two-stage model.

3.1.1 Continuous Claim Model

We looked at continuous EBE claimers who were in the SOI data from 2003 to 2008. This limited the number of observations used to build the model. We used available variables from the Form 1040 and Schedule A in the imputation target years and the three closest known years (TY2006-2008) to build a linear regression model using the log scale EBE claim. This approach worked best for TY2005 and worst for TY2003 which was the most distant from the known EBE values.

[12] http://www.bls.gov/data/inflation_calculator.htm

Table 2: Continuous Claim Multi-Year Variable Linear Regression Model

Label	*2003*	*2004*	*2005*
Intercept	--	--	--
Line Item #1	0.538	0.513	0.460
Line Item #2	1.271	1.353	0.993
EBE claim in 2006	1.141	1.148	1.201
EBE claim in 2007	N/A	N/A	0.376
Binary Indicator #3	-0.670	-0.531	-0.486
Binary Indicator #4	-0.334	-0.454	-0.403
R-Square	0.353	0.501	0.542
Adjusted R-Square	0.352	0.500	0.541
Number of observations used	2,510	2,510	2,510

Source: SOI Individual Sample Data Tax Years 2003 –2007

3.1.2 Two Stage Model for Continuous and Intermittent Claimers

Including intermittent claimers allows us to include a no-claim state in the HMM which brings the modeled behavior closer to the available real world states. For this model we focused on same year variables. We reserved 15% of the data for final model testing for each year. We split the remaining data into estimation (70%) and validation sets (30%). This modeling approach had a couple advantages: (1) we were able to include more observations (most people are not sampled continuously for the SOI sample), and (2) we avoided possibly inducing co-linearity in the HMM vectors by having the early imputed EBE values be dependent on later known values.

We also improved the model by using a two-step estimation process. The first step was a logistic regression focused on estimating the probability of an EBE claim. The majority of returns with positive Net Miscellaneous have an EBE claim but some do not. The second step was linear regression to estimate the log of EBE claim magnitude similar to the continuous claim model. The claim magnitude was modeled using only positive EBE claims. The variables in each year were similar in magnitude and sign. By multiplying the two predictions together we had a combined EBE estimate with a range that included the zero values.

For both the logistic and claim magnitude regression we found that generally the same line items and indicator variables were significant across the three imputation years.

Table 3: Logistic Results

Label	*2003*	*2004*	*2005*
Intercept	--	--	--
Line Item #1	0.720	0.812	0.829
Line Item #2	0.492	0.438	0.634
Ratio #1	2.200	3.108	2.336
Line Item #3	-0.138	-0.326	-0.269
Line Item #4	-0.350	-0.342	-0.348
Line Item #5	-0.397	-0.397	-0.371
Line Item #6	-0.197	-0.246	-0.283
Line Item #7	-0.162	-0.152	-0.053
Line Item #8	0.430	0.384	0.296
Percent Concordant	92.3	92.4	93.4
Percent Discordant	7.7	7.6	6.6
Number of observations used	14,448	14,921	20,059

Source: SOI Individual Sample Data Tax Years 2003 –2005

For our purposes a false positive imputation was worse than a false negative one so the $\hat{p}$ cut-off for predicting the presence of an EBE claim was increased from 0.5 to 0.75. Several $\hat{p}$ cut-offs were tested; 0.75 was the point which best balanced decreasing the false positives against increasing the missed true values.

The estimated proportion correctly predicted ranged from 0.88-0.89 for the three years in the weighted Estimation dataset and 0.91 for all three years in the weighted Final Test set.

