Hadtudományi Szemle
19. évfolyam (2026) 1. szám 47–63. • DOI: 10.32563/HSZ.2026.1.3

Domonkos Balázs Pászthory[1]

Multivariate Regression Analysis of Personnel Fluctuation in the Land Forces

Abstract

Personnel fluctuation in military units is widely considered both inevitable and a necessity for organisational interoperability. The ideal scale and frequency, however, is a matter of ongoing academic and institutional discourse. The purpose of this study is not to assess the impact of fluctuation on effectiveness, but rather to uncover its internal patterns and drivers. As they are shaped by tradition and collective experience, the armed forces often elude simple statistical models. This study applies linear and multivariate regression analysis to an eight-year dataset (2017–2024) of an infantry battalion within the Hungarian Defence Forces. Rather than seeking to account for the full variance of fluctuation, the objective of the analysis is to capture a defined and major portion of it, thus offering a data-driven, disciplined and objective overview of fluctuation within the land forces.

Keywords: retention, fluctuation, multivariate regression, human resource management, statistical analysis

Introduction

Personnel fluctuation is mainly considered a natural by-product of organisational functions. Academic theories emphasise both its advantageous, inevitable and detrimental effects, balancing its positive and negative interplay across dimensions such as organisational scope, workflow range and fluctuation rate.[2] This publication does not intend to linger on the influence of fluctuation on effectiveness but rather focuses on the drivers and structure of fluctuation.

For this purpose and the overall goal of utility, I chose an infantry battalion of the Hungarian Defence Forces as the sample for my analysis. Comprising 1,136

[1] PhD student, Faculty of Military Science and Officer Training, Doctoral School of Military Sciences.

[2] Lee 2018: 522–523; Abelson–Baysinger 1984: 331–341.

observations, eight variables and a longitudinal sample spanning eight years, this dataset may very well be considered representative of all combat battalions and potentially the broader constitution of the Land Forces. I compiled the dataset utilising the personnel organisational chart of the battalion I served in, with the permission of the brigade commander, relying only on past data (2017–2024) and capturing a snapshot of each year's mid-point (August–September).

My objective with this publication is primarily to provide a data-driven linear regression analysis of personnel fluctuation within an infantry battalion – with its focus on structure and underlying patterns, which I believe have not been published recently. Additionally, this study concentrates on the following specific aims:

- Determine the scope of fluctuation: By establishing the presence and extent of personnel changes within the battalion over an eight-year period.
- Identify key variables: Through examining the dataset to uncover statistically significant variables which drive personnel turnover, while also quantifying their relative impact.
- Develop robust multivariate models: I will construct 1–3 regression models which explain the largest variance in fluctuation, highlighting the interaction of critical drivers.
- Deliver practical insights: I shall present the outcomes in a manner that informs a more profound understanding of personnel fluctuation and emphasise its patterns without moving into policy recommendations.

My hope, by pursuing these objectives, is to contribute to academic discourse as well as to the practical realm of human resource management within the armed forces, through combining statistical rigour with real-world relevance.

Methodology

As the 1848–1849 Hungarian revolutionary general Artúr Görgey observed: "In mere reasoning and even in observation itself, one can be deceived in many ways about reality."[3] Similarly, it takes more than a mechanical approach to data to comprehend all the factors behind soldiers' decisions. Military organisations – unlike many market-driven institutions – are shaped by history, tradition and collective experience. In accordance with the foundations of cultural anthropology, I subscribe to the notion that understanding personal goals may only be fully grasped through holistic and cultural lens. As no statistical model can completely capture these features, most attempts to do so risk reducing human behaviour into mere numbers.

In spite of these limitations, I find statistical tools to remain valuable through revealing patterns otherwise unnoticed. This also counters subjective perspectives built through personal experiences. These, while insightful, are inevitably shaped by cognitive biases and thus overgeneralisation. By employing quantitative methods, I do not seek to fully explain individual decisions but rather account for a significant

[3] Than 1893: 172.

portion of their variance and identify trends which emerge consistently across the dataset. In doing so, I aim to engage the dynamics of military service from multiple angles. By recognising the limitations of explanatory power, I focus on what can be measured while remaining aware of those that cannot be measured.

My preferred method, multivariate regression, builds upon the foundations of linear regression. The latter assumes a linear relationship between variables, with changes in the independent variable resulting in a predictable and gradual shift in the dependent variable. An example could be a soldier's decision to request a unit transfer theoretically influenced by rank. With enough observations, a linear regression could model this by showing that as rank increases, so does the probability of transfer rise or fall at a consistent rate. This is measured partly by regression coefficients, indicating the strength and direction of rank's effect on the likelihood of fluctuation.

As personal decisions are rarely influenced by a single factor, my use of multivariate regression extends this approach to incorporate multiple independent variables. This process not only captures their combined influence on personnel fluctuation but also reveals their interaction within the model. This allows for a more precise and practical representation of how multiple factors collectively shape outcomes. I am using the statistical software R to facilitate this with its Quarto extension and multiple libraries (ggplot2, rio, dplyr, tidyr, knitr). I also rely on R for data manipulation and visualisation.

For the context of this paper, fluctuation refers to the movement of personnel in and out of the observed military unit, including transfers, resignations and retirements. Thus it reflects the changes in the composition of the force over time, capturing both voluntary and involuntary departures. I have chosen this (fluctuation) as the dependent variable of the analysis, with Battalion as our relevant unit of analysis, as the Hungarian Defence Forces' current operational ambition level prioritise tactical cohesion at the Battalion level and below, including Battalion Combat Teams.

