

Static PID-5 and ESI

Corrado Caudek

Le misure “basali” corrispondenti al questionario ESI.

```
# Read and process 'esi_bf' data
esi_bf <- rio::import(
  here::here(
    "data",
    "processed",
    "esi_bf.csv"
  )
) |>
dplyr::distinct(user_id, .keep_all = TRUE) |> # Keep only distinct user_id
dplyr::select(user_id, esi_bf) # Select relevant columns

# Read and process 'pid5' data
pid5 <- rio::import(
  here::here(
    "data",
    "processed",
    "pid5.csv"
  )
) |>
dplyr::distinct(user_id, .keep_all = TRUE) |> # Keep only distinct user_id
dplyr::select(user_id, starts_with("domain_")) # Select domain variables

# Merge 'esi_bf' and 'pid5' data by user_id
df <- left_join(esi_bf, pid5, by = "user_id")

# Define list of user IDs with careless responding
user_id_with_careless_responding <- c(
  "ma_se_2005_11_14_490",
  "reve20041021036",
  "di_ma_2005_10_20_756",
  "pa_sc_2005_09_10_468",
  "il_re_2006_01_18_645",
  "so_ma_2003_10_13_804",
  "lo_ca_2005_05_07_05_437",
  "va_ma_2005_05_31_567",
  "no_un_2005_06_29_880",
  "an_bo_1988_08_24_166",
  "st_ma_2004_04_21_426",
  "an_st_2005_10_16_052",
  "vi_de_2002_12_30_067",
  "gi_ru_2005_03_08_033",
```

```

"al_mi_2005_03_05_844",
"la_ma_2006_01_31_787",
"gi_lo_2004_06_27_237",
"ch_bi_2001_01_28_407",
"al_pe_2001_04_20_079",
"le_de_2003_09_05_067",
"fe_gr_2002_02_19_434",
"ma_ba_2002_09_09_052",
"ca_gi_2003_09_16_737",
"an_to_2003_08_06_114",
"al_se_2003_07_28_277",
"ja_tr_2002_10_06_487",
"el_ci_2002_02_15_057",
"se_ti_2000_03_04_975",
"co_ga_2003_10_29_614",
"al_ba_2003_18_07_905",
"bi_ro_2003_09_07_934",
"an_va_2004_04_08_527",
"ev_cr_2003_01_27_573"
)

# Filter out users with careless responses
df1 <- df[!(df$user_id %in% user_id_with_careless_responding), ]

# Read EMA data and rename 'subj_code' to 'user_id'
ema_raw <- readRDS(
  here::here(
    "data",
    "raw",
    "ema",
    "ema_data_scoring.RDS"
  )
) |>
dplyr::rename(
  user_id = subj_code
)

# Merge EMA data with filtered main data
df2 <- left_join(df1, ema_raw, by = "user_id")

# Verify number of unique users
length(unique(df2$user_id))

[1] 429

```

Compliance

Escludiamo i soggetti che hanno risposto a meno di 10 notifiche.

```

# Conta quante risposte EMA ha fornito ciascun soggetto
user_counts <- df2 %>%
  group_by(user_id) %>%
  summarise(n_responses = n()) %>%

```

```

ungroup()

# Tieni solo i soggetti con almeno 10 risposte
valid_users <- user_counts %>%
  filter(n_responses >= 10) %>%
  pull(user_id)

# Filtra il dataframe originale
df2 <- df2 %>%
  dplyr::filter(user_id %in% valid_users)

length(unique(df2$user_id))

[1] 379

