JANUARY 2026

# Outcomes Not Optics: Canadian carbon markets need bold reform to be effective

## Canadian Climate Institute response to federal government 2026 Benchmark Review

**Dave Sawyer and Dale Beugin**

## Executive summary and recommendations

As Canada modernizes industrial carbon pricing, this submission to Environment and Climate Change Canada (ECCC) on the 2026 benchmark review examines whether the federal benchmark framework can distinguish systems that merely function from those that deliver outcomes of equivalent stringency.

We find that the updates proposed in the consultation materials are insufficient to ensure robust carbon markets and strong industrial carbon pricing outcomes.

While the recent Canada-Alberta Memorandum of Understanding (MOU) provides important context, benchmark equivalency should be assessed independently of bilateral commitments, including pipeline development, ensuring that industrial carbon pricing delivers stringent outcomes on its own merits.

Using new modelling and analysis aligned with ECCC's consultation materials, we assessed 57 policy-relevant scenarios for Alberta's TIER system. We also tested scenarios in Ontario and British Columbia, concluding the Alberta results are robust across systems, although for brevity we only report the Alberta results in this document. Outcomes are evaluated in 2030 using the effective marginal credit price (EMCP) rather than the headline market credit price. EMCP captures the marginal incentive firms face across all compliance pathways and allows us to test whether systems that pass proposed benchmark diagnostics deliver the $130-per-tonne stringency outcome referenced in the Canada-Alberta MOU.

The results are consistent and robust. Most scenarios pass benchmark tests yet fail to deliver stringency equivalent to a $130-per-tonne EMCP. We identify a clear path forward for improving benchmark assessment and large-emitter trading system (LETS) design grounded in evidence on the effects and trade-offs of key design features. These include benchmark tightening, interventions such as government buy-backs or procurement, price floor escalation relative to the ceiling, credit and offset use limits, banking and expiry rules,

and direct investment credits. Some of these features help stabilize and sustain investment signals while others systematically dilute them.

Based on this analysis, we offer four recommendations.

1. **Define the outcome standard clearly.** Define—within both the updated federal benchmark and the MOU—a $130-per-tonne effective marginal credit price (EMCP) as the benchmark's output-based pricing system (OBPS) stringency criterion that systems must achieve by 2030. Equivalency should be assessed against the marginal incentive firms face rather than the average market price for credits. Anchoring equivalency on EMCP clarifies the required investment signal and supports the use of price corridors, where floors rise as a share of the ceiling to sustain stringency.

2. **Preserve flexibility in system design, conditional on outcomes.** Provide provinces with flexibility in how they achieve a $130 EMCP, subject to a small set of non-negotiable conditions required for equivalency. These include progressively tightening benchmarks, a price floor and ceiling that move over time, and limits on compliance pathways that dilute marginal incentives. Within these constraints, provinces should retain discretion over specific instruments and implementation choices.

3. **Anchor equivalency in transparent analytics.** Ensure that benchmark assessments are underpinned by transparent and credible analytics capable of estimating EMCP. This requires data on how firms comply, which compliance pathways are used, and the effective prices firms face. Without this information, equivalency cannot be credibly assessed.

4. **Track performance over time.** Require ongoing performance tracking using observed market and compliance data. Transparency in credit markets and compliance behavior should be used to assess whether systems continue to conform with the minimum national stringency standards over time rather than relying on one-time design assessments.

Taken together, these recommendations support a benchmark framework that verifies equivalency based on outcomes rather than optics while maintaining flexibility in provincial system design and strengthening confidence that industrial carbon pricing delivers federal climate objectives.

## About the benchmark review

Canada's large-emitter trading systems are at a crossroads. Decisions on stringency and design made in 2026 will shape outcomes through 2030 and well beyond. Getting these decisions right matters. Industrial carbon pricing, delivered through LETS, remains the single most important policy lever for reducing emissions from Canada's highest-emitting sectors.

The Institute has written extensively on this issue, including through its independent assessment of carbon pricing, the accompanying roadmap, and a range of analytical work diagnosing weaknesses in existing systems, their impacts, and options to strengthen them. This work points to a consistent pattern. LETS are opaque, rely on outdated design choices, and have been systematically weakened by provinces over time, leading to material underperformance.

