Central European Journal of Operations Research manuscript No.
(will be inserted by the editor)

Estimating the strength of poker hands by Integer Linear Programming techniques

Marcello Dalpasso · Giuseppe Lancia

Received: date / Accepted: date

Abstract We illustrate how Integer Linear Programming techniques can be applied to the popular game of poker Texas Hold'em in order to evaluate the strength of a hand. In particular, we give models aimed at (i) minimizing the number of features that a player should look at when estimating his winning probability (called his *equity*); (ii) giving weights to such features so that the equity is approximated by the weighted sum of the selected features. We show that ten features or less are enough to estimate the equity of a hand with high precision.

Keywords Poker Texas Hold'em · Integer Linear Programming · Equity

1 Introduction

1.1 The game of poker

No-limit Texas Hold'em [5,6] (NLTH) is a poker variant that has gained huge popularity in the past few years. This game is played with a full deck of 52 cards by two up to nine players. The game develops in four phases. In the first phase (called the *preflop*), each player is dealt two private cards (i.e., known

M. Dalpasso
Dipartimento di Ingegneria dell'Informazione, University of Padova
Via Gradenigo 6/A, 35131 Padova, Italy
Tel.: +39-049-8277941
Fax: +39-049-8277699
E-mail: marcello.dalpasso@unipd.it

G. Lancia
Dipartimento di Informatica e Matematica, University of Udine
Via delle Scienze 206, 33100 Udine, Italy
Tel.: +39-0432-558454
Fax: ++39-0432-558499
E-mail: giuseppe.lancia@uniud.it

Table 1 The ranking of poker hands

Royal flush	[A♣]	[K♣]	[Q♣]	[J♣]	[T♣]	A,K,Q,J,T of the same suit
Straight flush	[T♡]	[9♡]	[8♡]	[7♡]	[6♡]	five consecutive cards of the same suit
4 of a kind	[Q♡]	[Q♢]	[Q♠]	[Q♣]	[3♡]	four cards of the same rank
Full house	[9♣]	[9♡]	[9♠]	[A♡]	[A♢]	three cards of the same rank and a pair
Flush	[K♡]	[J♡]	[8♡]	[5♡]	[4♡]	five cards of the same suit
Straight	[J♣]	[T♣]	[9♡]	[8♢]	[7♠]	five consecutive cards
3 of a kind	[K♡]	[K♢]	[K♠]	[8♣]	[5♡]	three cards of the same rank
Two pair	[J♡]	[J♢]	[8♠]	[8♢]	[2♡]	two separate pairs
Pair	[Q♠]	[Q♢]	[K♡]	[9♢]	[6♣]	two cards of the same rank
High card	[K♢]	[J♡]	[9♡]	[7♣]	[5♢]	unrelated cards ranked by the highest card

only to him). These cards are called his *starting hand*. Then there is a round of betting in which some players fold (i.e., abandon the game), while others remain in play. Then three cards are turned face-up. These cards are called the *flop*, and they are community cards, i.e., they can be used by all players still in play. After another round of betting, if two or more players are still in play, a single community card, called the *turn*, is turned face-up. After another round of betting, another single community card is turned, called the *river*. A final round of betting follows. Each player still in play eventually computes the best 5-cards hand obtained by combining his starting cards with the five community cards (i.e., the best 5 cards out of 7). The owner of the highest-score hand wins the pot, and the pot is split in case of a tie. The scores of all 5-cards poker hands are shown in Table 1 (where the strength of each hand decreases from top to bottom). Note that, by convention, the letter T is used to indicate the rank "10" when listing poker cards. Another convention of poker is to use the letter "s" to indicate two cards of the same suit (e.g., [A♣][2♣] and [A♡][2♡] can be both indicated as `A2`s), while the letter "o" is used to indicate two cards of different suit, called two "offsuit" cards (e.g., [A♣][2♡] can be indicated as `A2`o).

Each starting hand has a potential to become, eventually, the winning hand after all community cards have been turned up. For some starting hand this potential is greater than for others. The potential of all starting hands has been assessed in many books and by extensive use of computer programs. Since in NLTH the four suits have all the same value, the $\binom{52}{2}$ starting hands can be reduced to only 169 possibilities: 13 pairs, 78 non-pairs suited, and 78 non-pairs offsuit. It is now well-known that the best preflop starting hand is a pair of aces, followed by a pair of kings, while `72`o (seven-deuce offsuit) is considered to be the worst starting hand. There are many poker books and journals reporting tables of 169 entries listing the preflop strength of each starting hand.

After the flop has been exposed, the strength of each hand might drastically change, depending on how well the starting hand combines with the flop. Unfortunately, it is highly impractical to create tables ranking the strength of each starting hand h for each flop f, since there are $\binom{52}{2} \times \binom{50}{3} = 25,989,600$ such combinations. The subject of our paper is to investigate some possible ways to compute the strength of a hand once a particular flop has been exposed.

1.2 The concept of Equity

Equity of hand vs hand. When only two players remain in play, we say that they are *heads up*. The flop is a crucial point in a hand, when many times two players remain heads up. The flop is crucial since it is the time when most of the cards are exposed and the players try to estimate the potentiality of their hand with only two cards still to come.

Suppose a particular flop $f = \{c_1, c_2, c_3\}$ has been exposed (where the c_i are three specific cards), and that two players, A and B, are heads up. A holds $\{a_1, a_2\}$, while B holds $\{b_1, b_2\}$. What is the probability p_w that, after the turn c_4 and the river c_5 have been exposed, A will hold the winning hand? Let us call p_w such probability. It is easy to compute p_w by a simple algorithm: We loop over all possible values for $\{c_4, c_5\}$ (there are $\binom{45}{2} = 990$ such pairs). For each pair we record if A wins or if the hand is a tie. Let n_w be the number of wins for A and n_t the number of ties. Since in case of a tie the pot is split between the winners, for each tie we assign 1/2 win to A and 1/2 to B, so that $p_w = (n_w + \frac{1}{2}n_t)/990$. The probability for a hand to be the winner after all cards have been exposed, given some cards that have already been exposed, is called the hand's *equity*. A hand has a preflop equity, a flop equity, a turn equity and a river equity. In this paper we are particularly concerned with the flop equity, that is the most crucial (and difficult to estimate) in the course of the play. Many important decisions are based on good estimates of the flop equity of a hand. In particular, *commitment* decisions (i.e., decisions that put at stake all of our chips) can be based on the so called *pot-odds* which, in turn, are based on our equity. Loosely speaking, if the money that we can gain (i.e., the pot) is (or is expected to become) v_P, and it costs us v_u to play, and our equity is E, playing is profitable if $E > v_u/(v_u + v_P)$ and unprofitable otherwise.

Equity of hand vs range. In the game of poker it is very difficult (or, more likely, impossible) to pinpoint the opponent's starting hand to just one possibility. In practice, all one can do is to formulate an educated guess on a set of hands (the fewer, the better) that his opponent might be holding in a specific situation. In the poker jargon, such a set of hands is called a *range*.

