Recovering Credible Trade Elasticities from Incredible Trade Reforms—Using AI!

Joseph B. Steinberg*

January 2026

1 Introduction

The trade elasticity is one of the most central objects in international trade. It governs how bilateral trade flows respond to changes in trade costs, and underpins welfare calculations, counterfactual exercises, and policy evaluation. (Arkolakis et al., 2012). In both empirical and quantitative work, this elasticity is typically interpreted as the long-run response of trade to an unanticipated and permanent change in tariffs—what Alessandria et al. (2025, henceforth AKKRS) call a *canonical reform*. However, canonical reforms do not exist in the data: real-world trade-policy shocks are transitory and often anticipated in various ways. Consequently, standard reduced-form econometric approaches cannot identify the response to a canonical reform. AKKRS argue that one can nevertheless recover the canonical trade elasticity elasticities by interpreting the data through a structural model of trade dynamics in which trade policy follows a Markov process.

This note explores whether canonical elasticities can instead be recovered by an artificial intelligence (AI) model trained on simulated data from non-canonical reforms, and then used to extrapolate to a canonical reform that it has never observed during training. Concretely, I ask whether a neural network trained solely on stochastic, transitory policy changes can learn enough of the underlying structural mapping to correctly predict the long-run response to a deterministic, permanent reform. This approach complements the model-based identification strategy in AKKRS by replacing explicit functional-form assumptions calibrated to firm-level microdata with a learned representation extracted solely from aggregate (at the country-product level) trade data.

This research is motivated by recent debates about whether AI can be used to circumvent the Lucas Critique (Lucas, 1976). Lucas cautioned that policy-invariant structural relationships cannot generally be inferred from historical reduced-form correlations. Recent

*joseph.steinberg@utoronto.ca, University of Toronto and NBER.

discussions about the impact of AI on economic research have revisited this issue, asking whether sufficiently large AI models trained on sufficiently expansive historical datasets can internalize the economic structure to extrapolate correctly about novel shocks or policy changes.

This experiment provides a controlled testbed for this question. Here, the underlying structural model and its response to a canonical reform are known, providing a clean benchmark that can be used to assess an AI model's performance. The central test is whether a neural network can learn the deep structure of the model from a limited subset of possible policy environments and extrapolate correctly to a qualitatively different regime located far away from this subset in parameter space. Preliminary results suggest that the answer yes, but with strong caveats. When the training data include extremely persistent (but not perfectly so) reforms, the model learns to approximate the canonical elasticity reasonably well. However, predictions trained on only short-lived reforms dramatically understate the true long-run elasticity.

The structural model is a version of **?**'s model of exporter life cycle dynamics where trade policy follows a two-state (i.e. high and low tariffs) Markov process with switching probability $1 - \rho$. Higher and lower values of ρ correspond to more or less persistent trade reforms; a canonical reform is an MIT shock with $\rho = 1$. The AI model is a Long Short-Term Memory (LSTM) model, a type of Recurrent Neural Network (RNN).

I first use the structural model simulate datasets for values of ρ ranging from 0.6 to to 0.99. I then train the AI model to predict the dynamics of trade following a tariff change using successively more expansive groups of training datasets, starting with only fairly transitory reforms (ρ between 0.6 and 0.85), and gradually introducing more and more persistent reforms. I use each version of the trained AI model to predict the outcome of a canonical reform, which is not included in any of the training datasets.

When trained only on transitory reforms ($\rho \leq 0.85$), which AKKRS show mimic "within-regime" changes in MFN tariff rates that constitute the vast majority of the historical trade data, the AI model underpredicts the long-run response to a canonical reform by more than 50%. When reforms with $\rho = 0.9$ are included, the AI's prediction is about 25% smaller than the ground truth, and when reforms with $\rho = 0.95$ are added, the bias falls to less than 10%. When values of $\rho = 0.98$ and up are included, the AI gets very close to matching the structural model's canonical elasticity, although there are few, if any, such reforms in the historical record.

