



Instructor guide: economic model building

Laboratory implementation, theoretical foundations,
experimental evidence, and variants

Behavioral and Experimental Economics Research Lab
Department of Economics



Revised edition | October 2026

<https://beer-lab.org/>

How to use this guide

This guide supports preparation, implementation and discussion of the BEER Lab Economic Model Building family. It serves economics and related courses without assuming programming or laboratory experience.

The operational route explains access, configuration and screens. Foundations develop variables, formulas, information and evaluation. Extensions and worked exercises support a debrief using classroom results.

Each language covers the basic edition and thirteen extensions. This is an individual activity with no matching or waiting. Common elements appear once and each extension develops what changes.

Author

This guide was prepared by Stanislao Maldonado, research professor in the Department of Economics at Tecnológico de Monterrey, Monterrey Campus. For questions or suggestions about its content, please contact the author at stanislao.maldonado@tec.mx.

For the instructor

This guide follows the English-language interface. Spanish configurations use the same logic. Never project individual codes or connect outcomes to names, student identification numbers, or seats in the room.

Contents

How to use this guide	i
1 What the activity teaches and how to access it	1
1.1 A question before an equation	1
1.2 Operational summary	2
1.3 From the laboratory website to a session	2
2 Foundations and worked explanations	8
2.1 Variables, assumptions and functional form	8
2.2 Solving the stable reference	8
2.3 Training, testing and revision	9
2.4 Error and points, step by step	10
2.5 Designing an informative comparison	10
3 Preparing and conducting the session	11
3.1 Decisions before class	11
3.2 What participants see	12
3.3 Common operational problems	12
4 Extensions and teaching comparisons	18
4.1 Selecting a comparison	18
4.2 Moderate and high outcome variation	18
4.3 A nonlinear relationship	19
4.4 A relevant condition	19
4.5 Interaction between variables	19
4.6 Correlated initial data	20
4.7 Prediction outside the initial range	20
4.8 An environment change	20
4.9 A condition without an effect	21
4.10 A complexity cost	21
4.11 An experimentation cost	21
4.12 Demand and production applications	22
5 Reading the report and using the data	23
5.1 Revelation and denominators	23
5.2 Formulas, trajectories and scatter	23
5.3 Download, privacy and reproduction	24
6 Class plans, exercises and closing discussion	27
6.1 A short session and a revision session	27
6.2 Exercises with answer guidance	28
6.3 Assessment and final checks	29

List of Figures

1.1	Laboratory website: CLASSROOM leads to teaching materials.	4
1.2	Teaching directory: locate the family card and choose the intended language. . . .	5
1.3	Isolated Demo: this excerpt contains model-building configurations.	6
1.4	Rooms contains four rooms belonging to this family.	6
1.5	English basic room: choose the protocol before creating a session.	7
3.1	Session creation: choose the protocol and number of people.	13
3.2	Welcome: students read and confirm the laboratory rules.	14
3.3	Resources: four initial observations, permitted trials and the meaning of ECU. . .	14
3.4	Comprehension: feedback explains the incorrect calculation and retains other answers.	15
3.5	Exploration: the chosen quantity determines which trials are executed.	15
3.6	Registration: the preview shows predictions while test outcomes remain hidden. . .	16
3.7	Individual result: compare errors, costs and predictions with revealed outcomes. .	16
3.8	Explicit completion with total and cycle scores.	17
5.1	Initial report: progress and download are available while aggregates remain locked.	25
5.2	Histogram: counts of individual cycles grouped by new-test error.	25
5.3	Trajectory: distinguish accumulated training data from the current cycle's tests. .	26
5.4	Scatter: distance from the diagonal indicates the direction and size of errors. . . .	26

What the activity teaches and how to access it

1.1 A question before an equation

An economic model is a deliberately incomplete representation of a situation. Its usefulness depends on the question it answers, the mechanisms it retains and the conditions under which its conclusions remain defensible. This activity lets students experience that relationship before receiving a collection of finished equations. They observe outcomes from an artificial system, choose additional information, register a rule and then discover how well it predicts observations whose outcomes they have not yet seen. Model building becomes a sequence of decisions that can be examined, rather than a label attached to an equation after the fact.

The family does not measure social preferences or pit two players against one another. Participants work independently, need no partner and do not wait for other students to make decisions. Any positive whole number of participants is admissible. Cycles provide opportunities to formulate and revise an explanation; they are not rounds of strategic interaction. The basic edition therefore offers a short supplied-data exercise, a one-cycle exploration with chosen trials and a three-cycle revision sequence. Instructors can use these protocols in introductory economics, methodology, microeconomics, macroeconomics or quantitative methods, choosing the learning objective rather than inheriting an unnecessary matching structure.

The initial setting is a virtual glider. A design setting is called X , a binary condition is called Z and simulated distance is called Y . Distance is measured in laboratory units. It is not a measurement in metres from a real aircraft, nor the output of a validated aerodynamic model. The setting makes the distinction between an object and a representation tangible. Demand and production extensions change the economic meanings of the variables while preserving the procedure of observation, formulation and evaluation. Introduce those comparisons after students understand how to register a rule and read its predictions.

Theoretical foundation

A useful rule needs a question and a domain of application. Matching an observed outcome does not establish that the underlying mechanism has been discovered. The activity teaches this distinction; it does not, by itself, certify general research ability or diagnose an individual psychological bias.

1.2 Operational summary

Purpose	Construct, register, evaluate and revise models using artificial data.
Experimental unit	Individual participant; results summarized by person and cycle.
Basic edition	Supplied data, one-cycle exploration and three-cycle revision.
Extended edition	Thirteen extensions concerning information, functional form, environment, costs and economic applications.
Decision	Choose up to two trials per cycle, one of five formulas and the coefficients used by that formula.
Outcome	Up to 100 ECU per cycle, depending on predictive error and announced costs; no initial monetary endowment.
Topics	model building, experimental design, predictive validation, omitted variables, structural change, complexity and parsimony.
JEL	A22, B41, C51, C52, C53, C91.
Participants	Any positive whole number; three people in the default demonstration.
Suggested duration	25–35 minutes for one cycle with discussion; 45–60 for revision and a comparison.

These durations are instructional estimates, not a timer enforced by the software. A class unfamiliar with algebra will need practice substituting numbers and distinguishing a coefficient from an observation. A more advanced class can spend less time introducing the screen and more time selecting informative experiments. The same exercise need not teach estimation, causality, forecasting and philosophy of science simultaneously. Select a principal question and reserve other connections for the transfer discussion. This makes the activity manageable without artificially limiting the number of extensions available to colleagues with different teaching goals.

ECU stands for *Experimental Currency Units*. In this family ECU express performance points, not an initial budget that can be spent on goods. No conversion into real money has been established. The informational resources are four initial observations and, when the protocol permits, up to two additional trials per cycle. A participant does not receive an initial balance of 100 ECU: 100 is the maximum gross score generated in a cycle. Explain this distinction explicitly, especially if students have already played the Dictator or another allocation game. The common unit of account does not imply that the resource structure is the same.

1.3 From the laboratory website to a session

Start at <https://beer-lab.org/>. The Spanish CLASE section leads to <https://teaching.beer-lab.org/>; the English navigation uses CLASSROOM. Find the Economic Model Building card and check its language, edition and objective. The navigation and screenshots in this guide were checked on 16 October 2026. The directory is an entry point; decisions are recorded in the game instance, not in the general teaching directory. An instructor should distinguish that public catalogue from the private administrative controls before sharing a link with a class.

Stable student routes are https://modelbuilding.beer-lab.org/room/modelos_es and https://modelbuilding.beer-lab.org/room/models_en for the basic editions. Extended routes are https://modelbuilding.beer-lab.org/room/modelos_extendido and https://modelbuilding.beer-lab.org/room/models_extended. Instructors enter through https://modelbuilding.beer-lab.org/room/modelos_es and https://modelbuilding.beer-lab.org/room/models_en.

[//modelbuilding.beer-lab.org/demo](https://modelbuilding.beer-lab.org/demo). Each room offers only configurations from its edition and language. A basic room has three options; an extended room has fourteen, comprising the revision reference and thirteen extensions. Changing the portal language does not translate a session that has already been created.