Federal oversight has also been inconsistent. The federal government has been reluctant to impose the backstop where provincial systems fall short, most notably as Saskatchewan zeroed out its industrial carbon price in 2025. At the same time, there has been limited effort under the federal backstop to align key design elements across systems. Better alignment would support linked credit trading and reduce competitive distortions between facilities operating in different jurisdictions, supporting the objective of one Canadian economy. It would also help shield Canadian exports from rising border carbon tariffs, including the EU Carbon Border Adjustment Mechanism.

Against this backdrop, the federal government has launched a process to modernize large-emitter trading systems. The first track is regulatory and technical. The accelerated federal benchmark review is intended to update the policies and methods used to assess whether proposed or operational provincial systems meet minimum federal stringency standards, avoiding the need for the federal government to impose its own OBPS under the backstop. At the same time, the federal government is updating its OBPS, which will shape the parameters against which provincial systems are assessed for conformity.

The second track is political and bilateral, centred on negotiations between Canada and Alberta. These negotiations are anchored in a Memorandum of Understanding (MOU) that links oil pipeline development to commitments by Alberta to strengthen its TIER program. The significance of this second track is clear. An effective marginal credit price of $130 has been identified in the MOU as the benchmark for stringency, although the definition and timing for achieving the stringency criterion were not specified.

## Just what are we measuring?

The consultation paper and the January 2026 benchmark slides are best read as a set of diagnostic principles and associated tests, intended to assess performance across different LETS designs.

Taken together, however, the consultation materials reveal an unresolved tension between the stated stringency criterion referenced in the MOU and the diagnostic tests proposed to assess systems. The federal industrial carbon pricing benchmark framework is intended to ensure that provincial systems deliver incentives equivalent to the federal OBPS. By establishing minimum national stringency standards, the benchmark seeks to ensure that regulated facilities face comparable incentives to reduce emissions and invest in low-carbon technologies.

The proposed ECCC benchmark tests are primarily designed to assess whether systems are structured to maintain a carbon pollution price signal with integrity over time. They focus on market balance, credit availability, and banking dynamics to assess whether the price signal can be sustained and whether systems can continue to function as designed. These considerations are necessary to ensure market operation and compliance feasibility. However, they do not, on their own, determine whether the signal reaches a level consistent with an outcome-based OBPS stringency criterion on a defined timeline.

This distinction matters. The relevant question is not simply whether systems adopt the minimum national carbon price (MNCP) schedule, but whether the price signal delivered by the system achieves the intended outcome. As with the benchmark's treatment of cap-and-trade systems, equivalency should be assessed based on outcomes, not solely on adherence to design features or price paths.

In this submission, we therefore focus on equivalency of stringency. Specifically, we assess whether systems deliver a price signal consistent with federal ambitions and whether scarcity is sufficiently durable to support investment decisions at that level. This framing does not prescribe system design. It clarifies the objective function that the benchmark must satisfy if equivalency is to reflect outcomes rather than system mechanics.

Under this interpretation, the two-track modernization process for LETS converges as follows. The MOU frames the ambition by defining a $130-per-tonne-by-2030 minimum effective credit price (MECP) as the OBPS stringency criterion. While negotiated with Alberta, this criterion establishes the reference point for assessing minimum national stringency. The benchmark review then determines whether provincial systems can deliver that stringency through their chosen design features, as assessed through diagnostic tests.

The benchmark does not set ambition; it determines whether ambition survives implementation. When diagnostics emphasize market operability rather than whether the price signal is sufficiently strong and durable, systems can pass benchmark tests while still falling short of outcome-based stringency under the OBPS and the MOU.

## How we approached the problem

The central question for the federal benchmark review is therefore not whether systems demonstrate liquid markets and compliance, but whether they deliver outcomes consistent with the MOU. In practice, this means asking whether LETS are sufficiently stringent to support a price signal consistent with a $130-per-tonne MECP. This distinction between market liquidity and stringency is a core principle that this submission revisits throughout.

