energies

MDPI

Article

State-Dependent Merit-Order Effects in a Hydro-Dominated Market: Conditioning Variable-Renewable Price Suppression on the Marginal Fuel Regime in New Zealand

Ahmed Hassan, Krishnamachar Prasad * and Jeff Kilby

School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand; ahmed.hassan@autuni.ac.nz (A.H.); jeffrey.kilby@aut.ac.nz (J.K.)
* Correspondence: krishnamachar.prasad@aut.ac.nz

Abstract

The merit-order effect (MOE), the suppression of wholesale prices by zero-marginal-cost variable renewable energy (VRE), is typically estimated as a single average coefficient per market. In a hydro-dominated system, the price-setting technology alternates between hydro and thermal plants, so the MOE should depend on the operating state. Using New Zealand nodal data, a baseline panel model first reproduces a published 2011–12 seasonal benchmark to within NZ$0.22/megawatt-hour (MWh). A model conditioning the VRE coefficient on reservoir storage (2015–24, 245 nodes, 28,827 node-month observations from approximately 42 million half-hourly records) passes every in-sample diagnostic (within R^2 = 0.81) yet fails out-of-sample and stability testing. The interaction adds no significant predictive value beyond a storage main effect (p = 0.15–0.22) and reverses sign across sub-periods (+58 to −51), an instability not attributable to collinearity. Replacing the storage level with the monthly thermal generation share, a dispatch-based proxy for the hydro-thermal regime, yields a stable, significant, out-of-sample-validated interaction (clustered t = 5.0, leave-one-month-out cross-validation gain +14.2%) that survives direct hydrological and demand controls, including a head-to-head test against a storage interaction. Identification rests on 120 monthly observations, since these regressors are system-wide. Price suppression is strongest in hydro-abundant months, at approximately NZ$5.8/MWh per 10% relative VRE increase, and attenuates toward zero as the thermal share rises, consistent with water-value pricing. A forward test on Jan 2025 to May 2026 data, which postdate model development, confirms the interaction's predictive value (forward RMSE improvement +33%, wild cluster bootstrap p = 0.047). Re-estimation on the 2011–12 panel returns a weakly identified, directionally opposite interaction (bootstrap two-sided p = 0.053). The validated state dependence therefore characterises the 2015–24 market regime rather than a structurally invariant relationship.

Keywords: merit-order effect; variable renewable energy; state-dependent price effects; hydro-thermal dispatch; water value; nodal pricing; out-of-sample validation; New Zealand electricity market; price cannibalisation

Check for updates

Academic Editor: Jay Zarnikau

Received: 19 June 2026
Revised: 28 July 2026
Accepted: 6 August 2026
Published: 10 August 2026

1. Introduction

The merit-order effect (MOE) describes how energy generation with near zero marginal cost, principally via wind and solar methods, displaces costlier plants and lowers the wholesale clearing price [1–3]. The effect is robustly negative on average and has been quantified across many markets [3–10]. Most empirical studies adopt specifications that estimate a single average price-suppression coefficient per market or period, implicitly treating the merit-order effect as constant across market conditions [1–6,8,9]. A smaller but growing body of literature relaxes this assumption by allowing effects to vary across the price distribution or operational states [7,10].

For Aotearoa New Zealand, treating the MOE as constant across market periods is questionable since it is a hydro-dominated market. When reservoirs are full, the marginal unit is typically flexible hydro priced at a low water value; when reservoirs are depleted, the marginal unit is thermal, or hydro priced at a high water value, reflecting scarcity [11–13]. Because the merit-order effect of an extra megawatt-hour of variable renewable energy (VRE) depends on the cost and the slope of the supply curve at the operating point, the effect should itself vary with the system state. Recent evidence confirms that the MOE is state-dependent, varying with renewable penetration, demand, and season across several markets [7,9]. The New Zealand government and grid-operator projections from the Ministry of Business, Innovation and Employment (MBIE) [14] and Transpower [15] anticipate substantial growth in wind and solar capacity over the coming decade, sharpening the operational salience of how the MOE interacts with the hydro-thermal regime.

This paper is an empirical study of the New Zealand electricity market that makes a model validation contribution to the merit-order literature. State dependence of the MOE is not in itself new. Effects varying with penetration, demand, and season are documented across several markets [7,9]. What has not been done is to condition the merit-order coefficient on the hydro-thermal marginal fuel regime. The methodological point is that the choice of conditioning variable must be validated out-of-sample, because a natural alternative can pass every conventional in-sample test and still fail.

Three steps organise the analysis. First, a credible estimation pipeline is established by replicating the published seasonal benchmark of [1] on the identical 2011–12 nodal panel to within NZ$0.22/MWh. Second, the most direct extension for a hydro-dominated market, conditioning the VRE coefficient on reservoir storage (Model B), is estimated in full. Out-of-sample and stability tests show that it fails despite excellent in-sample behaviour, and the failure mode is itself a substantive finding. Third, the storage level is replaced with the monthly thermal generation share, a contemporaneous dispatch-based index of the marginal fuel regime. The resulting state-dependent coefficient is stable, significant, and validated out-of-sample, including on 2025–26 data that postdate model development, and it is interpretable through the water-value logic of hydro-thermal dispatch.

2. Related Work and Theoretical Framework

2.1. New Zealand Studies

Wen et al. [1] provide the most directly comparable nodal evidence to the New Zealand electricity market. They report seasonal merit-order estimates that differ between wet and dry seasons, using a spatial econometric model on 2011–12 data. A related study [16] estimates spatial wind effects on New Zealand prices. That work establishes that seasonality matters, but it treats each season as a separate bucket rather than modelling the continuous state driver.

Suomalainen et al. [17] document the wind–hydro correlation structure that underlies seasonal price behaviour. Agent-based and market power analyses [18–20] reproduce New Zealand prices only when hydro storage and water values are tracked explicitly, which directly motivates conditioning the MOE on the hydro-thermal state. Poletti and Staffell [21] study a fully renewable New Zealand system, Macedo et al. [22] examine nodal price behaviour in the insular New Zealand market, Wang et al. [23] analyse extreme price responses to policy shocks, and Kapoor et al. [24,25] address price forecasting. None of these estimates the merit-order coefficient as a function of the marginal technology on observed nodal data.

2.2. International Evidence and the Single-Coefficient Norm

The international literature on the MOE is extensive. The foundational analysis of Sensfuß et al. [26] established the empirical merit-order framework in the German market using an agent-based simulation. Subsequent ex post empirical studies have refined the methodology and quantified the effect across many systems. Cludius et al. [3] document a robust negative coefficient for Germany 2008–16, and Antweiler and Muesgens [27] analyse the long-term merit-order effect of renewables in European markets, extending this work to energy-only markets dominated by intermittent renewables with grid-scale storage [28]. Hirth [29] introduces the complementary concept of the market value of variable renewables, linking the MOE to the value capture of VRE itself.

For hydro-thermal systems most analogous to New Zealand, Acar et al. [30] estimate the merit-order effect for wind and run-of-river hydroelectricity on Turkish electricity prices, while Benhmad and Percebois [31] and Martin de Lagarde and Lantz [32] provide further German market evidence on the price impact of solar and wind feed-in. For the Australian National Electricity Market, recent panel evidence is available from Csereklyei et al. [5] and Mwampashi et al. [6]. Zarnikau et al. [4] and Ajanaku and Collins [10] report on the Electric Reliability Council of Texas (ERCOT) and the wider US market, Halttunen et al. [8] on a global assessment of the merit-order effect and cannibalisation, López Prol et al. [9] on the California cannibalisation effect, and Tselika [7] on a European hourly panel, and Newbery [33] connects the MOE to storage economics under intermittent generation.

Across these markets, the price-setting technology is thermal for most hours, and consequently each study reports a single average effect or a per-absolute quantity. Where state dependence is examined, it is found to be material [7,9], yet it is rarely built into the structural coefficient itself and never conditioned on a hydro-thermal marginal state.

2.3. Hydro Price Formation and Water Value

In hydro-dominated systems, an equilibrium based on instantaneous marginal cost is inadequate. Stored water has no fuel cost but a substantial opportunity cost, namely the water value, which equals the shadow price of the reservoir balance constraint [11,13]. Prices in hydro markets are therefore governed by current and expected hydrological conditions [12].