The instructor signs in using credentials provided by the laboratory, selects a configuration, sets the number of participants and creates a new session. Demo is useful for exploring screens. A stable room is the address normally shared with the class. An individual link belongs to one participant in a particular session. The report is a private administration page. These four elements serve different functions and should not be substituted for one another when distributing links. In particular, do not project a list of participant codes together with identifiable decisions or give students an instructor link to reach their task.

This isolated family instance contains 34 configurations and four rooms. If Demo or Rooms displays other game families, stop preparation and contact laboratory coordination. Select the intended Spanish or English configuration before starting. Room names and configuration identifiers support administration; participant screens use task titles and rules, without internal production codes or treatment labels that reveal the hypothesis. Keep a record of the selected configuration so that exported results can be interpreted after the class. A screenshot of an administrative selection can be useful, but it is not a substitute for retaining the data file.

Checkpoint

Before students arrive, verify instructor access, four rooms from this family, the intended configuration and language, the participant count and a download from a fictitious session. Share the stable room with students and retain administrative links for the teaching team.

An individual link may stop working if its database has been reset: the session and participant must still exist in that database. Reopening the old address cannot recreate them. Create a new session and distribute its accesses. During teaching, do not reset databases or reuse links from earlier tests. Local previews also use a separate instance for each family; opening another family's address does not select a configuration within this one. Laboratory coordination manages public deployment and production credentials. A local successful test demonstrates the implementation, not that the proposed public hostname has already been deployed.

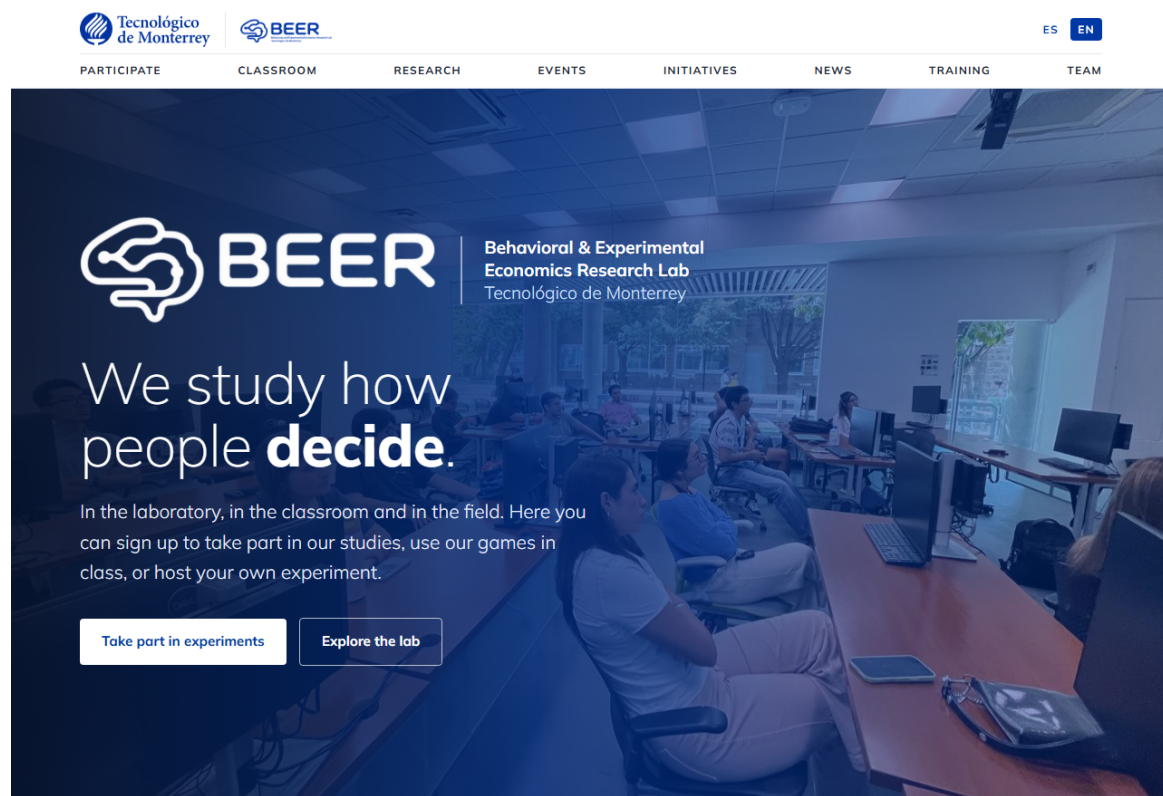


Figure 1.1: Laboratory website: CLASSROOM leads to teaching materials.

Tecnológico de Monterrey **BEER**

ES EN

PARTICIPATE CLASSROOM RESEARCH EVENTS INITIATIVES NEWS TRAINING TEAM

INSTRUCTORS AND STUDENTS

Classroom experiments

Experimental economics games for classroom activities and discussion. Open the game your instructor tells you.

Instructors: how to use these games in class or propose a new one → [instructor guide](#)
 Lab research: beer-lab.org/en/research

Search by title, topic, keyword or JEL code...

Topic ▾ JEL code ▾ 150 games

Electoral Accountability and Political Selection

Study how the prospect of remaining in office changes an officeholder's behavior and how information allows voters to choose between politicians with different types. Basic: One encounter or four encounters with fixed or reassigned groups. Extended: Eight treatments on continuation, information, types, office value, term limits, and audits.

electoral discipline political accountability political selection

corruption information political economy

JEL A22 · C92 · D72 · D73

Address: accountability.beer-lab.org

Aggregate Demand and Aggregate Supply

Understand how spending choices aggregate into demand and how equilibrium with supply determines output and prices. Basic: One period, ten periods with the same economy, or ten periods with reassigned economies. Extended: Demand expansion and contraction, supply shock, fiscal stimulus, capacity, expectations, and coordination.

aggregate demand aggregate supply price level potential output

macroeconomic expectations

JEL A22 · C92 · E12 · E31 · E32

Address: adas.beer-lab.org

Figure 1.2: Teaching directory: locate the family card and choose the intended language.

Demo

[ES] Construcción de modelos económicos — Básico — Entorno estable · 1 ciclo(s) · Datos suministrados	Construcción de modelos económicos / Economic Model Building — BEER Lab Seleccione una variante de este juego. Select a version of this game.
[ES] Construcción de modelos económicos — Básico — Entorno estable · 1 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Básico — Entorno estable · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Entorno estable · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Variación moderada en los resultados · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Variación alta en los resultados · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Relación no lineal · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Condición relevante · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Interacción entre variables · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Datos iniciales correlacionados · 3 ciclo(s) · Ensayos elegidos por el participante	
[ES] Construcción de modelos económicos — Extendido — Predicción fuera del rango inicial ·	

Figure 1.3: Isolated Demo: this excerpt contains model-building configurations.

Rooms

Current rooms:

[ES] Construcción de modelos económicos — Básico
[ES] Construcción de modelos económicos — Extendido
[EN] Economic Model Building — Basic
[EN] Economic Model Building — Extended

Figure 1.4: Rooms contains four rooms belonging to this family.

Room: [EN] Economic Model Building — Basic

Create a new session

Session Config

[EN] Economic Model Building — Basic — Stable environment · 1 cycle(s) · Supplied data

Number of participants

Individual decisions: any positive participant count is allowed.

Create

[Configure session](#) 

App sequence

model_building_game (3 rounds)

0 participants present

Start links

These URLs will stay constant for new sessions, even if the database is recreated.

Room-wide URL

Here is the room-wide URL anyone can use. Don't use this link when testing in multiple tabs on the same browser, because all tabs will be assigned to the same participant, using a cookie.

http://127.0.0.1:8188/room/models_en

Figure 1.5: English basic room: choose the protocol before creating a session.

Foundations and worked explanations

2.1 Variables, assumptions and functional form

A variable is a characteristic that can take different values. In the virtual glider, X is a setting chosen from the integers zero through ten. Z identifies condition zero or one. Y is the outcome to be predicted. Choosing X does not mean choosing Y directly: the system generates the latter through a rule that remains hidden during the task. This distinction introduces decision variables, environmental conditions and outcomes without assuming that students already know the terminology.

