STUDY ON ASYLUM SEEKERS IN EXPEDITED REMOVAL

As Authorized by Section 605 of the International Religious Freedom Act of 1998

SELECTED STATISTICAL ANALYSES OF IMMIGRATION JUDGE RULINGS ON ASYLUM APPLICATIONS, FY 2000-2003

FEBRUARY 2005

Patrick Baier
National Opinion Research Center (NORC)

TABLE OF CONTENTS

Preface

The *Study of Asylum Seekers in Expedited Removal* (the Study) was undertaken by the *U.S. Commission on International Religious Freedom* (the Commission) to respond to four questions posed by Congress in Section 605 of the *International Religious Freedom Act* (IRFA) of 1998.

Specifically, the Study is to determine whether immigration officers performing duties under section 235(b) of the *Immigration and Nationality Act* (INA) (8 U.S.C. 1225(b)) with respect to aliens who may be eligible to be granted asylum are engaging in any of the following conduct:

(A) Improperly encouraging such aliens to withdraw their applications for admission.

(B) Incorrectly failing to refer such aliens for an interview by an asylum officer for a determination of whether they have a credible fear of persecution (within the meaning of section 235(b)(1)(B)(v) of such Act).

(C) Incorrectly removing such aliens to a country where they may be persecuted.

(D) Detaining such aliens improperly or in inappropriate conditions.

The Study has several components, including direct data collection, thorough sample file reviews, direct observations of the removal process, and interviews with individuals seeking asylum. The study also made a systematic effort to review previous studies, notably the 2000 *General Accounting Office* (GAO) Study and to compile statistical tabulations that either already existed or which were requested by the Commission from administrative data available from the agencies involved in Expedited Removal.

The present report presents an analysis of administrative data tabulated by the Commission for the Study from the U.S. Department of Justice, *Executive Office for Immigration Review* (EOIR). The compilation and accompanying descriptive summaries were prepared under my general direction by Patrick Baier with assistance from Fritz Scheuren, Cory Fleming and Susan Kyle.

Let me also take this opportunity to again express my deep appreciation for the care, diligence, speed, and expertise of the EOIR staff led by Marta Rothwarf. These were Amy Dale, Kevin Chapman, Scott Rosen, and especially Isabelle Chewning, Brett Endress and Cecelia Espenoza.

Mark Hetfield
Immigration Counsel
U.S. Commission on International Religious Freedom
October 2004

1 Introduction

This report analyzes data about asylum applications, collected during the Fiscal Years 2000 to 2003 at fourteen (14) U.S. Immigration Courts: Atlanta, Chicago, Elizabeth (including Queens), Houston, Krome, Lancaster, Los Angeles, Miami, Newark, New York City, San Juan (Guaynabo), San Francisco, San Diego, and San Pedro. The report presents statistical summaries and highlights statistically significant differences in decisions on asylum applications, both across courts and among the judges at an individual court.

There are significant differences in the acceptance rates of asylum applications from court to court. However, the assignment of asylum cases to courts is clearly not random, but is determined by the applicant's port of entry. Whether the observed differences are due to the different profiles of asylum seekers at different courts, or whether other reasons are involved is subject for further research. Similarly the data shows differences in the decisions reached by individual judges at a court, but again this report refrains from interpreting these differences beyond a mathematical analysis of their statistical significance.

In Section 2 below, the data underlying these analyses is described. Section 3 determines the acceptance rates for asylum applications found at different courts. Section 4 is a brief introduction to Analysis of Variance (ANOVA), the main tool used for studying the effect of courts and judges on the outcome of asylum applications. This section is technical and not necessary for a broad understanding if the findings presented in this report. The following Section 5 applies these techniques to courts and judges. Concluding remarks are made in Section 6.

2 Expedited Removal Data

The expedited removal data for this report were provided as a collection of tables [4], one for each of the 14 immigration courts, which display summary data of decisions made by the individual judges at the court. The identities of the judges or of the applicants were not revealed on the tables, and only summary data were displayed.

The data used for this analysis comprise the time period of Fiscal Year 2000 through Fiscal Year 2003. The *Department of Homeland Security* (DHS) provided EOIR with a file of 40,694 credible fear receipts for the period October 1, 1999 through September 30, 2003. EOIR manipulated the file to eliminate duplicate records, and was left with a file of 40,206 records. Of these, EOIR was able to match 36,799 in its ANSIR system (91.5%) [4].

The categories shown in the tables classify the primary outcomes of these selected cases as follows:

1. Asylum granted
2. *Convention Against Torture* (CAT) withholding or deferral granted
3. Application for asylum or CAT relief withdrawn
4. Ordered removed
5. Adjustment of status granted

There were a few cases that did not fit into these categories; e.g., cases granted some other form of relief. Note also that these categories are not mutually exclusive; the same case may be counted in more than one category. For example an alien who withdraws an application for relief may subsequently file for another form of relief, or may be ordered removed by the Immigration Judge. For this reason some of these cases may be counted more than once in this table [4]. It was not possible to identify such multiple entries, and here lies a potential source of non-sampling error.

This report is concerned with the data from fourteen (14) courts.[1] Included are only the cases where the judge either granted asylum to the applicant, or a removal order was issued. Other outcomes, such as deferrals, withdrawals or adjustments of status were excluded. Removal Orders include

[1]The EOIR determined that providing the complete data from all immigration courts nationwide would be a too large task; hence a (non-random) sample of 14 courts was selected and provided for this study. This report does not aim to make any inferences to other courts not part of this study.

the decisions of: Deportation Order, Exclusion Order, Removal Order, Voluntary Departure Orders, and DHS Expedited Removal Orders affirmed by an Immigration Judge [4].