We will consider ranges in which all hands have the same probability of being the actual hand that a player holds. That is, if R is a range, then each hand in R is equally likely and has probability $1/|R|$. More realistically, we could have defined a range in such a way that different hands in R could have different probabilities (for instance, in a particular situation a player could hold either `AK` (ace-king) or `AQ` (ace-queen), but, given that we know he his somewhat conservative, it could be more likely than he has `AK` than `AQ`). Generalizing our arguments to ranges in which each hand has its own probability is straightforward, but it would require the knowledge of the probabilities with which a particular opponent is holding particular hands. For simplicity, we have chosen to consider here the case of uniform probabilities.

The opponent's range is a subset of all possible starting hands. Once the flop is exposed, the range is in fact a subset of a set of $\binom{47}{2}$ starting hands, namely all hands that do not include any of the cards that we hold or that belong to the flop. Assume player A holds a hand h, player B holds one hand from a specific range R, and a flop f has been exposed. The equity of h vs R on this flop, denoted by $E(h, f, R)$ is the probability that, after turning two more cards (turn and river) player A will be the winner.

Computing the equity of a hand vs a range with a computer is very easy, and there are many online sites that offer this service (see, e.g., [9]). It is just a matter of computing the average equity of h vs each hand $h' \in R$, as explained before. Of course this could be a demanding task (for instance, if R is the range "ATC" –any two cards– there are almost 1000 hand-vs-hand computations to be made). Needless to say, such computations are impossible to be made mentally at the poker table.

1.3 Paper results and organization

The main goal of this paper is to show that mathematical programming techniques [2, 7] can be applied to the game of poker to study the strength of a hand versus a range after the flop. Our objective was to come up with some relatively simple formulas that can potentially be computed mentally by a player at the table and give him his equity on any flop. In order to be simple, the formulas had to rely on as few "features" of the flop texture as possible. Our approach can be applied to any given hand h versus any given range R. In this paper we have considered some important cases for h such as strong pairs, strong non-pairs and medium suited connectors[1], and run them against some ranges taken from the literature. We define a set of n binary features that each flop may or may not possess (e.g., "is there an ace on the flop?", "is the flop all of the same suit?", "is there a pair on the flop?", etc.), so that to each flop there corresponds a binary n-vector. Then, by using Integer Linear Programming (ILP), we both select a subset of "few" features to look at and assign weights to the selected features in such a way that the equity of each flop can be estimated with high precision by the weighted sum of the features

[1] Two cards are said to be *connectors* if they are consecutive, and they can realize, potentially, up to six straights. If the cards are also suited, their strength increases due to the possibility that they will realize a flush. Therefore, suited connectors are considered fairly good starting hands by poker players.

Table 2 Sklansky groups

Group	Hands
1	AA, KK, QQ, JJ, AKs
2	TT, AQs, AJs, KQs, AKo
3	99, JTs, QJs, ATs, AQo
4	T9s, KQo, 88, QTs, 98s, J9s, AJo, KTs
5	77, 87s, Q9s, T8s, KJo, QJo, JTo, 76s , 97s, A9s, A8s, . . . , A2s, 65s
6	66, ATo, 55, 86s, KTo, QTo, 54s, K9s , J8s, 75s
7	44, J9o, 64s, T9o, 53s, 33, 98o, 43s , 22, K9s, K8s, . . . , K2s
8	87, A9o, Q9o, 76o, 42s, 32s, 96s, 85s , J8o, J7s, 65o, 54o, 74s, K9o, T8o
9	all remaining hands

possessed by the flop. In the remainder of the paper we will elaborate on this technique. We notice that this is the first time that ILP has been applied to the game of poker to derive simple formulas for estimating the equity of hands. The only other applications of ILP to poker in the literature are related to game-theoretical strategies for simple, specific, situations, such as deciding to either bet all chips preflop or to fold the hand [1,4].

The paper is organized as follows. In Section 2 we describe a set of important ranges taken from the literature, that will be used in our computational experiments. Furthermore, we describe the set of features that we have introduced to characterize each flop. In Section 3 we describe the ILP models proposed in order to accurately estimate the equities by using only a small number of features. The results of our computational experiments are presented in Section 4. Some conclusions are drawn in Section 5.

2 Ranges and features

2.1 Ranges

The community of poker players regards David Sklansky as the person who laid the mathematical foundations of the game of poker. In one of his pioneering books [6], he suggested the division of all starting hands into nine groups (today known as *Sklansky groups* [8]), in such a way that all hands in the same group have roughly the same strength, and are stronger than all hands in the groups that follow. For instance, group number 1 consists of the pairs AA, KK, QQ, JJ and of the unpaired suited hand AKs. In Table 2 we report all hand groups in Sklansy's classification.

From Sklansky groups we have derived four ranges, namely (i) *Range ultra-strong (RUS)* (Sklansky group-1 hands); (ii) *Range strong (RS):* (groups 1 and 2); (iii) *Range medium-loose (RML):* (groups $1, \ldots, 5$); (iv) *Range any-two cards (ATC)*: (all possible starting hands). These ranges are meant to represent the possible holding of a player *based on the preflop action*, i.e., immediately before the flop is exposed. In poker, given specific preflop situations, experienced players are able to assign certain ranges to their opponents. Let us look at some examples: (i) If player A raises the pot, player B re-raises and then player C puts in a third raise, then it is very likely that C has a *very* strong hand, something like a pair of aces or kings. (ii) If, in the previous situation, player C does not raise, but still elects to call, then it is likely that he has a quite strong hand. (iii) If everybody folds, and just two players (the blinds) remain in play without any raise, then each of them can literally have any two cards.

2.2 Flop features

The number of possible flops once a starting hand is known is $\binom{50}{3} = 19600$. Each flop has some peculiar characteristics. which we want to characterize by means of binary features. For instance, the flop [K♡] [Q♢] [8♣] "has one King", "has two cards 10 or above", "has three different suits", etc. The set of all possible features should be such that it hopefully contains a small subset which is sufficient to look at in order to compute the equity of any flop with high precision. Since a priori we have no idea about which features might be better than others in this respect, we included a lot of features, some of which have apparently little use or apply to very few flops. We have defined more than seventy different binary features, called $F_1 \ldots, ..., F_{73}$, that are described below.

Some of the features depend on the flop only, while others depend on the flop *combined with* the particular starting hand we hold. For instance "did we pair at least one of our starting cards?", or "do we have a draw[2] (straight and/or flush)?", etc.