A closely related, and for this paper decisive, observation is that reservoir levels alone are poor predictors of marginal prices. In Spain, hydropower set the marginal price for 52% of the hours of 2021, yet reservoir levels did not explain the high prices reached over that period [2]. The binding quantity is the inter-temporal energy constraint as priced by the water value, not the instantaneous stock. This is why conditioning on the storage level can fail even when storage clearly matters. A better approach is to use the direct, real-time signal of which type of plant is setting the price.

2.4. Theoretical Framework

In a marginal-pricing market, plants are dispatched in ascending order of marginal cost until supply meets demand, and the clearing price equals the marginal cost of the last dispatched unit. For interval t, the clearing price P_t may be defined as

$$P_t = MC(D_t - W_t) \tag{1}$$

where P_t is the clearing (wholesale) price at interval t, $MC(\cdot)$ is the supply (offer) stack viewed as a marginal-cost function, D_t is the system demand at the same interval, and W_t is the contemporaneous VRE generation. An increase in VRE shifts residual demand leftward, lowering the marginal cost of the last unit and hence the price [26]. The magnitude of that reduction depends on the local slope of $MC(\cdot)$ at the operating point, thus making the effect state-dependent.

In a hydro-thermal system, the stack has two regimes. When storage is abundant, the marginal unit is flexible hydro offered near a low water value (the water value is the worth of holding water back for later use; it is low when lakes are full). The relevant part of the stack is low and relatively flat. An extra unit of wind or solar then pushes out this cheap hydro and lowers the price sharply. When storage is scarce, the price is governed by the inter-temporal shadow value of water and by thermal plants at the top of the stack. An extra unit of wind or solar in a single period saves only a little water and does not relax the binding multi-period energy constraint, so its price impact is weak. The MOE should therefore be strong in the hydro-abundant regime and weak in the water value-constrained regime. This motivates conditioning the VRE coefficient on a system-state variable. Two candidates present themselves, namely reservoir storage level, which is a stock as far as New Zealand electricity market is concerned, and the share of generation supplied by thermal plants (a contemporaneous flow). The remainder of the paper shows how the choice between them is decisive.

3. Data and Methods

3.1. Data Sources and Estimation Panels

Data are drawn from the Electricity Authority Electricity Market Information (EMI) database [34]. Two panels are used (see Table 1), both derived from the same half-hourly records, namely nodal final prices for 48 trading periods per day and per-unit generation by fuel type at the same resolution. The 2011–12 panel matches the benchmark [1] exactly, using the same 10 nodes aggregated to hourly resolution. The 2015–24 panel aggregates approximately 42 million half-hourly nodal price observations to monthly node-level means, giving 28,827 node-month observations across 245 nodes with at least 60 months of data over 120 months. The EMI fulfils the same transparency role in New Zealand as the European Network of Transmission System Operators for Electricity (ENTSO-E) Transparency Platform does for European wholesale markets [35], providing standardised access to nodal prices, dispatch, and unit-level generation records.

Table 1. Characteristics of the two estimation panels.

Attribute	2011–12	2015–24
Source	EA EMI	EA EMI
Nodes	10	245
Period (months)	24	120
Raw half-hourly records	350,880	~42 million
Analysis observations	175,440 hourly	28,827 node-month
Mean VRE share	4.5% *	5.83%
Mean storage	0.58	0.535
Mean thermal share	24.6%	15.8%
Role	Benchmark	Estimation

* Wind only—utility-scale solar entered the New Zealand system much later.

3.2. Descriptive Patterns in Generation and Price

Annual shares of total generation by primary fuel type are shown in Figure 1, with values for 2011–12 and 2015–24 computed directly from the half-hourly Electricity Authority (EA) generation records used in this paper. Values for 2000–10 and 2013–14 are drawn from the Ministry of Business, Innovation and Employment (MBIE) Energy in New Zealand annual tables [36]. Hydro and geothermal together account for more than 70% of generation in most years [37]; the VRE share has risen from about 0.1% in 2000 to 9.4% in 2024. The figure also shows the 2015–24 analysis panel, which marks the estimation window used in Sections 5–7.

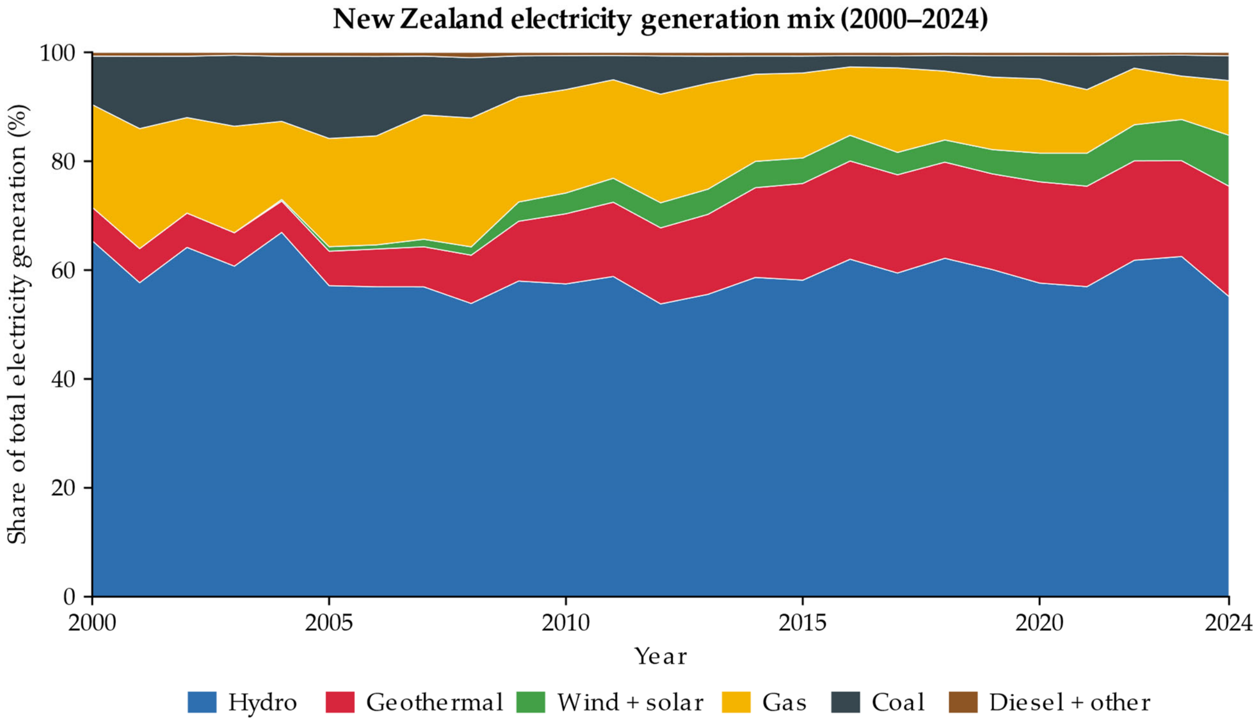

Figure 1. New Zealand electricity generation mix, 2000–24. Source: half-hourly Electricity Authority Electricity Market Information [34] and MBIE Energy in New Zealand annual tables [36].

The variable renewable share has grown markedly over the study period. Figure 2 shows the combined share of wind and utility-scale solar generation, which rose from an annual mean of 4.7% in 2015 to 9.4% in 2024 and exceeded 11% in a few individual months. Generation is predominantly from wind, with utility-scale solar emerging from early 2024. New Zealand has among the highest wind capacity factors in the world, at roughly twice the global average, which underpins the strength of the wind-driven merit-order effect examined here [38]. This near doubling of variable renewable penetration over the decade provides the variation in VRE share that the present analysis exploits.

Wholesale prices over the same period are summarised in Figure 3. The annual mean price across the market rose from NZ$68/MWh in 2015 to NZ$197/MWh in 2024, averaging approximately NZ$113/MWh over the decade. The pronounced peaks in 2021 (NZ$168/MWh) and 2024 (NZ$197/MWh) coincide with dry-year hydro conditions and constraints in gas supply, periods in which thermal plant set the marginal price more often. An independent review of market performance attributes the elevated prices of the most recent dry year primarily to reduced gas availability and the consequent reliance on diesel and other thermal generation [39]. These price movements motivate the state-dependent treatment of the merit-order effect developed below.