Fluctuation is measured through a variable called *Retention index,* which I have favoured for its ability to represent stability versus instability within the unit. With the dataset spanning from 2017 to 2024, it provides a multi-year basis for examining personnel trends. The Retention index captures each soldier's presence in the Battalion for a given year. It ranges from 1 to 8, with higher values indicating a greater Retention index, thus lower levels of fluctuation. For instance, a soldier with a Retention index of 8 has spent all observed years within the unit. As the dataset only captures a personnel status at a single mid-year snapshot – rather than tracking monthly changes – actual fluctuation may be even slightly higher than the recorded data suggest. However, I find this method to be sufficient for indicating broader, multi-year trends in personnel stability.

The dataset consists of 1,136 individual observations, each representing a soldier within the battalion over the period from 2017 to 2024. As I aimed to ensure anonymity and consistency, no names were used in the dataset, rather soldiers were identified through their unique Personal Staff Number (SZTSZ). Each observation in the dataset includes naturally measurable variables, which objectively outline the soldiers' attributes. These include age, sex (1: male, 0: female), status (1: career, 2: contract) and the soldier's hometown's distance to the unit (km) (in the dataset: *Distance to*

Hódmezővásárhely). In the case of the latter, whenever a temporary residence was recorded, I used that address to calculate the distance to the unit's location.

	Position	Position (Quantified)	ID Number	Rank	Rank (Quantified)	Status	Sex	Unit	Unit (Quantified)
85	távbeszélő	12	50015980	őrv.	2	0	0	Tő.tám. Szd.	4
86	századparancsnok	21	36073803	szds.	12	1	1	1. lszd.	1
87	századparancsnok-helyettes	20	50009028	fhdgy.	11	1	1	1. lszd.	1
88	vezénylő zászlós	19	30347902	zls.	8	1	1	1. lszd.	1
89	személyügyi altiszt	37	35267701	ftőrm.	7	0	1	1. lszd.	1
90	beosztott altiszt	38	29267602	ftőrm.	7	1	1	1. lszd.	1
91	beosztott altiszt	38	31190601	ftőrm.	7	1	1	1. lszd.	1
92	harcjárműirányzó	6	36490401	szkv.	4	0	1	1. lszd.	1
93	harcjárműirányzó	6	41125301	szkv.	4	0	1	1. lszd.	1
94	harcjárművezető	7	41490901	szkv.	4	0	1	1. lszd.	1

2017	2018	2019	2020	2021	2022	2023	2024	Retention Index	Age	Home	Distance_to_Hódmezővásárhely_km
1	1	0	0	0	0	0	0	2	33	Szentes	27
1	1	1	1	1	1	1	1	8	41	Szeged	23
1	0	0	0	0	0	0	0	1	36	Nyíregyháza	201
1	1	1	1	1	1	0	0	6	46	Szeged	23
1	1	1	1	1	1	1	1	8	42	Szeged	23
1	1	1	1	1	1	1	1	8	53	Battonya	55
1	1	1	1	1	1	0	1	7	44	Hódmezővásárhely	0
1	1	1	1	1	1	0	1	7	47	Kistelek	27
1	1	1	1	1	0	0	0	5	42	Szentes	27
1	1	1	1	1	1	0	1	7	37	Mezőkovácsháza	47

Figure 1: Quantified variables of the personnel dataset used for regression analysis
Source: compiled by the author based on Structured Personnel Dataset

The other – originally categorical – variables I have quantified using the following logic: rank, position and unit assignment were transformed into numerical values for structured comparison while preserving their underlying distinctions. The easiest case was provided by Rank, as it was assigned a numerical scale corresponding to its hierarchical position, ranging from 1 (private) to 14 (lieutenant colonel), where increasing values indicate a higher position within the military command structure. Position, although in multiple aspects correlated with rank, was quantified to capture a separate dimension – the nature of responsibilities ranging from practical field duties to theoretical assignments. The assigned values reflect this logic: rifleman and recon soldier commencing the most practical positions (1–6), followed by technical specialisations (7–16), small unit leadership (17–24) at the median, management and support roles (25–32) in the upper quartile, and command and staff positions (33–42) at the highest theoretical level, with chief of staff (42) occupying the highest value. Unit assignment was quantified to represent functional distinctions within the battalion. The numerical scale differentiates between combat roles and support elements: infantry companies (1), combat support companies and platoons (2), headquarters and staff (3), and HHC/logistics (4).

Through this approach I aimed to ensure that these interrelated variables capture distinguishable hierarchical, functional and operational aspects that each category represents, allowing each variable to contribute uniquely to the analysis of personnel retention.

Tracing the arc of retention

Before we get to the details of linear regression, I would like to broadly demonstrate the main characteristics of the data and the trends of fluctuation inside a Land Forces battalion. The mean of our dependent variable lies at 3.73, meaning that the 1,136 soldiers moving through the unit stayed for an average close to four years. The mean age of the personnel is around 34 years of age – modestly below the HDF average – which is consistent with the wider characteristics of a combat arms battalion.

Table 1: Summary statistics of selected variables

Metric	Retention index	Age	Distance to Hmvhely in km
Mean	3.73	34.38	61.85
Variance	6.05		
Standard deviation	2.46		
Number of observations	1,136		

Source: compiled by the author based on Structured Personnel Dataset

As we take a quick look at the density plot of the dependent variable, we find it being skewed to the left. This reflects more of a temporary snapshot than a long-term pattern. The reason being a significant wave of recruits, who have only spent a year within the unit in 2024. The more significant insight lies in the peak at six years – that being the most common length of time a soldier spends in the battalion. For future unit rotational planning, this finding holds practical value, as it makes three-year cohesion cycles not only viable but empirically grounded.