3 Acceptance Rates

This section provides summaries of the data used for the analyses in this report in Subsection 3.1. Displayed are the numbers of accepted and rejected asylum applications by court. Subsection 3.2 displays graphically the corresponding acceptance rates, together with confidence intervals around the estimates.

3.1 Summary data by court

The table below lists the 14 immigration courts which are part of this study. The data used are summarized by court. The more detailed data at the level of individual judges used for this analysis is available separately [3, 4].

Table 1. Data Summary by Court

Code	Court	Number of Judges	Asylum granted	Ordered Removed	Total
ATL	Atlanta, GA	12	37	516	553
CHI	Chicago, IL	9	103	494	597
ELZ	Elizabeth, NJ [2]	14	689	1407	2096
HOU	Houston, TX	8	48	236	284
KRO	Krome, FL	10	47	789	836
LAN	Lancaster, CA	5	46	241	287
LOS	Los Angeles, CA	53	202	548	750
MIA	Miami, FL	25	578	4676	5254
NEW	Newark, NJ	9	155	585	740
NYC	New York City, NY	64	1925	5386	7311
SAJ	Guaynabo, PR	12	* [3]	*	51
SFO	San Francisco, CA	37	333	374	707
SND	San Diego, CA	11	332	868	1200
SPD	San Pedro, CA	9	64	109	173

3.2 Acceptance rates by court

The acceptance rate p_i at court i is

$$p_i = \frac{a_i}{a_i + r_i}$$

where a_i is the total number of accepted cases at court i (from column "Asylum granted"), and correspondingly r_i are the total rejections (see column "Ordered removed").

Figure 1 shows estimated acceptance rates for the fourteen different immigration courts, together with two-sided 95% confidence intervals around the estimates. The vertical lines indicate the point estimates of p_i for the courts, while the horizontal bars are 95% confidence limits around the acceptance rates.

For example, the acceptance rate for Atlanta (ATL) is estimated to be $p_{\text{ATL}} = 6.6\%$, and from the size of the sample we can estimate the margin of error to be $\pm 2\%$. The overall average of 21.89% is displayed as a vertical line through the data. Newark is the court closest to this average.

The interpretation of this statistical data is as follows. If we assume that every immigration court consistently applies the same policies and procedures over time, and that any factors that may influence decisions (legal or administrative procedures, personnel appointments, political events, etc.) remain constant, we can model the decision made on an application as a "binomial variable" (a variable that has only two possible outcomes - success or failure). Such variables are completely determined by a single parameter: The probability of a "success" (acceptance of an asylum application). Under the assumptions outlined above we can treat the value of p_i (the success or acceptance rate at court i) as a characteristic of the court.

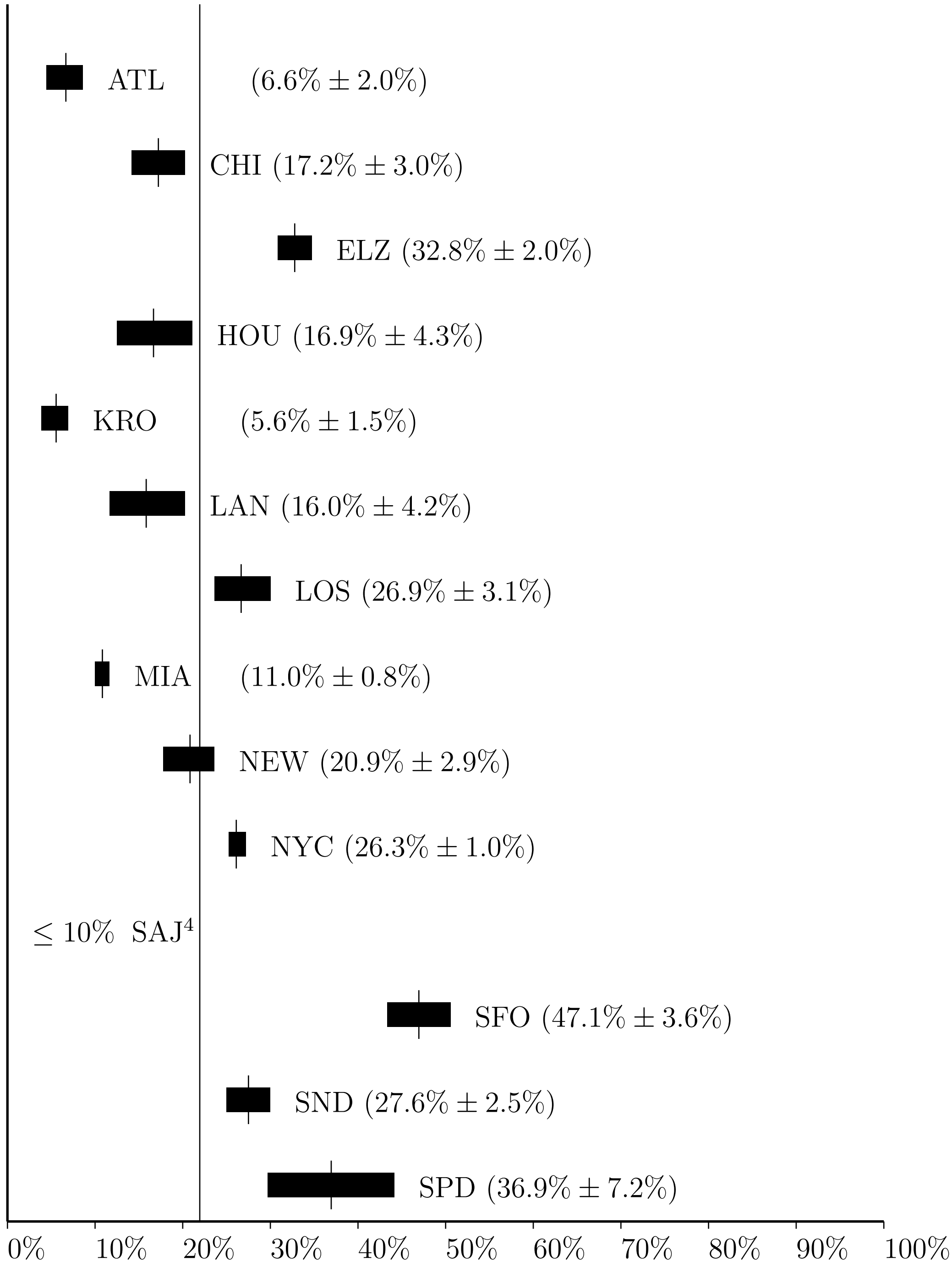
We have the (nearly) complete data for a fixed period of time and therefore we can directly compute the acceptance rate, at least for this period of