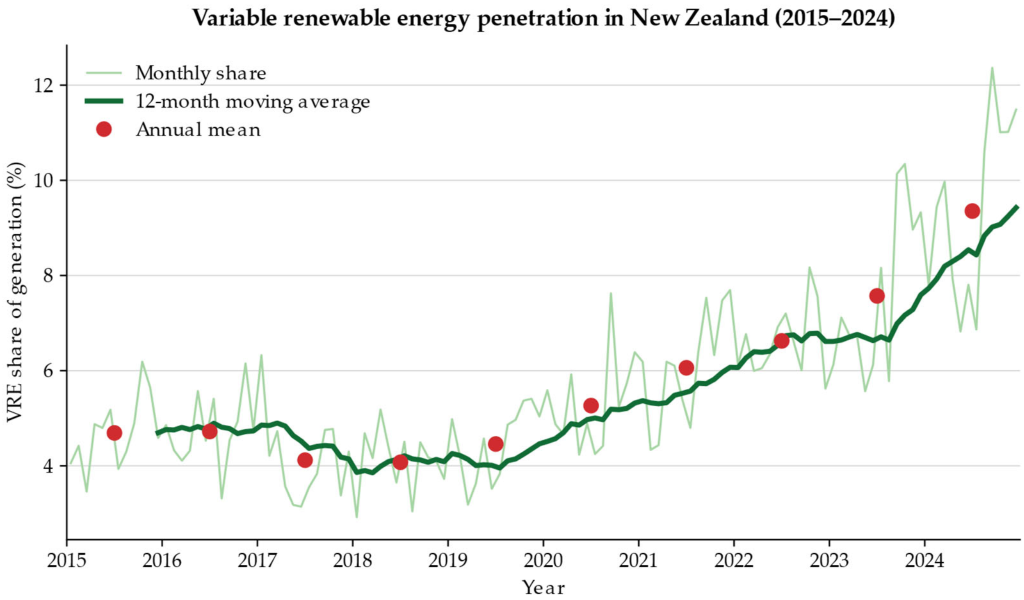

Figure 2. Penetration of variable renewable energy (combined wind and utility-scale solar) in the New Zealand electricity market, 2015–24, as a percentage of total generation. Source: Electricity Authority Electricity Market Information (EMI) [34].

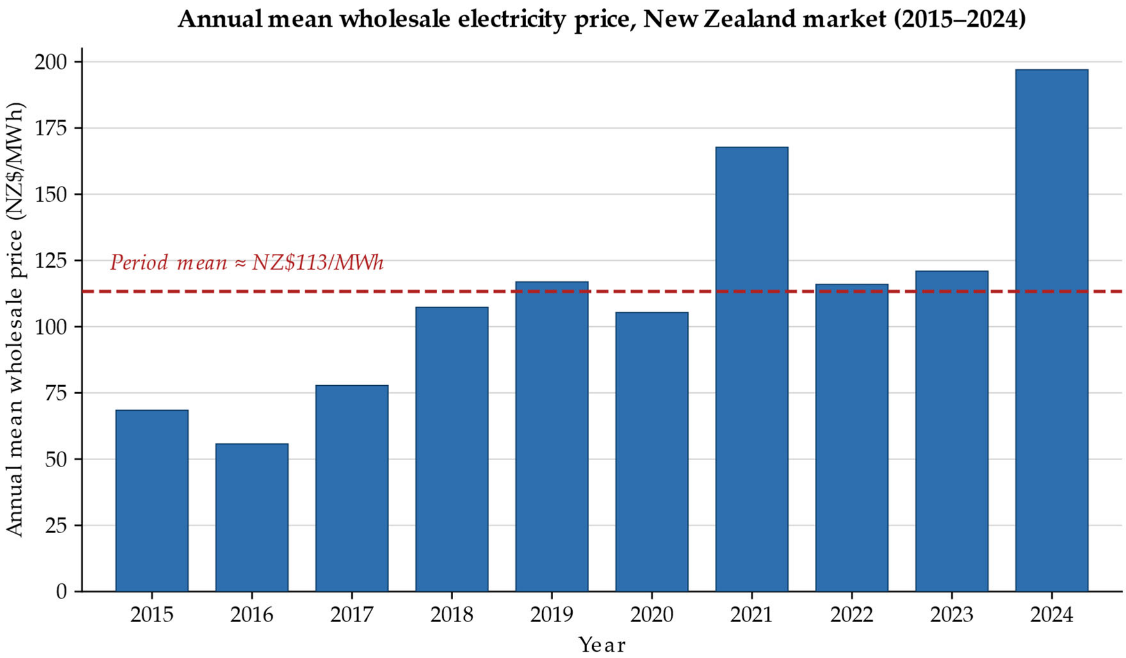

Figure 3. Annual mean wholesale electricity price across the New Zealand market, 2015–24 (NZ$/MWh). Source: Electricity Authority Electricity Market Information (EMI) [34].

The seasonal and interannual structure of price and renewable output is summarised in Figure 4 as year-by-month heatmaps. Prices are highest in the dry years 2021 and 2024 and in the autumn and winter months, when hydro inflows are low and thermal plants run more often. The renewable share is highest in spring and summer and rises steadily across the decade. These overlapping cycles in price, hydro availability, and renewable output are the conditions under which the merit-order effect is expected to change sign, motivating the state-dependent specification developed below.

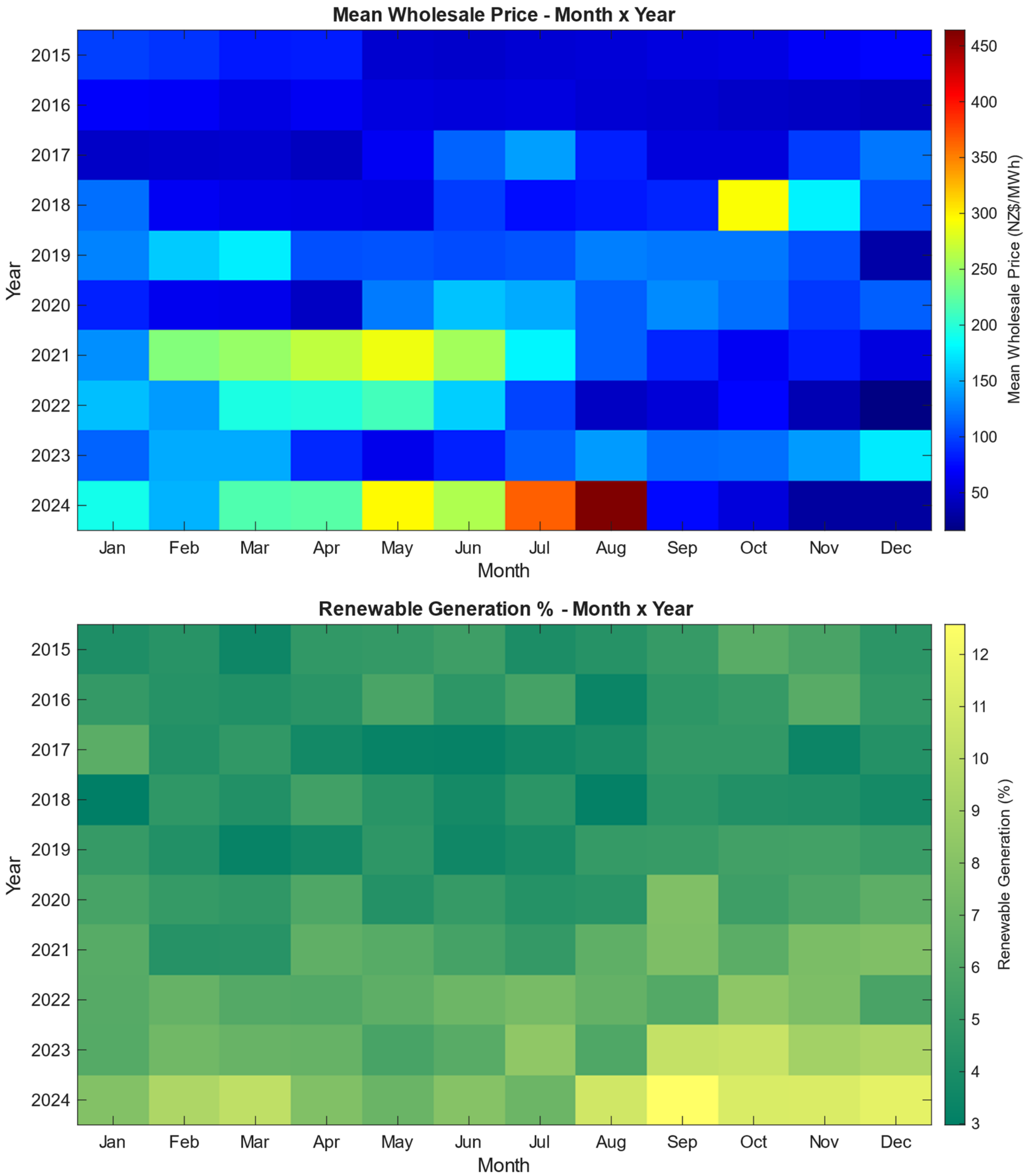

Figure 4. Year-by-month patterns in the New Zealand nodal market, 2015–24: (**top**) mean wholesale price (NZ$/MWh); (**bottom**) renewable generation share (%). Source: authors' computation from Electricity Authority EMI [34].

