Hunting in Pairs:
Testing for Coworker Effects in Professional Cricket*

Johan Fourie†

Abstract

I test cricket's belief that bowlers *hunt in pairs*. Using 1.36 million overs, I compare the same bowler alongside different partners within a match. Timing adjustments reduce coworker associations, but positive wicket-quality associations remain in T20. Better partner containment instead predicts fewer focal wickets in T20, conditional on wicket quality. Chronological validation and fixed-player history experiments distinguish improved own-output forecasts from coworker inference. Undirected and directional pair histories add little to long-form focal-output forecasts. The findings challenge a general positive partnership account and clarify what performance records establish about coworkers. The estimates remain conditional associations.

Keywords: coworker effects; assignment; team production; empirical Bayes; measurement; cricket
JEL codes: C18; D24; J24; M54; L83

*I thank the Cricsheet project and its contributors for the match records, and Krige Siebrits for helpful discussions. This paper was created with the help of Anthropic's Claude Code (Fable 5.1) and OpenAI's Codex. Cite this paper as: Fourie, Johan. 2026. "Hunting in Pairs: Testing for Coworker Effects in Professional Cricket." Working Paper, Department of Economics, Stellenbosch University.

†Department of Economics, Stellenbosch University. Email: johanf@sun.ac.za.

1 Introduction

A manager needs to know more than which workers produce the most. A colleague may help others produce, change the task they face or take an opportunity they might otherwise have completed. A successful pairing does not reveal which contribution occurred. Managers choose when people work together, and researchers rank those people using performance records of unequal length. How much does the measured coworker relationship depend on the task sequence and on the information used to describe the coworker?

Cricket gives this problem a familiar expression: bowlers are said to *hunt in pairs*. The claim is that one bowler helps the other produce. Stuart Broad describes restricting scoring at one end while his partner takes wickets (Yew, 2016). The pressure account predicts that such restriction makes further dismissals easier. I examine the general coworker claim using 1,355,340 eligible overs across five professional settings. Bowlers alternate ends, and ball-by-ball records identify the immediate partner, task sequence and credited output. Comparing the same bowler alongside different partners within a match links deployment to the information used to rank the coworker. Captain discretion leaves these associations open to selection.

When a pair works together changes the comparison. With worker-by-match and innings fixed effects alone, stronger measured partners predict fewer focal wickets in long-form cricket and more in shorter formats. Phase and workload controls reduce these associations. A weighted decomposition attributes most of the change to phase; workload partly offsets it in four settings. Timing removes little residual quality dispersion. The remaining slopes are close to zero in Tests, domestic cricket and one-day internationals, while positive associations remain in both twenty-over settings, consistent with the belief's wicket-quality prediction.

What the partner does matters too. Bowling has distinct performance dimensions: taking wickets and restricting runs. Earlier containment predicts the bowler's own later run rate conditional on wicket quality. Yet better partner containment accompanies fewer focal wickets in both T20 settings. Focal scoring also falls in franchise cricket, while the international run-rate estimate is imprecise. These findings challenge the simple pressure account in T20; the long-form comparisons leave its direction unresolved. Runs prevented can benefit the team without increasing the next bowler's dismissal credit.

More informative records improve forecasts of the bowler's own later wicket rate, but yield no precise improvement in fitted coworker contrasts or partial fit. Randomly shortening experienced players' histories while keeping their current matches fixed provides a separate information comparison, with only 33 Test partners. Pair records add little to the specified later focal-over forecasts, including when histories distinguish who bowls to whom. Better individual forecasts need not reveal more about another worker's output.

The paper contributes to four literatures: peer effects in team production, managerial allocation, measurement of worker quality and sports economics. In the first, Gould and Winter (2009) and Papps and Bryson (2019) show why productive roles and sequential opportunities matter in baseball; Arcidiacono, Kinsler and Price (2017) distinguish own productivity from helping teammates

produce in basketball. I compare two attributes of the same immediate coworker while observing the preceding task. Containment tests a production implication that a single wicket-quality ranking cannot assess.

The allocation literature shows why exposure needs explanation. Mas and Moretti (2009) establish the importance of visibility for supermarket peer effects, Chan (2016, 2018) connect hospital work to assignment systems and shift timing, and Minni (2026) links managers to reallocation and productivity. Cricket permits timing controls to enter while workers, matches and observations stay fixed. The resulting slope changes reveal the measurement consequences of deployment, complementing estimates of what managers do to productivity.

The third contribution concerns estimated worker quality. Empirical Bayes research distinguishes prediction from regression on a predicted attribute, especially when precision is informative about the worker (Chen, 2026; Chen, Gu and Kwon, 2025). I evaluate both uses of the same record and vary its length without changing current workers or outcomes. This also complements the separation of production and learning in Herkenhoff et al. (2024), whose model accommodates sorting and noisy wages. Changing the researcher's information supplies a comparison that career-stage differences cannot provide alone; it estimates neither skill accumulation nor a correction for measurement error.

The fourth contribution is to sports economics. Detailed sporting records permit scrutiny of economic hypotheses (Kahn, 2000; Palacios-Huerta, 2025); here they also test a belief within the industry itself. Nanavati and Nanavati (2024) identify bowling-partnership networks and propose selection applications. I examine the timing, performance dimensions and subsequent output relevant to such decisions. The findings challenge the presumption that successful bowling pairs establish a general positive coworker effect. Testing the mechanism behind a familiar partnership matters before similar reasoning informs assignments elsewhere.

Sections 2–4 describe production, records and methods. Results precede a discussion of experience, relationships and wider-team inputs.

2 Institutional setting

2.1 Bowling partnerships and team production

An over normally consists of six legal deliveries by one bowler. The next over comes from the other end, bowled by somebody else. For the focal bowler in over o, I define the immediate partner as the bowler in over $o-1$. A captain can preserve a pair or change either member between overs. The Hundred uses five-ball units and ten-ball end blocks; its partner is the bowler who completed the preceding ten-ball end. This rule respects the actual change of end instead of mechanically treating each five-ball unit as an alternating over.

A partner can change the task facing the focal bowler. In the pressure account of hunting in pairs, restricting scoring at one end induces the batter to take risks at the other, creating wicket opportunities. This is one testable prediction, rather than a complete account of partnership

value. A wicket can expose the next bowler to a new batter; different styles can make adaptation difficult; familiar partners can coordinate plans. Production technology matters for the direction of interaction (Gould and Winter, 2009). A wicket-taking partner can also remove a vulnerable batter whom the focal bowler might otherwise dismiss. Wickets and runs may therefore move differently, and credited wickets do not measure the team's total gain.

The bowling pair is embedded in an eleven-player fielding team. Fielders help create dismissals and prevent runs, the wicketkeeper participates in many outcomes and the captain directs the attack. Among innings with eligible observations, the mean number of deployed bowlers ranges from 5.30 in Tests to 6.07 in one-day internationals. These counts describe realised deployment, not the roster of every feasible replacement. The pair definition isolates a precisely timed exposure within that team. It does not assume that two bowlers constitute the entire production function.

Production unfolds sequentially. Each delivery changes the score, wickets remaining, fatigue, ball condition and the batters' objectives. The outcome measured here is production in the focal over, a short interval within that evolving process. The distinction matters for estimation: the current state helps the captain choose a bowler, but can also reflect what an earlier partner has done.

2.2 Coworker assignment within rosters

The captain allocates a fixed roster subject to availability, specialisation, fatigue and format rules. A bowler cannot normally deliver consecutive overs; one-day and twenty-over games also impose individual limits. These constraints do not randomise the realised pairing. They leave discretion over when to introduce a bowler, how long to retain him and which batter or match state he faces. Cricket therefore supplies repeated observations of deployment, rather than a comparison between freely chosen workforces of different average quality.

The closest workplace application is reassignment among available workers doing recurring tasks. Supermarket visibility (Mas and Moretti, 2009) shows why exposure must identify actual interaction. Hospital assignment systems (Chan, 2016) and shift timing (Chan, 2018) connect that exposure to when work occurs. Cricket observes immediate partner quality and task timing within one sequence. Checkout visibility, clinical quality and cricket's response to an opponent differ; these comparisons motivate an empirical question rather than transport a numerical effect.

Workforce composition is a broader decision. Holding average ability constant, Hamilton, Nickerson and Owan (2003) find more heterogeneous garment teams to be more productive, consistent with collaboration, learning or bargaining. Firms may accept lower current output to build future skills (Herkenhoff et al., 2024). Comparing overs by a player already selected for a match instead holds the roster fixed. Career concerns may still influence effort and deployment, but contracts, outside options and expectations are unobserved.

A scoreboard makes credited output visible while leaving important inputs unobserved. Tactical instructions, physical condition, effort and fielding contributions can affect both assignment and the outcome. The captain may have private information about them and may himself be uncertain.

Sorting can consequently run in either direction. The setting's advantage is that the timing and measurement comparisons can be made explicitly; output visibility alone does not bound assignment bias in less transparent workplaces.