As detailed in the previous section, net fluctuation based solely on overall personnel numbers can be plainly misleading: while the total number of soldiers in the battalion remained roughly stable throughout 2021, the actual turnover of individuals still reached 17% – a significant rate by most organisational standards. The first and last years of the sample stand as statistical outliers: 2017 excludes incoming fluctuation, while 2024 omits future withdrawals – making actual turnover in both years higher than what the figures alone indicate.

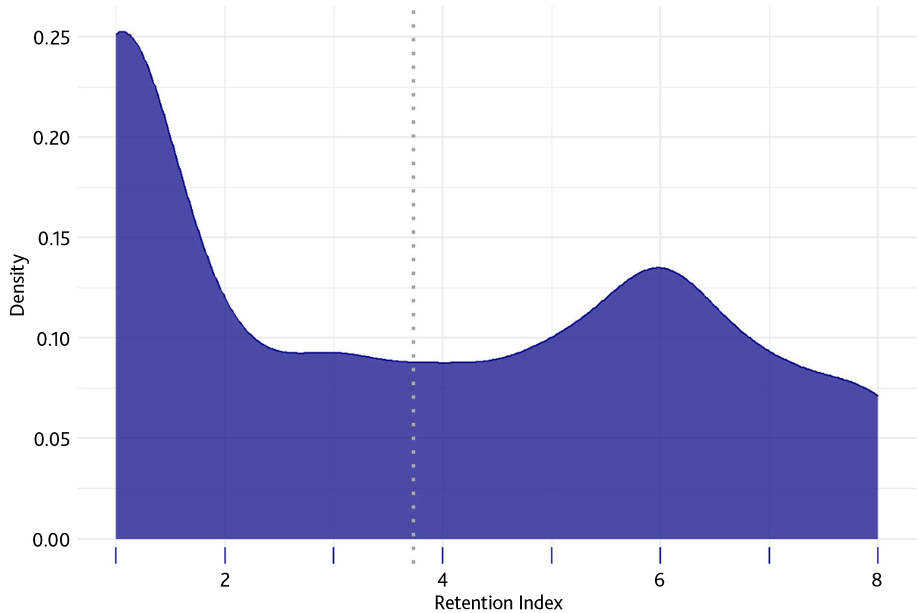

Figure 2: Density distribution of the Retention Index
Source: compiled by the author based on Structured Personnel Dataset

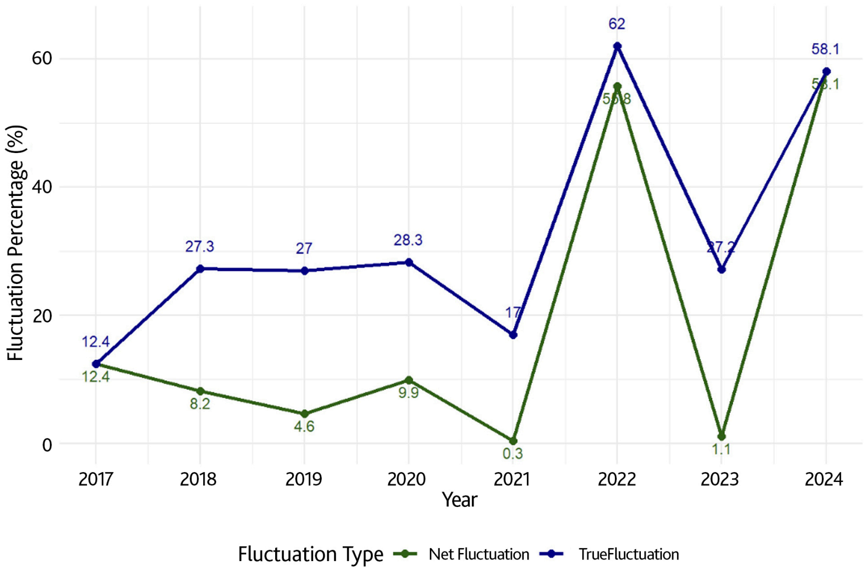

Figure 3: Yearly trends of net and true personnel fluctuation (2017–2024)
Source: compiled by the author based on Structured Personnel Dataset

The relatively stable years of 2018 and 2021 demonstrated a consistently high true fluctuation averaging around 27%. This puts into question the cohesion presumably reached during the 2018 NATO Combat Readiness Evaluation (CREVAL). The years 2017, 2019 and 2023 were generally deployment years. In 2022, the restructuring of the brigade is clearly visible on the graph, with the battalion losing about half of its organisational structure and nearly two-thirds of its NCO corps – reflected in a remarkable 62% personnel fluctuation. From a cohesion-of-force perspective, the continuity of the unit's built-up experience comes into doubt. The 2023–2024 period brought yet another revision: most of the battalion's mechanised positions were re-established, coupled with a focused recruiting effort directed by higher echelons. The training and integration of approximately 300 enlisted soldiers not only reshaped the composition of the force, but also peaked in another extremely high fluctuation value of 58%. The conclusion implies a natural question: can a unit function effectively with a yearly average fluctuation rate of 32.41% – one-third of its personnel changing annually – and is this figure representative of the wider Land Forces? (Regarding the wider Hungarian workforce as a comparative background, a 2018 special report on human resource management in the Hungarian Defence Forces notes that over half of companies report fluctuation rates exceeding 5% – considered high, while 13% report rates between 20% and 40%.)[4]