A functional form specifies how variables relate within a representation. The menu offers five possibilities: constant, linear, quadratic, two additive variables and an interaction. The constant is $\hat{Y} = a$. The linear formula is $\hat{Y} = a + bX$. The quadratic adds cX^2 . The additive formula is $\hat{Y} = a + bX + cZ$. The interaction adds dXZ to the additive formula. The hat denotes a prediction, not an observation. Participants choose the coefficients, each between -200 and 200 . The platform does not automatically estimate a regression or select a best-fitting specification.

The intercept a is the prediction when the remaining terms are zero. In the linear formula, b is the predicted change in Y when X rises by one. In the additive formula, c is the difference between conditions one and zero while holding X fixed. In the quadratic formula, c instead determines curvature. Thus the same coefficient name plays a different role depending on the selected formula. Before registering, students should read the complete expression and inspect the preview.

With an interaction, the difference between conditions is $c + dX$. There is no single effect of Z independent of X , unless $d = 0$. The marginal change associated with X also varies: it is b when $Z = 0$ and $b + d$ when $Z = 1$. This introduces heterogeneous effects without claiming that any observed difference in means establishes an interaction. The task requires relevant comparisons and a representation that generates concrete predictions. An instructor can demonstrate the difference between a statement such as “the condition matters” and a model specifying how much it matters at several settings.

2.2 Solving the stable reference

The artificial reference function is $Y = 20 + 4X$. Keep that information for the instructor until the debrief. Initial observations include $(X, Y) = (1, 24), (3, 32), (6, 44), (8, 52)$. Conditions alternate between zero and one but have no effect in this reference. From $X = 1$ to $X = 3$, the outcome rises by eight while the setting rises by two, giving slope $8/2 = 4$. Substituting the first observation yields $a = 24 - 4 = 20$. A linear model with those coefficients predicts outcomes exactly in the absence of noise. The calculation can be explained using differences before introducing regression terminology.

For $X = 5$, the prediction is $20 + 4(5) = 40$. A constant model equal to 40 matches that particular observation but fails elsewhere. At $X = 2$ it predicts 40 instead of 28; at $X = 9$ it still predicts 40 instead of 56. This example shows why matching an average or a central observation does not represent an entire relationship. It also separates the level of a function from its slope. Raising the intercept shifts all predictions by the same amount, whereas changing the slope alters differences between settings. Invite students to calculate both consequences rather than relying only on a diagram.

Instructors may solve the reference after collecting decisions or use a separate practice session to teach coefficient entry. Revealing the solution before the main exercise changes the activity into execution of a known rule. That may serve an introductory class, but the decisions can no longer be interpreted as spontaneous model search. Record additional oral information because it changes the meaning of responses even if the server configuration remains identical. The intended distinction is between teaching students how to operate the task and telling them which rule generates its data. The two interventions have different consequences for what can be learned from the resulting choices.

What participants will see

A minimal instruction is: “Observe what changes and what stays the same. Register a rule that calculates new outcomes. The platform evaluates predictions from that rule; you do not have to guess the name of a theory”. Do not announce the generating function before decisions.

2.3 Training, testing and revision

Training data comprise the four initial observations and all trials selected so far. They support construction and revision. Tests are three observations in the current cycle: their X and Z values appear when the model is registered, but their Y outcomes remain hidden. Evaluation calculates predictions from the registered formula and then reveals the outcomes. Students do not submit three unrelated forecasts. They submit one common rule that generates all three. That restriction makes internal consistency visible and distinguishes a model from a collection of disconnected guesses, even when both could be compared with the same realized outcomes.

The test design does not repeat as a set over the three cycles. Revealed outcomes are not automatically added to training, although participants can remember them and learn from them. Later tests are new observations, but not an external assessment insulated from all prior feedback. Moreover, in the active protocol a participant can choose trials matching test conditions. This is an allowed information strategy, not a system leak. Do not present results as a forecasting tournament with an untouchable secret validation set.

Revision is informative when linked to identifiable evidence. Changing every coefficient after a disappointing score does not establish that the cause of failure was recognized. Ask students which observation motivated the change and which prediction differs from their earlier rule. The export retains formulas, coefficients and tests by cycle so this comparison can be reconstructed. A lower error can also reflect an easier set of new tests. The trajectory is descriptive, not a causal measure of learning without an additional design.

2.4 Error and points, step by step

For each test, error is $e_j = \hat{Y}_j - Y_j$. A positive value means overprediction; a negative value means underprediction. The summary is

$$\text{RMSE} = \sqrt{\frac{e_1^2 + e_2^2 + e_3^2}{3}}.$$

Squaring prevents positive and negative errors from cancelling. Dividing by three gives the mean squared error. Taking its square root returns the measure to the original units of Y . The denominator is three tests within an individual cycle, not three students. Training error applies the same operation to all available training observations and can have a different denominator. Spell out each operation when participants are unfamiliar with the statistic rather than introducing it solely through its acronym.

If all errors equal five, RMSE equals five. Gross score is $\max\{0, 100 - 2(5)\} = 90$ ECU. If errors are zero, zero and fifteen, RMSE is $\sqrt{225/3} = 8.66$ and gross score is about 82.68 ECU. Although mean absolute error is five in both examples, the second earns fewer points because it concentrates a large mistake. The scoring rule therefore illustrates the consequences of choosing a loss function. It should not be described as the only correct way to value an economic forecast.

Net score subtracts chosen-trial costs and, when applicable, additional-parameter costs. A lower bound of zero applies again and the result is rounded to two decimal places. With gross score 90 and two trials costing three ECU each, net score is 84. Ordinary versions set both costs to zero. Total score sums active cycles, so the maximum is 100 for a short session and 300 for three cycles. Do not compare those totals without accounting for exposure. Distinguish a small error, a high gross score and a high net score; these coincide only when the cost structure permits it.

2.5 Designing an informative comparison

Choosing trials means deciding which evidence to obtain. Two observations with the same X and different Z help separate the condition from the setting. Two with the same Z and different X help examine the setting. Repeating exactly the same combination can check outcome variation when noise is present. Exploring an extreme may distinguish a line from a curve that agrees near the centre. None is universally superior: information value depends on which hypotheses remain plausible and the cost of obtaining an observation. An instructor should ask for that reasoning before judging a chosen experiment by its realized score alone.

Consider $20 + 4X$ and $20 + 8X - 0.6X^2$. At $X = 0$ both predict 20; at $X = 10$ they predict 60 and 40. A trial at zero does not distinguish them in the absence of other factors, whereas one at ten can. Asking for both predictions before choosing a trial turns “I will try something” into an explicit design argument. The practice can transfer to demand, production or policy evaluation without implying that this small simulator reproduces the identification problems of those settings. Students should identify both the shared reasoning and the additional obstacles that would arise in a real application.

Attention

Correlated initial observations do not, by themselves, turn an estimated slope into a causal effect. The simulator permits additional controlled conditions, while economic data can involve selection, constraints and equilibrium mechanisms absent here. Methodological transfer requires naming these differences.

Preparing and conducting the session

3.1 Decisions before class

First decide whether the goal is reading a relationship, designing information or revising a representation. For reading, choose supplied data: a small table and one model submission leave most of the class for interpreting coefficients. For design, choose one-cycle exploration: students add zero, one or two trials before registering their rule. For revision, choose three cycles: selected observations accumulate and the formula can change after feedback. The three basic protocols share the stable function, allowing operational complexity to change without introducing a strategic interaction. Their differences should be explained as information and revision opportunities, rather than as separate games with unrelated rules.

Prepare a test session with fictitious participants. Visit the welcome, instructions, comprehension and model form. Deliberately submit an incorrect comprehension answer to verify that an explanation appears beside that question and other answers remain available. The scoring explanation reconstructs $100 - 2 \times 5 = 90$; the trial explanation identifies the maximum in the selected protocol; the timing explanation reminds participants that registration precedes revelation. Corrections do not deduct points. This check should establish operational understanding rather than become an examination in statistics. If students cannot yet manipulate a square root, explain what the displayed error means and work through the simple equal-error example.