2 Structural model

I use the same structural model of trade dynamics as AKKRS and Steinberg (2025). It is a partial-equilibrium version of Alessandria et al. (2021)'s model of export participation over the firm life cycle. There is a mass of firms that differ in productivity and ability to export, both of which are subject to persistent idiosyncratic shocks. To begin exporting, firms must pay a large sunk cost, and they must pay a smaller cost to continue exporting in the future, which also gives a chance of achieving lower exporting costs. These features generate a large gap between the short- and long-run response to a change in tariffs and a very gradual adjustment process—but only if that tariff change is highly persistent.

Tariffs follow a process with two states, $\tau_L < \tau_H$, and a symmetric switching probability $1 - \rho$. Just as in AKKRS, I assume that there is a large number G of goods (think country-product pairs in the data), each with its own continuum of firms behaving as above, and that tariff shocks are independent across goods but common across all firms producing the same good. Given a value of ρ, I simulate the model for a large number T of periods, resulting in a ρ-specific panel dataset with dimension $G \times T$. I also use the model to perform the canonical trade reform: an unanticipated, perfectly persistent change in tariffs, i.e. an MIT shock with $\rho = 1$.

Figure 1 shows the dynamics of trade in the model following a change in tariffs for different values of ρ. The y-axis is the cumulative trade elasticity: the log change in exports between period $t - 1$ and period $t + h$ periods divided by the log change in tariffs over the same period. For stochastic (non-canonical) reforms with $\rho < 1$, I estimate the path of the trade elasticity using the local projections approach of Boehm et al. (2023). Note that this figure is the same as Figure 5b in AKKRS.

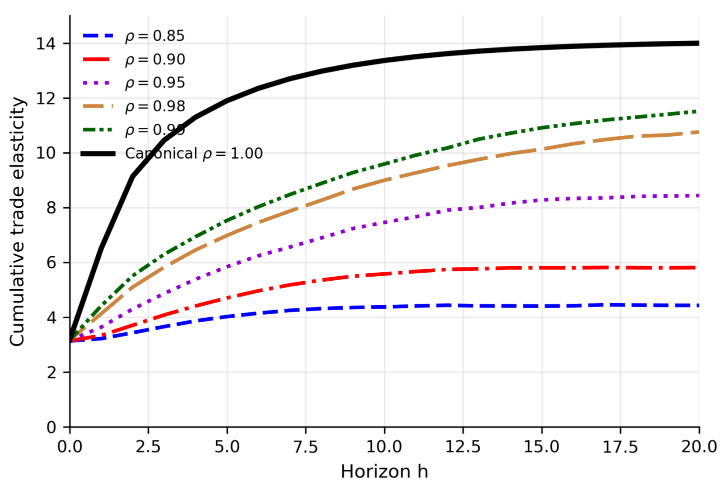

Figure 1: Trade elasticity dynamics in the structural model

By construction, all reforms have the same short-run trade elasticity of about 3.15 (which is the price elasticity of the importer's CES demand function), but there are large differences in long-run elasticities. The canonical long-run elasticity is about 14. At the other end of the spectrum, the most transitory reform in the experiment ($\rho = 0.6$) has a long-run elasticity that is not materially different from the short-run elasticity. A reform with $\rho = 0.85$, which AKKRS show behaves similarly to a change in MFN tariff rates, has a long-run elasticity of about 4.4 (about 30% of the canonical elasticity). A reform with $\rho = 0.95$, which AKKRS show roughly mimics the formation of a free-trade area, has a long-run elasticity of about 8.25 (about 60% of the canonical elasticity). Even highly persistent reforms with $\rho = 0.98$ or $\rho = 0.99$, which essentially do not exist in the data, have long-run elasticities of 11–11.5, which is materially lower (by about 15-20%) than the canonical elasticity.

3 AI model

Note: This is my first foray into AI and I know very little about model architecture. This section contains lightly edited output from GPT 5.2, which wrote the entirety of the code to implement the AI model. Corrections/suggestions are welcome!