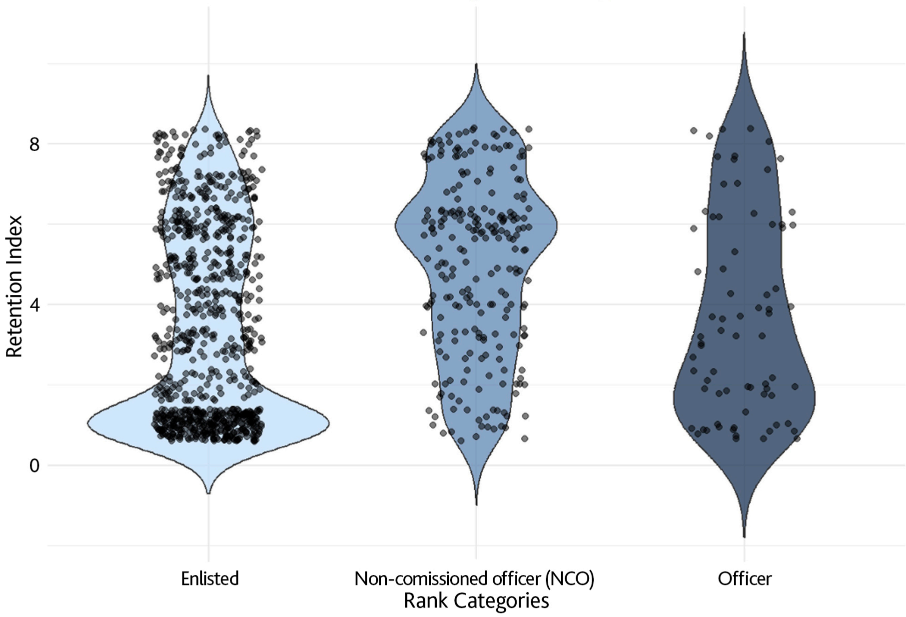

Figure 4: Distribution of the Retention Index across rank categories
Source: compiled by the author based on Structured Personnel Dataset

4 Jobbágy et al. 2018: 58.

Retention differentiated by rank – as visualised in a violin chart – holds both surprises and expectations confirmed. The large cohort of soldiers at the base of the Enlisted category represents new recruits, temporarily skewing retention data lower. NCOs – unsurprisingly – show the highest retention values, rightly functioning as the backbone of the unit, with the majority remaining at the battalion for 6–8 years within the observed period. Officers, on the other hand, average a Retention Index closer to two years, a value visibly shorter than what the new career strategy proposes.

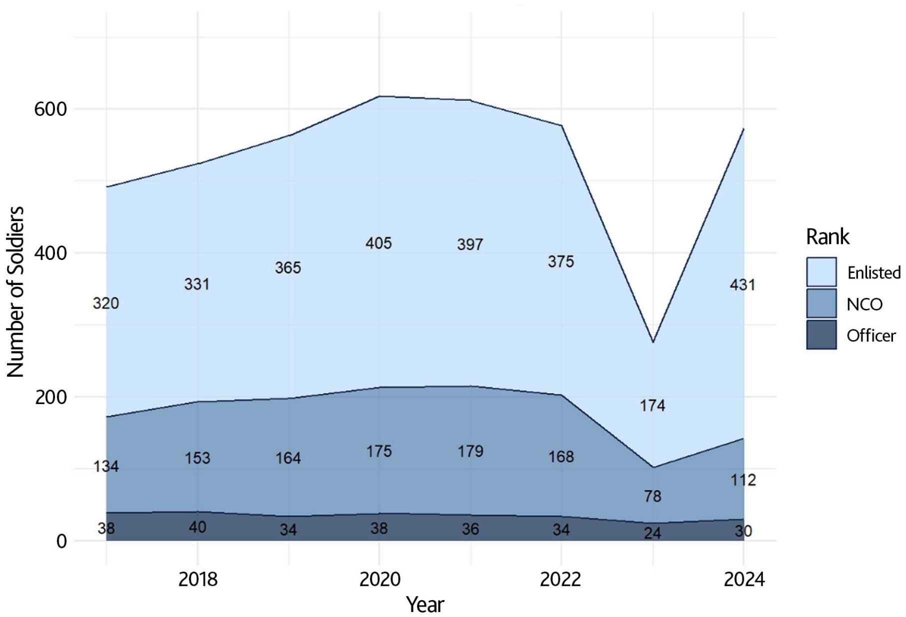

Figure 5: Changes in personnel distribution by rank over time (2017–2024)
Source: compiled by the author based on Structured Personnel Dataset

The overall strength of the Battalion followed a healthy upward trajectory up to 2022, defined by a stable NCO core and an expanding enlisted cohort. The formation of the new brigade within the garrison brought newly established organisational elements, manifesting in the dip of 2023, promptly reversed by 2024. A notable success at this level, as the rebound was built upon a new concept of recruitment and basic training. What lingers, however, is a marked shortage of NCOs and officers. While the analysis of personnel structure falls outside the scope of this paper, the pattern is worth noting. A rising enlisted ratio is not inherently negative; on the contrary, it may very well reflect a more operationally grounded force structure. Yet a question posed from a different perspective could also arise: how does a combat battalion experience a shortage of officers and NCOs, while the broader military structure tends towards a relative surplus of commissioned ranks – often concentrated in units with few or no enlisted personnel?