Decide whether calculators or a worksheet are permitted. The platform calculates predictions from coefficients, but does not discover the coefficients for students. For beginners, provide a blank formula, a column for substituting X and another for comparing with Y . If spreadsheets are allowed, apply the same rule to the whole group and record it. The menu does not require regression estimation: differences can identify a line and simple comparisons can justify revision. Additional tools change how the task is solved, so their availability belongs in the session record alongside configuration and oral instructions.

Plan how individual questions will be handled. A participant stuck in comprehension may have confused “before costs” with net score; another may have confused cycles with trials. Clarify the particular rule without revealing the generating function. Maintain quiet and prevent a solution from circulating between stations. If the activity is used collaboratively, run that as a separate exercise and declare the modification. The ordinary implementation records individual decisions, so its observations should not be described as team responses. Collaboration can be a valuable teaching choice, but it changes the unit of interpretation and should be intentional.

3.2 What participants see

The welcome presents the laboratory logo, institutional purpose and work rules. Participants confirm that they have read them. Instructions identify variables, resources and scoring. The resource box is in the body of the page, not merely a small corner display. It distinguishes initial observations, additional trials and possible points. Comprehension follows, and the selected protocol begins only after correct answers. There is no waiting page because decisions do not depend on another person. Instructors can monitor progress privately without asking faster participants to announce their formula or explain it publicly before others finish.

During exploration, the table displays available data and each row's source. Participants specify how many trials they want and complete the relevant fields. Zero trials means other fields are ignored; one means only the first is executed. Repeated or different conditions are allowed. Unused fields incur no costs. A chosen trial needs an integer from zero to ten for X and a valid value of Z . A field error can be corrected while retaining other entries. Ask students to check the chosen quantity before submitting: filling two combinations does not override a quantity of one.

The registration screen displays the table again, including newly executed trials. Participants choose a form and the coefficients used by it. Unused coefficients do not affect predictions or points. The preview shows three test conditions and calculated predictions; outcomes are still hidden. Students should check that the rule expresses their intention before continuing. Switching formulas can change what coefficient c means. A useful intervention is to ask them to read the complete expression silently before submission, while avoiding advice about which expression is preferable. This supports comprehension without supplying the answer to the modeling problem.

Cycle results present the formula, effective coefficients, both errors, costs and points. A table reconstructs the three test errors. Continuing opens the next cycle or the final screen, depending on the horizon. The final screen explicitly confirms completion and shows total and cycle scores. Repeated clicking is unnecessary: students should wait for the instructor's discussion. The instructor chooses when to reveal aggregates from the private report. A participant's result is available individually before the aggregate debrief, so ask students to retain their explanation rather than immediately compare scores with neighbours.

For the instructor

Before the first submission, remind students that they register one common formula for three tests. Do not recommend a formula. After submission, use the result table to clarify calculations; do not retrospectively change a decision because it produced a large error.

3.3 Common operational problems

If a participant cannot continue, inspect the marked field first. Decimals may appear with a comma in Spanish and a point in English; the preview recognizes that presentation. A required coefficient left blank is not zero. An unused coefficient may remain blank without changing the calculation. If a screen requests partners or an even participant count, it does not correspond to this individual activity. Check the family and configuration rather than artificially changing class size to satisfy a rule from another game. A participant-count error is not solved by introducing a partner who has no role in the intended protocol.

If the report remains locked, refresh it and examine pending cycles. Aggregate revelation requires every participant to complete the planned horizon. Someone who has left may prevent that condition; the lock does not mean completed decisions are absent. The private download still includes partial records. For an early discussion, use the export according to laboratory policy

without projecting individual codes. There is no automatic exclusion of slower participants. If a session must end early, document that fact and retain incomplete records rather than treating them as zero-score observations or silently dropping them without explanation.

The following images show distinct states. Form screenshots focus on relevant panels to preserve legibility. Numbers come from fictitious sessions, not empirical results from students. English participant images were captured from an English session. They are not Spanish screens with translated captions. Both languages share the template, mechanism and graphs. Administrative pages may retain oTree's common navigation labels, whereas participant-facing rules, field feedback and results follow the session language. This difference should not be mistaken for a partially translated decision interface.

Create session

Tip: consider creating your session in a [Room](#), for more stable start links that are not opened accidentally.

Session Config

[ES] Construcción de modelos económicos — Básico — Entorno estable · 3 ciclo(s) · Ensayos elegidos por el participante

Number of participants

3

Decisiones individuales: se admite cualquier número positivo de participantes.

Create

[Configure session](#) 

App sequence

model_building_game (3 rounds)

Figure 3.1: Session creation: choose the protocol and number of people.



Welcome to BEER Lab

The Behavioral and Experimental Economics Research Lab in the Department of Economics at Tecnológico de Monterrey is a space for learning about economic decision-making through controlled activities.

Before you begin

- Remain quiet and do not communicate with other participants.
- Work individually and privately; your results do not depend on other people.
- Follow the instructions on screen. Do not refresh, close or leave the page.
- If you need help, raise your hand and ask the staff privately.

☐ I have read these instructions and agree to follow them.

Continue

Figure 3.2: Welcome: students read and confirm the laboratory rules.

Resources and score

You receive four initial observations and up to two additional trials per cycle. You may choose zero, one or two trials. There is no initial monetary endowment and you do not need another person.

Each chosen trial costs 0 ECU. Each parameter beyond two costs 0 ECU per cycle. Zero coefficients still count if your formula includes them.

ECU means Experimental Currency Units. Here they express performance points; no conversion into real money has been established.

Planned cycles: **3**

Figure 3.3: Resources: four initial observations, permitted trials and the meaning of ECU.

What is the maximum number of additional trials you may choose per cycle?

If all three errors are 5, how many ECU do you earn before costs?

Your answer: 95. The correct answer is 90. $RMSE = \sqrt{((25 + 25 + 25)/3)} = 5$; $100 - 2 \times 5 = 90$ ECU before costs.

When can you see the three test outcomes?

☒ After registering the model

☐ Before registering the model

Figure 3.4: Comprehension: feedback explains the incorrect calculation and retains other answers.

Number of additional trials in this cycle (0, 1 or 2)

First trial

X value (integer from 0 to 10)

Condition Z

☐ 0

☐ 1

Second trial

X value (integer from 0 to 10)

Condition Z

☐ 0

☐ 1

Figure 3.5: Exploration: the chosen quantity determines which trials are executed.

Formula

Linear: $Y = a + bX$ ▼

Coefficient A (from -200 to 200) Coefficient B (from -200 to 200)

25 4

Coefficient C (from -200 to 200) Coefficient D (from -200 to 200)

0 0

Predictions to be registered

Outcome Y stays hidden until you continue.

X	Z	Prediction
2	1	33
5	0	45
9	1	61

Figure 3.6: Registration: the preview shows predictions while test outcomes remain hidden.

Linear: $Y = a + bX$

$a = 25, b = 4, c = 0, d = 0$

Training error (RMSE)	5
New-test error (RMSE)	5
Points before costs	90 ECU
Trial cost	0 ECU
Parameter cost	0 ECU
Net score	90 ECU

Predictions to be registered

X	Z	Prediction	Y	Prediction - outcome
2	1	33	28	5
5	0	45	40	5
9	1	61	56	5

Figure 3.7: Individual result: compare errors, costs and predictions with revealed outcomes.

Activity completed

Your responses have been recorded. Wait for the instructor's directions for the discussion. This screen marks the end of the activity.

Total score: 290 ECU

Cycle	ECU
1	90
2	100
3	100

Figure 3.8: Explicit completion with total and cycle scores.

Extensions and teaching comparisons

4.1 Selecting a comparison

Thirteen extensions retain the common access and registration procedure while changing an identifiable component of the problem. There is no need to use all of them in a course or produce a separate guide for every treatment. Select a reference and an extension that address the class question. Extensions use three cycles and active trials, so the appropriate basic reference is three-cycle revision when horizon and protocol should remain constant. Comparing a three-cycle extension with one-cycle observation confounds information, practice and treatment. The apparent simplicity of a menu comparison does not remove that design issue.