[2]Includes Queens, NY

[3]$*$ = five or fewer cases. According to DHS confidentiality policies, the actual number is not disclosed.

[4]The percentage corresponds to five or fewer cases and hence cannot be disclosed under DHS rules.

Figure 1. Court Acceptance Rates



time. However, it is common to consider such data as coming from a theoretical "super-population" of all possible asylum cases that might come to the court, and use the collected data to estimate what the acceptance rate for this hypothetical super-population would be.

Statistically, this amounts to treating the sample as a random sample from an infinite population to which we can make an inference (see [1]).

4 Analysis of Variance (ANOVA)

This section starts with a brief and informal introduction to Analysis of Variance (ANOVA) - by way of examples and without mathematical detail - in Subsection 4.1. ANOVA is discussed in more detail in most basic textbooks on mathematical statistics. A good (but by far not the only) reference is [2]. The following Subsection 4.2 introduces some nomenclature needed to put the judge data into the framework of ANOVA.

4.1 An informal introduction to ANOVA

Analysis of variance is a technique that enables the statistician to identify "effects" that cause the data to vary and to determine the significance of these effects. Mathematically, ANOVA is very similar to (linear) regression, but it can be more generally applied to categorical variables as well as numeric variables. The "downside" is that it does not allow for as simple a graphical illustration as the regression line in linear regression, and hence it is often perceived as somewhat abstract.

We think of the decision by an immigration judge on an asylum application as a "random variable." Of course, this is not to suggest that the judge would toss a coin to come to his decision, but just that we do not know the merits of a case and the considerations made by the ruling judge, hence, lacking this knowledge, the judge's decision appears like a random variable to us.

However, we do have some basic information about the cases - the court where the case was submitted and which judge ruled on the case. In Figure

1 we see that acceptance rates differ significantly across courts. There are also differences in the acceptance rates of individual judges, as shown in the diagrams on pages 14 and 15. ANOVA allows us, loosely speaking, to quantify "how much of the overall variability in the decisions is accounted for by the court (and judge)."

The general idea is that the random variable of interest (the decision made on an asylum application) can be "modeled" by one or more known variables (the court and judge to which the case was referred), up to an error term. The better the model, the smaller the error term.

> **Example.** *Suppose the members of a national farmers' union report their annual corn crop yield per acre to their organization. The nationwide data is likely to display more variability than the data within a state or region, because relevant factors such as climate, temperature, soil condition etc. are relatively uniform within a small region, but not across geographically distant regions. Hence we expect to find different regional averages in the reported data, and the data within a region will be more tightly centered around its local (regional) mean or average. In other words, the data within a single region will have smaller variance than the nationwide data, because the effects of regional differences on the data are factored out.*

ANOVA is a mathematical procedure of decomposing the variance into a component that comes from the model (variance between regions) and an error component (part of the overall variance that cannot be explained by regional differences).

If more auxiliary information is available about the population (e.g. farming methods employed by a farmer), it might be possible to strengthen the model and "explain" an even larger component of the variance. Once the best fitting model is found, the remaining "error component" is the remaining uncertainty or variability about which we cannot make any predictions.

4.2 Application to the asylum data

We have a population of $N = 20,839$ asylum cases (the total number of cases where a decision was made to either grant asylum or order the removal of an applicant; other cases, such as deferrals, were excluded).

The population is divided into courts ($n = 14$ courts) and within courts it is further divided by judge, where every judge is at only one court and the number of judges per court varies.

In order to avoid unstable estimates for judges who hear only very few cases, at each court the judges with the fewest cases were combined and treated as one single judge, so that each "judge" had at least 14 cases, but still retaining at least two judges per court to be able to look at effects across judges.[5]

The following table shows how this collapsing was done. We use

$$
\begin{aligned}
n_o &= \text{Original number of judges} \\
n_c &= \text{Number of judges collapsed} \\
n_f &= \text{Final number of judges after collapsing} \\
n_1 &= \text{Minimum number of cases per judge after collapsing} \\
n_2 &= \text{Upper bound on the number of cases per judge} \\
& \text{below which judges were combined}
\end{aligned}
$$

[5]The judges were ordered by their number of processed cases. The judges with the fewest cases were combined and treated as one single judge for the analysis. The general rule for determining how many judges should be collapsed was to use the minimum number of judges so that both the combined "pseudo judge" and all the remaining judges had 14 cases or more. In some cases, the next smallest judge was included in addition if this judge would have otherwise been an outlier with respect to the number of processed cases; in other words, if the range of the judge sizes could be reduced by expanding the collapse. While this decision rule is somewhat *ad hoc*, it was deemed to be appropriate for brining the data into a form where a meaningful analysis of judge and court effects could be performed.