Two further patterns appear, and they are shown in Figure 5. Price volatility, measured by the within month standard deviation, tracks the same dry-year and winter peaks as the price level. The decade average price profile across the 48 daily trading periods shows the familiar morning and evening peaks, confirming that the intraday shape of price is stable even as monthly levels vary widely. Three monthly state variables are constructed. VRE share R is the wind plus solar percentage of generation (sample mean 5.83%). Fractional hydro storage S is the national stored energy relative to capacity (mean 0.535). The marginal fuel share THERM is the monthly percentage of generation from thermal plants, namely gas, coal, and diesel, computed from the unit-level generation files (mean 15.8%, range 2.8%

to 27.2%). THERM is low in wet, hydro-dominated months and high in dry or high demand months and is therefore a direct contemporaneous index of the price-setting regime.

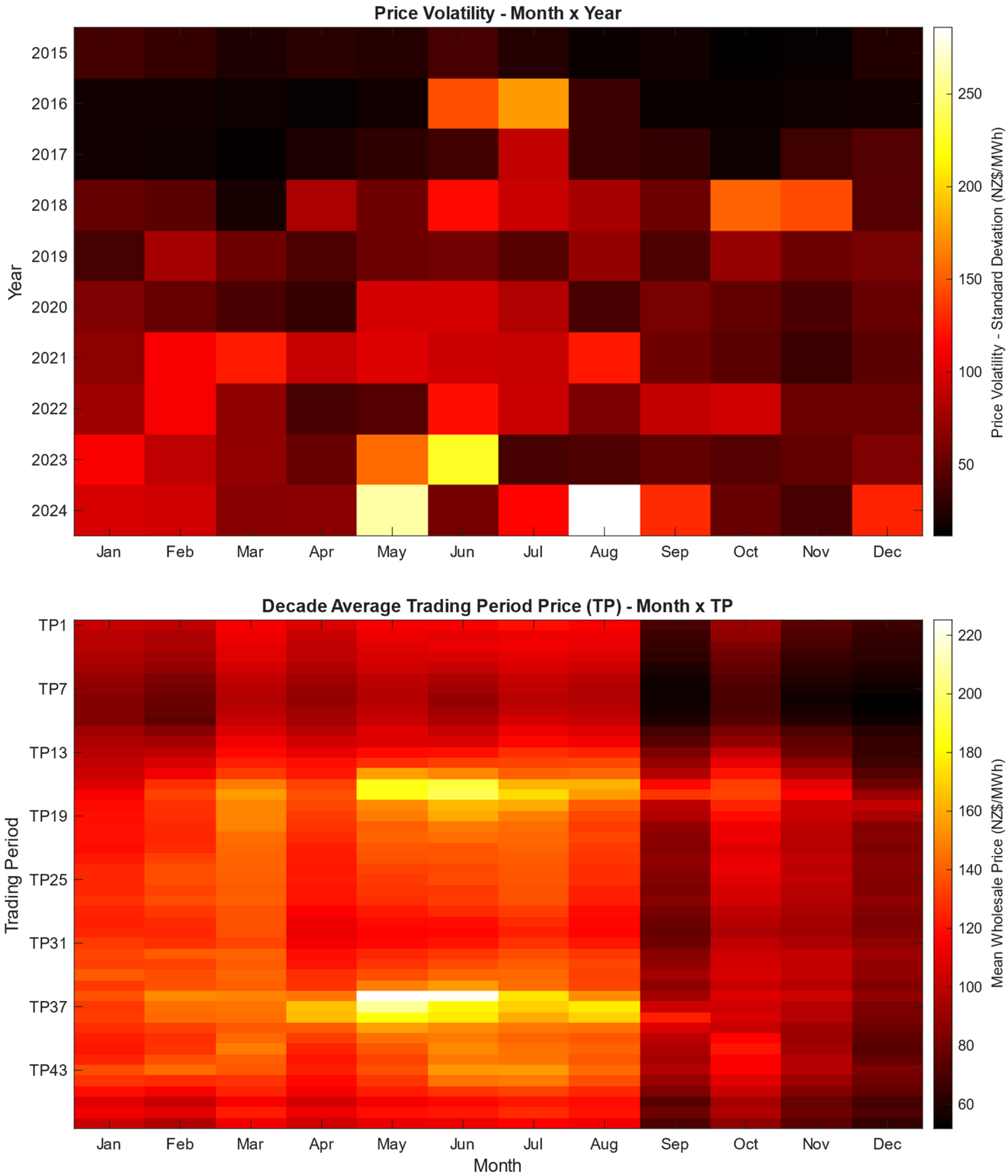

Figure 5. Further year-by-month patterns, 2015–24: (**top**) price volatility (within-month standard deviation, NZ$/MWh); (**bottom**) decade-average price profile by trading period and month (NZ$/MWh). Source: authors' computation from Electricity Authority EMI [34].

3.3. Model Specifications and Notation

Three model specifications are estimated. Model A is the baseline merit-order panel with VRE share only, Model B (HC-MOM) conditions the VRE coefficient on reservoir storage level (see Section 5), and the marginal fuel-conditioned merit-order model (MF-MOM) conditions it on the thermal generation share (see Section 6). Inference uses month-clustered standard errors throughout, supplemented in robustness checks by HC1 (White–Hinkley estimator) and Driscoll–Kraay standard errors. Out-of-sample validation uses leave-one-month-out and blocked time folds, a temporal hold-out (estimate on early months, predict on late months), and five-fold cross-validation on randomly held-out folds. Predictive accuracy is measured by the root-mean-square error (RMSE) in NZ$/MWh.

For clarity, we standardise notation across specifications, as follows. Model A denotes the baseline two-way fixed-effects panel with VRE share only (see Equation (2)). Model A1 refers to this same baseline estimated on the 2015–24 panel. Model A2 extends A1 by adding the conditioning variable as a main effect (reservoir storage or thermal share, depending on the specification). Model B denotes the interaction model in which the VRE coefficient is conditioned on the state variable (see Equation (3) for storage, Equation (5) for thermal share).

For the marginal fuel specifications in Section 6, we use the labels A2T (baseline plus thermal share) and BT (A2T plus the VRE $\times$ thermal share interaction) to distinguish them from the storage-conditioned models. These naming conventions are used consistently in all tables and out-of-sample comparisons.

A note on identification is warranted, because the VRE share R, storage S, and thermal share C are system-wide variables common to all nodes within a month. The time controls are therefore deliberately specified as additive year and calendar-month fixed effects, rather than saturated year-month effects. A saturated year-month specification, with one constant for each of the 120 sample months in the 2015–24 panel, would be perfectly collinear with any system-wide monthly regressor and would prevent identification of β, β_{int}, and γ altogether. Under the additive specification, the coefficients on the system-wide variables are identified from month-to-month variation, not absorbed by the year and seasonal dummies; that is, from within-calendar-month variation across years (a wet July vs. a dry July, for example) and from within-year variation across the seasonal cycle beyond its average pattern. The year effects additionally absorb the secular 2015–24 trends in prices and VRE penetration, so the interaction coefficient is not identified from the common upward drift of both series.

This constraint deserves emphasis. Although the estimation panel contains 28,827 node-month observations, R, S, and C each take a single value in any given month, so the effective independent variation identifying the system-wide coefficients is 120 monthly observations, not 28,827. The node dimension delivers precision for the node-specific intercepts and supports the spatial analysis (see Section 7 for details), but it does not multiply the identifying variation for the system-wide terms. Month-clustered standard errors are used throughout for this reason, and all inference on those coefficients should be read against an effective sample of 120. Re-estimating Equation (5) on the 120 national monthly means, with the same additive time controls and no node dimension, returns an interaction of +1.82 against +1.81 in the nodal panel, confirming that the two carry the same information.