The features depending entirely on the flop are:

1. **Individual broadway[3] cards:** The flop has one/at least two/zero/ aces (features $F_1/F_2/F_3$, respectively). The same for Kings (features F_4, F_5, F_6), Queens (F_7, F_8, F_9), Jacks (F_{10}, F_{11},F_{12}) , Tens (F_{13}, F_{14}, F_{15}).
2. **Suits:** The flop has one/two/three/ suits ($F_{16}/F_{17}/F_{18}$). The flop has two suits with a suited ace (F_{19}).
3. **Ranks:** The flop has three cards of the same rank (F_{20}), exactly two cards of the same rank (F_{21}) or three cards of different ranks (F_{22}).
4. **Paint[4] cards:** The flop has one/at least two/three/ paint cards ($F_{23}/F_{24}/F_{25}$).
5. **Broadway:** The flop has zero/one/two/three/ broadway cards ($F_{26}/F_{27}/F_{28}/F_{29}$).
6. **Cards size:** The flop has three medium (i.e., in $\{\mathtt{7}, \ldots, \mathtt{T}\}$) cards ($F_{30}$). The flop has three small (i.e., in $\{\mathtt{2}, \ldots, \mathtt{6}\}$) cards ($F_{31}$). The flop has three

[2] A *draw* is a hand that has the potential to become either a straight or a flush by the river. For example, we have a flush draw if our hand combined with the flop contains four cards of the same suit, so that one more card of that suit will give us a flush. When we cannot complete our point by the turn (i.e., we need two good cards in sequence), the draw is called a *backdoor* draw.

[3] *Broadway* = card from Ten to Ace.

[4] *Paint* = card from Jack to Ace.

medium-small cards (F_{32}). The flop has three small cards all of different rank $\leq$ 5 (F_{33}).

7. **Straight:** The flop has cards in straight, i.e., three consecutive cards (F_{34}). The flop has cards in straight with a 1-gap, i.e., with ranks r, $r+1$, $r+3$ or r, $r+2$, $r+3$ (F_{35}). The flop has cards in straight with a 2-gap (F_{36}). The flop has cards in straight with a 3-gap (F_{37}). The flop is such that no player can have a straight or make a straight by the turn (F_{38}).

Some of these features are better illustrated by means of examples. For instance, the flop [K♡] [8♣] [3♢] possesses the features F_3, F_4, F_9, F_{12}, F_{15}, F_{18}, F_{22}, F_{23}, F_{27}, and F_{38}, while the flop [6♣] [6♡] [2♡] has the features F_3, F_6, F_9, F_{12}, F_{15}, F_{17}, F_{21}, F_{26}, F_{31}, F_{32}, and F_{37}.

The features that depend on the flop combined with our starting hand are:

1. **Flush draws:** The flop is not monochromatic and our hand+flop has 3 cards of one suit (F_{39}). Our hand+flop has 4 cards of one suit (F_{40}). The flop is monochromatic but we have no flush draw (F_{41}). The flop is bichromatic but we have no flush draw (F_{42}). We have a made flush (F_{43}).
2. **Made hands:** We have 4-of-a-kind or full house (F_{44}). We have 3-of-a-kind/at least 3-of-a-kind/ using at least one card in starting hand (F_{45}/F_{46}). We have paired one of our cards (F_{47}). We have at least two pair (F_{48}). We have top/middle/bottom/ pair, i.e., we paired the highest/middle/bottom card in the flop ($F_{49}/F_{50}/F_{51}$). We have two pair or 3-of-a-kind and the flop has three/two/ ranks (F_{52}/F_{53}).
3. **Overcards[5]:** The flop contains zero/one/at least two/ overcards to our highest card ($F_{54}/F_{55}/F_{56}$). We have one/two/ overcards to the highest card in the flop (F_{57}/F_{58}).
4. **Straight:** We have a straight (F_{59}). We have an open-ended straight draw[6] with/without/ the bottom end (F_{60}/F_{61}). We have an inside/double inside/ straight draw[7] (F_{62}/F_{63}). We have a backdoor straight that can be complete in one/more than one/ ways (F_{64}/F_{65}). We have no draws to either a straight or a flush (F_{66}).
5. **Draw combinations:** We have an inside straight draw and zero/one/two overcards to the flop in our hand ($F_{67}/F_{68}/F_{69}$). We have an inside straight draw, a backdoor flush draw, and the flop has ≥ 2 suits (F_{70}). We have an inside straight draw and a (non-backdoor) flush draw (F_{71}). We have a (non-backdoor) flush draw and one/two/ overcards to the flop (F_{72}/F_{73}).

[5] A card X is an *overcard* to Y if it is a higher card than Y

[6] A straight draw is *open-ended* if our hand plus the flop contain four consecutive cards, say r, $r+1$, $r+2$, $r+3$. For an open-ended straight draw, there are 8 cards that can completed the straight by the turn. If r is in our starting hand, we say that we hold the bottom end. This is a dangerous situation, since a card of rank $r+4$ by the turn gives us a straight, but it could give our opponent a higher straight if he holds $r+5$.

[7] An *inside straight draw* consists of four of the five cards needed for a straight, but missing one in the middle. For an inside straight draw, there are 4 cards that can complete the straight by the turn. A *double inside straight draw* is a draw made of four non-consecutive cards, such that there are 8 cards that can complete the straight by the turn. For example if we hold 59 and the flop is 367, any 4 or 8 will give us a straight.

With respect to the features F_{39}–F_{73}, let us consider the following examples. Assume our starting hand is [K♣] [T♡]. If the flop is [J♡] [9♣] [7♣] then the following features are true: F_{39}, F_{54}, F_{57}, F_{63}. If the flop is [A♣] [Q♣] [T♣] then the following features are true: F_{40}, F_{46}, F_{51}, F_{55}, F_{62}, F_{67}, F_{71}.

3 ILP for computing equities

In this section we describe our Integer Linear Programming models for computing the equity of a certain hand h versus a certain range R. The actual hands and ranges utilized are detailed in Section 4.

We describe two ILP models, one relative to the maximum error, and one to the average error of the estimated equity with respect to the true equity. Our models are parametrized with parameters $\bar{N}$ and $\bar{E}$. More specifically, we can (i) fix $\bar{N}$ to be the maximum number of features that can be used, and then minimize the maximum (or the average) error, or (ii) fix $\bar{E}$ to be a threshold for the accepted maximum (respectively, average) error and minimize the number of features sufficient to obtain an error within the threshold.

There are $m = 19600$ flops $f^1, \ldots, f^m$. Each flop f^i does or does not possess each of n binary features F_j, and we denote by $F_j(f^i) \in \{0, 1\}$ the absence/presence of a particular feature in a given flop.

The first step consists in considering each flop f^i in turn, and computing the exact equity $e^i := E(h, f^i, R)$ of h vs R (we did this by a `C#` code that we developed, but the step can also be done by resorting to online sites that compute equities, such as [9]). Moreover, we check all features on f^i, thus obtaining a binary vector $b^i = (b^i_1, \ldots, b^i_n)$ with $b^i_j = F_j(f^i)$. At the end, we have a matrix B of m rows and $n + 1$ columns (the last column contains the equities $(e^1, \ldots, e^m)^T$). The matrix B is (part of) the constraint matrix of our linear programs, to be defined.