Table 2. Judges and number of cases per judge.

Court	n_o	n_c	n_f	n_1	n_2
ATL	12	9	4	43	15
CHI	9	2	8	22	20
ELZ	14	9	6	52	14
HOU	8	2	7	26	20
KRO	10	6	5	28	12
LAN	5	2	4	17	12
LOS	53	29	25	15	10
MIA	25	2	24	19	18
NEW	9	2	8	46	45
NYC	64	24	41	18	11
SAJ	12	11	2	15	5
SFO	37	21	17	14	7
SND	11	2	10	53	46
SPD	9	8	2	72	63

(See Table 1 for the court abbreviations.) The original number of judges n_o can be determined from n_c and n_f as

$$n_o = n_c + n_f - 1.$$

5 Sum of squares decomposition

The decomposition of the sum of squares, the basic mathematical procedure underlying ANOVA, is carried out in Subsections 5.1 and 5.2. Subsection 5.3 derives the mean square errors and Subsection 5.4 the F-ratios. Mean squares are normalized by the degrees of freedom to make them comparable and determine the significance of effects from the quantiles of a standard F-distribution.

5.1 Court effects

We label population elements (applications) by triples (i, j, k) where $1 \leq i \leq n = 14$ is the court, $1 \leq j \leq n_i$ is the j-th judge at court i (where there are a total of n_i judges), and $1 \leq k \leq n_{ij}$ is the k-th case heard by judge j at court i.

Let x_{ijk} be a binomial variable with $x_{ijk} = 1$ if case (i, j, k) is accepted and $x_{ijk} = 0$ otherwise.

In total, 4562 out of 20839 cases have been accepted, which gives us an overall acceptance rate of

$$p = \frac{\sum\limits_{i,j,k} x_{ijk}}{\sum\limits_{i,j,k} 1} = \frac{4562}{20839} = 21.89\%.$$

Let p_i be the acceptance rate at court i and p_{ij} the acceptance rate for judge j at court i.

$$\begin{aligned} p_i &= \frac{\sum\limits_{j=1}^{n_i} \sum\limits_{k=1}^{n_{ij}} x_{ijk}}{\sum\limits_{j=1}^{n_i} n_{ij}}, \\ p_{ij} &= \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk} \end{aligned}$$

Note that the overall acceptance rate is just the overall average or mean of the "flag" variable x, and the acceptance rate for a court is accordingly the average or mean over just that court, and similarly by judge.

$$p = \bar{x}_{\cdots}, \quad p_i = \bar{x}_{i,\cdot\cdot}, \quad p_{ij} = \bar{x}_{ij,\cdot} \quad .$$

A one-way analysis of variance (using the stratification by court only) can be carried out by decomposing the total sum of squares (squared differences from the mean), given in (1) below.

$$SS_{\text{Tot}} = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - p)^2 \tag{1}$$

The variance of x is the total sum of squares divided by the degrees of freedom ($df = N - 1$, where N is the total number of cases), or roughly the average (rather than total) square deviation.

$$\text{Var}(x) = \frac{SS_{\text{Tot}}}{N - 1}.$$

However, the algebra is simpler if we decompose the sum of squares of x, rather than the variance.

Recall the definitions

$$SS_{\text{Court}} = \sum_{i=1}^{n} n_i(p_i - p)^2 \tag{2}$$

$$SS_{\text{Err}} = \sum_{i=1}^{n} \left(\sum_{j,k} (x_{ijk} - p_i)^2 \right) \tag{3}$$

We then have

$$SS_{\text{Tot}} = SS_{\text{Court}} + SS_{\text{Err}}.$$

This differs from the common look of the ANOVA formulas only in the fact that the inner summation in (3) is indexed jointly by (j, k); however, if we relabeled the pairs (j, k) by a single variable, say m, the above would just reduce to the known formulas for one-way ANOVA. This would be the standard one way ANOVA using only the court to model responses, but not the judge.

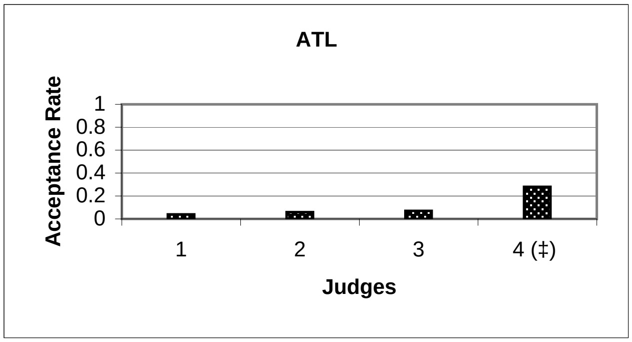
A calculation of this simple one-way model yields

