

Faster means earlier? A panel data analysis of running speed on Marathon finish times

Reg Monkey¹, Reg Donkey¹

¹ Department of Regression, Crazy Thursday University

Despite the widespread popularity of marathon running, the quantitative relationship between pacing strategy and final performance remains under-explored in empirical economic literature. While physiological studies suggest a link, potential endogeneity issues—such as unobserved runner ability and environmental conditions—pose challenges to causal identification. This study utilizes a rich panel dataset of 5,000 participants from major urban marathons to estimate the impact of average running speed on finish times. Employing a runner fixed effects model and controlling for a comprehensive set of covariates including weather conditions and gear quality, we find a robust negative relationship between speed and finish time. Specifically, a 1 km/h increase in average speed corresponds to a significant reduction in total duration. These findings provide empirical support for the efficiency hypothesis in endurance sports and offer actionable insights for athletic training optimization.

Running Speed | Finish Time | Tautology | Fixed Effects | Marathon

Introduction

Marathon running has evolved from a niche athletic pursuit into a mass participation phenomenon with significant economic implications. Understanding the determinants of performance is crucial not only for athletes but also for event organizers and the broader sports industry. While traditional coaching relies on physiological metrics, a rigorous econometric

Significance

While the broader scientific community has long operated under the assumption of a physical link between velocity and temporal duration, this paper provides the first econometric validation of this phenomenon using observational field data. By successfully isolating the causal effect of running speed from potential confounders such as shoe technology and meteorological conditions, we bridge the gap between theoretical physics and empirical social science.

Our findings offer critical policy implications for athletic training: they suggest that resources currently allocated to peripheral factors (e.g., aerodynamic socks) may be better redirected toward the development of core velocity maintenance. Furthermore, this study establishes a methodological precedent for the application of fixed effects models to deterministic systems, demonstrating that even the laws of nature can benefit from robust standard errors.

analysis of performance determinants offers a more structured approach to strategy optimization.

Existing literature has extensively documented the physiological limits of human endurance. However, the statistical dynamics of how running speed translates directly into finish times have received surprisingly little attention from a causal inference perspective. The primary challenge lies in isolating the effect of speed from confounding factors. For instance, runners who maintain a higher speed might possess superior genetic endowments or utilize advanced equipment, leading to omitted variable bias.

This paper aims to bridge this gap by applying panel data techniques to marathon performance data. We investigate the following research question: **Does increasing average running speed causally reduce the finish time, holding other factors constant?** By addressing potential endogeneity and employing robust standard errors, we provide the first large-scale econometric validation of this fundamental relationship.

Data and methods

Data Source

The dataset comprises detailed records from five major city marathons (e.g., New York, London, Tokyo) spanning 2018-2022. The final sample includes 5,000 runners after excluding incomplete records and outliers (e.g., wheelchair division participants).

Variables

The dependent variable is Finish Time ($Time_{it}$), measured in minutes. The independent variable is Average Running Speed ($Speed_{it}$), calculated as total distance (42.195 km) divided by finish time. To account for exogenous shocks, we include three sets of control variables: (1) Temperature ($^{\circ}C$), humidity (%), and wind speed (km/h); (2) Shoe price index (adjusted for inflation), shirt fabric type (cotton vs. synthetic); (3) Age, gender, and prior marathon experience. Empirical Strategy. We estimate the following specification using Ordinary Least Squares (OLS) with runner fixed effects:

$$Time_{it} = \alpha + \beta \cdot Speed_{it} + \gamma \cdot X_{it} + \mu_i + \epsilon_{it}$$

Where μ_i captures time-invariant individual heterogeneity (e.g., innate aerobic capacity), and X_{it} represents the vector of time-varying controls. The parameter of interest, β , is expected to be negative, indicating that higher speed reduces time.

Results

Main results

As reported in Table 1, the coefficient for Average Speed remains remarkably stable across all model specifications. In Model 3, a 1 km/h increase in running speed is associated with a 5.10-minute decrease in finish time ($p < 0.01$). The high R-squared (0.99) suggests that the model explains almost all the variance in the dependent variable, validating the specification.

Robustness checks

To ensure the validity of our findings, we conducted a battery of robustness tests. The results are summarized in Table 2.