The mechanisms below are BEER Lab adaptations for teaching construction and evaluation of representations. The guide's methodological sources provide general motivation; none of these is an exact replication of a published paper-airplane treatment. Predictions describe the artificial system, not estimated regularities of student behaviour. A choice compatible with an explanation does not establish that the participant used that reasoning. Ask for arguments and treat class observations as the beginning of analysis. Where a mechanism is shared between two configurations, say so rather than presenting a new name as independent experimental variation.

4.2 Moderate and high outcome variation

The central relationship remains $20 + 4X$, with a uniform disturbance between -3 and 3 , or between -10 and 10 , depending on configuration. Outcomes are rounded to two decimals. Repeating a trial can produce different values. The conditional mean is the reference prediction; even a formula with the central coefficients usually has nonzero realized error. The aim is to distinguish temporary variation from a changed relationship. One observation off the line is not automatic proof that the representation is useless. Conversely, a good fit to a noisy observation does not prove that a new coefficient captures a stable mechanism.

Examine the RMSE histogram, prediction-versus-outcome scatter and coefficient revisions. Greater dispersion is compatible with more noise, but a small sample may reverse the ordering between sessions. Ask whether repeating conditions or changing every element would better assess stability. Repetition better isolates variation under unchanged conditions. Class evidence illustrates learning from imperfect observations; it does not identify subjective beliefs or estimate uncertainty aversion. This extension is especially useful for explaining why a correct central rule can make an inaccurate individual prediction and why evaluating it requires considering more than one realized outcome.

4.3 A nonlinear relationship

The function becomes $Y = 20 + 8X - 0.6X^2$. Its negative quadratic coefficient makes the increment associated with another unit of X decline with the setting. A line can approximate one region and fail elsewhere. For example, outcomes are 20 at $X = 0$, 45 at $X = 5$ and 40 at $X = 10$. The outcome does not rise indefinitely. The available quadratic form can represent the relationship exactly without noise. Students can therefore distinguish choosing an adequate form from choosing adequate numerical coefficients; both are needed for a successful representation.

Ask students to compare a line and a curve at the same setting before seeing the outcome. In the report, look for substantial test error despite apparently reasonable initial fit and note when the selected form changes. Moving to a quadratic is not automatically an improvement: coefficients must also be suitable. Remaining with a line need not imply failure to recognize curvature; it may be a deliberate approximation. Connect domain of application, simplicity and consequences of mistakes rather than turning the formula's name into a universal correct answer. A model is assessed relative to the question and the range where it is used.

4.4 A relevant condition

The function is $Y = 20 + 4X + 12Z$. Holding the setting fixed, changing the condition from zero to one raises the outcome by twelve. A line in X omits this difference, while the additive form can represent it. Comparing observations that differ in both X and Z does not directly isolate either component. Chosen trials make it possible to hold one fixed and change the other. That operation is the teaching focus, rather than memorizing the expression "omitted variable". Encourage students to articulate the comparison in ordinary language before translating it into coefficients.

At $X = 5$, the prediction is 40 in condition zero and 52 in condition one. Ask what evidence someone who attributes all differences to the setting would need. Examine inclusion of Z , its coefficient and test error; a count of model names alone does not show whether the effect was understood. An interaction model can also be exact with a zero additional term. This illustrates how one representation can contain another as a special case. It does not establish that adding terms always improves learning or that a larger specification is always justified by the data available.

4.5 Interaction between variables

Now $Y = 20 + 2X + 12Z + 3XZ$. The condition changes both level and response to the setting. With $Z = 0$ the slope is two; with $Z = 1$ it is five. At $X = 2$, the difference between conditions is eighteen; at $X = 8$, it is thirty-six. A fixed effect of twelve cannot capture both. This extension explains why a comparison made at one setting is insufficient to establish a homogeneous effect. Students should identify the level at which a statement about a condition is being made and calculate how that statement changes elsewhere.

Propose four observations: two settings in each condition. That structure compares slopes. Participants have limited additional trials and must use initial observations carefully. Examine interaction choices in the report and coefficients in the export, then ask for a prediction at an unexamined combination. A flexible formula's good performance is compatible with recognizing interaction; it does not establish competence in identifying causal interactions from observational economic data. The simulator directly controls relevant conditions, an advantage often absent in empirical work. Make this difference explicit when connecting the exercise to policies whose effects vary across populations.

4.6 Correlated initial data

The true rule is additive, but high initial X values appear with $Z = 1$ and low values with $Z = 0$. That association makes it tempting to assign to one variable part of a difference related to the other. There is no perfect mathematical identity between them; the problem is an initially unbalanced comparison. Additional trials can break the association by choosing a low setting in condition one or a high setting in condition zero. The simulator allows that intervention directly, without the practical restrictions that might prevent it in a real market or policy setting.

Ask which observation would distinguish two explanations that describe the initial table similarly. Examine chosen trials, not just average points. Lower test error after crossed conditions is consistent with better separation of components. Nevertheless, this is not a policy experiment with random assignment among citizens. Distinguish control inside the simulator, voluntary selection of trials and causal identification in an external population. Hypothesis-search references motivate the discussion, not an individual bias diagnosis. A useful written response states both the competing explanations and the outcome each predicts for the proposed trial.

4.7 Prediction outside the initial range

This configuration uses the quadratic relationship but restricts initial observations to settings one through four. First-cycle tests use values above four, and subsequent cycles evaluate other combinations in the high range. A line that approximates the beginning may fail when transported elsewhere. Participants may choose trials anywhere from zero to ten. Extrapolation therefore refers to the initial range: it ceases to be extrapolation relative to all available training data if a participant explores higher settings before registering the model. This distinction matters for interpreting both errors and sensible information choices.

Do not label every test error an extrapolation failure without examining selected trials. Ask for the smallest and largest setting observed before registration and locate tests inside or outside that interval. The accumulated table in the export reconstructs this information. Discuss what evidence an authority should collect before applying a policy at an unprecedented scale. The economic connection concerns transporting predictions, not estimating policy effects with a virtual aircraft. A student who explores the evaluation range may avoid the problem intentionally; that is evidence about design choice rather than a reason to reclassify the implementation as broken.

4.8 An environment change

The stable reference receives an upward shift of fifteen units from the second cycle onward. The program announces an environment update without revealing its magnitude. Earlier observations remain visible with cycle labels, while new trials use the current environment. A model with intercept 20 and slope four was exact in cycle one; afterwards an intercept of 35 predicts the new central relationship. The slope is unchanged. Accumulated training data therefore contain observations from two regimes. Students must consider the relevance of an observation's date, not merely its position in a table.

Training RMSE pools both regimes, whereas tests assess only the current one. An updated model can predict well and still fit older data poorly. Do not call this overfitting without examining temporal composition. Ask which part of the rule should change and which might be retained. The update is announced, so the task does not reproduce an unexpected break or measure unaided change detection. Its value lies in discussing stability of assumptions and selective revision. Instructors know the programmed shift, but should not use that private knowledge to judge a

plausible student conjecture as unreasonable before enough evidence was available.

4.9 A condition without an effect

This extension retains $Y = 20 + 4X$ and makes the irrelevant-condition question explicit in instructor selection. Mechanically it matches the reference: the label introduces neither an additional effect nor a distinct sample merely by being different. Its purpose is curricular, organizing a comparison with a relevant condition or an interaction. It should not be treated as an independent intervention that changes outcomes. Participant screens do not announce that Z has no effect. The hypothesis remains something that can be examined through observations rather than something supplied as an instruction.

Inspect nonzero condition coefficients and whether they change during revision. An initial coincidence may support a conjecture that later evidence corrects. Without noise, suitable comparisons reveal unchanged outcomes when only Z varies. Do not conclude that a redundant variable always worsens predictions: a zero coefficient gives the same function. Equivalent representation and costly complexity are different issues, separated in the next extension. The methodological lesson is to justify variables by their function rather than their availability in a dataset. A thoughtful explanation can acknowledge that an apparently irrelevant condition may matter in an untested environment.