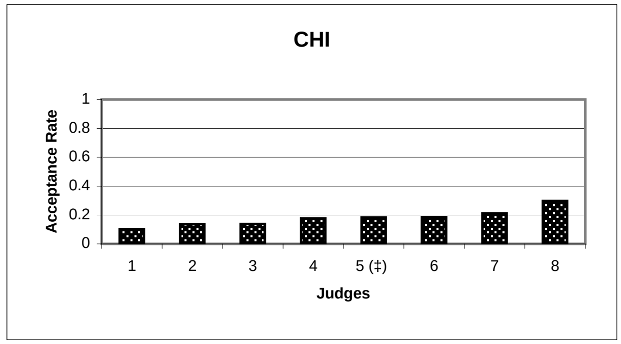
$$\begin{aligned} SS_{\text{Tot}} &= 3563.30 \\ SS_{\text{Court}} &= 196.03 \\ SS_{\text{Err}} &= 3367.27. \end{aligned}$$

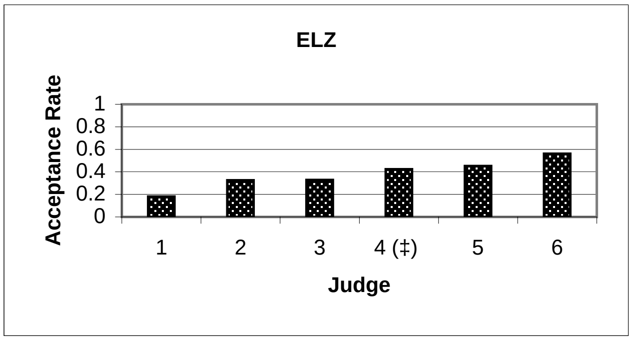
5.2 Judge effects

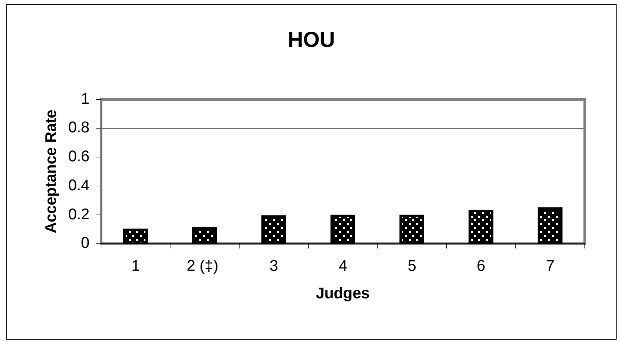
We cannot carry out a two-way ANOVA here that would include the judges as independent variables, because the second stratifier, judge, is only defined within a court. Hence we need to use a nested effect to incorporate the judges. First, we display graphically the different acceptance rates for each judge, separately by court.

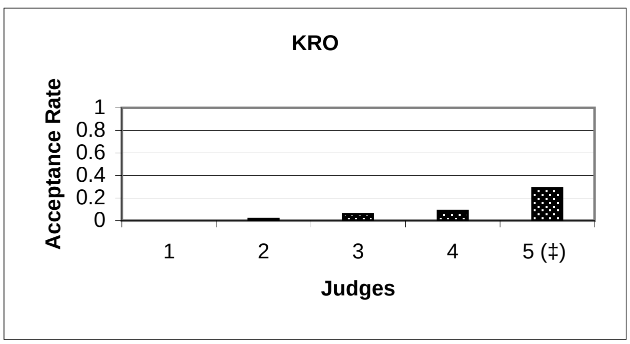
The judges are ordered by ascending acceptance rate. The "combined" judges are identified by a (‡). Their ordered ranks are as follows: Atlanta (4), Chicago (5), Elizabeth (4), Houston (2), Krome (5), Lancaster (1), Los Angeles (16), Miami (1), Newark (5), New York (27), Guaynabo (2), San Francisco (3), San Diego (1), San Pedro (1).

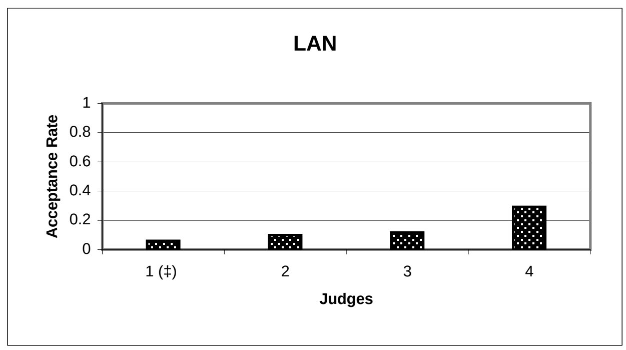

ATL
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1
2
3
4 (‡)
Judges

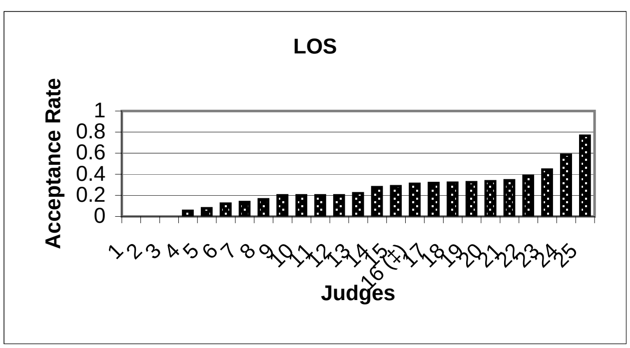

CHI
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5 (‡)
6
7
8
Judges

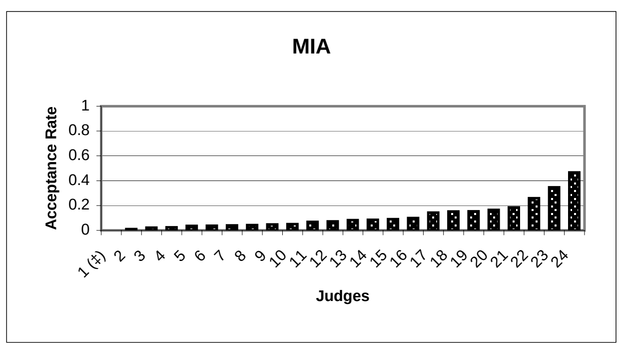

ELZ
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1
2
3
4 (‡)
5
6
Judge


HOU
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1
2 (‡)
3
4
5
6
7
Judges