```

Generate negative instant mood

```

# Costruisce una misura media dell'affetto negativo momentaneo

# Seleziona solo le colonne rilevanti (per velocità)
items <- c("sad", "angry", "happy", "satisfied")

# Imputa i missing (1 solo imputazione, dato che i NA sono pochi)
imputed <- mice(df2[, items], m = 1, maxit = 10, seed = 123)

iter imp variable
1 1 sad angry happy satisfied
2 1 sad angry happy satisfied
3 1 sad angry happy satisfied
4 1 sad angry happy satisfied
5 1 sad angry happy satisfied
6 1 sad angry happy satisfied
7 1 sad angry happy satisfied
8 1 sad angry happy satisfied
9 1 sad angry happy satisfied
10 1 sad angry happy satisfied

# Estrai il dataset imputato e sostituisci le colonne originali
df2_imputed <- complete(imputed)
df2[, items] <- df2_imputed[, items]

df2 <- df2 %>%
  mutate(
    happy_reversed = 100 - happy, # Scala 0-100
    satisfied_reversed = 100 - satisfied,
    neg_aff_ema = rowMeans(
      cbind(sad, angry, happy_reversed, satisfied_reversed),
      na.rm = TRUE
    )
  )

df3 <- df2 %>%
  dplyr::select(

```

```

    user_id,
    esi_bf,
    neg_aff_ema,
    starts_with("domain_"),
    pid5_negative_affectivity, pid5_detachment, pid5_antagonism,
    pid5_disinhibition, pid5_psychoticism
  )

df4 <- df3[!is.na(df3$domain_negative_affect), ]
length(unique(df4$user_id))

[1] 350

# Imputa i missing (1 solo imputazione, dato che i NA sono pochi)
imputed <- mice(df4, m = 1, maxit = 10, seed = 123)

iter imp variable
1 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
2 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
3 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
4 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
5 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
6 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
7 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
8 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
9 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p
10 1 pid5_negative_affectivity pid5_detachment pid5_antagonism pid5_disinhibition p

Warning: Number of logged events: 1

# Estrai il dataset imputato e sostituisci le colonne originali
df5 <- complete(imputed)

df5_scaled <- df5 %>%
  dplyr::mutate(
    # Applica la standardizzazione (scale) a tutte le colonne selezionate
    # tranne user_id. as.vector() è usato per assicurare che l'output sia un vettore.
    dplyr::across(
      c(
        esi_bf,
        neg_aff_ema,
        domain_negative_affect,
        domain_detachment,
        domain_antagonism,
        domain_disinhibition,
        domain_psychoticism,
        pid5_negative_affectivity,
        pid5_detachment,
        pid5_antagonism,
        pid5_disinhibition,
        pid5_psychoticism
      ),
      ~ as.vector(scale(.))
    )
  )

```

)

Nel caso presente:

esi_bf è costante entro ogni user_id. Le variabili come neg_aff_ema e le pid5_ sono tempovarianti. La variabile esi_bf è costante nel tempo per ciascun soggetto, quindi non è possibile usare dati a livello momentaneo (EMA) per predire variazione intra-soggettiva che non esiste nell'outcome.

Il modo corretto di affrontare la domanda di ricerca, preservando la validità statistica del confronto, è:

- Aggregare le variabili EMA (come neg_aff_ema, pid5_negative_affectivity, ecc.) a livello del soggetto.
- Usare media per il confronto.

Costruiamo dunque due modelli alternativi a livello soggetto:

- model_base_subject: senza interazioni,
- model_alt_subject: con le interazioni (tra tratti stabili e tratti EMA aggregati).

```
df_subject <- df5_scaled %>%
  group_by(user_id) %>%
  summarise(
    esi_bf = first(esi_bf),
    neg_aff_ema_m = mean(neg_aff_ema, na.rm = TRUE),
    pid5_neg_aff_m = mean(pid5_negative_affectivity, na.rm = TRUE),
    pid5_detach_m = mean(pid5_detachment, na.rm = TRUE),
    pid5_antag_m = mean(pid5_antagonism, na.rm = TRUE),
    pid5_disin_m = mean(pid5_disinhibition, na.rm = TRUE),
    pid5_psych_m = mean(pid5_psychoticism, na.rm = TRUE),
    domain_negative_affect = first(domain_negative_affect),
    domain_detachment = first(domain_detachment),
    domain_antagonism = first(domain_antagonism),
    domain_disinhibition = first(domain_disinhibition),
    domain_psychoticism = first(domain_psychoticism)
  )

model_base <- brm(
  esi_bf ~ 1 +
    domain_negative_affect + domain_detachment +
    domain_antagonism + domain_disinhibition + domain_psychoticism,
  data = df_subject,
  family = asym_laplace(),
  prior = c(
    prior(normal(0, 1), class = "Intercept"),
    prior(normal(0, 1), class = "b"),
    prior(exponential(1), class = "sigma")
  ),
  chains = 4,
  cores = 4,
  iter = 2000,
  seed = 123,
  backend = "cmdstanr",
  # algorithm = "meanfield",
```

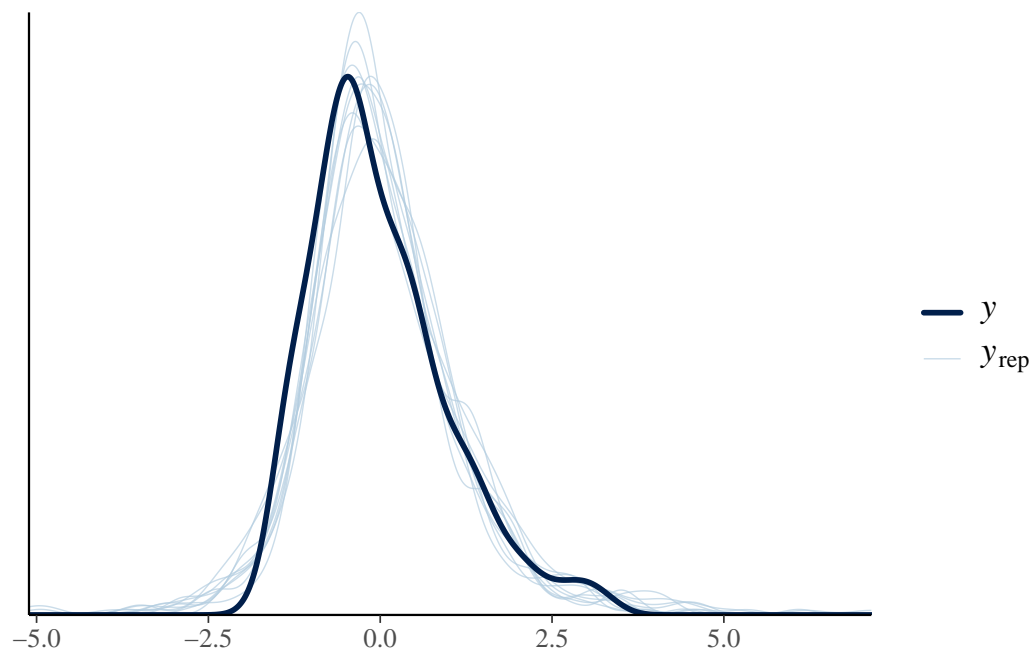
```

    save_pars = save_pars(all = TRUE)
  )

# Posterior predictive check for the baseline model
pp_check(model_base)

Using 10 posterior draws for ppc type 'dens_overlay' by default.

```



```

print(model_base)

Family: asym_laplace
Links: mu = identity; sigma = identity; quantile = identity
Formula: esi_bf ~ 1 + domain_negative_affect + domain_detachment + domain_antagonism + domain_disinhibition + domain_psychoticism
Data: df_subject (Number of observations: 350)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000

```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS
Intercept	-0.34	0.09	-0.52	-0.17	1.00	1798
domain_negative_affect	-0.13	0.06	-0.23	-0.02	1.00	2165
domain_detachment	-0.10	0.05	-0.21	0.00	1.00	2476
domain_antagonism	0.07	0.05	-0.03	0.18	1.00	2863
domain_disinhibition	0.25	0.06	0.13	0.37	1.00	2377
domain_psychoticism	0.17	0.07	0.04	0.31	1.00	2271