4. Baseline Model and Benchmark Replication

The baseline merit-order model (Model A) is a two-way fixed-effects panel regression, following the non-spatial benchmark from [1] of nodal price on VRE share:

$$P_{i,t} = \alpha_i + \beta\, R_{i,t} + \lambda_t + \varepsilon_{i,t} \tag{2}$$

where $P_{i,t}$ is the nodal price at node i in period t (in NZ$/MWh), and α_i is the node fixed effect at node i. In plain terms, a fixed effect is a separate constant fitted for each node that captures that node's own baseline price level. For example, Otahuhu node, near the Auckland load centre, sits at a higher baseline than a generation-rich southern node. Because each node keeps its own constant, the model only explains how prices move around that baseline. β is the average VRE coefficient (the merit-order effect being estimated); $R_{i,t}$ is the system-wide VRE share, which is common across nodes within a given t. λ_t denotes additive time controls. Specifically, it includes year fixed effects, with a separate constant for each calendar year in the estimation window and the first year omitted as the reference. It also includes calendar-month fixed effects, with a separate constant for each month of the year and January omitted as the reference. The year effects capture market-wide shifts shared by all nodes, such as fuel price changes and demand growth, while the calendar-month effects capture average seasonality. This additive year-and-calendar-month structure applies to the monthly 2015–24 panel. The hourly 2011–12 benchmark of Section 4, estimated within single years and seasonal subsamples, instead controls for node, hour-of-day, weekday, and season, while the cross-period re-estimation of Section 6.7 uses node and year effects with the calendar month state entering through the centred thermal share regressor. In each case, the time controls are chosen so that the conditioning regressors remain identified at that sample's resolution. Lastly, $\varepsilon_{i,t}$ is the idiosyncratic error term.

Throughout, $i = 1, \ldots, 245$ indexes a transmission node and t a month of the 2015–24 panel (for the benchmark, the eight season years of 2011–12). The nodal price is in NZ$/MWh, and the VRE share R is system-wide, common to every node within a month. It is expressed as a percentage of total generation, so the VRE slope β is in NZ$/MWh per percentage point of VRE share. The node and month terms are also in NZ$/MWh, and the error term has zero mean. The implied price effect of a 10% relative VRE increase, evaluated at the sample mean VRE share, follows from $\Delta P = 0.1 \times \beta \times \bar{R}$, where $\bar{R}$ is the sample mean VRE share. We adopt this convention from [16] for direct comparability across specifications and panels.

The model deliberately leaves out any hydro-share term. The reason is discussed in detail in Section 4.1. At each node, the fixed effect already absorbs the part of hydro generation that does not change over time. The part that does change moves together with the same seasonal and storage signals that drive the VRE share R. The two are therefore too closely linked for the regression to separate their effects reliably; their correlation runs from -0.46 to -0.67. The storage-conditioned model in Equation (3) (see Section 5 for details) and the marginal fuel model in Equation (5) (see Section 6 for details) take the same approach. Each one replaces the dropped hydro-share term with a more informative conditioning variable, namely the storage level S in Equation (3) and the thermal share C in Equation (5).

4.1. Autumn Diagnostic

A specification diagnostic preceded estimation. Including both VRE-share (only wind in this case) and hydro-share regressors produced wrong sign coefficients in 5 of 6 autumn months on the 2011–12 panel. Before adjusting the model, three competing hypotheses for that anomaly were examined.

Hypothesis H1 is that a single anomalous month is pulling the average. Hypothesis H2 is that the two regressors are sufficiently correlated that the regression cannot separately identify their contributions (classic multicollinearity). Hypothesis H3 is that reservoirs reached the spill threshold, forcing water to be discharged independently of wind output, which would break the normal economic relationship between wind and hydro generation.

The data refute H1 and H3 and confirm H2. The slopes of the raw bivariate price vs. wind plots are negative for all 6 autumn months, as shown in Figure 6. So, the underlying relationship is intact; the opposite sign appears only when both regressors are in the fit together. The wind-share/hydro-share correlation across the 6 months is in the range -0.46 to -0.67 (strongly negative) [17], the regime of severe multicollinearity in which ordinary least square (OLS) cannot reliably split the two contributions. Storage peaked at 0.82 in autumn 2011 and 0.53 in autumn 2012, below the spill regime (>0.85) in both years. The H2 confirmation justifies the standard remedy, viz., dropping the hydro-share regressor for collinearity, after which the autumn wind slopes recover their correct negative sign for all 6 months.

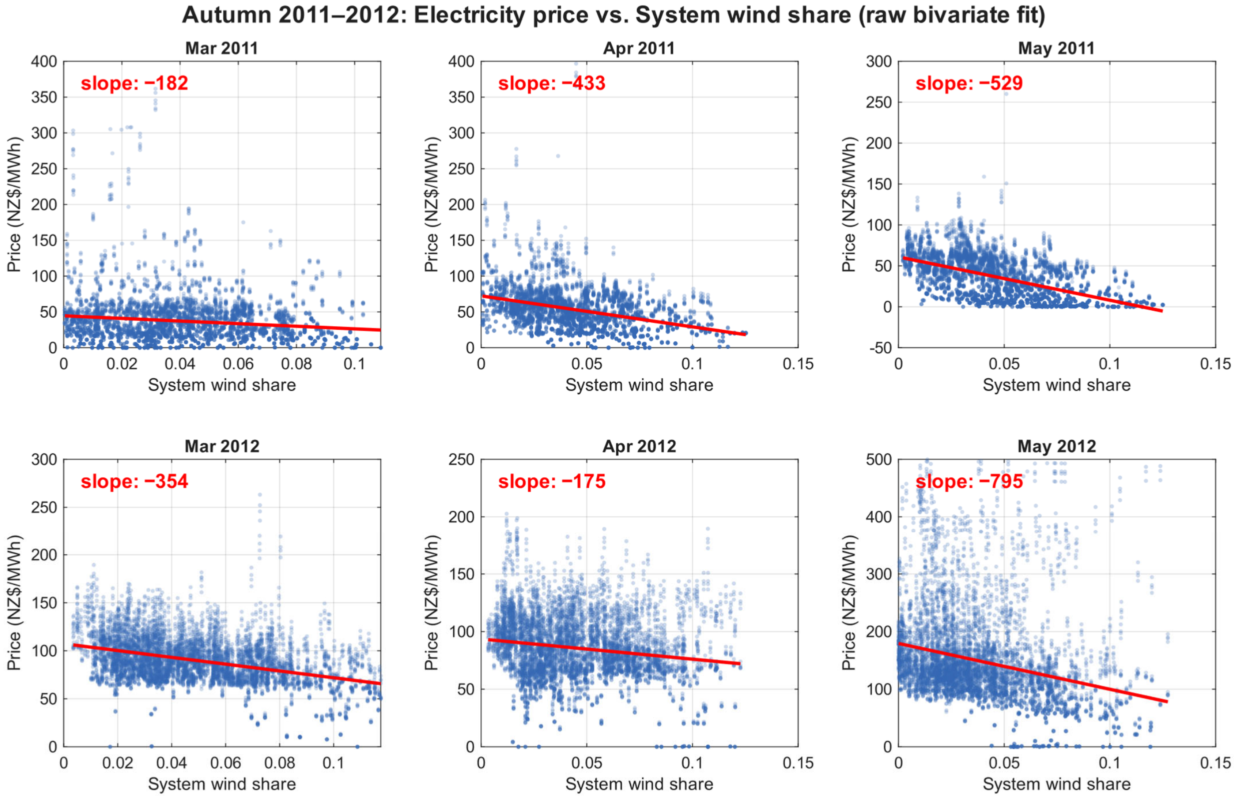

Figure 6. Raw nodal price plotted against system wind share for each of the 6 autumn months in 2011 and 2012, with no model adjustment.

4.2. Seasonal Match to the Benchmark

With the no-hydro specification, Model A reproduces the benchmark for all eight season year subsamples, as shown both in Table 2 and Figure 7. The sign of the implied price effect agrees in all eight, and the mean absolute gap to the published values is NZ$0.36/MWh across all eight, falling to NZ$0.22/MWh when the statistically insignificant summer 2011 subsample is excluded. These matching results establish that the estimation pipeline is sound and that subsequent findings are not mere artefacts of data handling.

Table 2. Seasonal MOE: benchmark [1] vs. Model A (2011–12, NZ$/MWh).