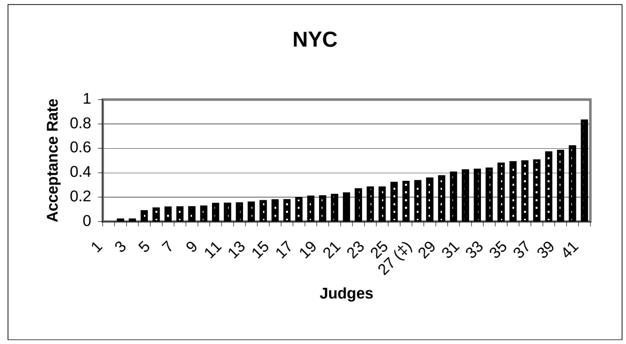

KRO
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5 (‡)
Judges

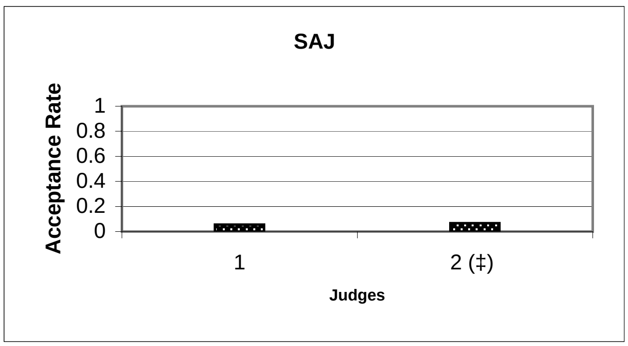

LAN
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 (‡)
2
3
4
Judges

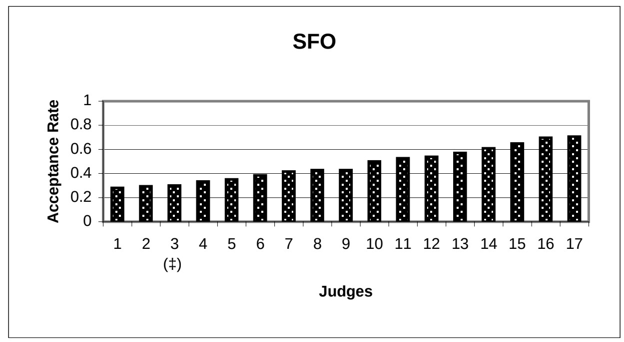

LOS
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16 (‡)
17
18
19
20
21
22
23
24
25
Judges

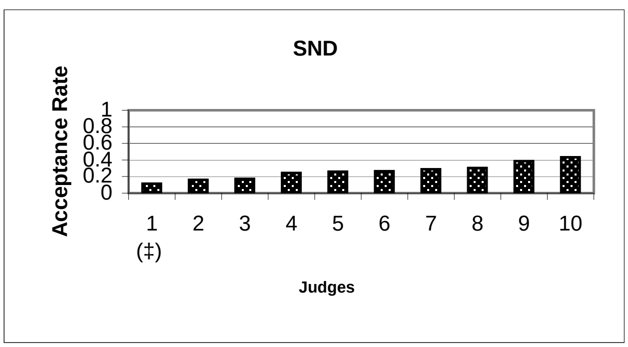

MIA
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 (‡)
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
Judges

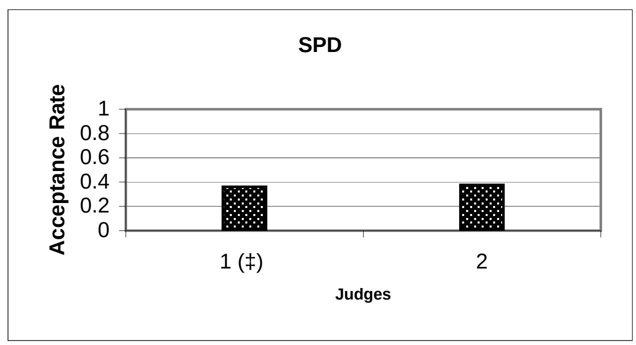

NEW
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 (‡) 6 7 8
Judges


NYC
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 (‡) 29 31 33 35 37 39 41
Judges


SAJ
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 2 (‡)
Judges


SFO
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 2 3 (‡) 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Judges


SND
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 (‡) 2 3 4 5 6 7 8 9 10
Judges


SPD
Acceptance Rate
1
0.8
0.6
0.4
0.2
0
1 (‡) 2
Judges

In order to grasp these differences mathematically, we restrict attention to the term in parentheses in the formula (3) for SS_{Err}. Note that within a fixed court (for fixed i), the judge at that court is a stratifying variable, and we can further decompose the sum of squares by splitting the term in parentheses into a model and an error term. Hence

$$\begin{aligned} SS_{\text{Err}} &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n_i} n_{ij}(p_{ij} - p_i)^2 + \sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - p_{ij})^2 \right) \\ &= \sum_{i=1}^{n} \left(SS_{\text{Judge},i} + SS^{(2)}_{\text{Err},i} \right) \\ &= SS_{\text{Judge}} + SS^{(2)}_{\text{Err}} \end{aligned}$$