To operationalize this question, we assess whether the diagnostic tests proposed by ECCC can distinguish between systems that merely operate and systems that deliver on the stringency criterion. Our analysis focuses on identifying false positives, where systems can pass benchmark checks but fail the stringency criterion, and where additional diagnostics could better inform the stringency assessment.

Specifically, we ask:

- Do systems that pass the proposed federal tests deliver equivalent outcomes?
- Which indicators reveal stringency and which generate false positives?
- What indicators distinguish markets that survive from markets that drive investment consistent with the stringency criterion?

To answer these questions, we modelled 57 policy-relevant scenarios for Alberta's TIER system using a modular simulation approach aligned with the structure of ECCC's consultation materials. Findings were also tested across LETS designs in Ontario and British Columbia. Outcomes are assessed in 2030, assuming the stringency criterion is to be met by that date.

The modular approach allows us to stress-test the major policy questions considered in the consultation materials, including coverage, signal integrity, tightening, banking, and compliance pathways such as credit limits and expiry. This allows us to isolate the impact of individual design choices, such as the benchmark ratchet rate, while holding other system features constant. This, in turn, enables combinations of scenarios to generate a wide range of stringency outcomes, allowing us to assess how the proposed tests perform, identify the main drivers of stringency, and develop a set of recommendations foundational to the benchmark review.

## Defining the stringency objective

We move beyond headline market prices as a measure of stringency. In multi-pathway LETS, the relevant marginal incentive facing firms is the weighted incentive across all available compliance options, including trading, offsets, fund payments, investment credits, and banking. We capture this incentive through the effective marginal credit price, which reflects the marginal cost firms face when deciding whether to abate or comply through alternative compliance pathways.

This distinction is material. Firms with emissions limits and higher abatement costs, for example, are more exposed to ceilings, credit limits, and alternative compliance options than to observed market prices alone. As a result, market prices can appear similar across systems while underlying incentives differ substantially.

Across the modelled scenarios, systems with nearly identical market prices exhibit wide variation in MECP, ranging from below $50 to above $130 per tonne. In 2030, scenarios with the same $100 market price show differences of up to 12 Mt in annual abatement. Market price alone is therefore an incomplete and potentially misleading indicator of policy stringency and investment incentive.

## Do systems that pass the benchmark deliver equivalent outcomes?

The core finding from our assessment is straightforward. The benchmark tests assess whether the market works, not whether it will reach the required level of stringency on time.

Figure 1 shows that, across the 57 scenarios, the benchmark tests have a high probability of returning false positives:

- **High test pass rate:** 84 per cent of modelled scenarios (48 of 57) pass all four federal design tests.
- **False positives (pass test, fail stringency):** 77 per cent of those passing systems (37 of 48) fail to deliver a $130 EMCP by 2030.
- **Limited $130 attainment:** Only 11 scenarios both pass the tests and achieve the MOU target.

This result is robust across variations in price ceilings, floors, banking rules, and credit limits. Tests of net demand, market balance, and static banking metrics are useful for determining

whether a market is operational. They are not sufficient for determining whether a system meets a given stringency requirement.

The implication is not that these tests should be abandoned. They should be treated more as a screening tool for market liquidity than a diagnostic for policy stringency.

**Figure 1: The diagnostic gap persists across all four federal tests**

| Test | Pass rate | Of passing, % achieving $130 | Diagnostic |
|---|---|---|---|
| Net demand >0 | 100% | 19% | Trades could happen |
| Expiry test <3 years | 91% | 21% | Will credits expire and be worthless? |
| Magnitude >0.33 or 3 yrs | 95% | 20% | Can the market absorb the bank? |
| Buffer adequate (6%) | 84% | 23% | Is there a demand cushion? |
| **All tests** | **84%** | **23%** | **False positive risk** |

Source: LEMS modelling results, 54 scenarios

## Buffer magnitude and banking durability

ECCC's consultation materials provide an illustrative six per cent net demand buffer, expressed as a share of covered emissions. The Institute's roadmap following the Independent Expert Assessment recommended such a buffer, and we remain supportive of it.

Our modelling confirms that this threshold plays a necessary role. A six per cent buffer prevents market failure and ensures liquidity. It is not sufficient, however, to deliver the $130 EMCP required for outcome equivalency.