Season	State	Wen ΔP [1]	Our ΔP	$\lvert$Gap$\rvert$	Sign Agreement
Spring 2011	dry	−1.263	−1.555	0.292	✓
Summer 2011	wet	−0.040	−1.404	1.363	✓ *
Autumn 2011	wet	−0.829	−1.248	0.419	✓
Winter 2011	wet	−1.236	−1.077	0.159	✓
Spring 2012	wet	−1.340	−1.061	0.279	✓
Summer 2012	dry	−1.314	−1.352	0.038	✓
Autumn 2012	dry	−0.613	−0.858	0.245	✓
Winter 2012	dry	−0.428	−0.518	0.090	✓

* Summer 2011 is the only subsample the benchmark reports as insignificant. State labels follow [1].

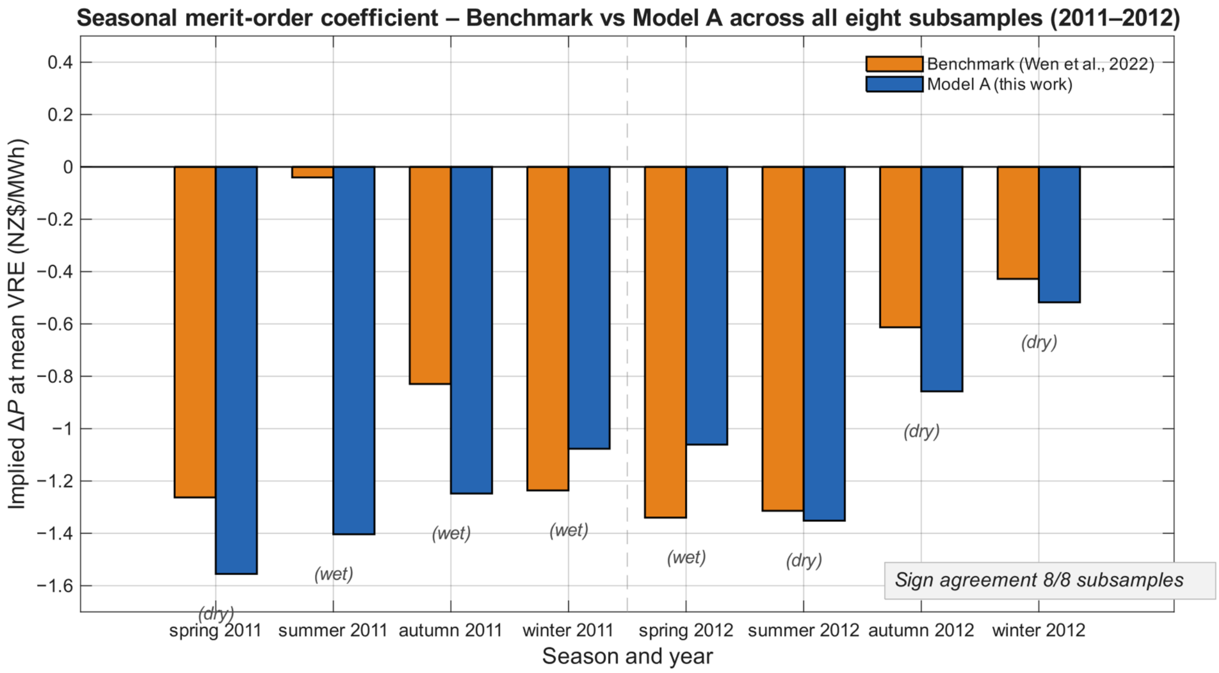

Figure 7. Seasonal implied price effect on the 2011–12 panel [1].

5. Storage-Conditioned Merit-Order Model

With benchmark replication established, the framework is extended to the 2015–24 panel. The most direct extension of the merit-order framework to a hydro-dominated market conditions the VRE coefficient on reservoir storage. This is Model B, in which storage S summarises the hydrological state and a VRE-by-storage interaction lets the price suppression effect differ between wet and dry conditions. The section is a falsification exercise as much as an estimation exercise. The model is evaluated in three stages, viz., in-sample fit and robustness, incremental out-of-sample value over a storage main effect, and temporal stability. A model that passes the first stage while failing the latter two may not be read as a stable merit-order relationship.

5.1. Model Specification

Centring storage at its sample mean, the storage-conditioned model is

$$P_{i,m} = \alpha_i + \beta_{base}\, R + \beta_{hyd}\, (\tilde{R} \cdot \tilde{S}) \; + \; \zeta\, \tilde{S} + \lambda_m + \varepsilon \tag{3}$$

$$\beta(S) = \beta_{base} + \beta_{hyd}\, (S - \bar{S}) \tag{4}$$

where $P_{i,m}$ is the nodal price at node i in month m, α_i is the node fixed effect (as in Equation (2)), β_{base} is the VRE coefficient evaluated at mean storage, R is the system-wide VRE share in month m, and β_{hyd} is the storage interaction coefficient (the rate at which the VRE coefficient changes with storage). $\tilde{R} = R - \bar{R}$ and $\tilde{S} = S - \bar{S}$ are the centred VRE share and storage level, respectively, where a bar denotes a sample mean, S is the national fractional hydro storage, and ζ is the main effect of storage on price at mean VRE share. λ_m denotes the additive year and calendar month fixed effects defined in Section 4, and ε is the idiosyncratic error. A negative β_{hyd} would mean the MOE strengthens as storage rises, viz., the suppression effect is larger in wet months than in dry ones.

Equation (3) is the estimated regression. Equation (4) re-expresses its coefficients as the implied VRE merit-order coefficient, $\beta(S)$, which is the quantity of interest. Storage S is the national fraction of usable reservoir capacity and is therefore a dimensionless variable bounded between 0 and 1. Accordingly, β_{base} is measured in NZ$/MWh per percentage point of VRE share, while β_{hyd} gives the change in that coefficient per unit of fractional storage. The storage main-effect coefficient ζ is measured in NZ$/MWh per unit of fractional storage. Because S is centred, β_{base} is the VRE coefficient at mean storage.

5.2. In-Sample Evidence

Three nested specifications are estimated on the 2015–24 panel (Table 3). Model A1 (VRE share only) yields a slope of −15.13, an implied NZ$8.80/MWh reduction per 10% relative VRE increase at the sample mean share, but it explains only 41% of the month-to-month variation in price. Adding storage as a main effect (Model A2) raises R^2 to 0.76 and flattens the VRE slope to −8.24, confirming storage as a strong price driver in its own right. Adding the VRE-by-storage interaction (Model B) raises R^2 to 0.81. The implied VRE coefficient at mean storage is −12.73, and the interaction is large and precisely estimated ($\beta_{hyd} = -47.22$, month-clustered t = −4.22). An F-test of Model B against A2 rejects the restriction $\beta_{hyd} = 0$ decisively (F = 8243). The implied wet–dry contrast is sharp and economically intuitive. At high storage ($S = 0.7$), the implied coefficient is −20.52, a NZ$11.94/MWh reduction per 10% relative VRE increase. At low storage (S = 0.2), the implied coefficient is +3.09, so the suppression effect is absent and marginally reversed (a nominal NZ$1.80/MWh). The wet–dry asymmetry is therefore NZ$13.74/MWh, considerably sharper than a season-by-season split [1]. Judged on in-sample evidence alone, the interaction would appear a real and sizeable improvement.

Table 3. Nested decomposition on the 2015–24 panel.

Model Specification	β	R^2	ΔP (NZ$)	Interpretation
A1 (VRE only)	−15.13	0.41	−8.80	Overstates; absorbs storage effect
A2 (A1 + Storage)	−8.24	0.76	−4.79	Understates; omits $R{\cdot}S$ interaction
B (A2 + $R{\cdot}S$)	−12.73	0.81	−7.41	Interaction model; evaluated at mean storage

ΔP = implied price effect of a 10% relative VRE increase at sample mean $\overline{R}$ = 5.83%. Model B has $\beta(S = 0.7) = -20.52$ (ΔP = −NZ$11.94) and $\beta(S = 0.2) = +3.09$ (ΔP = +NZ$1.80).