I use a sequence model based on a Long Short-Term Memory (LSTM) encoder, which is a gated recurrent neural network designed to summarize time-ordered data into a state vector. An LSTM is well-suited to this environment, as the data are naturally organized as sequences (past histories plus a partially revealed future tariff path up to horizon h), and the the level of exports serves as an endogenous state variable that implicitly embeds firms' beliefs about the stochastic process that governs tariffs for their particular good. More generally, LSTMs work well in Markovian or near-Markovian contexts because they update a latent "memory" state period-by-period and can learn to carry forward persistent, slow-moving information without needing to explicitly attend to all past observations at once.

Relative to a plain (Multi-Layer Perceptron (MLP), the LSTM can learn nonlinear interactions across lags and compress them into a state representation that is reusable across horizons. Relative to Transformers, LSTMs typically train more stably in moderate-length sequences and are less parameter-hungry. Transformers can be more expressive for long-range dependencies, but they often require more data/compute and careful regularization to avoid overfitting or learning spurious patterns, especially when the true data-generating process is already close to a low-dimensional state-space model—as is the case in this setting. In short, the LSTM's "state compression" bias matches closely the actual structural setting: firms behave as if they track a small number of latent sufficient statistics, and the LSTM is explicitly built to approximate that kind of recurrent state update.

Each training example is indexed by a "tariff change anchor," a period t in which good g's tariff changes from τ_H to τ_L or vice versa, and a prediction horizon $h \leq H$. The features are organized into 'the past" (sequences of both exogenous and endogenous variables) and "the future" (only exogenous variables). The "past" is a fixed-length window of length P including the following:

- log tariffs, $\{\log \tau_{t-k}\}_{k=0}^{P}$;
- one-period changes in log tariffs, $\{\Delta_0 \log \tau_{t-k}\}_{k=0}^{P}$;
- log exports, $\{\log X_{t-k}\}_{k=0}^{P}$;
- one-period changes in log exports, $\{\Delta_0 \log X_{t-k}\}_{k=0}^{P}$;
- the discrete state indicator $\{s_{t-k}\}_{k=0}^{P}$;
- and the number of consecutive periods in the current state, $\sum_{k=0}^{P} \mathbb{1}_{\{s_{t-k}=s_t\}}$.

The "future" is simply the path of tariff changes from $t+1$ to $t+h$, $\{\Delta_0 \log \tau_{t+k}\}_{k=1}^{H}$. The target is the cumulative log change in exports from $t-1$ to $i+h$: $y = \Delta_h \log X_t := \log X_{t+h} - \log X_{t-1}$. Importantly, I do not let the model see tariff persistence ρ; it must infer it from the joint dynamics of tariffs and trade around each anchor.

I train five versions of the model. The first uses only the simulation files from fairly transitory tariff processes with $\rho \in \{0.7, 0.8, 0.85\}$, which AKKRS show behave similarly to what they call "within-regime" tariff changes: changes in MFN tariffs for countries that have MFN status from one period to the next. The second version adds $\rho = 0.9$, which they show behaves similarly to the average tariff change when a country switches from one trade-policy regime to another (e.g., moves from Non-Normal Trade Relations to MFN, forms a Free Trade Area, or gains access to the GSP program). The third training run adds $\rho = 0.95$, which is similar to the most persistent tariff changes observed in the historical record. The fourth and fifth runs add $\rho = 0.98$ and $\rho = 0.99$, respectively, which are even more more persistent.

In each training run, anchors are selected using an "empirical budget" that matches the actual distribution of tariff changes across the ρ values that are included in a that run. Specifically, I use a fixed fraction (default 20%) of eligible tariff switches per simulation file. This ensures that even when the model is allowed to see extremely persistent tariff changes by including simulation files for $\rho \geq 0.95$, the model is not artificially induced to focus on the anchors from these files. This approach forces the model to reckon with the fact that the vast majority of the real-world data is dominated by transitory tariff changes. Horizon coverage is limited to $H = 20$, which is more generous than most empirical studies but still keeps the supervised dataset size manageable. To reduce compute while still identifying the shape of responses, I train on a grid of horizons $h \in \{0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20\}$, which empirically preserves the curve while lowering sample explosion.