Linear regression

Before turning to our main multivariate regression models, I would like to broadly present the individual regression values of the independent variables. The p-values, t-statistics, coefficients and R-squared values all indicate a statistically significant relationship between Fluctuation (measured through the Retention Index) and the variables of *Status, Rank, Age, Position* and *Distance to the Unit.* While the variable *Unit* carries modest statistical significance, its low R-squared describes negligible variation (0.6%) of the dependent variable. Notably, *Sex* holds no statistical relevance and does not exhibit any meaningful linear relationship with Fluctuation. R-squared values show us the proportion of variance (in Retention) explained by each variable. Consistent with the theoretical framing in the Introduction, no single factor explains the majority of retention variance – but several account for a meaningful portion of it. *Age* stands out, independently explaining 23.8% of the variation.

Table 2: Univariate linear regression results by independent variable

Variable	P-value significance	R^2 (%)	Coeffi-cient	Coefficient interpretation
Status	*** (< 2e-16)	5.2	1.4453	Changing from contracted to career increases retention by ~1.45 years
Rank (Quantified)	*** (< 2e-16)	6.0	0.2201	Each rank level higher increases retention by ~2.5 months
Age	*** (< 2e-16)	23.8	0.1248	Each additional year of age increases retention by ~1.5 months
Position (Quantified)	*** (< 2e-16)	5.8	0.0658	More theoretical/command roles increase retention by ~3.5 weeks per level
Distance to Hmvhely	*** (< 2e-16)	6.5	–0.0096	Each km farther from the unit decreases retention by ~3.5 days
Unit (Quantified)	** (p = 0.0075)	0.6	0.1658	Moving from a combat unit to a combat support or staff element increases retention by ~2 months
Sex	• (p ≈ 0.096)	0.2	0.4383	Gender does not have a statistically significant relationship with fluctuation

Note on statistical significance:
*** $p < 0.001$ – *Very highly statistically significant*
** $p < 0.01$ – *Highly statistically significant*
* $p < 0.05$ – *Statistically significant*
• $p < 0.1$ – *Marginally significant*
Source: compiled by the author based on Structured Personnel Dataset

As mentioned before, coefficients indicate the strength and direction of the linear relationship. In statistical terms, a one-unit increase in Retention – equivalent to a soldier remaining at the same unit for one additional year – is associated with a measurable change in the independent variable. For instance, a change in *Status* from contracted (0) to career (1) corresponds to a 1.44-year increase in Retention. Another noteworthy coefficient is *Rank,* where each level of advancement predicts an additional 2.5 months of retention. As for *Age,* being a year older correlates with 1.5 months longer service at the battalion. While the effect of *Distance* might seem modest at first, one should take into consideration that it is measured on a much broader numerical scale than binary variables like *Status.* Even though the individual coefficient appears less impactful, it may very well carry greater statistical weight when considered along other factors. This leads us to multivariate models, built on the five predictors found to be highly statistically significant.

Multivariate regression

Multivariate regression allows us to evaluate the interplay of chosen factors, demonstrating their 'true' effect on fluctuation within a complex and noisy organisational trend.[5] For the last phase of the study, I have constructed ten multivariate regression models, each incorporating three independent variables. (Table 3 serves to present a transparent selection process.)

Table 3: Summary table of ten multivariate regression models

Model	Predictors	P-value Statistical significance level	Adj. R^2 (%) 0.00–1.00 Variance explained by the model, higher values indicate better explanatory power	Multicollinearity (VIF) Degree of predictor correlation 1–5 = Low > 5 = High > 10 = Severe
Model 1	Status	< 0.001***	0.279	1.06
	Age	< 0.001***		1.10
	Distance to the Unit	< 0.001***		1.04
Model 2	Status	< 0.001***	0.256	2.18
	Age	< 0.001***		1.29
	Rank	0.002**		2.59
Model 3	Status	< 0.001***	0.252	1.72
	Age	< 0.001***		1.25
	Position	0.036*		2.02

5 Izenman 2013: 159–160.

Model	Predictors	P-value Statistical significance level	Adj. R^2 (%) 0.00–1.00 Variance explained by the model, higher values indicate better explanatory power	Multicollinearity (VIF) Degree of predictor correlation 1–5 = Low > 5 = High > 10 = Severe
Model 4	Status	0.016*	0.131	2.13
	Distance to the Unit	< 0.001***		1.00
	Rank	< 0.001***		2.13
Model 5	Status	< 0.001***	0.127	1.72
	Distance to the Unit	< 0.001***		1.01
	Position	< 0.001***		1.72
Model 6	Status	0.042*	0.067	2.19
	Rank	0.070 •		3.33
	Position	0.018*		2.68
Model 7	Age	< 0.001***	0.266	1.33
	Distance to the Unit	< 0.001***		1.05
	Rank	0.055 •		1.28
Model 8	Age	< 0.001***	0.265	1.28
	Distance to the Unit	< 0.001***		1.04
	Position	0.152		1.24
Model 9	Age	< 0.001***	0.237	1.29
	Rank	0.793		2.70
	Position	0.553		2.66
Model 10	Distance to the Unit	< 0.001***	0.129	1.01
	Rank	< 0.001***		2.63
	Position	0.074 •		2.64

Source: linear regression code executed in R based on the Structured Personnel Dataset