5.3. Internal Robustness

The estimated interaction survives the standard internal checks, summarised in Table 4. Permuting the storage series across months is a placebo that destroys the temporal alignment between storage and prices. It collapses the coefficient to −1.19 and removes its significance, so the baseline estimate is not a spurious timing artefact. Winsorising the top and bottom 1% of prices preserves the sign at a similar magnitude (−37.21). Dropping the ten most influential nodes leaves the estimate essentially unchanged (−47.20), and three standard error treatments (HC1, month-clustered, and Driscoll–Kraay) agree that it is well separated from zero. Nor is the interaction a numerical artefact of collinearity. Computed on the within-panel variation that identifies the coefficient, the variance inflation factor for the interaction ranges from 1.01 to 1.19 across rolling windows, with a condition number of 1.3, far below conventional thresholds. The interaction is therefore a genuine in-sample pattern. It is not driven by outliers, influential nodes, standard error choice, random timing, or collinearity. The difficulty, developed next, lies elsewhere. The signal is present, but it is not stable.

Table 4. Storage model: in-sample robustness (β_{hyd}).

Test	β_{hyd}	SE	t-Statistic
HC1 SE (baseline)	−47.22	0.79	−59.7
Month-clustered SE	−47.22	11.18	−4.22
Driscoll–Kraay SE	−47.22	9.87	−4.78
Winsorised 1/99	−37.21	0.49	−75.3
Drop top 10 nodes	−47.20	0.81	−58.5
Placebo (permute *S*)	−1.19	27.73	0.05

5.4. Out-of-Sample Predictive Value of the Storage-Conditioned Model

At first sight, the out-of-sample evidence also favours the model. Against the VRE-only baseline A1, Model B reduces prediction error by 43.4% in five-fold cross-validation and by 35.0% in a temporal hold-out (Table 5). These comparisons, however, answer the wrong question, because they credit the interaction with the predictive value of storage itself. The comparison that isolates the structural claim is Model B against A2, which already contains the storage main effect. Under that comparison, the support weakens decisively. The interaction improves the cross-validated prediction by 11.9% ($p = 0.15$) and the temporal hold-out by 5.0% ($p = 0.22$), and neither gain is distinguishable from zero. One caveat applies to these five-fold figures. Storage and VRE share take a single value per month, so randomly held-out node-months leave the same regressor values in the training set, and the design flatters every model in the table. It flatters them by a similar amount, so the comparison between them is preserved, and the verdict on the interaction rests on the stability evidence below rather than on predictive accuracy alone. Reservoir storage is an important determinant of price levels, but the VRE-by-storage interaction adds no significant predictive value once the main effect is included.

Table 5. Model B out-of-sample performance.

Design	A1	A2	B	A1→B Gain	A2→B Gain	*p* for A2→B
Five-fold CV	57.54	37.01	32.59	43.4%	11.9%	0.15
Temporal hold-out	110.50	75.61	71.83	35.0%	5.0%	0.22

Gains are percentage reductions in RMSE (NZ$/MWh) relative to the stated baseline.

5.5. Coefficient Stability and Diagnostic Summary

Model B is re-estimated based on rolling 36-month windows (Figure 8). The interaction does not drift smoothly. Windows centred around 2016 give positive estimates of roughly +35 to +50, windows centred around 2023 give strongly negative estimates between about −47 and −72, and the cross-window standard deviation is 57.4. A split-sample comparison gives the same verdict, with +58.21 over 2015–2019 against −50.76 over 2020–24, each precisely estimated within its own subsample. The pooled full-sample coefficient (−47.22) and the mean of the rolling estimates (approximately +25) even disagree in sign. This is possible only because the pooled regression weights sub-periods with opposing dynamics. The instability is not explained by rising VRE penetration, as the cross-window correlation between the rolling estimates and window-average penetration is only +0.17 (Figure 9). Table 6 collects the full diagnostic record. The storage interaction is real in-sample but time-varying in sign and magnitude. This is the signature of a conditioning variable whose mapping to the price-setting regime itself shifts over time, and no single time-invariant coefficient characterises it.

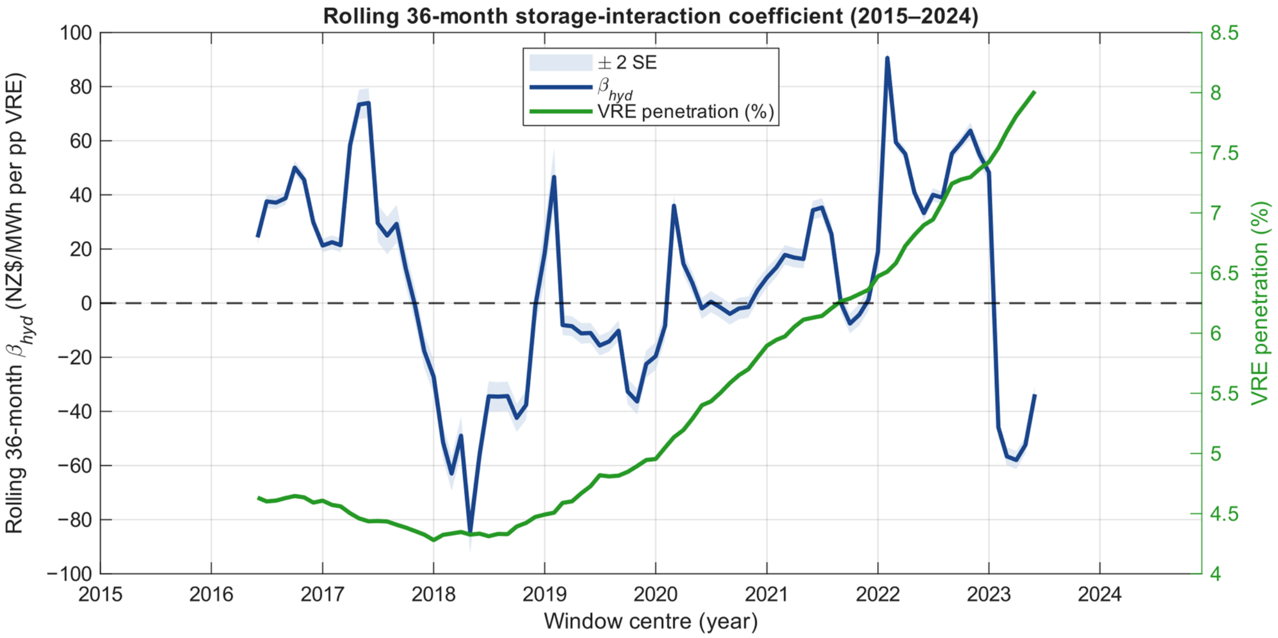

Figure 8. Rolling 36-month storage-interaction coefficient on the 2015–24 panel.

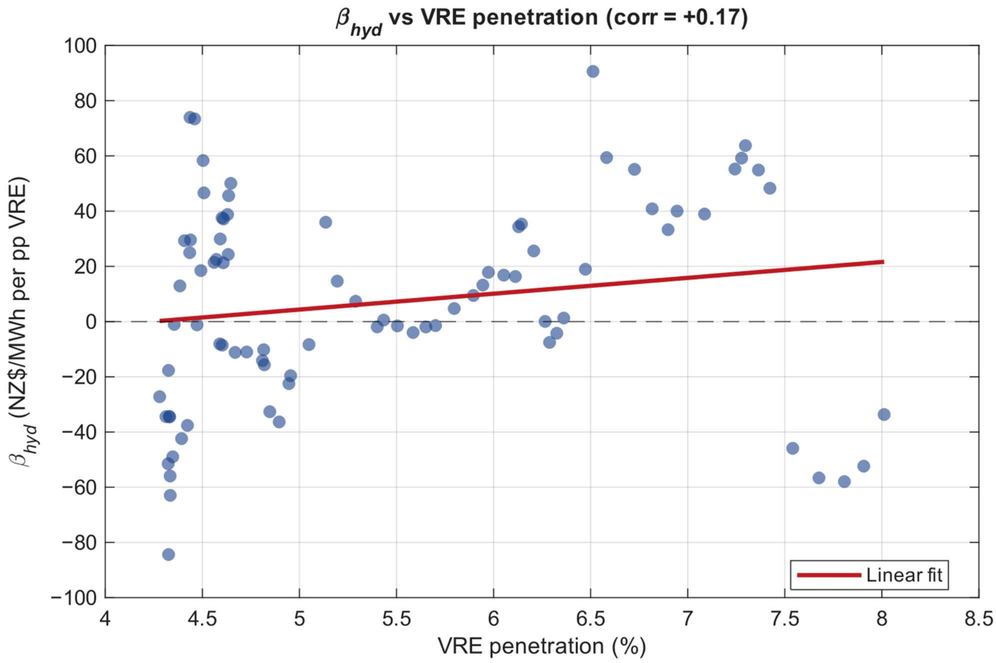

Figure 9. Rolling 36-month storage interaction vs. VRE penetration.