Below is a comparison of the true and predicted values between the estimating data and holdout sample:

Table 4: Logistic Results

Confusion Matrix for Estimation dataset (where $\hat{p}$ >=.75 = EBE Claim)

2003			2004			2005		
	Predicted			*Predicted*			*Predicted*	
True	0	1	*TRUE*	0	1	TRUE	0	1
0	10%	6%	0	11%	5%	0	11%	5%
1	6%	78%	1	6%	78%	1	7%	78%

Confusion Matrix for Final Test data (where $\hat{p}$ >=.75 = EBE Claim)

2003			2004			2005		
	Predicted			*Predicted*			*Predicted*	
True	0	1	*True*	0	1	*True*	0	1
0	11%	5%	0	10%	4%	0	11%	4%
1	4%	80%	1	5%	81%	1	5%	80%

Source: SOI Individual Sample Data Tax Years 2003 –2005

Our continuous claimant regression for our first HMM fit provided candidate predictors for the stage two magnitude regression. To these predictors we added a ratio and several binary indicators.

Table 5: Stage Two Magnitude Regression Results

Label	*2003*	*2004*	*2005*
Intercept	--	--	--
Line Item #1	0.518	0.481	0.489
Line Item #2	1.440	1.416	1.355
Ratio #1	2.214	2.216	2.521
Binary Indicator #1	0.146	0.218	0.182
Binary Indicator #2	-0.171	-0.170	-0.169
Binary Indicator #3	-0.132	N/A	N/A
Binary Indicator #4	0.359	0.226	0.393
R-Square	0.698	0.665	0.685
Adjusted R-Square	0.697	0.665	0.685
Number of observations used	6,083	6,280	9,714

Source: SOI Individual Sample Data Tax Years 2003 –2005

To evaluate the composite two-step model we compared the root mean square error for the training and holdout (test) sample. The results, as with the logistic alone, were similar between the two sets. We used the composite (two stage) model to impute missing EBE values for the second HMM fit.

Table 6: Root Mean Square Error Table

Training Data		Holdout Data	
2003 Log Scale RMSE	2.40	*2003 Log Scale RMSE*	2.25
2003 RMSE	$4,213	*2003 RMSE*	$4,026
2004 Log Scale RMSE	2.29	*2004 Log Scale RMSE*	2.39
2004 RMSE 2004	$4,816	*2004 RMSE 2004*	$4,357
2005 Log Scale RMSE	2.30	*2005 Log Scale RMSE*	2.35
2005 RMSE	$4,444	*2005 RMSE*	$4,392

Source: SOI Individual Sample Data Tax Years 2003 –2005

3.2 Emissions Analysis

3.2.1 HMM fits for continuous EBE claimants only

For our first HMM fit, we used NRP data to estimate the distribution of one year changes in EBE claims, among the population of continuous claimers of EBE. NRP data includes audit results for a stratified sample of all individual income tax returns.[13] Among the NRP cases audited for EBE compliance, there appeared to be partial separation between compliant and non-compliant cases, so we used the location and spread parameters from these distributions of one year differences in logged EBE values as our initial values for describing emissions for the two Markov states.