3 Data and variable construction

3.1 Sample construction

The analysis uses structured ball-by-ball records from Cricsheet. The five settings are men's Test cricket, domestic multi-day cricket, one-day internationals (ODIs), twenty-over internationals (T20Is) and franchise T20. Domestic coverage includes the County Championship, Sheffield Shield, Plunket Shield, Bob Willis Trophy and Sri Lanka's Major League Tournament. The franchise category contains the Hundred, handled with its end-block rule. Coverage dates and archive depth differ across settings and competitions. Table 1 reports the resulting samples.

Table 1: Sample characteristics and partner experience

Setting	Matches	Eligible overs	Regression overs	Median prior balls
Test	885	279,535	279,155	2,677
Domestic multi-day	2,162	603,854	602,898	2,784
ODI	2,547	215,093	214,042	998
T20I	3,442	117,946	112,490	269
franchise	3,900	138,912	133,447	828

Notes: The unit is a focal over, or a five-ball unit in the Hundred. Matches and eligible overs precede fixed-effect singleton removal. Regression overs are the timing-specification sample. Prior balls are the partner's recorded legal balls before the focal match date, with T20 histories pooled across international and franchise matches. Counts describe archive coverage rather than complete careers.

An over is eligible when it has legal deliveries, has an identifiable preceding partner unit in the same innings, and has different focal and partner bowlers. I exclude super overs, mixed-bowler units and units whose partner would be defined from a mixed-bowler unit. First overs cannot enter. The eligible corpus contains 1,355,340 observations; removing fixed-effect singletons leaves 1,342,032 in the timing regressions. Appendix A gives coverage dates and construction details.

The primary outcome is bowler-credited wickets divided by legal balls. Credited dismissals on illegal deliveries, such as a stumping on a wide, enter the numerator. Retired hurt is not a dismissal. Focal runs per legal ball provide a secondary outcome, counting total runs including byes, leg-byes and penalties. Wickets and runs capture different aspects of performance and neither is a complete measure of a bowler's contribution.

3.2 Measures of bowler quality and experience

Let W_{jt} and B_{jt} be bowler j's recorded wickets and legal balls in matches dated strictly before focal match t. Test, ODI and domestic histories are separate; T20I and franchise histories pool the two T20 settings. Dates refer to match starts; the history assumes earlier recorded matches finish before the player's next appearance. The same-date exclusion prevents information from the focal match entering a player's history. It also avoids using match ordering that the archive cannot reliably establish within a date.

For prior mean m_s and strength k, measured wicket quality is

$$q_{jt}(m_s, k) = \frac{W_{jt} + km_s}{B_{jt} + k}. \tag{1}$$

The formula pulls a bowler's recorded wicket rate towards a common rate. At $k = 300$, a 300-ball record receives the same weight as the prior; longer records receive more weight. Pooling information this way is a familiar use of empirical Bayes in labour economics (Walters, 2024). Although the formula has a posterior-mean interpretation under a common-mean conjugate model, here it is a constructed forecast whose calibration can be tested. Constant ability and exchangeable deliveries are not imposed as facts. The descriptive full-period regressions use $k = 300$ and the setting's wicket rate among eligible overs for m_s. This mean is a retrospective calibration constant using the full period, while individual histories exclude the focal date. The forecasting exercises fit both the prior mean and strength on earlier data.

I retain quality in wickets per legal ball throughout the main results. A raw slope of 0.10 means that a 0.01 increase in the quality index is associated with 0.001 more focal wickets per ball. This unit is common across settings and history subsamples. Within-setting standardisation is useful for describing scale, but changing a standard deviation after restricting the sample would confound a slope comparison with its units.

Recorded experience is especially uneven. Median partner histories are 2,677 balls in Tests, 2,784 in domestic cricket and 269 in T20Is. These differences combine career exposure, format schedules and incomplete coverage. I use common absolute history bins of 0–299, 300–999, 1,000–2,999 and at least 3,000 balls. A bowler with a short observed record need not be a new professional. That is one reason for separately thinning histories while holding the workers themselves fixed.

4 Empirical strategy

4.1 Coworker output and assignment

A coworker can help produce today and help build skills for tomorrow. These are separate margins in Herkenhoff et al. (2024); a manager's assignment decision determines the exposure through which either can operate. A compact framework keeps the three objects distinct. Let a_{jt} denote productive skill, e_{jt} effort, S_{to} the production state and R_t the available roster. Current output,

assignment and subsequent skill evolve as follows:

$$y_{ito} = F_s(a_{it}, e_{ito}, a_{jt}, e_{jto}, S_{to}, R_t, H_{ij,t}) + u_{ito}, \tag{2}$$

$$j = \pi_s(R_t, S_{to}, Z_{to}), \tag{3}$$

$$a_{i,t+1} = a_{it} + L_s(H_{ij,t}, a_{jt}, \text{training}_{it}) + \nu_{it}. \tag{4}$$

The first equation allows the partner, the wider team and their working relationship $H_{ij,t}$ to affect current output. The second says that the captain chooses the partner using the roster, state and private signals Z_{to}. The third allows that relationship to change future skill. The analysis measures an association in current production, conditional on deployment; it does not estimate the learning function. The quality index q describes past outcomes and supplies an imperfect signal of productive skill.

For focal bowler i, match m, innings n and over o, the timing regression is

$$y_{imno} = \beta_s q_{jt} + g_s(o) + \rho_s h_{imo} + \alpha_{im} + \delta_{mn} + \varepsilon_{imno}, \tag{5}$$

where $g_s(o)$ contains five-over phase indicators and h_{imo} is the focal bowler's own over number in the innings. Focal-bowler-by-match effects α_{im} absorb fixed player-match attributes; match-innings effects δ_{mn} absorb shared innings conditions. Regressions weight observations by legal balls and cluster standard errors by focal bowler and match.

The regression compares a bowler with himself within a match as measured partner quality changes, accounting for the included timing variables. Interpreting its slope causally would also require the remaining variation in partner quality to be unrelated to unobserved determinants of the focal outcome. Private signals, batter matchups and tactical responses make this a demanding restriction. The selection problem remains even when the peer measure precedes the outcome, within the identification concerns discussed by Angrist (2014). Timing controls show how the observed comparison changes; they do not establish that restriction.

I compare fixed effects alone, timing controls and an additional state adjustment. State adjustment includes wickets in hand and score rate at the start of the over. Although observed before the current outcome, these variables may transmit effects of earlier partners. Workload and duration can also respond to earlier performance. The nested estimates consequently describe different conditional comparisons; adding controls is not an automatic progression towards a total causal effect.

To locate the timing change, I use a weighted coefficient decomposition (Gelbach, 2016). Each added control is projected on quality and the baseline regressors; its quality coefficient times its coefficient in the full outcome model contributes to the baseline-minus-full slope difference. Summing separately over phase indicators and workload gives an exact, order-independent accounting on common fitted observations. Residual quality dispersion and concentration describe the variation supporting that comparison.

4.2 Sensitivity to the prior mean

For fixed k, define $s_{jt} = W_{jt}/(B_{jt} + k)$ and $r_{jt} = k/(B_{jt} + k)$. Then

$$q_{jt}(m_s, k) = s_{jt} + m_s r_{jt}. \tag{6}$$

The prior loading r is the weight placed on the common rate, and is largest for short histories. It is observed whenever history length is known. Adding it to equation (5) gives

$$y_{imno} = \beta_s q_{jt} + \gamma_s r_{jt} + g_s(o) + \rho_s h_{imo} + \alpha_{im} + \delta_{mn} + \varepsilon_{imno}. \tag{7}$$

For any setting-constant change $m_s \mapsto m_s + c$, replacing γ_s by $\gamma_s - \beta_s c$ leaves fitted values unchanged. The coefficient β_s is invariant, provided the sample, k, quality units and other regressors are unchanged and the relevant columns have rank. This is a linear reparameterisation, not a new estimator. Appendix B states the corresponding partial-regression expression.

The check isolates one source of sensitivity. It does not remove error in W, time variation in ability, dependence between skill and history length, or endogenous assignment. It also changes the conditioning set: the regression compares quality conditional on its prior loading. If prior means vary across history categories, their matching category-specific loadings must enter. If a quality interaction is estimated, the loading needs the corresponding interaction. Prior strength k changes both columns and is not covered by the invariance.

A posterior mean can nevertheless be the right regressor. For a linear outcome model $y = \beta a + X'\eta + \epsilon$, if $q = E[a \mid D, X]$ and $E[\epsilon \mid D, X] = 0$, then $E[y \mid D, X] = \beta q + X'\eta$. In words, a skill forecast conditional on the relevant information and controls can recover the conditional output relationship when the remaining error is unrelated to that information. A forecast based only on an incomplete record need not meet these conditions. Chen, Gu and Kwon (2025) explain why regression on shrinkage requires more than a generic measurement-error argument, while Chen (2026) studies heterogeneity related to estimation precision. The empirical task here is to examine which features of the quality measure matter in practice, keeping prior location distinct from prior strength and record length.

At fixed k, estimating a setting-constant prior location adds no first-stage uncertainty to the invariant slope: it is algebraically independent of that location. This result does not correct uncertainty about latent skill. Partner/match and pair/match clustering, log prior balls and all four history-bin restrictions provide sensitivity checks.