Of these, five models (Models 1–5) showed consistently statistically significant results with p-values under 0.05. The variable Position – quantified to capture the level of practicality – initially showed significant power in univariate models, but had its influence diminished once Age was included. In fact, Position derived its explanatory power from its correlation with Age, as older soldiers are naturally more likely to hold administrative or command-related positions. Thus, Position functioned as a quasi-proxy for Age in models that lacked the latter. Several models yielded high predictive strength, underscoring the robustness of the dataset. I chose two to focus on in particular, both for the magnitude of their effects and the causal depth offered by them.

```r
# Model 1: Status, Age, Distance_to_Hódmezővásárhely_km
model_1 <- lm(`Retention Index` ~ Status + Age + Distance_to_Hódmezővásárhely_km, data = dataset
summary(model_1)
```

```
Call:
lm(formula = `Retention Index` ~ Status + Age + Distance_to_Hódmezővásárhely_km,
    data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2261 -1.4775 -0.2909  1.6549  6.5970

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                      0.2541373  0.2533386   1.003    0.316
Status                           0.8231612  0.1650903   4.986 7.12e-07 ***
Age                              0.1085809  0.0067615  16.059  < 2e-16 ***
Distance_to_Hódmezővásárhely_km -0.0066176  0.0009653  -6.856 1.16e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.089 on 1132 degrees of freedom
Multiple R-squared:  0.2813,    Adjusted R-squared:  0.2794
F-statistic: 147.7 on 3 and 1132 DF,  p-value: < 2.2e-16
```

Figure 6: Multivariate linear regression results for Model 1

Source: linear regression code executed in R based on the Structured Personnel Dataset

Model 1 – consisting of *Age, Status* (career or enlisted), and *Distance to the Unit* – stood out in terms of adjusted R^2 (0.2794), as it described close to 30% of the variation within the Retention Index. The p-values ($p < 0.001$) were highly statistically significant for all three predictors. Where I was paying particular attention was the Variance Inflation Factor (VIF), coming in consistently low and indicating no multicollinearity. That is, the variables were not strongly correlated and described separate portions of variance. With that we can note that higher values in Age and Status, combined with lower Distance, jointly contribute to the lowest fluctuation within the unit. This is the least surprising, yet statistically most potent result. Additional insights can be compiled through the coefficients produced by this model.

Status, as a binary variable, yields a measurable positive effect on retention. Career soldiers statistically remain with the unit for 10 months longer (0.82 years) than their enlisted peers. While the difference is relevant in the context of cumulative cohesion, it is smaller than expected given the institutional weight typically attached to professional status.

Distance to the Unit has a measurable, though surprisingly modest effect on retention in the multivariate model. The negative coefficient of –0.0066176 predicts that a soldier co-located in the brigade's town (0 km) remains with the unit only about 3.7 months longer than a colleague living in a major town 47 km away, across county lines. While statistically significant, the difference is far less substantial than anticipated and generally aligns with my practical experience. In reality, once a soldier commits to a unit and is accepted into its community, geographic proximity tends to be secondary to group identity and professional trajectory. This evidence supports current Hungarian military planning: building one central, professional military base per region (e.g. Southern Great Plain) in addition to developing a more widespread distribution of territorial Defence units (one regiment per county).

Age, on the other hand, emerges as the most influential individual variable in the model. Its coefficient of 0.1086 indicated that each additional year of age corresponds to approximately 1.3 months longer retention. The cumulative effect is also notable: a 40-year-old soldier will remain in the unit for about 2.2 years longer than a 20-year-old peer, as observed within our 8-year research period. This underscores the stabilising role age plays in personnel cohesion. The age-retention relationship likely captures both career anchoring and reduced workforce mobility among older soldiers, thus both voluntary and structural patterns.