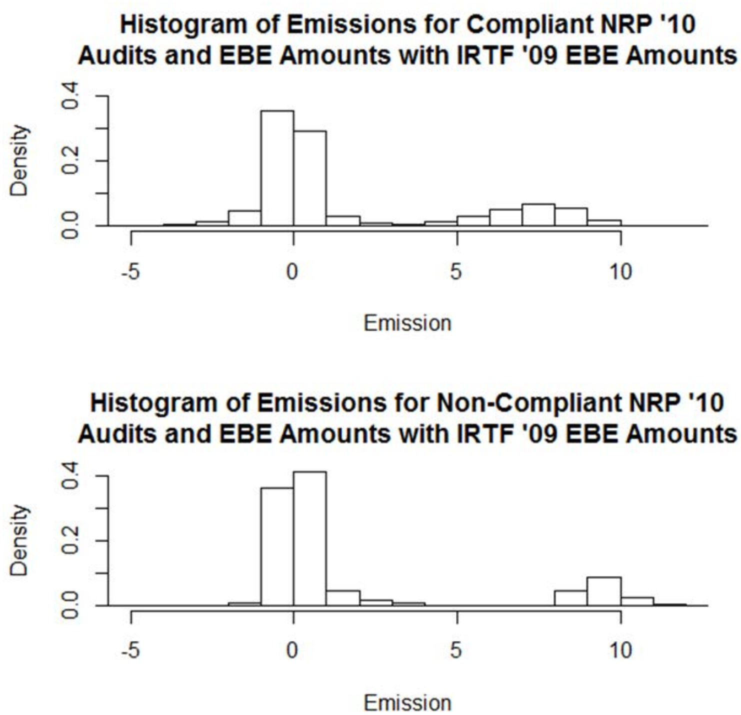

Figure 1: Histograms of calculated emission function on NRP TY2010 audits.

3.2.2 HMM fits including intermittent EBE claimants

Since the no-claim state is not hidden, we specify an initial emissions probability density that is concentrated in a displaced interval where the initial densities the other two Markov states are close to zero. In addition, as we built the input data, we set observed emissions for no-claim years to values at the center of this displaced interval. Doing so made it unlikely that the fitting algorithm would even temporarily infer that a no-claim emission belonged to any but the no-claim Markov state.[14]

[13] Under NRP protocols, all line items on a return are potentially subject to audit, even items that do not necessarily meet selection criteria for routine tax audits.
[14] The final fit preserved this large displacement of the no-claim emission distribution.

Using operational audit data from TY2006-2010 we estimated a logistic regression of audit change vs. no-change on tax return characteristics.[15] This logistic regression provided the basis for formulating the emission function for compliant and non-compliant emissions and for selecting initial parameters for these states' emission distributions. We compared histograms of the fitted values for the compliant and non-compliant cases in our logistic regression data set. The two distributions could be described as similar but slightly different mixtures of Normal distributions. The compliant distribution had a smaller expectation than the non-compliant, but the variances of both distributions were large enough that the densities had considerable overlap. We used the parameters of these empirical distributions as initial values for *mhsmm*. The initial emission distribution for the no-claim state was defined to have extremely little overlap with the other two initial densities, centered well to their left and with a small spread.[16]

When there was a claim, the emission function was based on indicators for Interest and Ordinary Dividends, and Capital Gains and Losses, Total Itemized Deductions as reported on Form 1040, and Net Miscellaneous Deductions as reported on Schedule A.

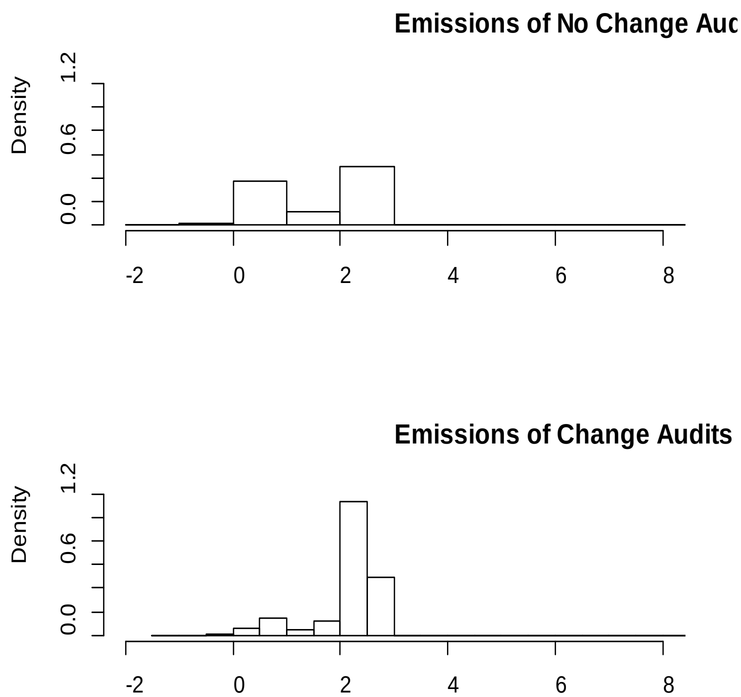

Figure 2: Histograms of calculated emission function on multi-year sets of no-change and change EBE audits.