4.3 Forecast validation and history length

For each setting, I partition distinct match dates chronologically: the first 60 per cent form training data, the next 20 per cent validation data and the remaining dates test data. Training data determine either one setting mean or means for the four absolute history bins. Candidate strengths are 100, 300, 1,000 and 3,000 balls. The candidate with the lowest legal-ball-weighted squared

prediction error for the bowler's own match wicket rate in validation data is selected. Its later test performance is compared with an always-reported anchor using a global training mean and $k = 300$.

Histories update from strictly earlier dates; hyperparameters stay fixed. Forecasts condition on observed appearance. Paired bowler/match loss intervals condition on trained rules; own-output calibration reports intercepts and slopes. Each rule then measures partners on identical later focal overs, with timing, fixed effects and matching history-bin loadings. Coworker outcomes do not select the rule.

Raw coefficients can change when a rule compresses quality. I therefore cross both rules with fixed effects alone, phase, workload and both timing controls on common fitted observations. Partial fit measures the quality column's contribution conditional on its loadings, and the complete partner block's contribution beyond task controls and fixed effects. For each observed partner change between successive eligible focal appearances within an innings, the fitted contrast applies the full quality-and-loading block to the new-minus-old partner attributes while holding the task fixed. I report its root mean square (RMS) and signed mean in wickets per 100 balls. These contrasts survive a change of coefficient units. Paired cluster-score intervals compare mean contrasts and residual losses across rules; RMS and partial-fit summaries are descriptive. Appendix B defines the calculations.

The history experiment fixes current players and matches. Tests and domestic cricket supply 430 and 2,262 later owner-matches with at least 3,000 recorded balls in the preceding 1,095 days; shorter formats lack sufficient support. For each owner-match, random orderings of whole preceding matches generate nested budgets reaching 300, 1,000 and 3,000 balls. One hundred repetitions, with $k = 300$ and the training mean fixed, are compared with the complete three-year window.

Whole-match sampling preserves historical clustering and permits budget overshoots. Every budget draws from the same calendar window on fixed current observations, with closely balanced mean record ages. For each finite budget I average fitted contrasts and losses across the 100 draws, then compare them with the complete window using paired scores. Averaging the scores before estimating variance preserves dependence across draws of the same workers. Variation across historical draws remains a separate sensitivity measure. The experiment changes noise, shrinkage and historical composition together; it estimates neither learning nor an attenuation correction.

4.4 Partner containment and task conditions

Let R_{jt} be the partner's total runs conceded on the same earlier dates used to construct wicket quality. Define containment as $c_{jt} = -(R_{jt}+k\nu_s)/(B_{jt}+k)$, so a larger value means fewer predicted runs per legal ball. I add c_{jt} to the prior-loading model, retaining q_{jt}, timing and the two sets of fixed effects. Both indices use $k = 300$ and setting-constant prior means. Because their denominators coincide, the same loading r_{jt} absorbs changes in either prior location. This is checked numerically. Historical runs include byes and leg-byes, so containment differs from the usual bowler-charged economy rate.

I validate containment on the same chronological owner-match splits as wicket quality. Training run-rate models include wicket quality and history loadings, with and without containment. Validation selects among the same global/history means and four strengths; the global 300-ball anchor remains reported. Each incremental comparison retains identical loadings. Later loss gains and conditional calibration assess whether containment supplies information about its owner's run rate beyond wicket quality. The primary comparison uses the anchor wicket rule; the selected wicket rule is a sensitivity.

The primary containment comparisons concern T20I and franchise cricket; both focal wickets and total runs are reported for all settings. A stacked regression interacts every coefficient and fixed effect with setting and outcome, retaining covariance across shared bowlers. Holm adjustment applies across the two T20 settings separately for each outcome. Effects use a 0.1-runs-per-ball containment difference, or 0.6 runs per six-ball over. Residual dispersion and actual partner changes locate this scale in the data. The pre-estimation 0.1-focal-wickets-per-100-balls reference has minimum detectable changes of 0.152 and 0.104 in the two T20 settings, leaving smaller associations hard to exclude.

The supplementary comparison excludes the Hundred and uses six-ball T20 matches. On identical observations I add the previous partner over's runs and all dismissals, then the logarithm of one plus each starting batter's prior legal balls faced in that innings. Earlier-over records construct these exposures; no focal-over deliveries enter them. Deterministic checks against raw deliveries verify the joins. These controls can be mediators or colliders. Changes across specifications describe conditional patterns and cannot identify pressure, sorting or a mediated share.

4.5 Individual and pair forecasting models

A pair history can predict output even when a simple lagged coefficient is imprecise. I therefore compare three focal-over forecasts in Tests and domestic cricket. The first uses the focal worker's individual record and task context. The second adds the partner's wicket quality and containment. The third adds the pair's prior wicket excess over the second model's predictions. Context includes phase, workload, wickets remaining and score rate. No current-match effect is fitted with later outcomes. All three models predict the same observations at the over's start.

The initial pair stock uses five expanding-date training-prefix forecasts, avoiding residuals fitted with their own outcomes. Stocks then update from strictly earlier dates. Besides an undirected stock pooling both members, I construct directed excess using only the focal bowler's earlier output alongside that partner. Their two directional numerators and exposures sum to the undirected totals before shrinkage. One extension adds both stocks to the individual-plus-partner model, with training-fitted multipliers and a shared strength selected from 300, 1,000, 3,000 and infinity. Infinity omits both. The undirected-only benchmark retains its own selection. This tests whether pooling directions conceals useful information; it does not exhaust possible relationships. The chronological periods have already been examined for other questions.

Paired loss reductions cluster by worker and match, with pair and match as a sensitivity.

Alongside all later overs, I report established pairs and observations with at least 300 usable balls in the focal direction. The 0.5 per cent reporting benchmark and intervals concern these fitted rules. Appendix E specifies the construction.

5 Results

5.1 Assignment timing and prior sensitivity

When a captain pairs two bowlers matters to the relationship we observe between their performances (Figure 1). Comparing a bowler with himself within a match initially suggests that stronger partners accompany fewer focal wickets in long-form cricket and more in shorter formats. Accounting for phase and workload moves the long-form and ODI relationships close to zero and reduces the positive T20 associations. The workers and observations have not changed. For the hunting-in-pairs claim, this distinction matters: a successful pairing can reflect when the bowlers meet as well as what they contribute to each other.

The stage of the innings accounts for more of this change than the focal bowler's workload in every setting. Workload partly offsets the phase contribution in most settings. A pairing observed early in an innings need not face the same opportunities as one observed later. The decomposition cannot establish whether captains select partners in response to those opportunities or earlier bowling helps create them.

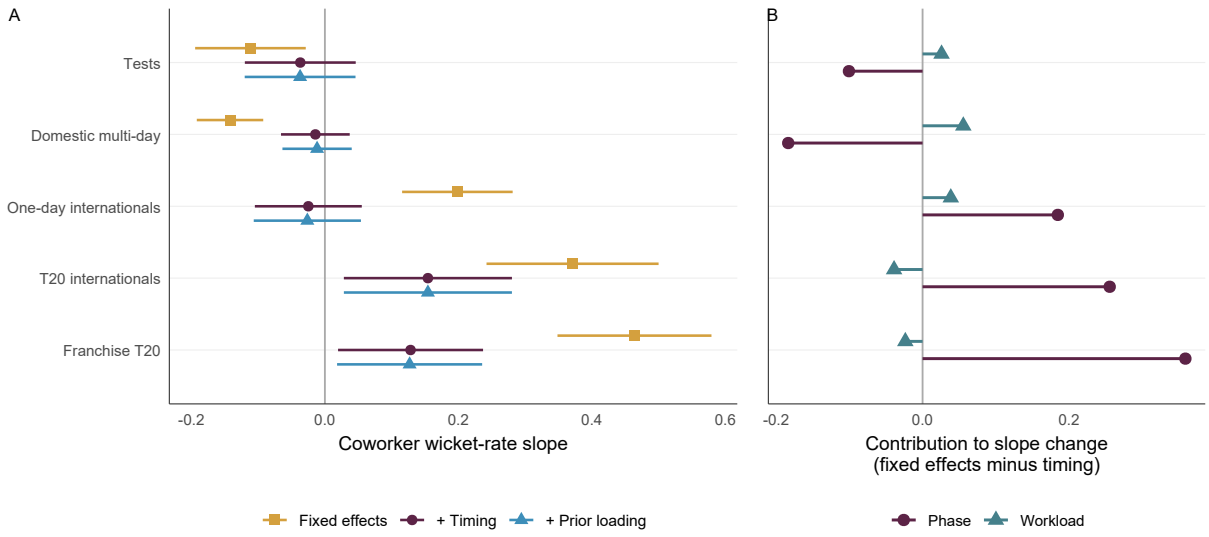

Figure 1: Coworker associations and assignment timing

Notes: A: raw slopes and 95 per cent intervals on common fitted overs; timing adds phase and workload to worker-match and innings effects. B: weighted phase and workload contributions to the fixed-effects-minus-timing change. Legal-ball weights; focal-bowler/match clusters.

