Goalies

Avery Bennett

2026-01-20

Introduction

In labour economics, one of the core questions asked is how increased labor input affects worker productivity. When workers are pushed too hard, performance can diminish and as a result, output falls below its potential. Economists have long studied this trade-off, and while standard models predict diminishing marginal returns due to fatigue, empirical identification is often challenging due to unobserved heterogeneity and and how tasks are assigned often depends on the worker's prior outcomes, creating potential endogeneity. Professional sports, however, offer a unique approach to studying this problem, with detailed data and distinct measures of output. This paper will study the causal effect of workload on productivity using professional hockey goaltenders, where labor (input) is measured by the number of starts and output is observed through save percentage. We will examine how additional starts affect goaltender performance while controlling for individual and team characteristics, providing evidence on how increased labor generates diminishing returns in a high-stakes, performance-based labor market.

What is the causal effect of the number of starts on goalie performance in terms of save percentage?

Controls

- Shots against
- 1 rest day
- 2 rest days
- 3 rest days
- IG: Incomplete Games When Starter
- Games played
- Playoff (looking to add binary)

Fixed Effects

- Goalie FE (height, weight, age, draft position, goalie coach, experience)
- Team FE

IV

- back-to-back games
- goalie partner injury (looking to add)

Datasets

```
library(readxl)
library(dplyr)
summary <- read_excel("Summary.xlsx", sheet= 1)
summary <- summary %>% select(Player, `Game Date`, GS, Team, SA, `Sv%`)
days_rest <- read_excel("Days Rest.xlsx", sheet= 1)
days_rest <- days_rest %>% select(Player, `Game Date`, `0 Days Rest`, `1 Days Rest`, `2 Days Rest`, `3 D
                                  `4+ Days Rest`)

goalie <- summary %>% left_join(days_rest, c("Player", "Game Date"))
goalie <- goalie %>% rename(b_to_b= `0 Days Rest`, one_day_rest= `1 Days Rest`, two_days_rest= `2 Days R
```

plm Model

```
library(plm)
library(stargazer)
pdat <- pdata.frame(goalie, index=c("player", "game_date"))
pdat <- pdat %>%
  mutate(
    starts= as.numeric(starts),
    save_percentage= as.numeric(save_percentage),
    one_day_rest= as.numeric(one_day_rest),
    two_days_rest= as.numeric(two_days_rest),
    three_days_rest= as.numeric(three_days_rest),
    b_to_b= as.numeric(b_to_b),
    shots_against= as.numeric(shots_against))

fit <- plm(save_percentage~starts+I(starts^2)+one_day_rest+two_days_rest+
             three_days_rest+shots_against, data=pdat, model="within")

stargazer(fit, header=FALSE, float=FALSE)
```

	Dependent variable:
	save_percentage
starts	−0.072***
	(0.007)
one_day_rest	−0.003
	(0.004)
two_days_rest	−0.002
	(0.004)
three_days_rest	0.0002
	(0.004)
shots_against	0.004***
	(0.0002)
Observations	2,762
R^2	0.143
Adjusted R^2	0.108
F Statistic	88.343*** (df = 5; 2654)
Note:	*p<0.1; **p<0.05; ***p<0.01

IV Model

```
library(momentfit)
```

```
## Loading required package: sandwich
```

```
library(metricsUW)
mod <- momentModel(save_percentage~starts+one_day_rest+
                     two_days_rest+three_days_rest+shots_against,
                   ~b_to_b+one_day_rest+two_days_rest+three_days_rest+
                     shots_against,
                   data = pdat, vcov= "CL",
                   vcovOptions = list(cluster= ~player))
fit2 <- tsls(mod)
printReg(fit2, stars=TRUE, strength=TRUE)
```

```
## \begin{equation*}
## \begin{split}
## \widehat{save_percentage} &=\underset{( 0.0464 ) ^{***} }{ 0.8285 }~-~\underset{( 0.0753 )   }{ 0.030
## \end{split}
## \end{equation*}
```

```
summary(fit2)
```

```
## Model based on moment conditions
## *********************************
## Moment type: linear
```

```
## Covariance matrix: CL
## Clustered based on: player
## Number of regressors: 6
## Number of moment conditions: 6
## Number of Endogenous Variables: 1
## Sample size:  2762
##
## Estimation:  One-Step, Just-Identified
## Sandwich vcov: TRUE
## coefficients:
##                    Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)      0.82854614  0.04644958 17.8375 < 2.2e-16 ***
## starts          -0.03646392  0.07529886 -0.4843 0.6282043
## one_day_rest     0.00174502  0.00434566  0.4016 0.6880121
## two_days_rest    0.00288383  0.00706548  0.4082 0.6831576
## three_days_rest  0.00538502  0.00407386  1.3218 0.1862192
## shots_against    0.00363638  0.00094539  3.8464 0.0001199 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  J-Test
##                 Statistics  df  pvalue
## Test E(g)=0:    2.1067e-23   0      NA
##
##
## Instrument strength based on the F-Statistics of the first stage OLS
## starts : F( 1 ,  2756 ) =  6.030377  (P-Vavue =  0.01412306 )
```