Across 57 modelled policy scenarios, the evidence shows that buffer size is a predictor of stringency, not just a liquidity check when net demand exceeds zero.

Tests of buffer magnitude show a clear pattern:

- Systems operating with small to moderate buffers (below 10 per cent) consistently pass federal design tests but never achieve a $130 EMCP.
- Systems with larger buffers begin to deliver stronger outcomes (10 to 30 per cent), though not consistently.

- Only systems with very large scarcity cushions (>30 per cent) reliably sustain price signals aligned with federal ambition.

This pattern highlights a structural risk in the current benchmark assessment framework. Systems operating close to the minimum buffer frequently appear compliant yet lack sufficient scarcity to drive investment consistent with the MOU. These systems are false positives, liquid, but not equivalent.

**Figure 2: Buffer magnitudes predict outcomes**

| Buffer magnitude | Pass rate (ECCC tests) | Hit $130 EMCP? | Characterization |
|---|---|---|---|
| Small to Moderate (<10%) | High | 0% | Liquidity only: Sufficient for trading but too weak for investment signals. |
| Large (10% - 30%) | High | Moderate | Transition: Sufficient to drive abatement in some sectors but lacks durability. |
| Very Large (>30%) | High | Consistent | Stringent: Provides the structural scarcity required to hit the $130 target. |

Source: LEMS modelling results, 54 scenarios

Another way to view this result is through net demand duration, which provides a direct test of bank durability and closely aligns with the magnitude test proposed by ECCC. A recalibrated magnitude test would focus on net demand duration, asking how long scarcity persists as compliance obligations draw down the credit banks. In the modelled scenarios, systems that deliver the $130-plus outcome are those with scarcities maintained across multiple compliance periods. In practice, this corresponds to net demand durations of roughly five years or more.

The policy implication is straightforward. The six per cent net demand buffer ensures a market can operate. Outcome equivalency depends on whether scarcity is substantial and durable. Buffer magnitude, and the durability of banking over time, determines whether markets merely function or deliver benchmark stringency aligned with the MOU.

**Figure 3: A 3-year drawdown is a liquidity screen, 5-year is a stringency screen**

| | Liquidity = | Stringency = | Outcome ($130 hit rate) |
|---|---|---|---|
| Magnitude test | 3-Year drawdown | 5-Year drawdown | 20% (3yr) vs ~100% (5yr) |

## Drivers of LETS stringency

The modelled scenarios allow us to isolate how individual design choices affect the EMCP, abatement, and total system cost. The results are unambiguous: Not all policy levers contribute equally to stringency.

- **Benchmark tightening is the dominant driver of stringency.** When benchmark stringency is varied in isolation, it produces the largest marginal effect observed in the modelling. Across tested scenarios, benchmark tightening generates more than a $120 range in EMCP outcomes, exceeding the impact of any other design element.

  This reflects a basic structural reality. Benchmark tightening is the only lever that directly creates scarcity by decreasing the emissions limit via benchmark tightening. As a result, it delivers the highest abatement outcomes and remains the most effective pathway to achieving the $130 EMCP required for outcome equivalency. No combination of secondary design features can compensate for an insufficiently stringent benchmark.

  In the modelled cases, benchmark tightening consistently produces the largest increase in system costs (roughly a six-fold increase above the modelled TIER baseline), but also high levels of emissions reductions (about 10 Mt). We also observe diminishing returns at higher stringency levels: costs continue to rise even as EMCP gains flatten when baseline performance benchmark tightening rates are double current TIER rates (one to four per cent differentiated by sector).

- **Direct investment credits systematically dilute stringency.** Across modelled scenarios, direct investment credits and similar mechanisms have a large and consistently negative effect on EMCP. Introducing investment credits reduces EMCP by roughly $60 on average, even under otherwise stringent benchmark settings.

  This erosion is not marginal. By allowing compliance that adds supply while simultaneously adding more abatement and reducing demand, investment credits bypass scarcity and weaken the price signal. In the modelling, scenarios with investment credits exhibit sharp declines in both EMCP and market prices, despite passing all federal design tests.