Once we had decided on an emission function we needed to create the emission data vectors for our three study groups. The three data files consisted of taxpayers who were

[15] This logistic regression differs from the one used for imputation.

[16] Recall that the no-claim state is considered to be a directly observable type of compliance, hence our decision to center it to the left of the compliant non-zero claim, which is already slightly to the left of the non-compliant non-zero claim.

audited for EBE in Tax Year 2003 and received an adjustment (the change group); taxpayers who were audited for EBE in TY2003 and received no adjustment (the no-change group); and taxpayers who were eligible for an EBE audit in TY2003, but did not receive one (the no-audit group). Observations began in returns for TY2005, at which point most taxpayers would know the outcome of their audits, and continued as far as TY2012. The string of observations for any taxpayer terminated if they were audited again or stopped filing altogether. Taxpayers were not dropped from the sample if they claimed no EBE for one or more years; instead, “no EBE claim” was defined as a third, technically non-hidden Markov state, with its emission normal around -10 with a small variance. The following table and histograms summarize the three study data files for the second HMM fit.

Table 6: Study Groups

Group	*Number of taxpayers*	*Number of observations*
Audit with Tax Change 2003	8,692	58,440
Audit with No Change 2003	2,384	16,750
No Audit 2003	16,269	111,182

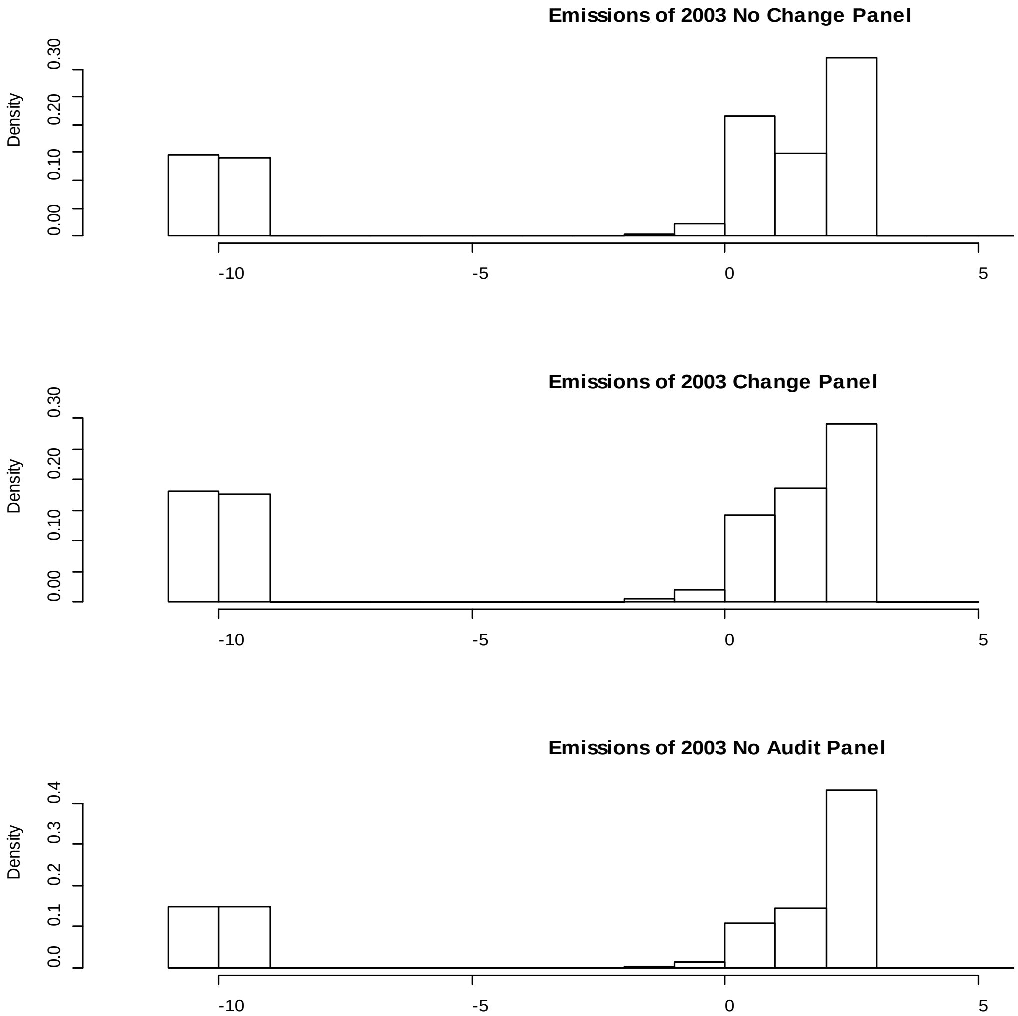

Figure 3: Histograms of calculated emission function for each analysis set.

4. Results

We fit two sets of HMM models, one set to only continuous EBE claimants, and the second set with intermittent claimants included. For both sets of fitted models we had three groups of subjects. The first group was a random sample of taxpayers with Net Miscellaneous Deductions who were not audited for EBE but whose imputed EBE claims for TY2003 met the criteria for possible non-compliance. The second consisted of taxpayers whose TY2003 EBE claims were audited but with no resulting tax change. The third consisted of taxpayers whose TY2003 EBE audits led to a tax change.

4.1 HMM fits for continuous EBE claimants

The emissions function for these models was a one year difference in logged EBE claims (equivalently the log of the ratio of two successive claims). The *mhsmm* R package gave the estimated transition matrices shown in Tables 7-9.