This sensitivity connects exposure (Mas and Moretti, 2009), task timing (Chan, 2018) and

managerial allocation (Minni, 2026) to the measurement of coworker contributions. The same available workers meet at different production stages, and observing those stages changes their estimated relationship within a worker's match. The decomposition cannot distinguish confounding from sequential production channels, or identify the return to changing the captain's choices.

The timing controls leave almost all the remaining variation in partner quality available for comparison. The change in the estimated relationship therefore cannot be attributed simply to running out of differences between partners. Those differences are unevenly distributed, however: a tenth of partners supplies roughly half or more of the variation used by the regression (Table A1). A large number of observed pairings is consequently not the same as a large number of equally informative comparisons.

Accounting separately for the weight placed on the common prior barely changes this pattern (Table 2). Positive associations remain in both T20 settings, whereas the long-form estimates remain close to zero with intervals allowing either sign. Differences in production technology could make coworker interactions vary across formats, as Gould and Winter (2009) suggest. But captains also deploy bowlers differently, and their recorded histories differ. The estimates cannot distinguish these explanations merely because each setting has its own coefficient.

Table 2: Coworker estimates by control specification

Setting	Timing	Prior loading	State adjusted
Test	-0.037 [-0.120, 0.046]	-0.037 [-0.120, 0.046]	-0.029 [-0.112, 0.054]
Domestic multi-day	-0.014 [-0.066, 0.037]	-0.012 [-0.064, 0.040]	0.036 [-0.017, 0.088]
ODI	-0.025 [-0.105, 0.055]	-0.026 [-0.107, 0.054]	0.018 [-0.063, 0.099]
T20I	0.154 [0.028, 0.280]	0.154 [0.028, 0.280]	0.167 [0.043, 0.291]
franchise	0.128 [0.020, 0.237]	0.127 [0.018, 0.236]	0.159 [0.051, 0.267]

Notes: Coefficients and 95 per cent confidence intervals. The dependent variable is focal wickets per legal ball; quality has the same units. Timing includes phase and focal over number. Prior loading adds $r = 300/(B + 300)$. State adjusted additionally includes log prior balls, wickets in hand and score rate at the over start. Samples are those in Table 1; legal-ball weights, focal-bowler-by-match and innings effects, and focal-bowler/match clusters apply. State can mediate earlier exposure.

The prior check matters most where the association is already small. Changing the common rate towards which short records are shrunk can alter its magnitude and sometimes its sign. Including the prior loading removes that particular sensitivity, but does not explain away the positive T20 associations. Alternative clustering and further history adjustment leave the main picture similar. Conditioning on the score and wickets remaining changes the estimates more, especially domestically. This is economically relevant because the state facing a bowler can be both a reason for his assignment and a consequence of earlier bowling.

5.2 Partner attributes and coworker output

The containment results run against the simple pressure prediction behind hunting in pairs. Holding partner wicket-taking quality constant, better containment accompanies fewer focal wickets in both T20 settings (Figure 2). A partner predicted to concede 0.6 fewer runs per six-ball over accompanies about 0.15 fewer focal wickets per 100 balls in internationals and 0.14 fewer in franchise cricket. Both confidence intervals exclude zero after adjustment for the two comparisons. Alternative partner-based clustering supports the same reading. The other settings leave the direction uncertain.

The containment measure also tells us something about the bowler whose record it describes. Even after accounting for wicket quality and history length, it improves forecasts of his later run rate in every setting (Table 3). It therefore captures a predictive attribute beyond wicket-taking. Choosing the best rule on validation matches does not guarantee that it will remain best: the selected Test rule performs worse later than the fixed anchor. That distinction matters. The record contains useful information about containment, while the preferred amount of shrinkage remains uncertain.

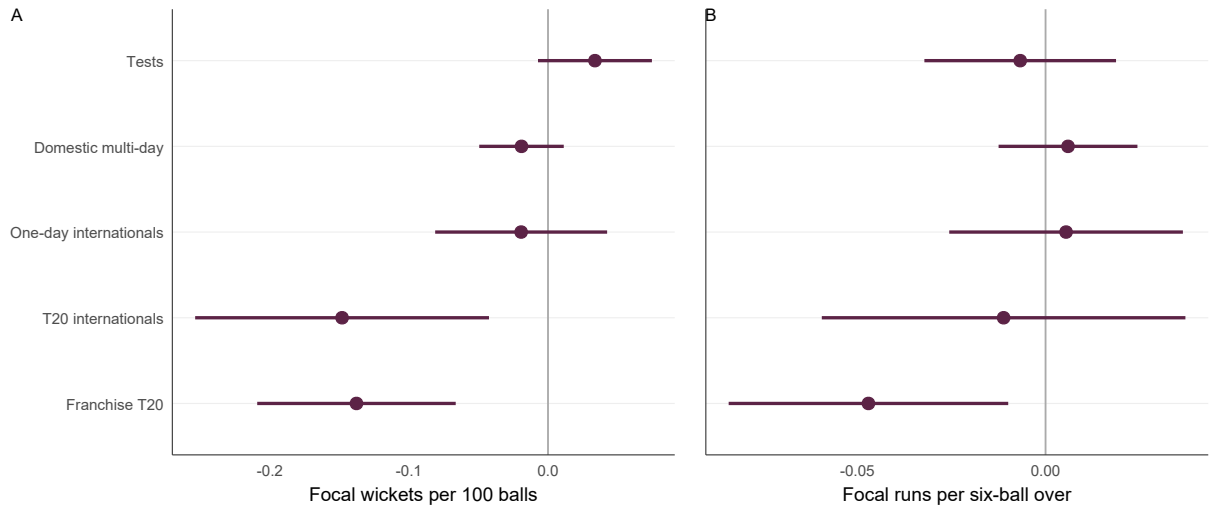

Figure 2: Partner containment and focal output

Notes: A: focal wickets; B: focal total runs. Effects use 0.6 fewer predicted partner runs per six-ball over, conditional on wicket quality, prior loading, timing and fixed effects. Bars are 95 per cent intervals; legal-ball weights and focal-bowler/match clusters. T20 estimates use the joint setting-and-outcome model.

The positive T20 wicket-quality associations are consistent with hunting in pairs on that dimension. In the model with both attributes, wicket quality still has a positive point estimate in each T20 setting, although the franchise interval includes zero; containment has a negative association with focal wickets. The challenge concerns the proposed pressure mechanism and its generality, not the existence of any positive partner association. This extends the own-versus-helping distinction in Arcidiacono, Kinsler and Price (2017). Restricting runs can benefit the team without increasing

the next bowler's credited wickets. That individual outcome cannot price the runs prevented or the allocation of dismissal opportunities.

Focal scoring helps distinguish two possible accounts. If containment redirects batting aggression towards the next bowler, scoring at his end might rise. If the pair instead accompanies more cautious batting or difficult scoring conditions, runs might fall at both ends. In franchise cricket, the same containment contrast predicts roughly 0.05 fewer focal runs per six-ball over; its interval excludes zero after adjustment for the two settings. The international estimate allows either direction. Franchise evidence is therefore compatible with suppression at both ends, while the international comparison remains unresolved. The reported containment contrast is sizeable relative to the variation remaining after controls. Observed partner changes often reach that size, but it should not be treated as a typical or universally feasible substitution.

Observing the task inherited by the focal bowler does not restore the simple positive pressure account. On the same six-ball T20 observations, adding the preceding over's runs and dismissals weakens the negative containment associations; the international interval then includes zero. Accounting also for how long the starting batters have been at the crease leaves both associations negative, with intervals excluding zero. These controls describe circumstances through which a partner might matter, but those circumstances also reflect batting responses and captain choices. The changes cannot be read as the share of an effect transmitted through each channel.

These patterns narrow the production account. The positive cross-end pressure prediction is unsupported by focal wickets, and franchise run rates point towards joint scoring suppression. Production complementarities and incentives in Papps and Bryson (2019), and technology-dependent interactions in Gould and Winter (2009), caution against a universal sign. Separating two partner attributes and observing the inherited task makes that distinction testable here. Defensive role combinations, conditions and batting tactics remain alternatives; neither technological substitution nor a loss in team output is identified.

5.3 Forecast performance and coworker estimates

A better description of a partner's own ability need not clarify his relationship with the focal bowler. Validation favours stronger shrinkage in every setting: large observed differences between workers are not always reliable differences in expected performance. Table 3 and Figure 3 compare later forecast gains with coworker results. As in Arcidiacono, Kinsler and Price (2017), information about own production and contributions to others have different roles.

Selected rules reduce own forecast error by about two per cent in long-form cricket and less in shorter formats. Yet forecasts still exaggerate differences in expected performance: calibration remains incomplete. Domestic selection also reaches the strongest shrinkage allowed by the grid. These are better forecasts among the candidates considered, without establishing accurate latent skill or an optimal rule.