  Investment credits and similar mechanisms lower compliance costs substantially, but they also reduce EMCP and materially reduce abatement at facilities. In the scenarios, introducing direct investment credits reduces costs by roughly two-thirds and cuts abatement by more than half. The cost savings are therefore not a productivity improvement but rather a dilution of policy stringency.

- **Price floors sustain the incentive.** Floor escalation has a positive but more moderate direct effect on EMCP than benchmark tightening. When varied in isolation, escalating floors increase EMCP by roughly $40 across scenarios.

  While not a primary driver of stringency, floor escalation plays a critical stabilizing role. Dynamic floors eliminate prolonged low prices and prevent benchmark-driven scarcity from collapsing with oversupply. In this sense, floors are defensive rather than offensive tools. They sustain and protect the investment signal created by benchmarks, even if they cannot generate that signal on their own.

  Escalating price floors materially increases EMCP at modest incremental cost. In the modelled floor scenarios, EMCP rises by roughly $40 for a relatively small cost increase (about 50 per cent above the TIER baseline).

- **Tighter credit limits can add to stringency.** Tighter offset, bank, and fund credits deliver a modest boost to ECMP, typically in the range of $10 to $15. These limits function as guardrails by ensuring that compliance obligations must be partially met through abatement effort. This helps prevent the price signal created by benchmark tightening from being diluted by low-cost compliance options.

  Tighter limits on offsets and banked credits can rapidly increase system costs by pushing firms toward higher-cost compliance credits or ceiling fund payments. However, this increased cost is accompanied by higher expected levels of abatement, highlighting a critical trade-off between total system cost and emission reductions.

- **Bank expiry rules affect timing, not ambition.** Credit expiry rules influence how quickly scarcity emerges, rather than the ultimate level of stringency. By shortening expiry limits, the drawdown of accumulated banks is accelerated.

  In the modelled scenarios, systems with tightened benchmarks but no expiry rules often take several additional years to reach a $130 EMCP as older credits continue to be used for compliance. When expiry rules are in place, this adjustment happens more quickly.

  Expiry rules do increase system costs by reducing the value of banked credits but they also improve the durability and reliability of the price signal. Their role should therefore be transitional. They shape how quickly markets move toward stringency but do not determine whether the stringency criterion is ultimately met.

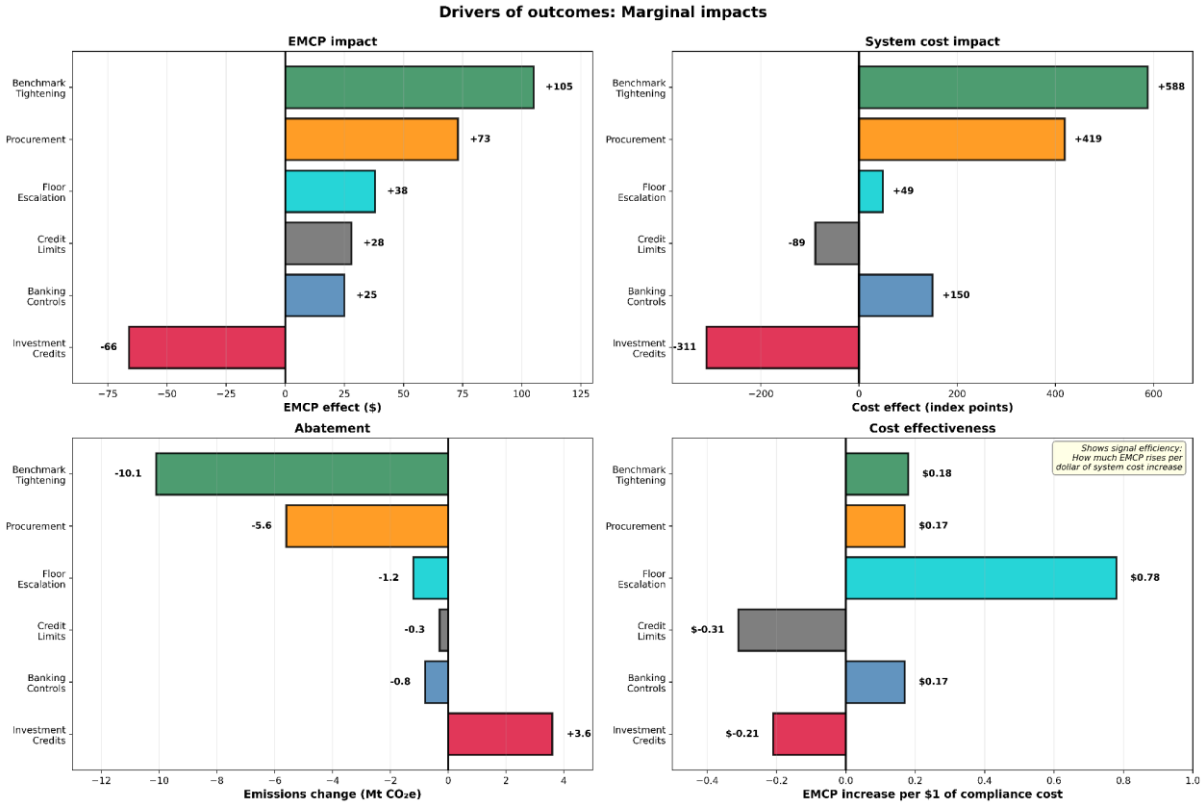
- **Market stability mechanisms.** Measures such as market stability reserves contribute modestly to EMCP outcomes by limiting surplus accumulation and accelerating drawdown. Their role should be transitional. They help markets move more quickly toward a stringent state but do not determine the ultimate level of stringency.