4.10 A complexity cost

Moderate noise is added along with a cost of two ECU per parameter beyond two. A constant uses one, a line two, quadratic and additive forms three, and an interaction four. Costs are therefore zero, zero, two, two and four ECU. A parameter counts even if its submitted coefficient is zero. The charge concerns the registered structure rather than the number of nonzero coefficients. It is announced before decisions and is not AIC, BIC or an estimated statistical criterion. Explain the convention directly so students do not infer a statistical procedure that the program does not perform.

Compare this treatment with moderate outcome variation, which retains the same noise without charging for parameters. A more complex formula must reduce error sufficiently to offset its cost. Reducing RMSE by two units improves gross points by four; an interaction relative to a line consumes the entire difference through its four-point charge. Compare error and net score because a lower score need not mean a worse prediction. This extension supports discussion of parsimony and objectives, but the penalty is a teaching convention. The resulting ranking does not prove that a true model should always use the fewest parameters. Rather, it makes explicit how a choice depends on both predictive performance and a stated cost of complexity.

4.11 An experimentation cost

Moderate noise is added and each selected trial costs three ECU. Participants may choose zero, one or two, costing zero, three or six. Initial observations are free. Costs apply separately in each cycle and reduce its gross points, subject to a zero lower bound. There is no borrowing, debt or transferable budget. Use moderate outcome variation as the comparison: it retains the same noise without trial costs. The noiseless reference is not an otherwise identical control. Identify any other configuration difference. The fact that a treatment has an informative title does not make all of its differences disappear from an empirical comparison.

An RMSE improvement of one unit is worth two gross points. A three-ECU trial must generate enough expected improvement to compensate; its realized outcome cannot reconstruct the participant's initial expectation. Ask what uncertainty each trial was intended to resolve and whether repetition supplied something different from a new setting. The software does not calculate the expected information value implied by subjective beliefs. The report shows decisions and performance; economic interpretation separates known cost, uncertain benefit and retrospective evaluation. A trial that turned out unhelpful may still have been a defensible choice when the uncertainty was unresolved.

4.12 Demand and production applications

In demand, X is price per unit, Z denotes absence or presence of a campaign and Y is quantity demanded. The artificial function is $100 - 4X + 12Z$. Holding the campaign fixed, one more ECU of price reduces demand by four units; activating the campaign shifts the relationship upward by twelve. The outcome is not profit and participants do not choose a profit-maximizing price. They construct a rule for predicting quantities. This distinction prevents the extension from being described as an incomplete version of the monopoly game already available elsewhere in the teaching platform.

In production, X is input units and Y is output, with $20 + 8X - 0.6X^2$. The relationship permits discussion of diminishing marginal product and, at the high end, a negative effect of additional input. No output price, economic input cost or firm optimization problem is programmed. Ask what extra data are required to move from a production relationship to a profit choice. Both extensions explicitly transfer a methodological tool to economic content. Their numbers are artificial designs rather than estimates of real markets or technologies, so claims about actual elasticities or productive capacity would go beyond the exercise.

Literature and the adaptation's scope

Geerling, Mateer and O'Roark use paper airplanes to teach model construction and evolution. This implementation retains comparison between representations and outcomes through an original digital setting with artificial functions and explicit scoring. It does not replicate the physical experiment or attribute its thirteen extensions to that article. Klahr and Dunbar examine search across hypotheses and experiments, a useful perspective for discussing informative choices. Box provides a framework for revision and parsimony. These sources motivate methodological aspects; the platform's numerical examples and specific mechanisms are original teaching adaptations.

The class is not a replication of published learning outcomes. Improving scores do not establish the general effectiveness of a pedagogy; a failure to improve does not invalidate economic modeling. Evaluating learning would require appropriate before-and-after measures, a comparison design and a record of instruction. The activity instead supplies decisions, errors and arguments for substantive discussion. Its immediate value is that students can identify what they represented, what they assumed and what evidence would change their view. This is a concrete objective that can be assessed without making claims the classroom design cannot support.

Reading the report and using the data

5.1 Revelation and denominators

The private report begins with participants, completed individual cycles and pending cycles. A session with three people and three cycles expects nine records. If one person has finished and the others have not started, three are complete and six pending; this is not a class of nine students. Aggregate revelation becomes available when all planned cycles are complete. Refreshing checks progress without changing decisions. Hide conceals aggregates again on the screen; it does not delete data or reverse a revelation students have already seen. Explain that distinction when operating a shared classroom projector.

The central unit is person-cycle. Overall mean RMSE pools completed cycles with equal weight. In a three-cycle session, one person can contribute three related observations. That sample size is not a count of independent people. Each individual RMSE summarizes three tests; training error summarizes a quantity that can grow with chosen trials. The report also presents results separately by cycle, each with its own denominator. Keep these units distinct when making statements about the class. A precise sentence identifies both what was averaged and how many completed records supplied that average.

The histogram groups test error into zero, greater than zero through two, greater than two through five, greater than five through ten, and greater than ten. Bars show counts of individual cycles, not percentages or probability densities. Their sum should match completed cycles. Intervals differ in width and the last is open-ended, so bar area is not a continuous probability. The grouping makes concentration and dispersion easy to discuss; use the export for other cutoffs. A low average can coexist with a long tail of large mistakes, which is why the distribution contributes information beyond the headline mean.

5.2 Formulas, trajectories and scatter

The model table counts forms selected in each cycle. Rows display completed records and their distribution across five possibilities. It does not count true models or the percentage who understand a theory. Different formulas can produce identical predictions when some coefficients are zero. Conversely, two participants selecting linear can have very different errors. Combine this table with coefficients and outcomes rather than treating a menu name as an automatic sophistication measure. A useful debrief asks what changed in a model's predictions when its form changed, not merely whether its label became more complex.

The trajectory has training error in grey and test error in blue. The horizontal axis is cycle, not clock time. The vertical axis uses Y units. A table provides values and counts by cycle. A gap invites questions about data composition, noise and environment. In the environment-change

treatment, training pools old and current data, so a reversal of the usual gap is not a plotting error. Explain composition before suggesting a mechanism. Also avoid interpreting a line segment as continuous learning between submissions: the graph connects a small number of discrete recorded states for readability.

The scatter places observed outcome horizontally and prediction vertically. The diagonal represents equality. Points above it overpredict and points below it underpredict. Each point is one of three tests from a completed individual cycle. Nine completed cycles yield 27 points, though some may overlap. Transparency helps reveal concentration but does not allow every coincident point to be counted visually. No individual codes appear in the graph. An apparent slope different from one can suggest compressed or expanded predictions without identifying the reason. To examine a particular model, return to the exported formula rather than guessing its coefficients from the cloud.

The simulator reference appears only in the revealed instructor block. It shows the central function and disturbance range, plus the shift where applicable. It supports calculation and feedback after decisions; do not project it early. The report neither automatically combines sessions nor estimates causal differences between treatments. Comparing groups requires identified session exports and documentation of what changed, how participants were assigned and what each group was told. Descriptive classroom comparisons can still be useful, provided their limitations are explicit and differences are not attributed to one mechanism without considering the others.

5.3 Download, privacy and reproduction

The report button downloads a CSV for the selected session. It is long format: one row per participant and active cycle, including pending records. Columns contain language, edition, treatment, protocol, completion status, trials, form, effective coefficients, errors, costs and points. Two JSON columns retain training observations and tests. These reconstruct what information was available and what outcome was evaluated. Identifiers support longitudinal ordering; they should not appear in slides beside particular decisions. Retain the original file before making classroom summaries so an accidental spreadsheet edit does not become the only surviving copy.

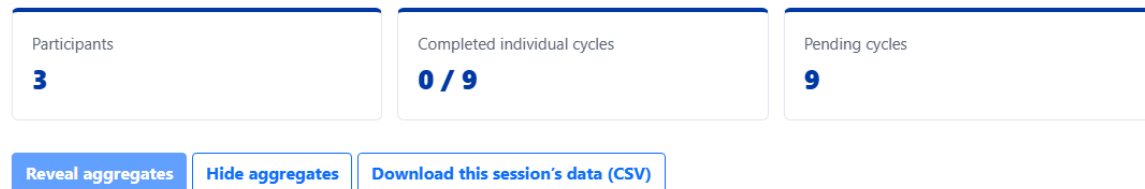
Missing values differ from zeros. An observed zero score is valid; an empty score means the model is incomplete. With supplied data, trial count may be blank because that page is not used; protocol distinguishes it from an active choice of zero trials. Calculate means only over completed records and state the criterion. Do not replace every missing value with zero when opening the file in a spreadsheet. That would distort both averages and progress. Similarly, a coefficient recorded as effectively zero may have been unused by the selected form rather than deliberately estimated as zero; interpretation requires the form as well as the numerical column.