Table 1 | Fixed Effects Estimates of Running Speed on Finish Time

Dependent Variable: Finish Time (mins)	Model 1 (Baseline)	Model 2 (Weather Controls)	Model 3 (Full Specification)
Average Speed (km/h)	-5.12* (0.02)	-5.11* (0.02)	-5.10* (0.03)
Temperature		0.15	0.14
Shoe Price Index			-0.02
Constant	215.4***	218.6***	220.1***
R-squared	0.99	0.99	0.99
N	5,000	5,000	5,000

Note: Robust standard errors in parentheses. *** p<0.01.

Alternative Functional Forms. First, we tested for non-linearities by including a quadratic term for speed ($Speed^2$). The quadratic term was statistically insignificant, suggesting that the relationship between speed and time is strictly linear within the observed range of human physiology. This confirms that the standard linear specification is appropriate for modeling the physical laws of motion.

Sub-sample Analysis. We examined whether the effect holds across different sub-populations. (1) We split the sample into sub-3-hour runners and sub-5-hour runners. The negative coefficient remained significant in both groups, suggesting that the laws of physics apply equally to both professionals and novices. (2) We then isolated runners wearing “Brand A” shoes versus “Brand B” shoes. The effect of speed on time remained consistent, indicating that marketing claims do not override the fundamental kinematic relationship.

Placebo Test. We constructed a “placebo” independent variable by randomly shuffling the speed values among runners. As expected, the relationship between this randomized “fake speed” and finish time yielded a coefficient statistically indistinguishable from zero ($\beta = 0.01, p > 0.5$). This confirms that the original finding was not driven by spurious correlation or statistical artifacts.

Excluding Outliers. To address concerns that extreme cases (e.g., runners who crawled or were carried by stray dogs) might drive the results, we trimmed the top and bottom 1% of the finish time distribution. The coefficient for speed remained virtually unchanged ($\beta = -5.09$).

Table 2 | Robustness Checks Summary

Specification	Coefficient on Speed	Standard Error	Observations
Baseline (Model 3)	-5.10*	(0.03)	5,000
With Quadratic Term	-5.08*	(0.04)	5,000
Elite Runners Only	-5.15*	(0.05)	500
Placebo (Randomized Speed)	0.01	(0.05)	5,000
Trimmed Sample (1%)	-5.09*	(0.03)	4,900

Note: *** p<0.01.

Discussion

The empirical evidence strongly supports the hypothesis that running speed exerts a negative causal effect on marathon finish time. This finding aligns with the fundamental principles of kinematics but provides necessary nuance by quantifying the effect within a multivariate framework. The stability of the coefficient after controlling for shoe price is particularly noteworthy; it suggests that while equipment may play a role, the kinetic input (speed) is the primary driver of temporal output.

Our findings have direct implications for training protocols. Coaches and athletes often focus on complex interval training or nutrition strategies. However, our econometric analysis highlights that the most statistically significant lever for reducing finish time is the maintenance of a higher average velocity. Consequently, training resources should be reallocated to focus on velocity enhancement exercises.

This study is limited by its focus on distance-based speed. Future research could explore the impact of “perceived speed” versus “actual speed” or investigate the non-linear effects of acceleration at specific race intervals. Furthermore, the potential reverse causality—whereby a desire to finish early drives speed—warrants investigation using instrumental variables, though current data supports a unidirectional flow.

References

- [1] Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- [2] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- [3] Greene, W. H. (2012). *Econometric Analysis*. Pearson Education.
- [4] Jones, A. M., & Carter, H. (2000). The effect of endurance training on parameters of aerobic fitness. *Sports Medicine*, 29(6), 373-386.
- [5] Berndt, E. R. (1991). *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley.
- [6] Smith, R. L., & Womack, J. P. (2015). The impact of sock color on marathon performance: A randomized controlled trial. *Journal of Trivial Sports Science*, 12(4), 112-128.
- [7] Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*. Royal Society.
- [8] Runner, F., & Sprinter, G. (2018). Does wanting to finish early make you run faster? Evidence from sleeping habits. *Journal of Behavioral Athletics*, 5(2), 45-60.
- [9] Foster, C., & Lucia, A. (2007). Running economy: the forgotten factor in elite performance. *Sports Medicine*, 37(4-5), 316-319.
- [10] Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Cengage Learning.
- [11] Smart, A. L., & Dumb, B. (2022). Can money buy speed? An instrumental variable approach using lottery winners. *Review of Athletic Economics*, 8(3), 220-235.
- [12] Cheuvront, S. N., & Kenefick, R. W. (2014). Am I drinking enough? Yes, no, and maybe. *Journal of the American College of Nutrition*, 33(4), 332-338.