The model has n binary variables $y_1, \ldots, y_n$. Each variable y_j can allow (if $y_j = 1$) or forbid (if $y_j = 0$) the use of the corresponding feature. Furthermore, there are real variables x_j, for $j = 1, \ldots, n$. Each variable x_j is the weight that we associate to a feature. The weight is meant to represent how much equity we gain (if $x_j > 0$) or we lose (if $x_j < 0$) when the feature F_j is present. Since x_j represents a probability, we want $|x_j| \leq 1$.

Let ϵ be a variable representing the maximum error (i.e., difference in absolute value between the estimated equity and the true equity of a flop). We obtain the following ILP for minimizing the maximum error, given a budget $\bar{N}$ on how many features can be used altogether:

$$\texttt{MAXERR}(\bar{N}) := \min \epsilon \tag{1}$$

subject to

$$-y_j \leq x_j \leq y_j \qquad \forall j = 1, \ldots, n \tag{2}$$

$$\sum_{j=1}^{n} y_j \leq \bar{N} \tag{3}$$

$$e^i - \epsilon \le \sum_{j=1}^{n} b_j^i x_j \le e^i + \epsilon \qquad \forall i = 1, \dots, m \tag{4}$$

$$\epsilon \ge 0,\ x_j \in \mathbb{R},\ y_j \in \{0,1\} \qquad \forall j = 1, \dots, n. \tag{5}$$

The objective function (1) calls for minimizing the error ϵ. Constraints (2) are "activation" constraints: when a variable y_j is 1, the weight x_j of the corresponding feature can be non-null, while when $y_j = 0$, x_j must be null. Constraint (3) says that we can use at most $\bar{N}$ features. Constraints (4) force that, for each flop f^i, the estimated equity differs from the true equity (in excess or in defect) by at most ϵ. The model has $2(n+m)+1$ constraints and $2n+1$ variables, n of which are integer.

From the above model, it is easy to derive a model in which we minimize the number of features sufficient to stay within a certain maximum error. We simply need to use the objective $\texttt{FEAT\_MAXERR}(\bar{E}) := \min \sum_{j=1}^{n} y_j$ under constraints (2) and (4) in which the variable ϵ has been replaced by the constant $\bar{E}$.

We now turn to the model for minimizing the average error. In this model we have real variables ϵ^i, for $i = 1, \dots, m$, that represent the error in estimating the equity of each specific flop f^i. The model calls for the minimization of $\texttt{AVGERR}(\bar{N}) := \frac{1}{m} \sum_{i=1}^{m} \epsilon_i$ under constraints (2), (3) and

$$e^i - \epsilon_i \le \sum_{j=1}^{n} b_j^i x_j \le e^i + \epsilon_i \qquad \forall i = 1, \dots, m \tag{6}$$

with variables $\epsilon_i \in \mathbb{R}^+$, $x_j \in \mathbb{R}$ and $y_j \in \{0,1\}$. The model has $2(n+m)+1$ constraints and $m+2n$ variables, n of which are integer. Again, we can obtain from the above a model for minimizing the number of features sufficient to stay within a certain average error. We simply need to use the objective $\texttt{FEAT\_AVGERR}(\bar{E}) := \min \sum_{j=1}^{n} y_j$ under constraints (2), (6) and with the constraint

$$\frac{1}{m} \sum_{i=1}^{m} \epsilon_i \le \bar{E}. \tag{7}$$

4 Computational results

In this section we describe the computational experiments that we ran in order to assess the effectiveness of our ILP models. Out of the 169 possible starting hands, we chose a subset representative of some significant categories, such as:

- *Very strong/strong Pairs*: `AA`, `KK`, `QQ` and `JJ`
- *Medium/low pairs*: `99` and `55`
- *High, suited, cards*: `AK`s, `AQ`s and `KQ`s
- *Medium suited connectors*: `JT`s and `98`s

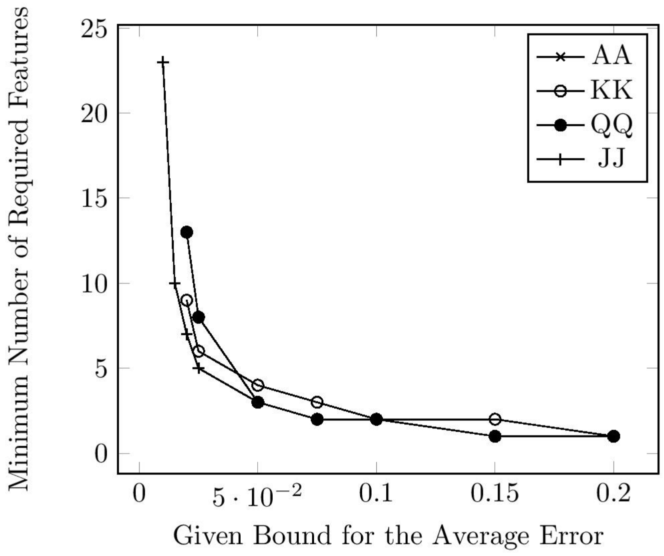

Fig. 1 Strong pairs vs. RUS.

Each hand was run against the four ranges that were described in Section 2, namely, RUS (Range Ultra-Strong, i.e., hands in Sklansky group 1), RS (Range strong, i.e., hands in Sklansky groups 1 and 2), RML (Range Medium-Loose, i.e., hands in Sklansky groups $1, \ldots, 5$), ATC (Range Any-Two Cards, i.e., all possible hands).

In this paper we focus on the minimization of the *average* error. As a matter of fact, the minimization of the maximum error is not as important, nor as robust, as that of the average error. The problem is due to the presence of outliers (i.e., very few flops on which it is difficult to estimate the equity with a small error). Clearly a player does not care about a handful of flops on which he might be very wrong in estimating the equity, as long as on most of the flops (i.e., on average) his estimate is quite accurate. Furthermore, these outliers seem unavoidable, at least with the features we used, and spending a lot of computation time with the goal of minimizing the error on the outliers does not appear to be justified. For more results on the maximum error goal, we refer the reader to the preliminary version of this paper, appeared in the SOR proceedings [3], in which we minimized the maximum error for `JJ` and `AKs` against the above four ranges. Finally, a note about the error percentages is in order. In the poker community, equities are indicates in percentage points, rather than as probabilities. So, an equity of 0.3 is usually referred to as a 30% equity. In accordance with the tradition, we use the percentage point (i.e., 1%) as the unit of measure for equity, so that, in estimating equities, also the error is reported in percentage points. Therefore, when we say that we estimated an equity with an error of, e.g., 2%, we mean an *absolute* error rather than