Tail_ESS

Intercept	1618
domain_negative_affect	2484
domain_detachment	2921
domain_antagonism	2694
domain_disinhibition	2612
domain_psychoticism	2084

Further Distributional Parameters:

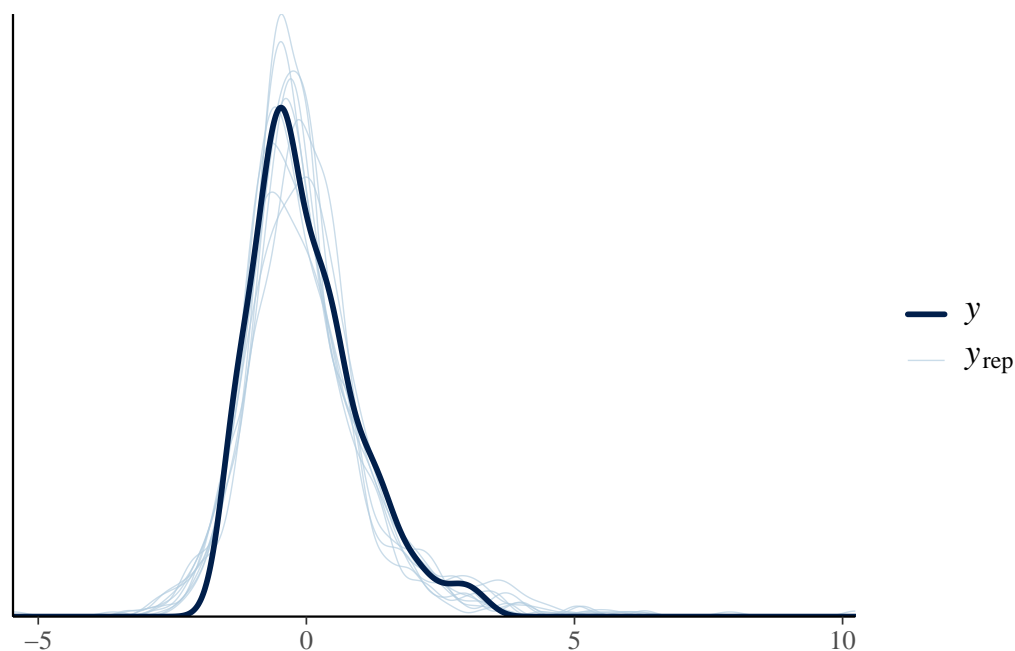
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.33	0.02	0.29	0.37	1.00	2285	2487
quantile	0.38	0.04	0.31	0.45	1.00	1745	1809

Draws were sampled using `sample(hmc)`. For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
model_alt <- brm(
  esi_bf ~ 1 +
    domain_negative_affect * pid5_neg_aff_m +
    domain_detachment * pid5_detach_m +
    domain_antagonism * pid5_antag_m +
    domain_disinhibition * pid5_disin_m +
    domain_psychoticism * pid5_psych_m,
  data = df_subject,
  family = asym_laplace(),
  prior = c(
    prior(normal(0, 1), class = "Intercept"),
    prior(normal(0, 1), class = "b"),
    prior(exponential(1), class = "sigma")
  ),
  chains = 4,
  cores = 4,
  iter = 2000,
  seed = 123,
  backend = "cmdstanr",
  # algorithm = "meanfield",
  save_pars = save_pars(all = TRUE)
)

pp_check(model_alt)
```

Using 10 posterior draws for ppc type 'dens_overlay' by default.