The session Data page also offers the native wide file, one row per participant with columns for three technical periods. In a one-cycle session, later periods exist structurally but are not active decisions. The report and custom CSV traverse only configured cycles. The global export page may combine several sessions; distinguish it from the session-specific link. Keep configuration, date, oral instructions and original data before producing an anonymized teaching version. These records are especially important if a colleague later wants to compare a supplied-data session with active exploration or to reproduce an example from the guide.

BEER LAB

Instructor report: model building

Stable environment · Participant-selected trials



Aggregates become available once everyone completes the planned cycles. Refresh the report to check progress. The download also includes pending records.

Long file for the selected session: one row per person and active cycle; includes data and tests in JSON, coefficients, costs and status. Empty fields indicate pending responses, not zeros. Download the native wide file from this session's Data page.

Figure 5.1: Initial report: progress and download are available while aggregates remain locked.

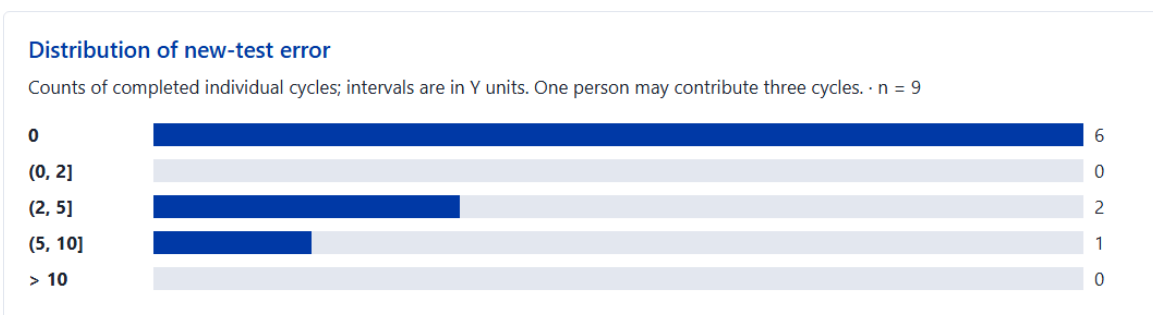


Figure 5.2: Histogram: counts of individual cycles grouped by new-test error.

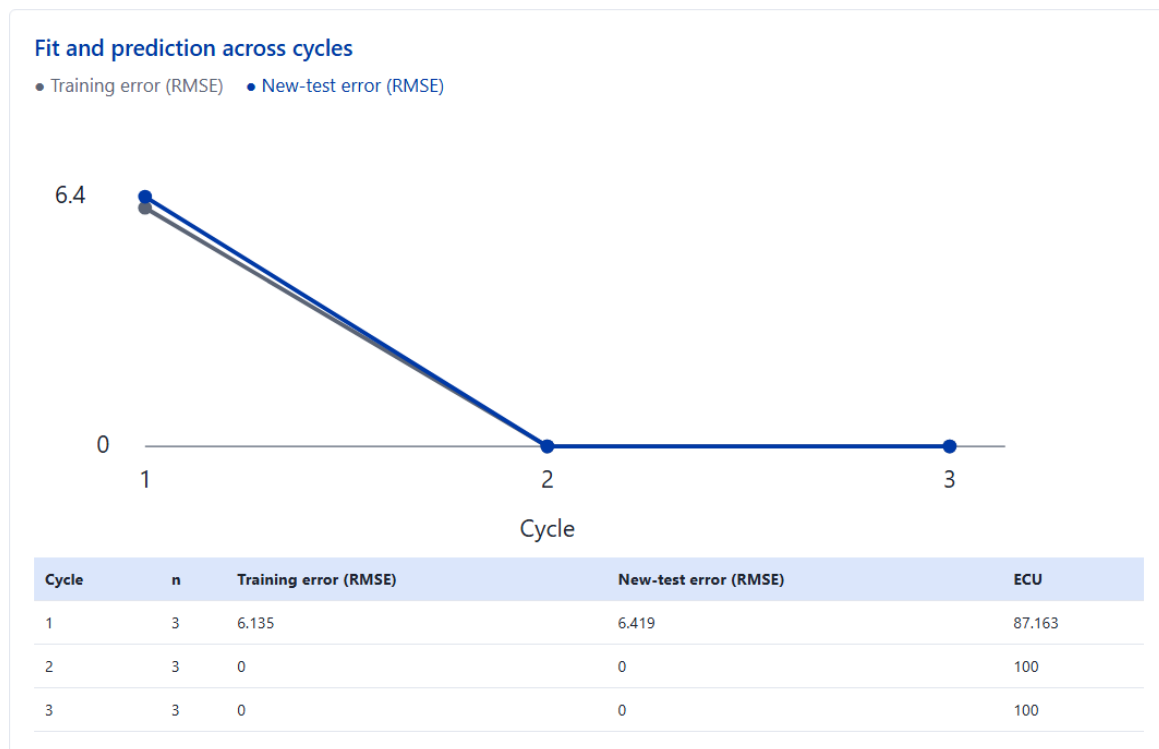


Figure 5.3: Trajectory: distinguish accumulated training data from the current cycle's tests.

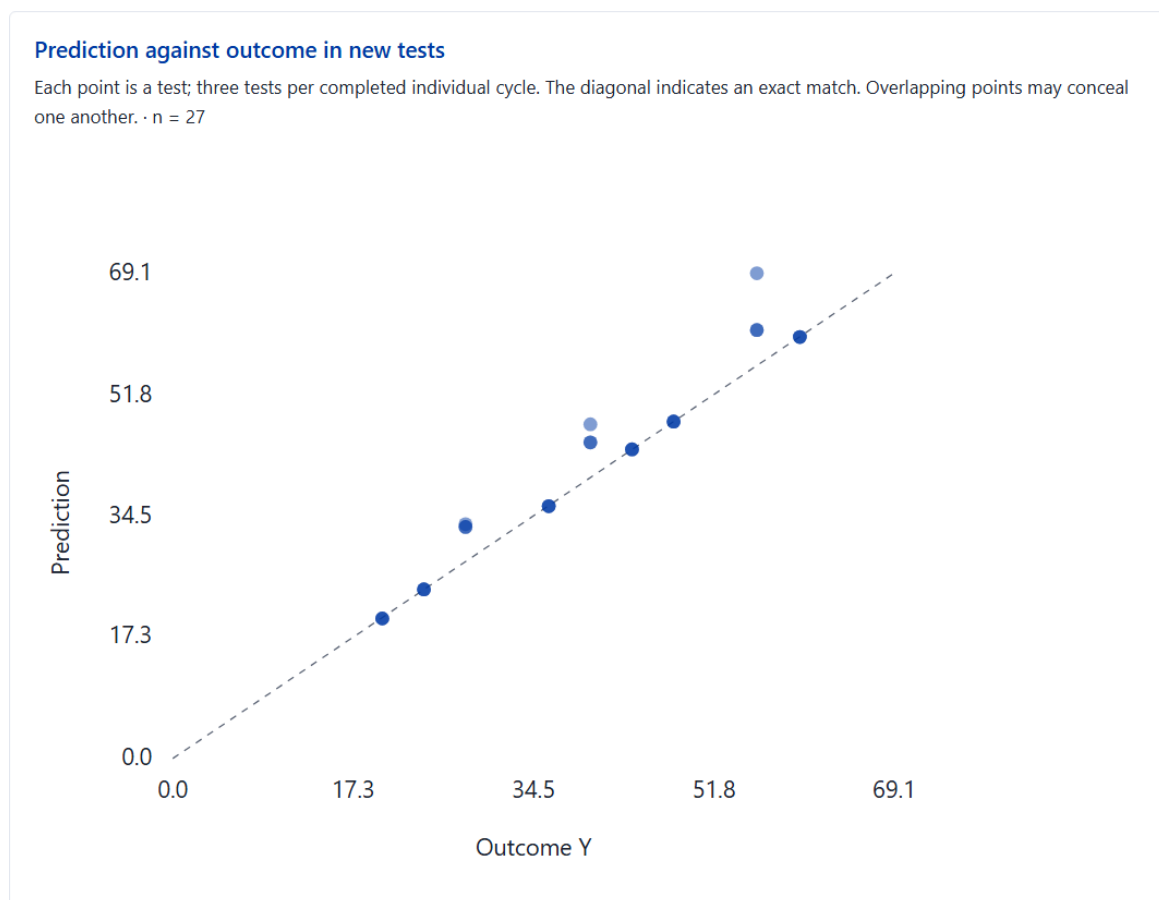


Figure 5.4: Scatter: distance from the diagonal indicates the direction and size of errors.