The coworker comparison yields no similarly clear improvement. A rule that draws quality

Table 3: Out-of-sample forecast performance

Setting	Owner-matches	Wicket-rule gain (%)	Containment gain (%)
Test	2,044	2.12 [1.06, 3.18]	5.25 [1.84, 8.66]
Domestic multi-day	7,418	2.04 [1.29, 2.79]	11.58 [9.30, 13.87]
ODI	6,504	0.62 [0.27, 0.96]	8.50 [6.28, 10.72]
T20I	10,407	0.41 [0.18, 0.63]	6.50 [5.14, 7.86]
franchise	10,938	0.25 [0.06, 0.43]	8.53 [7.15, 9.91]

Setting	Focal overs	Undirected gain (%)	Directed addition (%)
Test	50,984	0.017 [-0.002, 0.037]	-0.008 [-0.018, 0.001]
Domestic multi-day	154,530	0.000 [-0.005, 0.006]	-0.002 [-0.009, 0.005]

Notes: Upper: later own-match MSE gains from the selected wicket rule over its anchor, and from adding selected containment conditional on wicket quality and matching loadings. Lower: undirected history over individual-plus-partner attributes, then the directed extension over the unchanged undirected model. Brackets are conditional 95 per cent intervals; legal-ball weights and worker/match clusters.

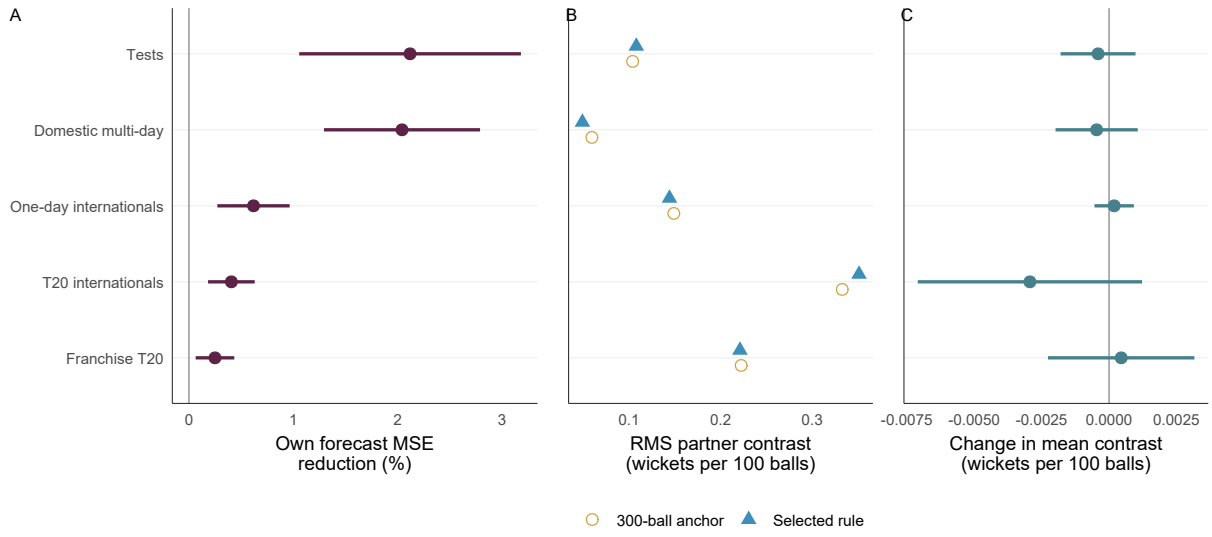

Figure 3: Forecast performance and coworker estimates

Notes: A: own forecast loss gain. B: RMS full-block contrasts on common observed partner changes. C: paired change in mean contrast, selected minus anchor. Coworker fits include timing, fixed effects and history-bin loadings. Bars in A and C are conditional 95 per cent intervals; B is descriptive. Legal-ball weights; bowler/match clusters.

estimates closer together can produce a larger coefficient simply because a unit of the new index represents a different comparison. Figure 3 instead asks how much fitted focal output changes across the same observed partner substitutions, using the whole partner-related part of the model. These contrasts remain similar in magnitude under the two rules, and paired differences in their signed averages are imprecise in every setting. Because positive and negative changes can cancel in an average, the figure also reports their root mean square, which describes their size without that cancellation.

Neither rule explains much of the focal output left after timing and fixed effects, and changing the rule produces no precise improvement in that fit. Timing remains consequential under either measure: comparing similar task stages reduces the apparent contribution of partner information. The assignment and measurement results therefore belong together. A better forecast of a colleague's own output does not remove the need to understand when he is encountered, and a larger coefficient on a more compressed index need not mean a stronger coworker relationship.

This gives practical content to Chen, Gu and Kwon (2025): a useful forecast need not recover a relationship with underlying ability. Even own-output calibration changes when match conditions enter (Table A4). Using unconditional calibration to rescale a coworker coefficient therefore requires further assumptions about ability and measurement error. The forecasting gains validate a particular use of the record.

5.4 Predictive value of pair histories

Knowing how the pair performed together adds little to the specified forecasts once individual and partner records are included (Table 3). The Test gain from undirected history is less than 0.02 per cent, with an interval including zero; the domestic gain is smaller still. Separating who bowls to whom does not produce a precise improvement. The selected directed model performs slightly worse than the undirected benchmark in Tests, although that difference is uncertain. All these intervals remain well below the reporting benchmark. Pooling the two directions therefore does not account for the limited gains in this particular forecasting exercise.

The result concerns prediction of a noisy individual over. It should not be compared directly with the larger gains for a bowler's whole-match forecast, where aggregation smooths some of that noise. Nor does it establish that relationships have little value. The linear history measures may miss useful combinations of styles or tactics, and their fitted multipliers must transfer from shorter training records to a better-informed later benchmark. The intervals condition on those fitted forecasting rules.

Bowling-partnership networks motivate selection applications (Nanavati and Nanavati, 2024). Evaluating those applications also requires asking what a pair's record predicts about later tasks. Here, restricting attention to established pairs leaves the small average gain intact. This limits the support that these particular forecasts offer for hunting in pairs. It does not measure knowledge exchange: Sandvik et al. (2020) changes opportunities for workplace advice experimentally, whereas cricket histories observe joint output. Even a successful pair forecast could reflect persistent

opponents, venues or assignments rather than a benefit from retaining the relationship.

5.5 Performance-history length and coworker estimates

Keeping more of the same players' histories improves own forecasts in both long-form settings (Figure 4). Nobody acquires additional skill in this comparison: current players, matches and outcomes stay fixed while the researcher sees more of their past. Historical matches are drawn from the same recent window, and their average age remains closely balanced across budgets. The improvement therefore reflects the information supplied by a fuller record, rather than a comparison between workers at different career stages or between older and more recent evidence.

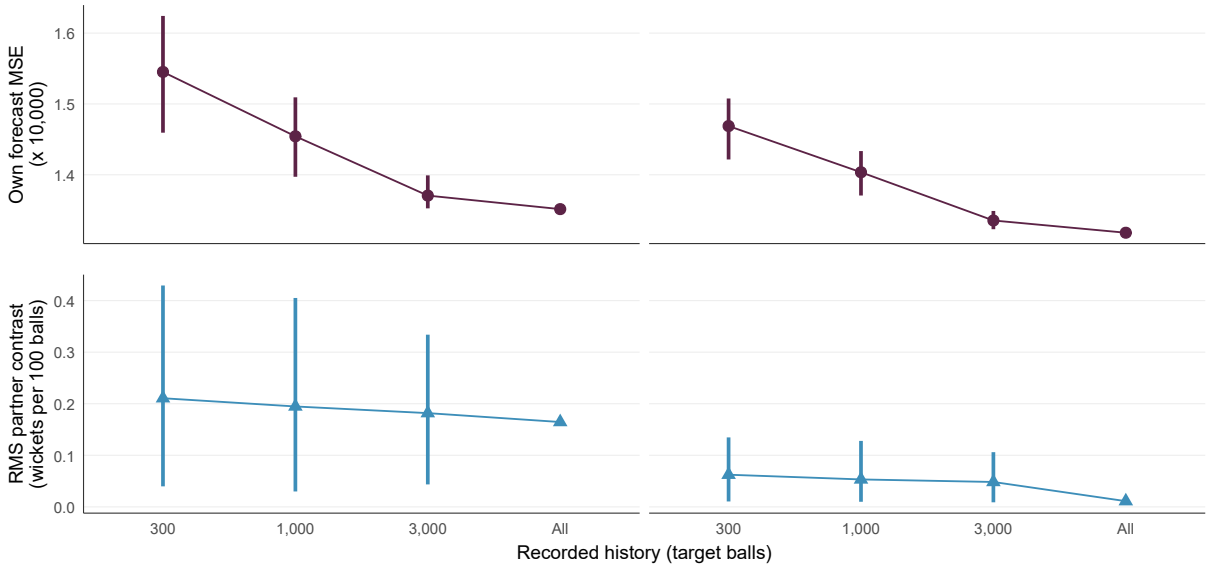

Figure 4: Forecasts and fitted coworker contrasts by history length

Notes: Tests left; domestic multi-day right. Own-output MSE above; RMS full-block partner contrasts below, on identical partner changes across budgets. Points average 100 whole-match draws; bars show their central 95 per cent, not sampling intervals. Current samples contain 33 Test and 148 domestic partners. All uses the complete three-year window.