```
print(model_alt)
```

```
Family: asym_laplace
Links: mu = identity; sigma = identity; quantile = identity
Formula: esi_bf ~ 1 + domain_negative_affect * pid5_neg_aff_m + domain_detachment * pid5_detach_m
Data: df_subject (Number of observations: 350)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat
Intercept	-0.48	0.13	-0.73	-0.24	1.00
domain_negative_affect	-0.05	0.07	-0.19	0.08	1.00
pid5_neg_aff_m	-0.11	0.09	-0.28	0.06	1.00
domain_detachment	-0.12	0.05	-0.23	-0.01	1.00
pid5_detach_m	0.14	0.08	-0.03	0.30	1.00
domain_antagonism	0.04	0.06	-0.07	0.16	1.00
pid5_antag_m	-0.01	0.09	-0.20	0.17	1.00
domain_disinhibition	0.22	0.07	0.09	0.35	1.00
pid5_disin_m	-0.03	0.10	-0.23	0.17	1.00
domain_psychoticism	0.18	0.07	0.04	0.31	1.00
pid5_psych_m	-0.04	0.13	-0.28	0.21	1.00
domain_negative_affect:pid5_neg_aff_m	0.02	0.07	-0.12	0.14	1.00
domain_detachment:pid5_detach_m	-0.07	0.05	-0.16	0.04	1.00
domain_antagonism:pid5_antag_m	0.17	0.08	0.03	0.32	1.00
domain_disinhibition:pid5_disin_m	-0.00	0.08	-0.16	0.16	1.00
domain_psychoticism:pid5_psych_m	-0.07	0.07	-0.21	0.07	1.00
	Bulk_ESS	Tail_ESS			
Intercept	1184	1863			
domain_negative_affect	2016	2433			
pid5_neg_aff_m	2132	2583			
domain_detachment	2579	2527			

pid5_detach_m	2863	2189
domain_antagonism	2521	2748
pid5_antag_m	1892	2301
domain_disinhibition	2601	2486
pid5_disin_m	2592	2305
domain_psychoticism	3039	2978
pid5_psych_m	1849	2813
domain_negative_affect:pid5_neg_aff_m	2811	2499
domain_detachment:pid5_detach_m	2979	2433
domain_antagonism:pid5_antag_m	2785	2518
domain_disinhibition:pid5_disin_m	2249	2662
domain_psychoticism:pid5_psych_m	1911	2829

Further Distributional Parameters:

	Estimate	Est.Error	1-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.31	0.03	0.25	0.36	1.00	1339	1753
quantile	0.33	0.05	0.24	0.42	1.00	1097	1537

Draws were sampled using `sample(hmc)`. For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
loo0 <- loo(model_base, save_psis = TRUE)
loo1 <- loo(model_alt, save_psis = TRUE)
loo_compare(loo0, loo1)
```

	elpd_diff	se_diff
model_base	0.0	0.0
model_alt	-7.7	4.6

```
bayes_R2(model_base)
```

	Estimate	Est.Error	Q2.5	Q97.5
R2	0.1189947	0.03102776	0.05999484	0.1809222