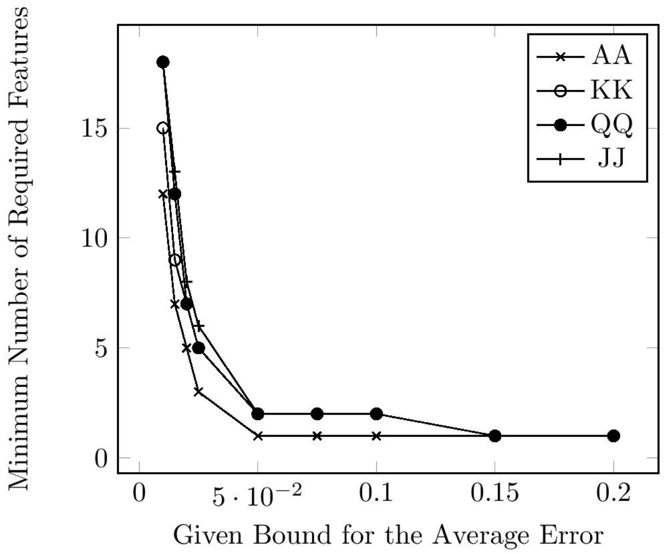

Fig. 2 Strong pairs vs. ATC.

a *relative* error. That is, if e is the estimate and t is the true equity, it is $|t - e| = 2\%$.

The experiments have required running a total of 680 ILPs, to compute AVGERR($\bar{N}$) and FEAT_AVGERR($\bar{E}$) for various values of $\bar{N}$ and $\bar{E}$. The LP solver used was CPLEX 12.5 on an Intel® Core™ i3 CPU M 350 @ 2.27 GHz with 2.8 GB of RAM. If the running time to solve an ILP exceeded a maximum limit of 2 hours, then the computation was aborted and the best solution found so far was returned (this happened in about 14% of the cases).

In Figure 1 we show how the error and the number of features relate to each other when estimating the equity of strong pairs vs RUS. The x-axis corresponds to an upper bound to the average error, while the y-axis corresponds to an upper bound to the number of features used. A point (x, y) in the plot means that it is possible to achieve an average error $\leq x$ by using at most y features. We see how a very small average error, i.e., around 4%, can be obtained by looking at only five features of the flop, while, with about ten features, one can achieve an average error between 2% and 3%. It is also very surprising that by using only three features one can achieve an error of 10% on average. Figure 2 shows the same type of results when running strong pairs against the range ATC, Figures 3 and 4 illustrate the same type of results when running suited high cards against the ranges RUS and ATC, while Figures 5 and 6 illustrate the results of running medium-strength hands, such as suited connectors and small pairs, against the ranges RUS and ATC. As a general comment, we can see how estimating an equity vs ATC is slightly

Table 3 Results for test hands vs. the four ranges. Objective is minimization of the number of features needed to estimate the equity with an average error of at most 2.5%.

hand	range	N	1st fea.	2nd fea.	3rd fea.	4th fea.	5th fea.
AA	RUS	8	0.29 (F_6)	0.29 (F_9)	0.29 (F_{12})	0.24 (F_{24})	-0.17 (F_{41})
	RS	8	0.96 (F_{49})	0.74 (F_{58})	-0.19 (F_{41})	-0.11 (F_{24})	0.10 (F_{15})
	RML	*15*	0.80 (F_{54})	-0.17 (F_{41})	0.13 (F_{44})	0.11 (F_{32})	-0.11 (F_{45})
	ATC	3	0.85 (F_{54})	0.10 (F_{48})	-0.08 (F_{35})		
KK	RUS	6	0.63 (F_{49})	0.45 (F_{50})	0.29 (F_6)	0.18 (F_9)	0.18 (F_{12})
	RS	7	0.97 (F_{32})	0.88 (F_{23})	0.80 (F_{24})	-0.31 (F_6)	-0.14 (F_{41})
	RML	*13*	0.67 (F_{32})	0.64 (F_{23})	0.59 (F_{24})	0.44 (F_5)	0.32 (F_4)
	ATC	5	0.88 (F_2)	0.79 (F_1)	0.83 (F_3)	-0.14 (F_{41})	0.12 (F_{47})
QQ	RUS	5	0.73 (F_7)	0.24 (F_9)	0.19 (F_{12})	-0.07 (F_{55})	-0.01 (F_{42})
	RS	6	0.77 (F_{47})	0.26 (F_9)	0.20 (F_{58})	0.09 (F_{12})	0.08 (F_{15})
	RML	*11*	0.58 (F_{12})	0.55 (F_{10})	0.47 (F_{11})	0.25 (F_{48})	0.18 (F_3)
	ATC	5	0.94 (F_{47})	0.77 (F_9)	-0.13 (F_{41})	0.04 (F_{58})	-0.02 (F_{42})
JJ	RUS	6	0.79 (F_{11})	0.65 (F_{10})	0.12 (F_{62})	0.11 (F_3)	0.10 (F_6)
	RS	9	0.44 (F_{47})	0.22 (F_3)	0.19 (F_6)	0.10 (F_{40})	-0.09 (F_{41})
	RML	8	0.77 (F_{47})	0.54 (F_{12})	0.18 (F_3)	-0.12 (F_{41})	-0.09 (F_7)
	ATC	6	1.00 (F_{11})	0.94 (F_{10})	0.75 (F_{12})	-0.13 (F_{41})	-0.06 (F_{35})
99	RUS	7	0.63 (F_{48})	0.16 (F_{40})	0.12 (F_{62})	0.10 (F_{49})	0.09 (F_6)
	RS	*17*	0.61 (F_{51})	0.55 (F_{50})	0.51 (F_{49})	-0.28 (F_2)	-0.25 (F_1)
	RML	9	0.39 (F_{47})	0.23 (F_3)	0.17 (F_{12})	0.14 (F_6)	0.14 (F_9)
	ATC	5	0.74 (F_{54})	0.71 (F_{55})	0.66 (F_{56})	0.22 (F_{47})	-0.12 (F_{41})
55	RUS	7	0.69 (F_{47})	0.15 (F_{40})	0.12 (F_{62})	0.09 (F_3)	0.09 (F_6)
	RS	*19*	-0.53 (F_{20})	0.51 (F_{47})	0.20 (F_3)	0.18 (F_{40})	0.17 (F_6)
	RML	*23*	0.68 (F_{22})	0.45 (F_{44})	-0.43 (F_{45})	0.34 (F_{21})	0.20 (F_{61})
	ATC	12	0.62 (F_{21})	0.52 (F_{22})	0.40 (F_{20})	0.37 (F_{47})	0.18 (F_{61})
AKs	RUS	11	0.51 (F_{43})	0.38 (F_{54})	0.34 (F_{49})	0.24 (F_{47})	-0.15 (F_{66})
	RS	*11*	0.77 (F_{54})	0.70 (F_{43})	-0.33 (F_{16})	-0.27 (F_{58})	-0.25 (F_{42})
	RML	*21*	0.53 (F_{61})	0.44 (F_{54})	0.41 (F_{43})	0.32 (F_{49})	0.10 (F_{40})
	ATC	7	0.62 (F_{54})	0.57 (F_{43})	0.33 (F_{49})	0.16 (F_{73})	-0.14 (F_{16})
AQs	RUS	*18*	0.44 (F_{43})	0.42 (F_{47})	0.29 (F_1)	-0.24 (F_{52})	0.21 (F_{54})
	RS	*16*	0.73 (F_{47})	-0.60 (F_2)	0.48 (F_7)	0.30 (F_9)	0.20 (F_{40})
	RML	*26*	-0.62 (F_{52})	0.49 (F_{48})	0.43 (F_1)	0.39 (F_7)	0.22 (F_3)
	ATC	*10*	0.60 (F_{54})	-0.58 (F_{52})	0.37 (F_{48})	0.30 (F_1)	0.27 (F_7)
KQs	RUS	*14*	-0.91 (F_{52})	0.81 (F_{48})	0.34 (F_4)	0.30 (F_{40})	0.27 (F_7)
	RS	*14*	0.77 (F_{47})	0.50 (F_4)	0.30 (F_{40})	0.28 (F_7)	0.23 (F_{61})
	RML	*17*	0.57 (F_{48})	-0.42 (F_{46})	0.36 (F_{53})	0.35 (F_{45})	0.28 (F_{40})
	ATC	*13*	0.59 (F_{21})	0.51 (F_{22})	0.47 (F_{48})	-0.36 (F_{46})	0.31 (F_{45})
JTs	RUS	11	0.72 (F_{48})	-0.42 (F_{52})	0.29 (F_{40})	0.22 (F_{61})	0.18 (F_{49})
	RS	12	0.70 (F_{47})	0.27 (F_{40})	0.20 (F_{61})	0.19 (F_{10})	0.18 (F_{60})
	RML	*15*	0.61 (F_{47})	0.37 (F_{10})	0.26 (F_{40})	0.23 (F_{12})	0.22 (F_{49})
	ATC	*15*	0.51 (F_{47})	0.41 (F_{18})	0.38 (F_{17})	0.34 (F_{10})	0.34 (F_{13})
98s	RUS	12	0.70 (F_{48})	0.27 (F_{40})	0.22 (F_{61})	0.20 (F_{60})	0.19 (F_{45})
	RS	*40*	0.55 (F_{48})	0.33 (F_{44})	0.27 (F_{40})	0.25 (F_{45})	0.24 (F_{53})
	RML	*16*	0.66 (F_{48})	0.40 (F_{53})	0.38 (F_{45})	-0.35 (F_{46})	0.27 (F_{40})
	ATC	*15*	0.62 (F_{47})	0.53 (F_{52})	-0.41 (F_{46})	0.40 (F_{53})	0.38 (F_{45})