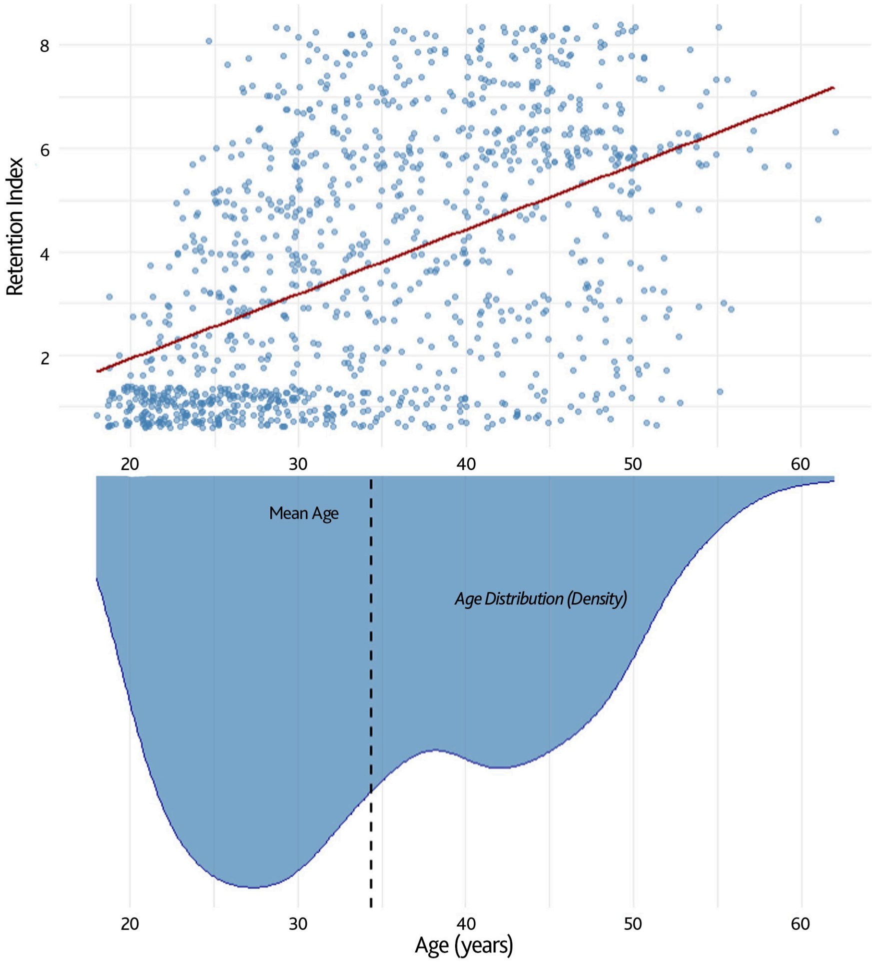

Figure 7: Scatterplot of Retention Index by age

Source: compiled by the author based on Structured Personnel Dataset

Even though the scatterplot appears widely scattered at first glance, the linear relationship between age and retention is both present and robust. So what about the intuitive assumption: the passing of years simply offering more opportunity for older soldiers to serve? While that may sound reasonable, in a longitudinal sample where average yearly fluctuation exceeds 30%, this assumption fails to hold. What practical conclusion may emerge from this correlation? Primarily, it confirms the renewed focus on winning over the cohort of soldiers between the ages of 18 and 34. Beyond that age, the data visibly stabilises, with the majority of personnel serving 6 to 8 (and in reality often more) years in the same unit. This does not imply holding on to unproductive personnel or lowering standards; as a combat arms unit is clearly not for everyone. Nor should we aim for zero fluctuation – letting some individuals go remains inevitable.

On the contrary, offering enlisted soldiers viable career paths past the age of 40 would likely reduce early departures and capitalise on the strong positive correlation between age and retention. A potential parallel could be drawn with the U.S. Army's E-4 rank (Specialist) – a position for soldiers who do not enter the Non-Commissioned Officer track but whose technical proficiency brings substantial value in maintenance, training, storage or supply positions. To put it plainly: if we can retain soldiers beyond the age of 34, the likelihood of self-initiated fluctuation will decrease significantly.

```r
# Model 2: Status, Age, Rank (Quantified)
model_2 <- lm(`Retention Index` ~ Status + Age + `Rank (Quantified)`, data = dataset)
summary(model_2)
```

```
Call:
lm(formula = `Retention Index` ~ Status + Age + `Rank (Quantified)`,
    data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0437 -1.5139 -0.4043  1.7028  5.4137 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.487954   0.235183  -2.075  0.03823 *  
Status               1.307432   0.240311   5.441 6.51e-08 ***
Age                  0.127686   0.007439  17.163  < 2e-16 ***
`Rank (Quantified)` -0.117867   0.037050  -3.181  0.00151 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.122 on 1132 degrees of freedom
Multiple R-squared:  0.2581,	Adjusted R-squared:  0.2562 
F-statistic: 131.3 on 3 and 1132 DF,  p-value: < 2.2e-16
```

Figure 8: Multivariate linear regression results for Model 2
Source: linear regression code executed in R based on the Structured Personnel Dataset

Model 2, though slightly behind in explanatory power (Adjusted $R^2 = 0.2562$), offers perhaps the most surprising insight. The three independent variables I have applied were *Rank, Status* and *Age,* which combined explained 25.6% of the variation in retention. As a result, when controlling for professional Status and Age, Rank demonstrates a negative coefficient of –0.1178 with Retention index. This portrays that advancing one rank level corresponds to a shorter service of approximately 1.4 months in the multivariate model. This data directly contradicts the widespread belief that higher rank is correlated with longer service within the unit. Despite the fact that Rank showed a modest positive effect on retention in univariate regression, this effect reverses in a more "real-world" and holistic setting. Thus, the data implies that a Captain (OF-2) statistically remains at the unit roughly one year less than a Private (OR-1 / Őrvezető). This finding challenges organisational concepts and indicates that, in the absence of stabilising factors such as career status and age, hierarchical advancement does in fact accelerate fluctuation.

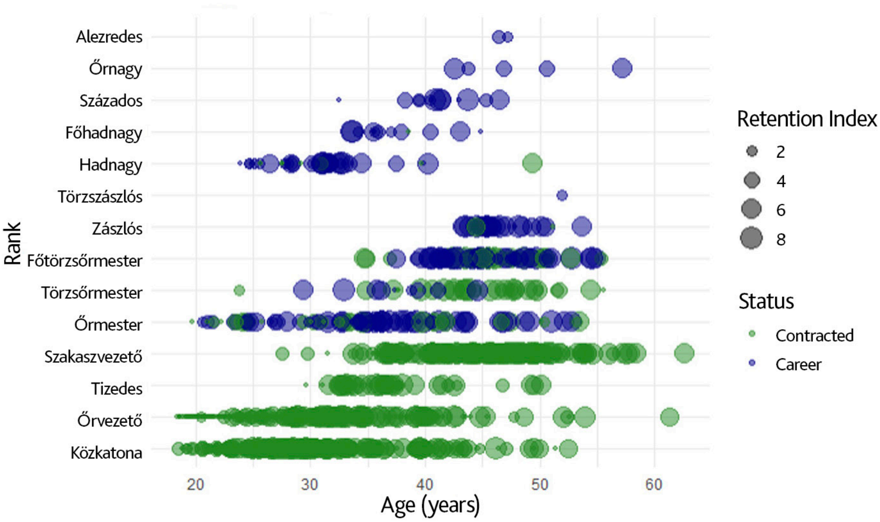

Figure 9: Bubble chart of age, rank and retention patterns

Source: compiled by the author based on Structured Personnel Dataset

As for practical implications, this offers a sobering takeaway. While officer mobility – whether due to institutional career progression or voluntary transfer – may contribute to the trend, it cannot fully account for the sharply negative correlation between Rank and Retention. In conclusion, this paper must take into consideration that higher rank typically corresponds with higher educational level and greater responsibility. Despite the understandably more mobile career path of officers, higher rank – higher fluctuation is an alarming trend, which may generally appear due to underlying drivers:

higher unpredictability compared to working conditions of civilian sector university degree holders, a diminishing wage gap between lower and higher ranks (despite the gap in responsibility), and a counter-selection dynamic working against less conformist officers – or a combination of these factors.[6] Although I fully subscribe to the historic ethos of the military officer's creed and the elevated expectations it entails, the higher fluctuation rates among officers present a challenge the organisation must confront – with clarity and without self-deception.