Taken together, these results establish a clear hierarchy of levers. Benchmark tightening does the heavy lifting for equivalency attainment. Floor escalation and banking controls protect and stabilize the signal that benchmarks create. Credit and offset limits provide guardrails. Direct investment credits, by contrast, act as a dilution lever capable of neutralizing even aggressive benchmark tightening.

Figure 4 summarizes the marginal impact of key design levers on four outcomes: the effective marginal compliance price, total system cost, emissions reductions, and cost effectiveness. Each bar reflects the isolated effect of changing a single design element while holding other features constant, based on modular scenario analysis and using a Shapley decomposition.

The results reveal a clear hierarchy of levers. Benchmark tightening is the primary driver of stringency and emissions reductions, but at high cost. Floor escalation delivers the strongest return per dollar by sustaining scarcity at relatively low cost. Investment credits, by contrast, lower costs by weakening the policy itself, reducing EMCP and increasing emissions despite appearing cost-effective. The figure makes clear that not all design choices contribute equally to stringency—and some actively undermine it.

**Figure 4: Drivers of stringency: marginal impacts across design levers**




## The role of interventions in sustaining stringency

The benchmark consultation also raises the question of intervention mechanisms, recognizing that market outcomes may diverge from intended policy signals even when systems are deemed compliant. This is an important acknowledgement. Interventions are not a sign of system failure but are a recognition that banking, surplus accumulation, and price dynamics can weaken stringency over time.

In practice, interventions function as structural support for the price signal. They help ensure that the interaction of the price floor and ceiling continues to reflect intended ambition as conditions evolve. Where surplus supply emerges, intervention tools can accelerate bank drawdown, reinforce rising floors, and prevent prolonged periods of weak prices that undermine investment incentives.

Seen this way, interventions are closely linked to the durability of scarcity. They complement buffer requirements and floor escalation by managing how banking evolves through time. Without credible intervention mechanisms, systems risk remaining technically liquid while drifting away from the outcomes implied by federal ambition.

## Mapping federal benchmark tests to stringency diagnostics

Taken together, the modelling and analysis suggest that the federal benchmark tests are best interpreted as minimum viability screens. To assess equivalency against an OBPS stringency criterion, these tests need to be complemented by outcome-based diagnostics that directly evaluate whether scarcity is sufficient and durable. Table 4 below summarizes the outcome diagnostics implied by the evidence and how they relate to the policy objective of equivalency.

These diagnostics do not prescribe system design. They clarify whether a given system, regardless of design choices, can deliver the stringency criterion. Importantly, they reduce the risk that provincial systems will be certified as benchmark-compliant when, in practice, they are merely functional but misaligned with the required level of stringency.