```
bayes_R2(model_alt)
```

	Estimate	Est.Error	Q2.5	Q97.5
R2	0.1471981	0.03022991	0.08722664	0.2049597

Discussione dei risultati

Obiettivo dell'analisi

L'obiettivo di questa analisi era verificare se le **dimensioni di personalità patologica** misurate tramite il **PID-5** (in particolare, le cinque dimensioni di dominio) siano in grado di predire i punteggi individuali al composito **ESI_BF**, che rappresenta un indice di funzionamento adattivo o disfunzionale clinicamente rilevante. Poiché `esi_bf` è una misura stabile a livello individuale, sono stati esclusi dal modello tutti i predittori momentanei, concentrandosi esclusivamente su tratti stabili e su eventuali interazioni tra **dimensioni stabili** e **variabili EMA aggregate** a livello soggetto.

Modello base

Nel **modello base**, sono state considerate come predittori solo le **cinque dimensioni del PID-5**:

- Negative Affect
- Detachment
- Antagonism
- Disinhibition
- Psychoticism

I risultati indicano che:

- **Negative Affect** ($\beta = -0.13$, $CI_{95\%} = [-0.23, -0.02]$) e **Detachment** ($\beta = -0.10$, $CI_{95\%} = [-0.21, 0.00]$) sono **negativamente associati** al punteggio ESI_BF. Ciò suggerisce che soggetti con maggiore affettività negativa o tendenza all'isolamento riferiscono livelli inferiori di funzionamento adattivo.
- **Disinhibition** ($\beta = 0.25$, $CI_{95\%} = [0.13, 0.37]$) e **Psychoticism** ($\beta = 0.17$, $CI_{95\%} = [0.04, 0.31]$) mostrano invece una **relazione positiva** con ESI_BF. Questo risultato può apparire controintuitivo, ma potrebbe indicare che in alcuni contesti la disinibizione e tratti eccentrici siano collegati a strategie più attive di risposta o a bias di autovalutazione più positivi.

Il coefficiente di determinazione bayesiano (**Bayes R^2**) è pari a **0.12**, indicando che il modello spiega circa il **12% della varianza** interindividuale del punteggio ESI_BF.

Modello alternativo

Il **modello alternativo** ha aggiunto, per ciascun dominio del PID-5, un'interazione con la **controparte EMA aggregata** (ad esempio: `domain_antagonism × pid5_antag_m`), allo scopo di esplorare se l'effetto dei tratti stabili cambia in funzione delle fluttuazioni medie momentanee rilevate tramite EMA.

L'unico effetto chiaramente interpretabile nel modello alternativo è l'interazione:

- **Domain Antagonism × EMA Antagonism** ($\beta = 0.17$, $CI_{95\%} = [0.03, 0.32]$)

Questo risultato suggerisce che per i soggetti con livelli elevati sia di antagonismo stabile sia di antagonismo momentaneo, il punteggio ESI_BF tende ad aumentare. Potrebbe trattarsi di un sottogruppo di individui che, pur presentando tratti interpersonali problematici, mantengono una percezione soggettiva di sé come funzionanti (o resilienti), oppure di un artefatto dovuto a risposte difensive o sovracompensatorie.

Gli altri effetti principali e interazioni **non mostrano associazioni robuste**, con intervalli di credibilità che comprendono lo zero.

Il Bayes R^2 del modello alternativo è pari a **0.15**, un incremento modesto rispetto al modello base.

Confronto tra i modelli

Per confrontare i due modelli è stato utilizzato il criterio **LOO** (Leave-One-Out cross-validation), che stima la capacità predittiva fuori campione. I risultati sono:

Modello	elpd_diff	se_diff
Modello base	0.0	0.0

Modello	elpd_diff	se_diff
Modello alternativo	-7.7	4.6

L'aggiunta delle interazioni con le medie EMA non ha migliorato le prestazioni predittive: anzi, il modello alternativo presenta un ELPD inferiore, con un'incertezza che non permette di concludere con sicurezza che sia peggiore, ma suggerisce **nessun vantaggio sostanziale** rispetto al modello più parsimonioso.

Conclusioni

In sintesi:

- Le **dimensioni stabili del PID-5** spiegano una quota moderata della varianza nei punteggi ESI_BF.
- L'inclusione delle **interazioni con le misure EMA aggregate** non fornisce un guadagno informativo significativo e comporta un aumento della complessità del modello.
- Il **modello base risulta preferibile**: è più semplice, offre interpretazioni più robuste e ha prestazioni predittive comparabili (o superiori) rispetto al modello esteso.

Questi risultati indicano che, per l'esito considerato, le **caratteristiche stabili della personalità** sono i predittori più rilevanti, e le fluttuazioni momentanee aggregate non sembrano aggiungere informazioni predittive rilevanti **a livello interindividuale**.