What we measure may be numerical, but what we observe are human lives in transition. The profound experience of my decade-long tenure in leadership positions has taught me that, in their work environment, soldiers primarily seek long-term predictability and a sense of group identity through cohesive teams.

It is understandable that our mission set cannot remain static or non-dynamic, but that does not necessarily mean our organisations should mirror that volatility. Among other factors, task delegation through individualised personnel databases – so-called task-organised temporary units formed for every domestic and external deployment, whether for one week or six months – and the concurrent existence of multiple readiness formations (HDF, NATO and EU Battlegroups) all contribute to organisational fluctuation and the erosion of cohesion within units (at the platoon, company and battalion levels). On the personnel level and somewhat ironically, individual fluctuation (initiated by the soldier) is likely driven upward by organisational fluctuation, resulting in year-on-year rates between 30% and 60%.

While this paper remains grounded in the qualified dataset, its core contribution lies in quantifying retention and building multivariate models to understand its underlying drivers. Although concrete policy recommendations remain outside the scope of this study, I find there is place for cautious reflection on possible directions for future consideration based on the practical nature of the findings. If tactical formations are to maintain operational effectiveness, the first step is recognising personnel fluctuation as a challenge. The next is addressing its drivers – not through further modelling, but by designing interventions that reduce turnover and support stability at the unit level.

Recently, higher echelons have made visible efforts to allocate breathing room to so-called Pilot units and provide them with the time needed to build proficiency and cohesion – steps that are very welcome. Although fluctuation between units should be understood as context-dependent and related to rank – differing in scope for enlisted soldiers, non-commissioned officers and officers, and to some extent necessary – certain key terms emerge within the academic world: *controlled* and *synchronised*.[7]

Controlled, in terms of cycles: suppressing fluctuation for a defined period (e.g. three years), while providing a designated window for a coordinated, large-scale transfer phase – aiming for a controlled 30% fluctuation rate that takes individual preferences into account. (This supports interoperability and Land Forces-wide standardisation goals.)

[6] Jobbágy et al. 2018: 55–62.

[7] A good example is provided through the evolution of U.S. Army readiness systems: the Army Force Generation (ARFORGEN) model (2006–2016), followed by the Sustainable Readiness Model (SRM, 2017–2021), and more recently the Regionally Aligned Readiness and Modernization Model (ReARMM, 2021–). Linick et al. 2023: 1–5.

Synchronised, in terms of Land Forces brigades, regiments and their Mission Essential Task List generation – structured around a three-year cycle that culminates in a collective training event and an external deployment or allied reaction force assignment.

Although this concept is presumably already embedded within both the Army Planning Directorate (HVK Haderőtervezési Csoportfőnökség) and the Hungarian Defence Forces Joint Operations Command (Magyar Honvédség Összhaderőnemi Műveleti Parancsnokság), integrating *controlled* and *synchronised* fluctuation into rotation cycles would certainly yield strong results for internal cohesion – and, indirectly, for the soldier's sense of purpose and esprit de corps.

References

Abelson, Michael – Baysinger, Barry (1984): Optimal and Dysfunctional Turnover: Toward an Organizational Level Model. *Academy of Management Review,* 9(2), 331–341. Online: https://doi.org/10.2307/258446

Hausknecht, John – Rodda, Julianne – Howard, Michael (2009): Targeted Employee Retention: Performance-Based and Job-Related Differences in Reported Reasons for Staying. *Human Resource Management,* 48(2), 269–288. Online: https://doi.org/10.1002/hrm.20279

Izenman, Alan J. (2013): Multivariate Regression. In *Modern Multivariate Statistical Techniques. Regression, Classification, and Manifold Learning.* New York: Springer, 159–194. Online: https://doi.org/10.1007/978-0-387-78189-1_6

Jobbágy, Zoltán – Krizbai, János – Papp, Tibor (2018): *Tematikus Honvédelmi Jelentés. Haderő és humánerőforrás-gazdálkodás. Speciális Jelentés 2018.* Budapest: Nemzeti Közszolgálati Egyetem.

Lee, Shinwoo (2018): Employee Turnover and Organizational Performance in U.S. Federal Agencies. *The American Review of Public Administration,* 48(6), 522–534. Online: https://doi.org/10.1177/0275074017715322

Linick, Michael E. – Eckhause, Jeremy M. – Saum-Manning, Lisa – Orvis, Bruce R. – Abler, Andrea M. – Wallace, Sam – Pham, Phoebe F. – Baker, Sarah (2023): *Addressing the Friction Between the Army's People First Initiatives and Its Readiness Generation Process.* Santa Monica: RAND Corporation, RR-A2006-1. Online: https://www.rand.org/pubs/research_reports/RRA2006-1.html

Than, Károly (1893): Egy magyar hadvezér mint chemikus. *Budapesti Szemle,* 74(197), 161–180.