The raw Test coefficient rises as more history is retained, but the fitted differences between the same partner substitutions become smaller. Domestic fitted contrasts also shrink, while raw coefficients remain close to zero. This is the distinction that a list of slopes would miss: changing the record changes the scale of the measure as well as its information. A larger coefficient does not necessarily imply a larger difference in predicted focal output. The Test comparison also rests on only 33 measured partners, so the upward coefficient path is highly uncertain.

The paired comparisons provide no precise evidence that average signed coworker contrasts change with the history budget. Tests also leave differences in explanatory fit uncertain. Domestically, the shorter records produce tiny improvements in fitting current focal outcomes, with the two smallest budgets retaining statistical support after adjustment across settings. Yet those same

records make own forecasts worse. There is no contradiction: a model refitted to current outcomes can use random historical variation to fit noise. A small gain in explanatory fit need not be better information about economic relationships. These results distinguish the value of a record for forecasting from what its use in a coworker regression can establish.

Holding current people fixed separates this information comparison from career differences. The same observed record changes without changing current capability or assignment. This complements the production-learning distinction in Herkenhoff et al. (2024) and the precision concerns in Chen (2026). It does not measure how a player would perform with less accumulated skill, and an uncertain upward raw-slope path cannot identify classical attenuation.

6 Worker heterogeneity and team composition

6.1 Experience heterogeneity and common support

Immediate output can understate the value of a less experienced coworker. Proximity increases feedback to younger software engineers while reducing senior engineers' code output (Emanuel, Harrington and Pallais, 2026). Heterogeneous-team productivity (Hamilton, Nickerson and Owan, 2003) similarly distinguishes own current output from broader team contributions. Cricket permits a narrower comparison by the focal bowler's recorded experience.

Most long-form innings already combine bowlers with short and long recorded histories. Restricting the analysis to established players would omit much of the deployment decision a captain actually faces. I therefore allow the partner-quality relationship to differ for focal bowlers with fewer than 1,000 recorded prior balls, retaining the corresponding prior-loading interaction and adjustment for partner history. This compares immediate output among workers already selected for a match; it does not measure the return to recruiting or training less experienced players.

Table 4: Coworker estimates by focal-bowler experience

Setting	Experienced focal	Less experienced focal	Difference
Test	-0.035 [-0.127, 0.057]	-0.049 [-0.195, 0.098]	-0.014 [-0.175, 0.147]
Domestic multi-day	-0.025 [-0.084, 0.035]	0.022 [-0.081, 0.125]	0.046 [-0.072, 0.164]
ODI	-0.001 [-0.115, 0.113]	-0.054 [-0.170, 0.062]	-0.053 [-0.218, 0.112]
T20I	0.106 [-0.137, 0.349]	0.171 [0.027, 0.314]	0.065 [-0.213, 0.342]
franchise	0.268 [0.098, 0.437]	0.008 [-0.138, 0.155]	-0.259 [-0.487, -0.031]

Notes: Less experienced means fewer than 1,000 recorded pre-match balls. Each setting uses its complete timing sample. Entries are raw slopes and 95 per cent intervals from one interacted regression per setting; the difference is less experienced minus experienced. Models include quality and prior-loading interactions, log partner history, timing, focal-bowler-by-match and match-innings effects. Legal-ball weighting and focal-bowler/match clustering apply. Intervals are marginal, not multiplicity-adjusted.

Most settings provide little precise evidence that the coworker relationship differs by focal experience (Table 4). Franchise cricket is the exception: the positive association is concentrated

among more experienced focal bowlers, and the contrast persists after state adjustment. This is one of several comparisons and should not be treated as a discovery after accounting for all of them. It nevertheless cautions against attributing the positive short-format pattern entirely to junior workers receiving senior help. Immediate wicket production cannot reveal mentoring or the eventual return to employing a less experienced colleague.

Comparing partners with similar recorded histories also complicates a simple format story (Table A3). Within the common middle range, the international T20 estimate is negative and uncertain, whereas the franchise estimate remains positive with an interval excluding zero. Restricting history does not therefore produce a uniform short-format relationship. It changes the workers and matches represented as well as the precision of their estimates. The comparison exposes that dependence; it does not make the formats interchangeable or establish that their underlying relationships are equal.

A long record also reflects opportunities to keep working. In domestic cricket, bowlers with short recorded histories are less likely to appear again within the following year than those with long histories. These are observed appearances in an incomplete archive, not a model of every professional career. The distinction still matters: a forecast evaluated among workers who receive another assignment says less about those who disappear from the record. Managers' selection of future workers remains outside the own-output forecasting exercise.

6.2 Pair relationships and team inputs

Repeated partnerships could contain information that a quality ranking misses. Familiarity, past run suppression and earlier pair excess nevertheless have imprecisely estimated associations with focal output after state adjustment (Appendix D). These regressions ask about conditional persistence; the forecasting exercise asks whether a trained rule predicts later outcomes better. Both face the possibility that captains repeatedly choose a pairing for reasons that also persist.

Repeated exposure need not imply knowledge exchange. Structured coworker conversations generate persistent gains in Sandvik et al. (2020), unlike incentives for joint output alone. Cricket records exposure and output, but not those conversations. Familiarity can proxy for stable assignments; pair excess is noisy; state can absorb earlier productive channels. These features test measured persistence, with limited precision, rather than the value of communication, style complementarity or learning.

The rest of the team matters, but adding another average cannot separate its contribution here. Within a fixed roster, a stronger immediate partner mechanically lowers the average quality of the coworkers left in the other group. After fixed effects, the measures carry the same information in opposite directions. Separating their contributions requires variation in composition, a distinct interaction measure or a different design. Deployed-bowler counts alone do not supply it.

An injury might appear to interrupt the captain's chosen pairing from outside the ordinary assignment process. But requiring the focal bowler to continue and a replacement to complete the partner's next regular-slot over leaves only 92 long-form events. Comparing focal performance be-

fore and after these interruptions is too imprecise to adjudicate the main associations (Appendix F). Continued deployment also selects the focal sample, while fatigue and match state may affect injury timing. The sequence thus offers neither the precision nor the assignment argument needed to emulate more informative coworker-absence designs (Hoey, Peeters and van Ours, 2023).

7 Conclusion

Evidence remains limited on when a successful pairing reflects one worker helping another, and how much the measured relationship depends on assignment and the records used to describe quality. Cricket's belief that bowlers hunt in pairs gives that uncertainty a concrete form. This paper uses alternating tasks and dated histories to examine how the pairing's timing, the partner's attributes and the available performance record shape the evidence for coworker contributions.

Timing changes the associations substantially, even within the same bowler's match and with most residual quality variation retained. Positive wicket-quality associations remain in both T20 settings, consistent with that dimension of hunting in pairs. Better partner containment, however, accompanies fewer focal wickets there and lower focal scoring in franchise cricket. The simple pressure prediction is challenged in T20; long-form comparisons leave its direction uncertain. More informative records improve own-output forecasts without revealing a clearer coworker relationship; retaining more history can raise a raw coefficient while reducing fitted partner contrasts. Pair histories add little under the specified forecasting rules. These comparisons do not identify what would happen if a captain assigned a different partner.

For team-production research, the relevant attribute and outcome must be explicit. Gould and Winter (2009) relate interactions to technology, Papps and Bryson (2019) examine complementarities and incentives, and Arcidiacono, Kinsler and Price (2017) distinguish own production from helping others. Cricket compares two partner attributes while observing the task left for the next worker. Runs prevented can benefit the team without giving that worker additional dismissal credit. A single quality ranking conceals this distinction.

For managerial-allocation research, the stage at which coworkers meet belongs in the description of exposure. Visibility matters in Mas and Moretti (2009), hospital assignment and shifts shape work in Chan (2016, 2018), and managers improve allocation in Minni (2026). The decomposition shows how deployment enters the measured relationship: phase matters under either forecasting rule. It does not separate sorting from productive channels transmitted through the evolving task.

The measurement contribution compares uses of the same record. The regression conditions in Chen, Gu and Kwon (2025) and precision heterogeneity in Chen (2026) explain why useful forecasts need not recover relationships with latent skill. Holding current workers fixed makes the information comparison visible without adding experience, complementing Herkenhoff et al. (2024). Limited pair-history forecast gains likewise do not negate returns to knowledge exchange (Sandvik et al., 2020): shared output does not reveal advice or learning.

The sports-economics contribution is to test an established industry belief with the detailed

records that make sport useful for economic inquiry (Kahn, 2000; Palacios-Huerta, 2025). Alongside evidence of bowling-partnership networks (Nanavati and Nanavati, 2024), the analysis asks whether the proposed benefit appears in the relevant output and whether pair records improve later forecasts. It challenges the presumption that successful bowling pairs establish a general positive coworker effect. Particular partnerships may still help through channels these comparisons cannot value.