easier (i.e., it requires a smaller number of features) than estimating it against RUS.

In Table 3 we give detailed results for the problem in which $\bar{E} = 2.5\%$ and the number of features sufficient to achieve an average error within $\bar{E}$ was minimized. For each hand in each data set there are four rows, corresponding to the four ranges. The column labeled "N" reports the solution value, i.e., the number of features sufficient to have an average error of at most 2.5%. When the value is written in italic, it means that it is not the exact value, but the best value obtained by CPLEX within the maximum time limit of 2hrs. The next five columns report the five most important features in the solution, sorted by the absolute value of their weight. For each feature we list its weight (rounded to multiples of 0.01) and its identifier. For instance, to compute the equity of `AA` vs. ATC with an average error of at most 2.5% it is sufficient to look at three features (F_{54}, F_{48} and F_{35}). The first of them, however, is always true for `AA`, since the flop cannot have any overcard. Therefore, here is the, very simple, mnemonic formula for a player holding `AA`: *"Start with a base equity of* 85%. *If at the flop you have at least two pairs, add 10%, and if the flop has a 1-gap straight, subtract* 8%*"*.

In Table 4 we give detailed results for the problem in which $\bar{N} = 10$ and the average equity error when using at most ten features was minimized. The table is organized in the same way as Table 3, only that this time the third column is labeled "E", and reports the solution value, i.e., the average error in equity estimation when using at most 10 features. When the value is written in italic it means that it is not the exact value, but it is the best value obtained by CPLEX within the maximum time limit of 2hrs. We notice that this set of problems turned out to be quite more difficult to solve to optimality than the problems in which $\bar{N}$ was minimized. At any rate, when the computation was aborted, the gap between the best solution found and the best lower bound was very small (in the order of 0.1%). Therefore, for all practical purposes with respect to human players adopting the weights to estimate equities, these solutions are as good as optimal. The results show how ten features are quite enough to estimate the equity with high precision. The best estimate happens for `AA` vs ATC, where the error is only 1.1%. For 12 cases out of the 44 pairs (hand, range), the average error is lower than 2%, while for 30 cases it is lower than 3%.

With respect to the complete set of 73 features that we considered, we observed the following facts: across all of our 680 tests,

- 60 features have been selected by at least 10 of the solutions,
- 31 features have been selected in at least 5% of the solutions,
- 16 features have been selected in at least 10% of the solutions,
- for 15 features the absolute value of the weight was greater than 0.1 in at least 5% of the solutions.

5 Conclusions

For human players, computing the equity of a hand vs a range on the flop is more of an art than a science, requiring mathematical, analytical, and also psychological skills. Being able to estimate the equity with high accuracy is what separates skilled professional players from the rest. Yet, the estimates can never be too accurate due to the high number of factors in play. In this paper we have shown how ILP can be used to select relevant features of the flop and weights to assign to such features so that one can approximate the equity of a hand versus a range by a weighted sum with few addends. This is the first time that Mathematical Programming techniques have been applied to the game of poker for estimating the strength of a hand.

It is very surprising, and unknown prior to this article, that by looking at only five characteristics of a flop, it is possible to estimate the equity of a hand vs any range with an average error of just around 3%. It is important to remark that an error of 3% is considered negligible. Poker is a very hard game of randomness and incomplete information, and humans look for simple rules of strategy. For instance, preflop, KQs vs 99 have equity of 46.5% and 53.5% respectively, but this matchup is usually considered a coin-flip (50-50). Similarly, poker books report that if we have either a flush draw or a straight draw on the flop, the probability of completing the draw is 1 in 3 while actually it is about 35% for the flush draw, and about 31% for the straight draw.