Table 7: Estimated Transition Matrix for Unaudited Continuous Claimants

	Compliant	*Non-Compliant*
Compliant	0.668	0.332
Non-Compliant	0.118	0.882

Source: Form 1040 and Schedule A filings 2003 - 2012

Table 8: Estimated Transition Matrix for Audit with No Change Continuous Claimants

	Compliant	*Non-Compliant*
Compliant	0.751	0.249
Non-Compliant	0.151	0.849

Source: Form 1040 and Schedule A filings 2003 - 2012

Table 9: Estimated Transition Matrix for Audit with Tax Change Continuous Claimants

	Compliant	*Non-Compliant*
Compliant	0.602	0.398
Non-Compliant	0.225	0.775

Source: Form 1040 and Schedule A filings 2003 - 2012

In all three groups the non-compliant state is “stickier” than the compliant. The probability of staying in the same state from one year to the next in the post audit period is higher for non-compliant taxpayers. Compared to the control group, both audited groups are less likely to remain in the non-compliance Markov state having once entered it, but the decrease is more pronounced for the tax change group. In this sense audits improve compliance, regardless of the outcome. On the other hand, the no-change audit group is less likely to remain in the compliance state having once entered it, even though the tax change group is more likely to do so. In this sense, the no-change audit has a perverse effect on future compliance. Exposing a seemingly compliant taxpayer to an unnecessary audit could worsen his future compliance.

After a tax change audit, the non-compliant Markov state becomes less “sticky”. After a no-change audit, the compliant Markov state becomes more “sticky”.

4.1.1 Sojourn distributions

Tables 10-12 show the computed probabilities for geometric distributions with the diagonal elements from the estimated transition matrices as the Bernoulli parameters. Our discussion refers to how long it takes a taxpayer to switch states. By this we mean how long it takes to make the *first* switch. Some taxpayers will of course switch back and forth between compliance states. The larger the one-year transition probabilities in the first row of the table, the more volatile state occupancy will be over a set period of follow-up years. In this sense, no-changed audited taxpayers are less volatile than unaudited controls, and those audited with tax change are more so.