Class plans, exercises and closing discussion

6.1 A short session and a revision session

For 30 minutes, allow five for variables and access, five for instructions and comprehension, seven for registration and thirteen for discussion. Choose supplied data or one-cycle exploration. Begin feedback with two concrete predictions and their calculations before introducing methodological categories. A short class is better served by a precise conclusion about slope or validation than a hurried list of every problem in modeling. If comprehension takes longer, reduce the extension comparison rather than omitting the meaning of ECU or what outcome is evaluated. Operational clarity is necessary for the substantive discussion to make sense.

For 55 minutes, allocate eight to preparation, twelve to the first cycle, ten to later cycles, fifteen to the report and ten to transfer. Use revision as the reference and reserve an extension for another session or comparison group. Repetition makes it possible to ask what was retained and why. Before revealing the true function, request two written sentences: the assumption that mattered most and the additional observation the student would seek. This small record preserves uncertainty that can disappear in retrospective explanations. It also gives quieter students a concrete contribution to the debrief without requiring public disclosure of their individual scores.

Feedback script

“First describe the result: how many cycles are represented and where are the errors? Then reconstruct one prediction. Which rule produced it? What evidence supports or challenges it? Finally identify what you would change and where you expect the rule to remain useful. Keep three tasks separate: description, interpretation and proposing another test”.

Do not make score ranking the centre of the class. A sensible information choice can turn out badly under noise, and an unexplained coincidence can earn a good score. Ask for verifiable arguments and construct a comparison using actual observations. Precision is one dimension of quality; mechanism, coherent assumptions and domain are others. Explain how an incomplete representation can still be useful and how a detailed representation may answer a particular question poorly. Students should leave with a procedure for assessing claims, rather than the impression that modeling means choosing whichever equation most recently earned the largest number.

6.2 Exercises with answer guidance

The exercises below separate calculation, information design and interpretation. Work through one on the board and assign another in writing. Request the operation used, a claim supported by the result and one limitation. Solutions guide the debrief; they should not be revealed before responses are collected.

A line and an additional observation

Give $(2, 28)$ and $(7, 48)$ under a fixed condition and request a prediction for $X = 10$. Slope is $(48 - 28)/(7 - 2) = 4$, intercept is $28 - 4(2) = 20$, and the prediction is 60. Ask what remains unproved: continued linearity beyond the observed range and irrelevance of the condition. A complete answer separates calculation within a chosen form from justification of that form. Request one trial confronting extrapolation and another examining the condition. They need not be the same trial because the two uncertainties concern different aspects of the representation.

Errors that cancel in an average

Suppose errors are $-6, 0, 6$. Mean signed error is zero, but RMSE is $\sqrt{72/3} = \sqrt{24} \simeq 4.90$, giving about 90.20 gross ECU. Balanced overprediction and underprediction are not accuracy. Draw the points relative to the diagonal and connect visual distances to squared errors. If someone obtains zero by summing first, identify the changed operation. This recovers the scoring rule without criticizing the person. An extension asks whether zero mean error would nevertheless matter for a different objective and why that objective would require a different interpretation.

Separating condition and setting

Two observations are $(X, Z, Y) = (2, 0, 28)$ and $(8, 1, 64)$. The difference is 36 but two elements change. Request another trial and a reason. Choosing $(2, 1)$ compares conditions at the same setting; the additive function gives 40, twelve above the first outcome. Choosing $(8, 0)$ also works and gives 52. Dividing 36 by six and assigning the whole difference to X is not justified. Explain which components were mixed in the original comparison and how the new observation separates them. The answer should specify the counterfactual comparison rather than merely naming a relevant-variable problem.

Updating without discarding everything

A model $20 + 4X$ was exact, then the result at $X = 5$ rises from 40 to 55. An announced update is compatible with intercept 35. One point does not establish that the slope remains four; another setting helps test that proposition. Ask what evidence distinguishes a parallel shift from a changed slope. Instructors know the programmed change, but participants must infer its magnitude. Do not use reserved knowledge to describe an alternative conjecture as irrational while it remains consistent with limited evidence. This exercise teaches selective revision and the difference between the designer's information and the decision-maker's information.

Complexity and the objective

A line has RMSE four and an interaction RMSE three. Without costs, their gross points are 92 and 94. With the complexity penalty, the line retains 92 while the interaction receives 90 because

it uses four parameters. The lowest-error representation does not have the highest net score. Ask which objective is being optimized and what changes when the penalty is zero. The solution recognizes an explicit trade-off, not a proof that every model-selection criterion should penalize parameters this way. It should also distinguish the structural parameter count from the number of nonzero submitted coefficients.

Transfer to an economic decision

In demand, at price five with a campaign, quantity is $100 - 4(5) + 12 = 92$. Does this calculation recommend the profit-maximizing price? No: costs, constraints and the firm's objective are missing. A quantity prediction can be built without solving an optimal choice. Ask students to add a unit cost verbally and describe the new equation required. This transfers modeling to economic decisions without attributing an unimplemented optimization problem to the application. It also shows why a well-predicted outcome may still be insufficient to answer the policy or business question that initially motivated collecting data.

6.3 Assessment and final checks

A simple rubric covers four dimensions: clarity of variables and units, coherence of the rule, justification of trials and interpretation of new evidence. Within each dimension distinguish an absent response, a partially supported response and an explanation with an example or explicit contrast. Platform points can accompany this assessment but do not replace reasoning. Announce grading criteria beforehand if the exercise contributes to a course grade. Avoid penalizing the same numerical mistake twice when a student explicitly identifies and corrects it during discussion. An accurate correction is itself evidence of understanding the evaluation procedure.

At closure, verify that students distinguish observation from prediction, trial from test, fit from transport, gross from net points, and parameters from variables. Request one claim supported by the data and another still unsupported. "The line worked under these conditions; another environment remains to be checked" is more informative than "models are false". The objective is to make bounded, revisable claims. That ability can be revisited in markets, policy and behaviour where information constraints become more demanding. The small artificial setting offers a clear starting point, while the final discussion should identify which real-world complications were deliberately left out.

Checkpoint

At the end: confirm completion, download the session-specific CSV and native wide file if needed, retain instructions and configuration, anonymize classroom materials and record recurring questions. Create a new session for another group. Do not delete a database to refresh a preview.

Bibliography

- Geerling, W., Mateer, G. D., and O’Roark, B. (2019). The View from 30,000 Feet: Using Paper Airplanes to Understand Economic Modeling. *Journal of Economics and Finance Education*, 18(1), 18–24. <https://research.monash.edu/en/publications/the-view-from-30000-feet-using-paper-airplanes-to-understand-econ/>.
- Klahr, D., and Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science*, 12(1), 1–48. [https://doi.org/10.1016/0364-0213\(88\)90007-9](https://doi.org/10.1016/0364-0213(88)90007-9).
- Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71(356), 791–799. <https://doi.org/10.1080/01621459.1976.10480949>.
- American Economic Association. JEL Classification System. A22, B41, C51, C52, C53 and C91, checked 16 October 2026. <https://www.aeaweb.org/econlit/jelCodes.php?view=jel>.