The baseline configuration is designed to balance expressiveness and robustness. The sequence encoder use 128 hidden dimensions and 2 layers with dropout rate 0.1. The past and future sequences are encoded separately and passed through a two-layer MLP trunk (width 256 with ReLU and dropout) followed by linear heads for the outputs. Training uses AdamW with learning rate 10^{-3}, weight decay 10^{-5}), batch size 2048, early stopping with patience 10, and a 85/15 train/validation split.

4 Results

Figure 2 shows the main result from this exercise. Each line represents the model's predicted cumulative trade elasticity following a canonical tariff cut, conditional on the range of ρ values observed during training. The predictions are compared to the true structural elasticity from the AKKRS model. As the training data incorporate more persistent policy regimes, the predicted canonical elasticity converges toward the structural benchmark. Models trained only on highly transient policies (e.g., $\rho \leq 0.80$) substantially understate the long-run elasticity, while those trained on $\rho \geq 0.95$ nearly match the true curve.

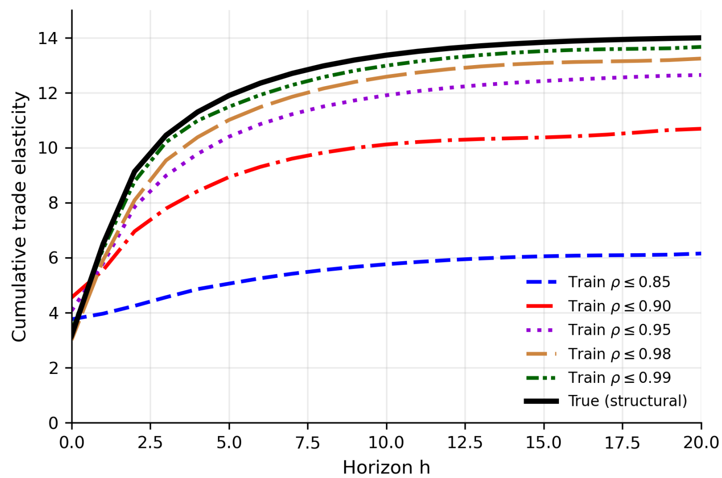

Figure 2: Trade elasticity dynamics in canonical reform as predicted by neural net

It is worth mentioning that even though most of the versions of the trained neural net substantially understand the long-run canonical elasticity, all versions do predict that the canonical reform will generate a larger response than any reforms seen during training. For example, yhe model trained only on $\rho \leq 0.85$ never sees reforms with long-run elasticities larger than about 4.2 but predicts the canonical reform's long-run elasticity is about 6.1, and the model trained on $\rho < 0.9$ never sees long-run elasticities larger than 5.8 but predicts a long-run canonical elasticity of more than ten. Thus, these versions of the model are still fairly successful in extrapolating far outside of their training data.

Overall, these results suggest that neural networks can partially internalize the deep structure of economic environments, allowing meaningful extrapolation from non-canonical to canonical regimes. However, as the Lucas Critique reminds us, such success hinges on whether the data contain enough variation to reveal the underlying structural laws. The next step is to train the same architecture on real-world data, using the empirical tariff and trade series from AKKRS Section 3 instead of the simulated data from the model.

References

Alessandria, G., Choi, H., and Ruhl, K. J. (2021). Trade adjustment dynamics and the welfare gains from trade. *Journal of International Economics*, 131:103458.

Alessandria, G. A., Khan, S. Y., Khederlarian, A., Ruhl, K. J., and Steinberg, J. B. (2025). Recovering credible trade elasticities from incredible trade reforms. Working Paper 33568, National Bureau of Economic Research. NBER Working Paper 33568. (Drafts updated subsequently.).

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

Boehm, C. E., Levchenko, A. A., and Pandalai-Nayar, N. (2023). The long and short (run) of trade elasticities. *American Economic Review*, 113(4):861–905.

Lucas, R. E. J. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

Steinberg, J. B. (2025). Trade policy ambiguity. Working Paper 34408, National Bureau of Economic Research.