When unaudited continuous EBE claimants start out non-compliant we estimate they remain so for many years; it takes six years before at least half of them have switched to the compliant state (the cumulative probability for the sixth follow-up year is 0.529). When unaudited continuous claimants start out compliant, over half of them have switched to non-compliance within two years (the cumulative probability for the second follow-up year is 0.554). Apparently, EBE audits are necessary to maintain compliance in the group of taxpayers whose EBE claims meet IRS audit selection criteria.

In contrast to the unaudited controls, we estimate that when a no-change taxpayer becomes non-compliant he returns to compliance within five years rather than the six years it takes for the unaudited group.[17] After a no-change audit we estimate that within three years, rather than two, he will have switched to non-compliant.

The tax change group has the shortest estimated sojourn times. We estimate that within three years over half (53%) of the initially non-compliant have switched to compliance, while within two years almost two thirds (64%) of the initially compliant have switched to non-compliance. In comparison to the unaudited, when a taxpayer's EBE audit has led to a tax change but he continues to claim EBE, his EBE compliance is estimated to be more changeable from year to year. This is not the case for the audited but no-changed continuous EBE claimant.

We noticed some unexpected behavior of higher powers of the transition matrices, which consist of k-step transition probabilities, where k is the power of the one-step matrix. The two audited groups' k-step matrices converged to very similar matrices within 16 steps, while the non-audited group's matrix converged to a different matrix with a more "sticky" non-compliant state. It is unrealistic to expect an audit impact to persist for as long as 16 years, so we would have expected all three k-step transition matrices to converge to a common matrix.[18] Using the "remain in current state" probabilities, the expected sojourn times for the three groups are given in Tables 10-12.

[17] IRS audits are not infallible, so some small percentage of no-change taxpayers may have been non-compliant. In addition, some compliant no-change taxpayers may immediately switch to non-compliance after their audit.

[18] To check whether our expectation of a common limiting k-step matrix was justifiable, before fitting our second set of HMMs (which allowed for a no claim state), we fit a directly observable Markov model to a two-state space, with the states being EBE claim and no EBE claim. The k-step transition matrices for the three groups took longer to converge, closer to 30 steps, and when they did it was the no-audit and audit no-change groups whose matrices resembled each other. The no-claim state was stickier for the audit tax change group than for the other two groups, both in the one-step and the long-term stable k-step matrices. We thus found the behavior of very long-term transition matrices estimated by HMM somewhat contradicted by the behavior of analogous matrices fit to a fully observable claim/no-claim Markov process.

Table 10: Estimated Sojourns for Un-audited Continuous Claimants

	Probability of move from non-compliant to compliant	*Cumulative Probability*	*Probability of move from compliant to non-compliant*	*Cumulative Probability*
1	11.8%	11.8%	33.2%	33.2%
2	10.4%	22.2%	22.2%	55.4%
3	9.2%	31.4%	14.8%	70.2%
4	8.1%	39.5%	9.9%	80.1%
5	7.1%	46.6%	6.6%	86.7%
6	6.3%	52.9%	4.4%	91.1%
7	5.6%	58.5%	2.9%	94.0%

Source: Form 1040 and Schedule A filings 2003 - 2012

Table 11: Estimated Sojourns for Audit with No Change Continuous Claimants

	Probability of move from non-compliant to compliant	*Cumulative Probability*	*Probability of move from compliant to non-compliant*	*Cumulative Probability*
1	15.1%	15.1%	24.9%	24.9%
2	12.8%	27.9%	18.7%	43.6%
3	10.9%	38.8%	14.0%	57.6%
4	9.2%	48.0%	10.5%	68.1%
5	7.8%	55.8%	7.9%	76.0%
6	6.7%	62.5%	5.9%	81.9%
7	5.6%	68.1%	4.5%	86.4%