The broader lesson applies most directly to deployment among available workers doing recurring, individually recorded tasks. Wage outcomes (Cornelissen, Dustmann and Schönberg, 2017) and heterogeneous-team gains (Hamilton, Nickerson and Owan, 2003) concern horizons and benefits that immediate wickets cannot value. Other industries need tests of what a colleague is expected to change, when exposure occurs and whose output should respond. Assignment policy also requires evidence on feasible alternative pairings and joint output. Randomising rotations while recording task state before assignment would help distinguish deployment from productive interaction; following workers afterwards could reveal persistent learning. Familiar partnerships motivate these tests. They cannot substitute for them.

Data availability

Source records are public at Cricsheet. The replication documentation identifies the derived panels and analysis scripts. Derived data and the code that reproduce all tables and figures are available from the author for replication, subject to the source terms.

References

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics*, 30: 98–108.

Arcidiacono, Peter, Josh Kinsler, and Joseph Price. 2017. "Productivity Spillovers in Team Production: Evidence from Professional Basketball." *Journal of Labor Economics*, 35(1): 191–225.

Chan, David C. 2016. "Teamwork and Moral Hazard: Evidence from the Emergency Department." *Journal of Political Economy*, 124(3): 734–770.

Chan, David C. 2018. "The Efficiency of Slacking Off: Evidence from the Emergency Department." *Econometrica*, 86(3): 997–1030.

Chen, Jiafeng. 2026. "Empirical Bayes When Estimation Precision Predicts Parameters." *Econometrica*, 94(2): 305–340.

Chen, Jiafeng, Jiaying Gu, and Soonwoo Kwon. 2025. "Empirical Bayes Shrinkage (Mostly) Does Not Correct the Measurement Error in Regression." Working paper, 24 March 2025.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg. 2017. "Peer Effects in the Workplace." *American Economic Review*, 107(2): 425–456.

Emanuel, Natalia, Emma Harrington, and Amanda Pallais. 2026. "The Power of Proximity to Coworkers." *Quarterly Journal of Economics*, 141(3): 1825–1870.

Gelbach, Jonah B. 2016. "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics*, 34(2): 509–543.

Gould, Eric D., and Eyal Winter. 2009. "Interactions between Workers and the Technology of Production: Evidence from Professional Baseball." *Review of Economics and Statistics*, 91(1): 188–200.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy*, 111(3): 465–497.

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M. Phillips. 2024. "Production and Learning in Teams." *Econometrica*, 92(2): 467–504.

Hoey, Sam, Thomas Peeters, and Jan C. van Ours. 2023. "The Impact of Absent Co-workers on Productivity in Teams." *Labour Economics*, 83: 102400.

Kahn, Lawrence M. 2000. "The Sports Business as a Labor Market Laboratory." *Journal of Economic Perspectives*, 14(3): 75–94.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review*, 99(1): 112–145.

Minni, Virginia. 2026. "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs." *Quarterly Journal of Economics*, 141(3): 1871–1920.

Nanavati, Praharsh, and Amit Anil Nanavati. 2024. "Bowlership: Examining the Existence of Bowler Synergies in Cricket." In *Complex Networks & Their Applications XII*. Vol. 1143 of *Studies in Computational Intelligence*, , ed. Hocine Cherifi, Luis M. Rocha, Chantal Cherifi and Murat Donduran, 124–133. Cham:Springer.

Palacios-Huerta, Ignacio. 2025. "The Beautiful Dataset." *Journal of Economic Literature*, 63(4): 1363–1423.

Papps, Kerry L., and Alex Bryson. 2019. "Spillovers and Substitutability in Production." IZA Institute of Labor Economics Discussion Paper 12252.

Sandvik, Jason, Richard Saouma, Nathan Seegert, and Christopher Stanton. 2020. "Workplace Knowledge Flows." *Quarterly Journal of Economics*, 135(3): 1635–1680.

Walters, Christopher R. 2024. "Empirical Bayes Methods in Labor Economics." In *Handbook of Labor Economics*. Vol. 5 of *Handbooks in Economics*, , ed. Christian Dustmann and Thomas Lemieux, 183–260. Amsterdam:Elsevier.

Yew, Oliver. 2016. "Are James Anderson and Stuart Broad England's Greatest Bowling Pair?" *Sky Sports, 9 June.*

A Data construction and estimation

The eligible sample covers Tests from 19 December 2001 to 28 June 2026; domestic cricket from 27 October 2012 to 19 June 2026; ODIs from 27 June 2002 to 16 June 2026; T20Is from 17 February 2005 to 2 July 2026; and franchise cricket from 18 April 2008 to 2 July 2026. Coverage within these dates is incomplete. The international and franchise archive contains 10,774 distinct matches across those four settings, and domestic data contain 2,162. A bowler's quality history can include earlier recorded matches whose overs do not qualify as focal observations.

The analysis uses R 4.5.2, `data.table`, `fixest` and `ggplot2`. Singleton removal is iterative. Intervals use 1.96 clustered standard errors. Two-way covariance adds worker and match components, subtracts their intersection, and multiplies by $G_{\min}/(G_{\min}-1)$, without a parameter-count correction. Indefinite full matrices are not repaired: reported scalar variances are positive, and joint-test contrast matrices are positive definite. Independent component and unit-change checks agree within 10^{-7}. Prior-shift checks reproduce slopes and standard errors within 10^{-8} and fitted values within 10^{-6}. Numerical loss rescaling is reversed before reporting.

Specifications were recorded after baseline evidence but before the respective extension estimates; this is not a preregistration. Replication files retain all settings, forecasting candidates, clustering sensitivities and thinning draws (seed 20260907).

Table A1: Regression support and partner switching

Setting	Partners	Focal matches	Switching	Residual SD	Top 10% share
Test	776	9,910	9,336	0.251	56.1
Domestic multi-day	1,511	26,215	24,527	0.256	50.0
ODI	1,449	29,438	26,469	0.353	60.2
T20I	3,134	35,092	29,948	0.426	59.5
franchise	1,603	40,496	37,404	0.453	58.4

Setting	Measured partners	Focal matches	Switching focal-matches
Test	33	1,149	439
Domestic multi-day	148	4,777	2,844

Notes: Upper panel: prior-loading regressions after singleton removal. Lower panel: eligible thinning observations before removal. Switching means at least two partners within a focal-match. Residual SD is in wickets per 100 balls after prior loading, timing and fixed effects; the last column gives the largest tenth of partners' share of weighted residual quality squares.

B Prior-mean invariance

The prior-mean check asks whether changing the common rate changes the part of measured quality used in the regression. Let M_X remove the included fixed effects and timing controls using weights

Table A2: Confidence intervals by inference method

Setting	Inference	Raw slope [95% interval]
Test	One-way match	-0.037 [-0.119, 0.045]
Test	Match wild-score, 9999 draws	-0.037 [-0.121, 0.044]
Domestic multi-day	One-way match	-0.012 [-0.065, 0.041]
Domestic multi-day	Match wild-score, 9999 draws	-0.012 [-0.064, 0.041]
ODI	One-way match	-0.026 [-0.106, 0.053]
ODI	Match wild-score, 9999 draws	-0.026 [-0.106, 0.053]
T20I	One-way match	0.154 [0.028, 0.281]
T20I	Match wild-score, 9999 draws	0.154 [0.031, 0.280]
franchise	One-way match	0.127 [0.021, 0.233]
franchise	Match wild-score, 9999 draws	0.127 [0.020, 0.234]

Notes: Same prior-loading models and samples as Table 2. One-way intervals cluster by match. Wild-score intervals use 9,999 Rademacher multipliers of match-level residualised coefficient scores (seed 20260908), with the match correction. These are score-multiplier intervals, not a re-estimated bootstrap-t.

collected in Ω. Write $\widetilde{s} = M_X s$, $\widetilde{r} = M_X r$ and $\widetilde{y} = M_X y$. Without a separate prior-loading control, the coefficient at prior mean m is

$$\widehat{\beta}(m) = \frac{(\widetilde{s} + m\widetilde{r})'\Omega\widetilde{y}}{(\widetilde{s} + m\widetilde{r})'\Omega(\widetilde{s} + m\widetilde{r})}.$$

Both numerator and denominator can change with m; neither a sign change nor monotonicity is guaranteed. After including r in X, $M_X r = 0$, so the expression no longer depends on m. This argument also applies to the corresponding clustered sandwich covariance under an exact nonsingular linear reparameterisation. It does not extend to changing k, the sample, standardisation or the conditioning variables.

For the rule comparison, write the partner block as $z = (q, r_1, \ldots, r_4)$ and its coefficients as $\widehat{\theta}$. At switch ℓ, the full contrast is $d_\ell = 100(z_{new,\ell} - z_{old,\ell})'\widehat{\theta}$, weighted by the new over's legal balls. Its signed mean and RMS use identical switches across rules. The same-phase subset gives a sensitivity. Complete-block partial fit is $100(1 - L_F/L_C)$, where L_F is residual MSE and L_C omits the partner block; quality-only fit retains the loadings. Full contrasts and complete-block fit are invariant to invertible block transformations; quality-only fit is invariant to quality rescaling and prior-location shifts.