By incorporating judges like this we can "model an additional component of the sum of squares" and hence reduce the error term further. We thus get

$$\begin{aligned} SS^{(2)}_{\text{Mod}} &= SS_{\text{Court}} + SS_{\text{Judge}} \\ &= SS_{\text{Court}} + \sum_{i=1}^{n} SS_{\text{Judge},i} \\ &= \sum_{i=1}^{n} \left(n_i(p_i - p)^2 + \sum_{j=1}^{n_i} n_{ij}(p_{ij} - p_i)^2 \right) \qquad (4) \\ SS^{(2)}_{\text{Err}} &= \sum_{i=1}^{n} SS^{(2)}_{\text{Err},i} \\ &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n_i} \sum_{k=1}^{n_{ij}} (x_{ijk} - p_{ij})^2 \right) \qquad (5) \end{aligned}$$

If we carry out these calculations for the given data set, we obtain

$$\begin{aligned} SS^{(2)}_{\text{Tot}} &= 3563.30 \qquad (6) \\ SS^{(2)}_{\text{Mod}} &= 371.78 + 196.03 = 567.81 \qquad (7) \\ SS^{(2)}_{\text{Err}} &= 2995.49. \qquad (8) \end{aligned}$$

To interpret this decomposition, we need to do further calculations with the sums of squares. This will take the rest of this section.

5.3 Mean squares

The reason for considering sums of squares rather than variances was that the formulas need not be adjusted for different "degrees of freedom"[6] associated with the model and error terms. However, as it stands, the absolute values of the model and error components do not give us direct information about the significance of effects.

We can calculate the mean squares for the model and the error by dividing the sum of squares through the appropriate degrees of freedom.

The simple one-way ANOVA ($df(\mathrm{Mod}) = 14 - 1 = 13$) has

$$MS_{\mathrm{Court}} = \frac{SS_{\mathrm{Court}}}{df(\mathrm{Court})} = \frac{196.03}{13} = 15.08 \tag{9}$$

$$MS_{\mathrm{Err}} = \frac{SS_{\mathrm{Err}}}{df(\mathrm{Err})} = \frac{3367.27}{20825} = 0.16. \tag{10}$$

The corresponding "root mean square error (RMSE)" is

$$\mathrm{RMSE} = \sqrt{MS_{\mathrm{Err}}} = \sqrt{0.16} = 0.40.$$

For the two-way model we get

$$MS^{(2)}_{\mathrm{Mod}} = \frac{SS^{(2)}_{\mathrm{Mod}}}{df^{(2)}(\mathrm{Mod})} = \frac{567.81}{162} = 3.51 \tag{11}$$

$$MS^{(2)}_{\mathrm{Err}} = \frac{SS^{(2)}_{\mathrm{Err}}}{df^{(2)}(\mathrm{Err})} = \frac{2995.49}{20676} = 0.14 \tag{12}$$

with root mean square error RMSE$= \sqrt{0.14} = 0.38$. Note the degrees of freedom are

$$df^{(2)}(\mathrm{Mod}) = \left(\sum_{i,j} n_{ij}\right) - 1 = 162,$$

[6] The "degrees of freedom" of a statistic, e.g. the mean or sum of squares of a sample, are roughly the number of input variables whose value can be varied freely without changing the statistic. For example, for an n-tuple $(x_1, \ldots, x_n)$ to attain a prescribed mean μ we can arbitrarily assign values to $n-1$ of the variables, say, $x_1, \ldots, x_{n-1}$. Then $x_n = n\mu - (x_1 + \ldots + x_{n-1})$ is determined by the requirement that the mean be μ. Hence this statistic, like the sum of squares, has $n-1$ degrees of freedom. More precisely, if the statistic can be given as a continuously differentiable function $f(x_1, \ldots, x_n)$ of its input variables, then $df = \dim(\ker(Df(x)))$ where x is a regular point of f (that is a point where the derivative $Df(x)$ has maximal rank).

and consequently

$$df^{(2)}(\text{Err}) = N - df^{(2)}(\text{Mod}) - 1 = 20676.$$

From the mean squares we can construct F-ratios which allow us to determine the significance of effects.

5.4 F-ratios

The F-ratio is commonly defined as

$$F = \frac{MS_{\text{Mod}}}{MS_{\text{Err}}}.$$

Under the null-hypothesis

$$H_0\colon p_{i1} = p_{i2} = \ldots = p_{in_i}$$

the F-ratio is a statistic whose distribution is an F-distribution. This distribution depends on two parameters, the degrees of freedom of the model and error terms. A large F-ratio (beyond the 95%-quantile of the F-distribution) would lead us to reject the null-hypothesis.

From our decomposition of the sum of squares into a court term, a judge term, and an error term, we get the following values

$$F_{\text{Court}} = \frac{MS_{\text{Court}}}{MS_{\text{Err}}} = \frac{15.08}{0.16} = 93.26 \tag{13}$$

$$F^{(2)}_{\text{Mod}} = \frac{MS^{(2)}_{\text{Mod}}}{MS^{(2)}_{\text{Err}}} = \frac{3.51}{0.14} = 24.19 \tag{14}$$

See (9) and (10) for (13), and (11) and (12) for (14). Both of those values are highly significant, even at a 99.9% confidence level, in other words, the p-values satisfy $p \ll 0.001$. This means that, given the observed data, we can be almost certain that the acceptance rates across courts, respectively judges (within and across) courts are not the same.