References

1. D. Billings, N. Burch, A. Davidson, R. Holte, J. Schaeffer, T. Schauenberg, and D. Szafron. Approximating game-theoretic optimal strategies for full-scale poker. In Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI), pp. 661–668 (2003)
2. S. P. Bradley, A. C. Hax and T. L. Magnanti. Applied Mathematical Programming. Addison-Wesley Publishing Company. Reading, MA (1977)
3. M. Dalpasso and G. Lancia. Computing the equity of a poker hand by Integer Linear Programming. In: Z. Stirn, J. Zerovnik, J. Povh, S. Drobne, and A. Lisec, eds.. Proceedings of the 12th Int. Symp. on Operational Research (SOR13). pp. 39-44, SDI-SOR, 2013.
4. P. B. Miltersen and T. B. Sorensen. A near-optimal strategy for a heads-up no-limit Texas Hold'em poker tournament. Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems (AAMAS), pp. 1160–1167 (2007)
5. D. Sklansky, The Theory of Poker. Two Plus Two Publications. Henderson, NV (1999)
6. D. Sklansky and M. Malmuth, Hold 'em Poker for Advanced Players. Two Plus Two Publications. Henderson, NV (1999)
7. L. Wolsey and G. Nemhauser. Integer and Combinatorial Optimization. John Wiley and Sons. New York, NY (1988)
8. `en.wikipedia.org/wiki/Texas_hold_'em_starting_hands#Sklansky_hand_groups`
9. `www.cardplayer.com/poker-tools/odds-calculator/texas-holdem`

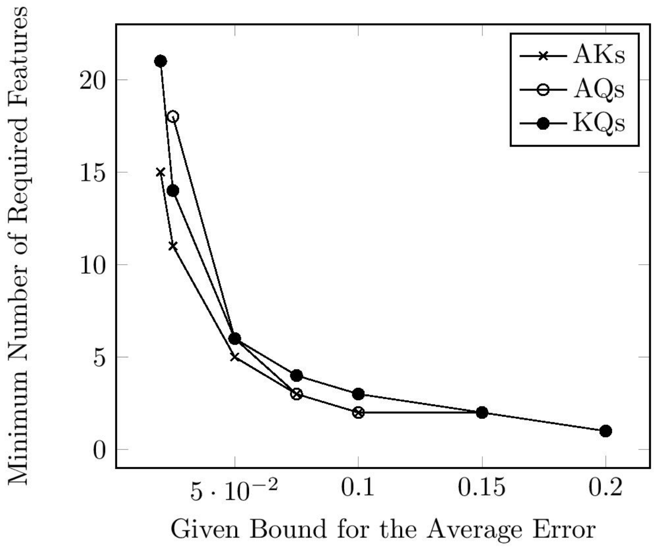

Fig. 3 Some strong-connector starting hands vs. RUS.

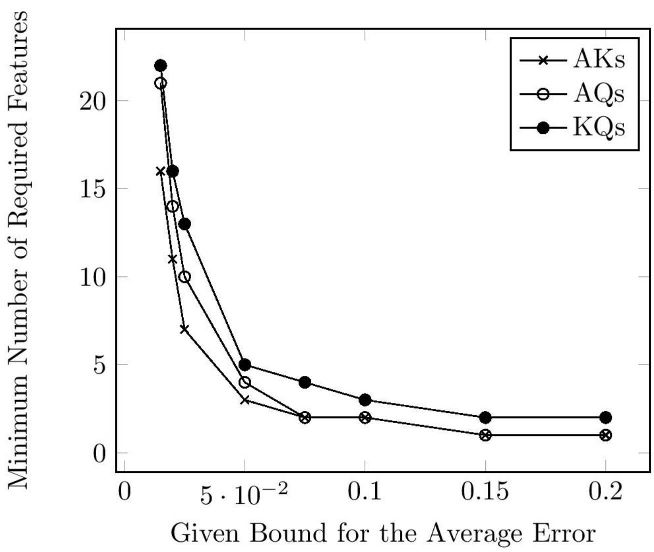

Fig. 4 Some strong-connector starting hands vs. ATC.

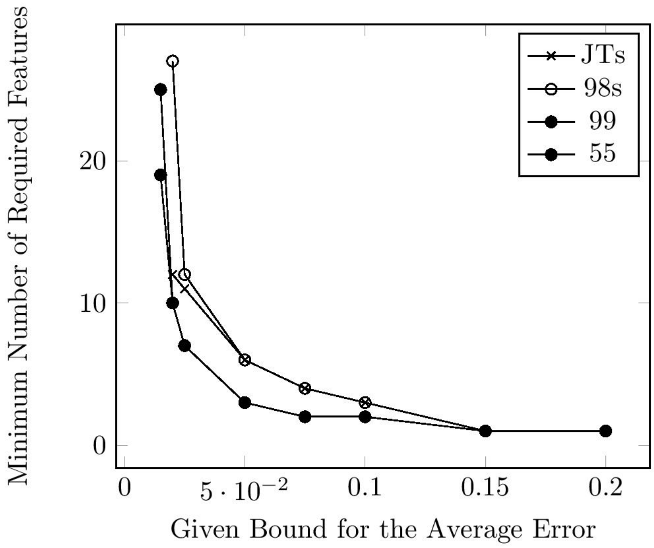

Fig. 5 Some medium-strength starting hands vs. RUS.

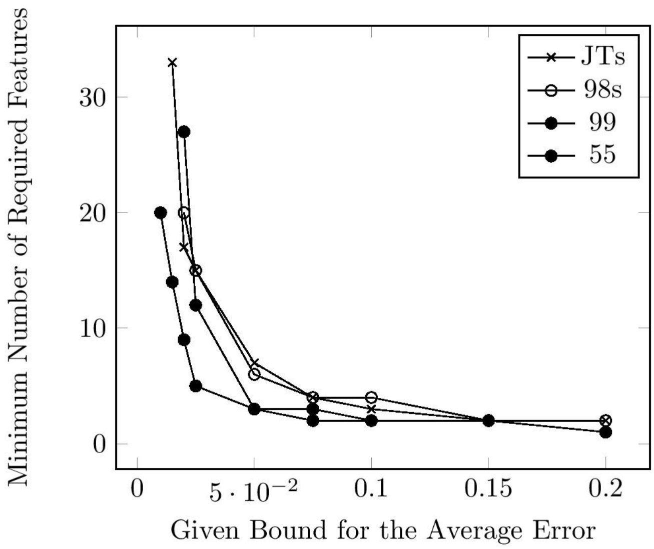

Fig. 6 Some medium-strength starting hands vs. ATC.