Paired mean-contrast scores retain cross-model covariance, conditional on observed switches. Loss-ratio scores include the denominator's variation; at the least-squares optimum the fitted-coefficient derivative of in-sample loss is zero. For thinning, squared residual losses and coefficient scores are averaged across draws before clustering. Redraws do not supply independent populations. These intervals condition on forecasting rules and observed histories. Holm families compare settings within a specification and contrast; RMS and partial-fit ranges receive no first-order intervals. Numerical reparameterisation verifies fitted values, contrasts and standard errors.

C History length and forecast calibration

Table A3: Coworker estimates by partner-history length

Setting	Prior balls	Partners	Regression overs	Raw slope [95% CI]
Test	0-299	765	37,874	-0.203 [-0.544, 0.138]
Test	300-999	440	40,310	0.046 [-0.301, 0.394]
Test	1000-2999	279	68,212	-0.002 [-0.242, 0.238]
Test	3000+	153	130,785	-0.071 [-0.257, 0.116]
Domestic multi-day	0-299	1,475	67,912	-0.048 [-0.305, 0.209]
Domestic multi-day	300-999	922	88,149	-0.107 [-0.266, 0.053]
Domestic multi-day	1000-2999	632	154,117	0.035 [-0.100, 0.170]
Domestic multi-day	3000+	353	287,956	-0.013 [-0.138, 0.112]
ODI	0-299	1,441	47,828	-0.200 [-0.491, 0.090]
ODI	300-999	689	53,438	0.001 [-0.230, 0.233]
ODI	1000-2999	360	67,212	-0.065 [-0.247, 0.117]
ODI	3000+	128	34,917	-0.116 [-0.427, 0.194]
T20I	0-299	2,964	53,890	0.202 [-0.024, 0.427]
T20I	300-999	888	25,739	0.372 [0.079, 0.664]
T20I	1000-2999	295	12,818	-0.209 [-0.661, 0.243]
T20I	3000+	57	2,961	-0.081 [-1.170, 1.008]
franchise	0-299	1,479	25,183	-0.035 [-0.416, 0.347]
franchise	300-999	683	27,435	-0.067 [-0.370, 0.237]
franchise	1000-2999	342	32,288	0.399 [0.122, 0.676]
franchise	3000+	87	10,925	0.394 [-0.179, 0.966]

Notes: Each bin is a separate regression of focal wickets per legal ball on raw partner quality, prior loading and timing, with focal-bowler-by-match and match-innings effects, legal-ball weights and focal-bowler/match clusters. All displayed counts use fitted observations after bin restriction and singleton removal; bin counts need not sum to the full sample. Retained row and partner sets are subsets of the full fit. Brackets are 95 per cent intervals. "Prior balls" describes the recorded partner history, not true career tenure.

Training ends on 9 February 2017 for Tests, 27 October 2021 domestically, 20 October 2016 for ODIs, 21 May 2023 for T20Is and 4 September 2021 for franchise cricket. Validation ends on 8 December 2021, 21 March 2024, 13 April 2022, 17 December 2024 and 21 February 2024, respectively. Evaluation dates follow strictly. Candidates share owner-matches and dates; histories may include earlier evaluation matches but never same-date or later matches.

D History thinning and lagged pair measures

Mean realised balls are 424, 1,123 and 3,120 in Tests and 393, 1,095 and 3,094 in domestic cricket; complete windows average 4,280 and 4,878. Record ages vary by less than one day across budgets. Independent random orderings for each bowler-match mean that overlapping windows may retain

Table A4: Forecast calibration with and without match effects

Setting	Rule	Intercept	Slope [95% CI]	With match effects
Test	anchor	0.0116	0.404 [0.197, 0.610]	0.226 [0.034, 0.418]
Test	selected	0.0075	0.649 [0.364, 0.934]	0.407 [0.126, 0.688]
Domestic multi-day	anchor	0.0098	0.438 [0.357, 0.519]	0.293 [0.208, 0.378]
Domestic multi-day	selected	0.0048	0.726 [0.552, 0.901]	0.599 [0.432, 0.765]
ODI	anchor	0.0132	0.541 [0.418, 0.663]	0.570 [0.438, 0.702]
ODI	selected	0.0063	0.811 [0.640, 0.981]	0.853 [0.661, 1.045]
T20I	anchor	0.0346	0.396 [0.237, 0.556]	0.430 [0.254, 0.606]
T20I	selected	0.0165	0.738 [0.453, 1.024]	0.815 [0.498, 1.132]
franchise	anchor	0.0286	0.474 [0.310, 0.638]	0.489 [0.325, 0.653]
franchise	selected	0.0175	0.699 [0.470, 0.929]	0.718 [0.490, 0.946]

Notes: Bowler-match regressions of realised wickets per legal ball on predicted quality with an intercept, using legal-ball weights and bowler/match clusters. The anchor has a training global mean and $k = 300$; the selected rule is chosen on validation data. Brackets are 95 per cent slope intervals conditional on the forecasting rule. The last column adds match effects on identical observations. Perfect unconditional linear calibration requires intercept zero and slope one. Sample sizes are in Table 3. These are forecasts conditional on observed appearance.

different historical matches. Information is varied per prediction occasion, rather than permanently erased from the archive.

Full-window raw slopes are 0.453 (SE 0.409) in Tests and 0.011 (SE 0.124) domestically, on 14,156 and 62,377 fitted overs. Population-sampling uncertainty differs from the historical-draw variation in Figure 4.

At 1,000 and 3,000 balls, paired signed-contrast differences from the complete window are 0.0040 $[-0.0163, 0.0243]$ and -0.0004 $[-0.0105, 0.0098]$ in Tests, and 0.0004 $[-0.0064, 0.0071]$ and 0.0004 $[-0.0021, 0.0028]$ domestically. Test residual-loss gains are 0.0087, 0.0069 and 0.0039 per cent across budgets, with intervals $[-0.0173, 0.0347]$, $[-0.0156, 0.0294]$ and $[-0.0131, 0.0209]$. The domestic 3,000-ball gain is 0.0018 per cent $[0.0001, 0.0036]$; its Holm-adjusted $p = 0.085$. All six finite-budget own forecasts have precisely higher loss than the full window. These paired comparisons average across the fixed draws before clustering.

Familiarity is $\log(1+P_{ij,t}/300)$, with P the undirected pair's prior eligible balls. Prior pair excess sums credited wickets minus balls times pre-match own quality, divided by prior pair balls plus 300 and standardised within setting. Historical quality priors are 0.018 in Tests and 0.020 domestically. Run suppression is negative standardised partner runs per ball. On 882,053 state-adjusted long-form overs, familiarity, suppression and pair-excess coefficients are 0.000153 $[-0.000160, 0.000465]$, -0.000036 $[-0.000188, 0.000116]$ and 0.000001 $[-0.000140, 0.000142]$. These lagged regressions condition on partner quality, prior loading and state; they are distinct from trained forecasting rules.

E Construction of pair-history forecasts

Training burn-in covers its first half of dates. Five later date blocks receive predictions fitted only on earlier prefixes, including prior means; burn-in retains zero pair features. Excess and usable balls accumulate at undirected pair-date level, excluding the focal date. Validation selects strengths 1,000 in Tests and 3,000 domestically. Each candidate has a training-fitted multiplier. Sensitivities use selected individual rules. Up to 53 of 50,984 Test forecasts are negative; linear predictions are not clipped.

The directional stock uses the same honest excesses but accumulates only the focal member's output. A zero-padded pair-date grid updates both directions strictly after each date; direction totals reproduce the undirected numerator and exposure exactly. The extension fits undirected and directed multipliers jointly on training outcomes, choosing a shared strength on validation loss. A relative singular-value tolerance of 10^{-10} defines a minimum-norm fit if that block lacks rank. Established-pair support uses all prior eligible balls; directional support uses usable balls after burn-in. Candidate losses, ranks and both clustering sensitivities accompany the code.

F Injury-event construction

An event starts when a source-recorded bowling injury interrupts the opposite-end partner's over k. The focal bowler must deliver $k-3$ and $k-1$, with the injured partner at $k-4$ and $k-2$, and return at $k+1$ and $k+3$. A different single bowler must complete at least six legal balls at $k+2$. This is a realised regular-slot replacement, not a pre-announced plan. Mixed replacement and outcome overs are excluded; raw deliveries verify every retained bowler and outcome. There are 38 Test and 54 domestic events, involving 79 focal bowlers and 92 matches.

The event outcome is $\Delta_e = (y_{k+1} + y_{k+3} - y_{k-3} - y_{k-1})/2$, with y the focal credited wickets per legal ball. Its unweighted event mean is 0.00725, with focal-bowler/match clustered SE approximately 0.0100. A two-sided 5 per cent normal test's approximate 80 per cent power threshold is $(1.96 + 0.842)\, SE \simeq 0.028$. The comparison is this pre/post change, not a quality slope or a purely post-replacement effect: $k+1$ precedes the replacement's full over. Continued focal deployment selects the sample, while fatigue and match state may affect the injury. The calculation establishes limited precision for this sequence, not an assignment experiment.

Data and code availability

Source records are public at Cricsheet. Replication documentation identifies the derived panels and analysis scripts. Derived data and code are available from the corresponding author during review and will be made available for replication, subject to source terms.