Finally, let us look at individual courts and the effect of judges by court. Note that we can write

$$MS^{(2)}_{\text{Mod}} = \frac{SS_{\text{Court}}}{df^{(2)}(\text{Mod})} + \sum_{i=1}^{n} \frac{SS_{\text{Judge},i}}{df^{(2)}(\text{Mod})}$$

$$= \frac{1}{df^{(2)}(\text{Mod})} \sum_{i=1}^{n} \left[n_i (p - p_i)^2 + \sum_{j=1}^{n_i} n_{ij} (p_i - p_{ij})^2 \right]$$

$$= \frac{df(\text{Court})}{df^{(2)}(\text{Mod})} MS_{\text{Court}} + \sum_{i=1}^{n} \frac{df(\text{Judge}, i)}{df^{(2)}(\text{Mod})} MS_{\text{Judge},i}.$$

We obtain the following values for the judge effects within individual courts:

Table 3. Sum of squares, mean squares, and F-ratios by court.

Court	$SS_{\text{Judge},i}$	df	$MS_{\text{Judge},i}$	$SS_{\text{Err},i}$	df	$MS_{\text{Err},i}$	F-Ratio
ATL	2.1706	3	0.7235	32.3538	549	0.0589	12.28
CHI	1.2779	7	0.1826	83.9516	589	0.1425	1.28
ELZ	18.6264	5	3.7253	443.8846	2090	0.2124	17.54
HOU	0.9452	6	0.1575	38.9421	277	0.1406	1.12
KRO	2.0980	4	0.5245	42.2596	831	0.0509	10.31
LAN	2.2894	3	0.7631	36.3378	283	0.1284	5.94
LOS	19.4101	24	0.8088	128.1846	725	0.1768	4.57
MIA	57.2373	23	2.4886	457.1761	5230	0.0874	28.47
NEW	6.8462	7	0.9780	115.6875	732	0.1580	6.19
NYC	238.3106	40	5.9578	1179.8334	7270	0.1623	36.71
SAJ	0.0013	1	0.0013	2.8222	49	0.0576	0.02
SFO	11.9307	16	0.7457	164.2249	690	0.2380	3.13
SND	10.6269	9	1.1808	229.5197	1190	0.1929	6.12
SPD	0.0096	1	0.0096	40.3141	171	0.2358	0.04
Total	371.7802	149	0.0025	2995.4921	20676	0.1449	24.19

There are only four courts, Chicago (CHI), Houston (HOU), Guaynabo (SAJ) and San Pedro (SPD), where the judge effects are insignificant, even at low confidence levels. That is, there is no indication in the available data that different judges accept asylum applications at different rates. In the two last cases, however, Guaynabo and San Pedro, all judges except the one with the largest number of cases were collapsed, so that only two "judges" were left to compare. Hence the judge effect is only of limited use since it essentially compares the judge with the largest number of cases against all others.

On the other side, the judge effects for five courts, Atlanta (ATL), Elizabeth (ELZ), Krome (KRO), Miami (MIA), and New York City (NYC), are highly significant, even at a 99.9% confidence level[7].

Note that the total sum of squares from this table plus the sum of squares from the one way ANOVA yields (see (5)) the sum of squares for the model using the nested judge effect.

$$\sum_{i=1}^{n} SS_{\text{Judge},i} + SS_{\text{Court}} = 371.78 + 196.03 = 567.81.$$

The corresponding degrees of freedom are

$$\sum_{i=1}^{n} df(\text{Judge}, i) + df(\text{Court}) = 149 + 14 - 1 = 162$$

and hence we can calculate the value

$$MS_{\text{Mod}}^{(2)} = \frac{567.81}{162} = 0.14,$$

as given in (7).

6 Conclusion

We observe that the overall variability in the decisions made on immigration and asylum applications can be modeled to some extent by the court where an application is processed and the judge handling it.

Obviously, great care is needed in drawing conclusions from the observed differences across courts since these may well be caused by differences in the

[7]For a judge who handles n_{ij} cases the number of granted applications is a $B(n_{ij}, p_{ij})$ distributed binomial variable. For large n_{ij} we can use the normal approximation to model this variable and test the null hypothesis $H_0: p_{i1} = \ldots = p_{in_i}$. In calculating the F ratios and their significance, we need to issue a note of caution that some of the judges (even after collapsing) had too few cases for their observed acceptance rate to attain normality. This makes the individual comparison of such judges to other judges and the calculation of type I and type II errors more difficult, especially since the standard error $\sqrt{\frac{p_{ij}(1-p_{ij})}{n_{ij}}}$ of a binomial variable depends on its mean p_{ij}. However, all courts were large enough overall so that a very large F-ratio is still a strong indicator for the failing of the null hypothesis.

applicant populations arriving at different courts, due to their geographic location and connection to global travel routes.

However, arguably, within a court the assignment of cases to judges may be "random" (in the sense that there is no association between the case itself and the judge whom it is assigned to). This would suggest that there should be no "judge effect." However, this is not supported by the data, and further research into the causes seems warranted.

References

[1] W.G. Cochran. *Sampling Techniques.* 3rd edition, Wiley, New York, 1977.

[2] R.V. Hogg, A.T. Craig. *Introduction to Mathematical Statistics.* 4th edition, Macmillan, New York, 1978.

[3] C. Fleming, S. Kyle, and F. Scheuren. Study of Asylum Seekers in Expedited Removal. *Statistical Report on Immigration Court Proceedings (FY 2000-2004).*

[4] U.S. Commission on International Religious Freedom. EOIR Tables - Judge Data, provided as a collection of Excel tables `EOIR_Tables_1.`n`-7(xxx).xls`