Source: Form 1040 and Schedule A filings 2003 - 2012

Table 12: Estimated Sojourns for Tax Change Continuous Claimants

	Probability of move from non-compliant to compliant	*Cumulative Probability*	*Probability of move from compliant to non-compliant*	*Cumulative Probability*
1	22.5%	22.5%	39.8%	39.8%
2	17.4%	39.3%	24.0%	63.8%
3	13.5%	53.0%	14.4%	78.2%
4	10.5%	63.5%	8.7%	86.9%
5	8.1%	71.6%	5.2%	92.1%
6	4.9%	76.5%	3.1%	95.2%
7	3.8%	80.3%	1.9%	97.1%

Source: Form 1040 and Schedule A filings 2003 - 2012

4.2 HMM fits for all EBE claimants

The emissions function for cases with an EBE claim was computed based on a logistic regression fit to known compliant and non-compliant claims. When there was no EBE claim the emissions was set to a large negative value. The fitted transitions for the three groups are given in Tables 13-15.

Table 13: Estimated Transition Matrix for Un-audited EBE Claimants

	Compliant	*Non-Compliant*	*No claim*
Compliant	0.755	0.047	0.198
Non-Compliant	0.109	0.745	0.146
No claim	0.146	0.076	0.778

Source: Form 1040 and Schedule A filings 2003 - 2012

Table 14: Estimated Transition Matrix for Audited with No-change EBE Claimants

	Compliant	*Non-Compliant*	*No claim*
Compliant	0.779	0.0525	0.1685
Non-Compliant	0.1175	0.778	0.1045
No claim	0.158	0.044	0.798

Source: Form 1040 and Schedule A filings 2003 - 2012

Table 15: Estimated Transition Matrix for Audited with Tax Change EBE Claimants

	Compliant	*Non-Compliant*	*No claim*
Compliant	0.7324	0.0383	0.2293
Non-Compliant	0.103	0.717	0.180
No claim	0.153	0.068	0.779

Source: Form 1040 and Schedule A filings 2003 - 2012

In all three groups, the no-claim state is the stickiest, and the non-compliant the least sticky, but the difference is noticeable only for the tax change group. No-change taxpayers are a bit less likely to migrate into the no-claim group than the unaudited, while tax change taxpayers are more like to do so. Migration into the compliant state is similar for all three groups (the first columns of each matrix look similar), but slightly lower for both audited groups, compared to the un-audited. Migration from compliant to non-compliant gets a little worse when the audit results in no tax change, and a little better when it results in a tax change, using the un-audited as a reference point.

We looked again at the behavior of the k-step transition matrices. This time it was the no-audit and audit with tax change cases that most closely resembled each other. After 20 steps the audit no-change matrix had uniformly higher transitions to the compliant state and lower transitions to the no-claim state, compared with the no-audit and tax change matrices.

5. Conclusions

Hidden Markov Models produce estimates of post-audit EBE compliance behavior that concur to some extent with the findings of other research on audit persistence.[1] We found some evidence that undergoing a no-change audit can adversely affect a taxpayer's compliance, but the impact is not large.

Acknowledgements

Thank you: Barry Johnson, Director SOI, for the data and support, Stephen Klotz of SB/SE Research Group 4 for the time to do this project, and Thi Nguyen of OCA/RAAS for portions of the analysis.

References

1. Beer, S, et al. "Audit Impact Study", TAS 2015 Annual Report to Congress, Vol 2. (2016).
2. DeBacker, Jason, et al. "Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting." (2015).
3. O'Connell, Jared, and Søren Højsgaard. "Hidden semi markov models for multiple observation sequences: The mhsmm package for R." *Journal of Statistical Software* 39.4 (2011): 1-22.
4. O'Connell, J., and S. Højsgaard. "mhsmm: Parameter Estimation and Prediction for Hidden Markov and Semi-Markov models for Data with Multiple Observation Sequences." *R package version 0.4.14* (2015).
5. Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77.2 (1989): 257-286.