Table 4 Results for test hands vs. the four ranges. Objective is minimization of the average error on equity estimation when using at most 10 features.

hand	range	$\bar{E}$	1st fea.	2nd fea.	3rd fea.	4th fea.	5th fea.
	RUS	*2.2%*	0.34 (F_6)	0.28 (F_9)	0.28 (F_{12})	0.24 (F_{24})	-0.18 (F_{41})
AA	RS	*2.0%*	0.45 (F_{54})	0.26 (F_6)	-0.20 (F_{41})	0.13 (F_9)	0.12 (F_4)
	RML	*2.8%*	0.69 (F_{54})	-0.17 (F_{41})	0.16 (F_{46})	0.11 (F_{26})	0.11 (F_{21})
	ATC	*1.1%*	0.93 (F_{54})	-0.14 (F_{41})	-0.12 (F_{34})	-0.08 (F_{35})	0.07 (F_{48})
	RUS	1.8%	0.45 (F_{18})	0.42 (F_{17})	0.33 (F_{16})	0.23 (F_{40})	0.18 (F_9)
KK	RS	*1.6%*	0.80 (F_{15})	0.76 (F_{15})	0.71 (F_{13})	-0.34 (F_6)	-0.14 (F_{41})
	RML	*2.6%*	0.96 (F_{32})	0.93 (F_{23})	0.88 (F_{24})	0.18 (F_{50})	-0.16 (F_6)
	ATC	*1.3%*	0.96 (F_{46})	0.81 (F_{53})	0.79 (F_{47})	-0.13 (F_{41})	-0.10 (F_{34})
	RUS	*1.4%*	0.70 (F_{17})	0.69 (F_{18})	0.61 (F_{16})	-0.47 (F_9)	0.21 (F_{12})
QQ	RS	*1.5%*	0.85 (F_{15})	0.77 (F_{14})	0.75 (F_{13})	-0.51 (F_9)	0.21 (F_{58})
	RML	*2.5%*	0.92 (F_{12})	0.90 (F_{10})	0.80 (F_{11})	-0.27 (F_9)	-0.13 (F_{41})
	ATC	*1.5%*	0.98 (F_{46})	0.80 (F_{53})	0.79 (F_{47})	-0.14 (F_{41})	-0.11 (F_{34})
	RUS	*1.8%*	0.69 (F_{49})	0.64 (F_{50})	0.47 (F_{51})	0.18 (F_{61})	0.13 (F_{40})
JJ	RS	*2.3%*	0.50 (F_{11})	0.45 (F_{10})	0.22 (F_3)	0.19 (F_6)	-0.09 (F_{41})
	RML	*2.1%*	0.95 (F_3)	0.87 (F_2)	0.78 (F_1)	-0.26 (F_{12})	-0.12 (F_{41})
	ATC	*1.7%*	0.97 (F_{48})	-0.92 (F_{20})	0.77 (F_{12})	-0.14 (F_{41})	-0.07 (F_{34})
	RUS	*1.9%*	0.72 (F_{49})	0.69 (F_{50})	0.56 (F_{51})	0.21 (F_{61})	0.16 (F_{40})
99	RS	*2.9%*	-0.55 (F_{20})	0.54 (F_{46})	0.20 (F_3)	0.16 (F_6)	0.15 (F_{61})
	RML	*2.1%*	0.38 (F_{46})	0.20 (F_3)	0.14 (F_{12})	0.12 (F_9)	0.11 (F_6)
	ATC	*1.8%*	0.78 (F_{11})	0.76 (F_{12})	0.71 (F_{10})	0.23 (F_{46})	-0.12 (F_{41})
	RUS	*1.9%*	-0.75 (F_{20})	0.69 (F_{48})	0.23 (F_{61})	0.15 (F_{40})	0.12 (F_{69})
55	RS	*3.1%*	0.53 (F_{48})	0.17 (F_{61})	0.17 (F_3)	0.15 (F_6)	0.12 (F_{40})
	RML	*3.4%*	0.36 (F_{48})	0.15 (F_3)	0.14 (F_{51})	0.13 (F_{12})	0.11 (F_9)
	ATC	*2.7%*	0.62 (F_{21})	0.51 (F_{22})	0.37 (F_{46})	0.14 (F_{62})	-0.11 (F_{42})
	RUS	*2.5%*	0.69 (F_{61})	0.66 (F_{43})	0.41 (F_{49})	0.30 (F_{73})	0.19 (F_9)
AKs	RS	*2.6%*	0.70 (F_{43})	0.33 (F_{49})	0.32 (F_{54})	0.17 (F_{73})	-0.11 (F_{16})
	RML	*3.1%*	-0.58 (F_{41})	0.49 (F_{61})	0.43 (F_{16})	0.39 (F_{54})	0.28 (F_{49})
	ATC	*2.0%*	0.62 (F_{54})	0.42 (F_{43})	0.41 (F_{61})	0.31 (F_{49})	0.15 (F_{73})
	RUS	*3.2%*	-0.78 (F_{52})	0.61 (F_{48})	0.39 (F_1)	0.24 (F_{40})	0.19 (F_{12})
AQs	RS	*3.1%*	0.54 (F_{43})	0.37 (F_{49})	0.28 (F_{40})	0.19 (F_{46})	0.18 (F_{50})
	RML	*3.3%*	0.60 (F_{54})	0.30 (F_{49})	-0.29 (F_{41})	-0.20 (F_{42})	-0.16 (F_{18})
	ATC	*2.4%*	0.56 (F_{43})	0.48 (F_{54})	0.29 (F_{49})	0.28 (F_{17})	0.23 (F_{50})
	RUS	2.9%	0.77 (F_{46})	0.33 (F_4)	0.30 (F_{40})	0.26 (F_7)	0.20 (F_{61})
KQs	RS	*3.0%*	0.74 (F_{46})	0.42 (F_7)	0.38 (F_4)	0.29 (F_{40})	0.18 (F_{57})
	RML	*3.3%*	0.55 (F_{17})	0.52 (F_{46})	0.36 (F_{18})	0.36 (F_4)	0.34 (F_7)
	ATC	*3.0%*	0.72 (F_{40})	0.63 (F_{18})	0.60 (F_{42})	0.51 (F_{16})	0.35 (F_{48})
	RUS	*2.6%*	0.48 (F_{48})	0.28 (F_{40})	0.26 (F_{46})	0.20 (F_{61})	0.11 (F_{62})
JTs	RS	*2.8%*	0.70 (F_{46})	0.33 (F_{49})	0.27 (F_{40})	0.19 (F_{60})	0.19 (F_{61})
	RML	*3.1%*	0.63 (F_{46})	0.45 (F_{10})	0.25 (F_{40})	0.25 (F_{13})	0.23 (F_{12})
	ATC	*3.3%*	0.79 (F_{46})	0.74 (F_{50})	0.49 (F_{56})	0.40 (F_{27})	-0.24 (F_{48})
	RUS	2.9%	0.65 (F_{48})	0.27 (F_{40})	0.23 (F_{61})	0.17 (F_{47})	0.15 (F_{60})
98s	RS	*3.8%*	0.56 (F_{48})	0.26 (F_{40})	0.22 (F_{61})	0.20 (F_{47})	0.16 (F_{53})
	RML	*3.7%*	0.55 (F_{48})	0.34 (F_{53})	0.32 (F_{47})	0.26 (F_{40})	0.22 (F_{32})
	ATC	3.2%	0.54 (F_{40})	0.54 (F_{48})	0.41 (F_{53})	0.38 (F_{47})	0.34 (F_{18})