**Figure 5: Stringency diagnostics**

| Diagnostic | What it assesses | Why it matters for equivalency | Evidence from modelling |
|---|---|---|---|
| **EMCP** | The weighted marginal incentive faced by firms across all compliance pathways | Captures the investment signal implied by the system, beyond headline market prices | Systems passing all federal tests often deliver EMCPs well below $130 |
| **Net demand buffer (magnitude)** | Depth of structural scarcity relative to system size | Distinguishes markets that merely function from those capable of driving investment | Buffers below ~10% don't deliver $130 EMCP; 30%+ buffers correlate with target success. |
| **Net demand duration (bank durability)** | How long scarcity persists as banked credits are drawn down | Aligns market signals with multi-year investment horizons | Systems delivering $130 EMCP maintain scarcity for roughly five years or more |
| **Price floor exposure** | Risk of prolonged price contact at the minimum | Identifies floor-locking and weak signals masked by market liquidity | Systems without floors consistently underperform on EMCP |
| **Investment credits (dilution)** | Extent of compliance outside the allowance market | Flags mechanisms that bypass scarcity and weaken marginal incentives | Investment credits (DICs) reduce EMCP by ~$60, neutralizing benchmark tightening. |

## Implications for the Benchmark Review

The evidence points to a clear set of findings for the federal benchmark review.

1. **The current benchmark tests are insufficient on their own.** They are effective at assessing whether markets can operate but not whether they deliver outcomes consistent with federal ambition. Outcome-based diagnostics are required to complement existing design screens.

2. **Equivalency must be defined in terms of outcomes.** In practice, this means verifying whether systems deliver an effective marginal credit price consistent with the $130 benchmark by 2030 rather than inferring stringency from market mechanics alone.

3. **Some design elements are foundational to stringency.** Benchmark tightening and a price corridor with a price floor and ceiling are non-negotiable components of any system claiming equivalency. Without these elements moving in tandem, other design features cannot compensate.

4. **Floor escalation relative to the ceiling is essential.** Floors that rise over time, anchored to the ceiling, prevent prolonged low-price lock-in and help sustain investment signals when surplus supply emerges.

5. **Investment credits and similar compliance pathways are actively harmful to stringency.** They weaken marginal incentives, erode effective prices, and generate false positives under benchmark diagnostics. Their use should be tightly limited or excluded from equivalency assessments.

6. **Intervention mechanisms must be available to manage banked oversupply and maintain system balance over time.** Without tools to address surplus accumulation, banking can undermine scarcity even in otherwise well-designed systems.

7. **Transparency is critical.** Equivalency assessment requires clear visibility into how firms comply, which pathways are used, and the effective prices they face. Without this information, equivalency cannot be credibly verified.

Taken together, these findings imply a benchmark framework that verifies equivalency based on outcomes while preserving flexibility in provincial system design.

**We recommend updating the benchmark to:**

1. **Define the outcome standard clearly.** Define—within both the updated federal benchmark and the MOU—a $130-per-tonne effective marginal credit price as the minimum outcome standard that systems must achieve by 2030. Equivalency should

be assessed against the marginal incentive firms face, rather than the average market price for credits. Anchoring the benchmark on EMCP clarifies the investment signal required to meet federal ambition while preserving flexibility in how provinces design and operate their systems.

2. **Preserve flexibility in system design, conditional on outcomes.** Provide provinces with flexibility in how they achieve a $130 EMCP, subject to a small set of non-negotiable conditions required for equivalency. These include progressively tightening benchmarks, a price floor and ceiling that move over time, and limits on compliance pathways that dilute marginal incentives. Within these constraints, provinces should retain discretion over specific instruments and implementation choices.

3. **Anchor equivalency in transparent analytics.** Ensure that benchmark assessments are underpinned by transparent and credible analytics capable of estimating EMCP. This requires data on how firms comply, which compliance pathways are used, and the effective prices firms face. Without this information, equivalency cannot be credibly assessed.

4. **Track performance over time.** Require ongoing performance tracking using observed market and compliance data. Transparency in credit markets and compliance behavior should be used to assess whether systems continue to conform with the minimum national stringency standards over time, rather than relying on one-time design assessments.