Table 6. Diagnostic summary for the storage-conditioned model.

Diagnostic	Result	Interpretation
In-sample fit	Within-R^2 rises to 0.81	Storage interaction improves fit
Placebo	Permuted storage gives a near-zero coefficient	Original signal is not random
Outlier checks	Estimate robust to Winsorisation and node removal	Not driven by extremes
Incremental CV over A2	+11.9%, $p = 0.15$	Interaction adds no significant predictive value

Table 6. *Cont.*

Diagnostic	Result	Interpretation
Temporal hold-out over A2	+5.0%, $p = 0.22$	Weak forward validation
Rolling windows	Coefficient changes sign	Not temporally stable
Collinearity checks	VIF $\approx$ 1.01–1.19	Instability is not numerical

5.6. Economic Diagnosis of Stock vs. Marginal Fuel

Why does conditioning on the reservoir level fail like this? The answer lies in what storage actually measures. Storage is a stock: a snapshot of how much water is in the lakes at the end of the month. What sets the marginal price, though, is a flow, namely which type of plant is running at the margin right now. The link from the stored-water level to the marginal technology is loose and indirect and changes over time. The same end-of-month storage level can be reached by drawing the lakes down quickly, so thermal plants are probably already setting the price, or by refilling them, so hydro is probably still on the margin. Wind and solar output, demand patterns, and how the hydro operators choose to release water all affect this link [40], and none of them stays constant across the decade.

This interpretation matches evidence from Spain, where hydropower frequently set the marginal price, yet reservoir levels alone did not explain the high prices observed in 2021–22. Prices were governed instead by the value of stored water and by the identity of the marginal unit [2]. The same mechanism explains the New Zealand result. Storage carries useful information, but it is not a stable proxy for the marginal technology, so tying the VRE coefficient to it produces a number that drifts with the loose stock-to-margin link rather than with any genuine property of the merit-order effect.

The remedy follows directly. We replace the indirect stock measure with a direct, same-month measure of which technology is on the margin. Storage fails because it is a stock proxy, whereas the thermal generation share is a contemporaneous flow proxy for the marginal fuel. That is the marginal fuel-conditioned model developed in Section 6.

6. Marginal Fuel-Conditioned Merit-Order Model

6.1. Motivation and Specification

We now replace the storage-based conditioning with thermal share, yielding the A2T and BT specifications. The diagnosis in Section 5 suggests that the conditioning variable should measure the contemporaneous price-setting regime rather than the reservoir stock. We therefore replace the storage level with the monthly thermal generation share, denoted C or THERM. THERM is the percentage of monthly generation supplied by gas, coal, and diesel units. It is not a direct observation of the marginal unit in every trading interval, but it is a dispatch-based proxy for the hydro-thermal regime. High-THERM months are more likely to be constrained by thermal costs or by water values, whereas low-THERM months are more likely to be hydro-abundant. This dispatch-based view aligns with fundamental price models that represent the clearing price through the marginal position of fuel-specific units in the supply stack [41]. The marginal fuel-conditioned model (MF-MOM) mirrors the storage specification, replacing the storage terms with the centred thermal share.

$$P_{i,m} = \alpha_i + \beta_{base}\, R + \beta_{int}\, (\tilde{R} \cdot \tilde{C}) \;+\; \gamma\, \tilde{C} + \lambda_m + \varepsilon \tag{5}$$

$$\beta(C) = \beta_{base} + \beta_{int}\, (C - \overline{C}) \tag{6}$$

where $P_{i,m}$, α_i, R, $\tilde{R}$, λ_m (the additive year and calendar-month fixed effects), and ε are as described in Equation (3), β_{base} is the VRE coefficient evaluated at mean THERM, and β_{int}

is the THERM interaction coefficient (the rate at which the VRE coefficient changes with the thermal share). $\tilde{C} = C - \bar{C}$ is the centred thermal shares, with C the monthly THERM share defined in Section 3, and γ is the main effect of THERM on price at mean VRE share. Because we do not observe the price-setting unit in every node-month, we treat THERM as a regime indicator rather than as an exogenous causal treatment. The coefficient β_{int} identifies how the VRE–price relationship varies across realised hydro-thermal dispatch states, not the causal effect of changing the thermal share with all else held fixed.

As before, Equation (5) is the estimated regression, and Equation (6) gives the implied VRE merit-order coefficient $\beta(C) = \beta_{base} + \beta_{int}\,(C - \bar{C})$ as a function of the thermal share. The thermal share C is expressed as a percent, so β_{int} is in NZ$/MWh per percentage point of VRE share per percentage point of thermal share, and γ is in NZ$/MWh per percentage point of thermal share. Because C is centred, β_{base} is the VRE coefficient at the mean thermal share, and $\beta(C)$ rises with C, so the merit-order effect strengthens when thermal plants set the price more often.

The reduced-form reading of THERM is what the robustness analysis in Section 6.3 supports and is deliberately more modest than a structural claim. We do not assert that exogenously raising the thermal share by one point would change the VRE coefficient by β_{int}. We show, rather, that the VRE price suppression effect varies systematically across the hydro-thermal dispatch states realised in New Zealand during 2015–24. This variation is associated with the marginal fuel regime and is not accounted for by reservoir level, its square, national demand, or a formal comparison with the storage-conditioned interaction. Because a valid external instrument for the thermal share is not available, meaning any candidate such as inflows, demand, or outages affects price through channels other than the marginal fuel, we do not pursue an instrumental-variables specification. The observable control battery of Section 6.3 is the appropriate test for a state-conditioned reduced-form relationship of this kind.

Two features of the half-hourly data support treating the regime as a monthly object. First, the within-month dispersion of the half-hourly thermal share is small relative to its movement across months. Our analysis shows that 78% of the variance of the half-hourly thermal share over 2015–24 lies between months rather than within them. The two most extreme months of the sample do not overlap at all (5th–95th percentile ranges of 1.4–4.9% in the wettest month, December 2024, against 22.9–32.4% in the driest, January 2018). Moreover, daily crossings of a 5% thermal share line are rare (median zero in both months). The hydro-thermal regime is therefore persistent at the monthly scale rather than switching intraday. Second, the monthly index maps monotonically onto the prevalence and depth of thermal operation. Months with THERM below 5% have thermal plants contributing more than 5% of generation in only about one trading period in eight. By contrast, months with THERM above 12% are thermal plants running in essentially every trading period, so the index measures how deeply the system dispatches into the thermal stack.

A frequency-sensitivity check addresses whether the state dependence is an artefact of monthly aggregation. Rebuilding the national series at weekly resolution over the same window gives 520 weeks, with a mean VRE share of 5.72%, a mean thermal share of 15.80%, and a mean price of NZ$113.6/MWh, closely matching the monthly aggregates and confirming that the two resolutions describe the same data. Re-estimating Equation (5) at weekly frequency, with year and calendar-month effects and standard errors clustered by calendar month, returns an interaction of +1.15 ($t = 4.85$). Sign and significance are retained, and the implied VRE coefficient in hydro-abundant conditions is essentially unchanged at -10.20 against -9.92 at a thermal share of 8.8%. The magnitude attenuates by about a third, which is the expected direction. This is because 78% of the variance in the half-hourly thermal share lies between months rather than within them; as reported above,

https://doi.org/10.3390/en19163747

a weekly conditioner carries more within-regime noise, and classical measurement error biases the interaction toward zero. The monthly specification is therefore the appropriate resolution for the regime being conditioned upon, and the weekly estimate can be read as a conservative lower bound.

Finally, estimated at the market's native half-hourly resolution within these two months, the VRE–price slope is negative and significant in both (−2.24, day-clustered t = −2.06, in the wet month; −5.54, t = −3.32, in the dry), so no positive instantaneous effect underlies the monthly estimates. The stronger instantaneous slope against a thermal margin, alongside the stronger monthly-level suppression in hydro-abundant states, is consistent with intertemporal hydro optimisation. That is, hydro smooths prices within the month but cannot relax the energy constraint across months, which is why the regime-conditioned effect is identified at the monthly horizon.

6.2. Estimated State Dependence

The full-sample MF-MOM estimates are reported in Table 7. With month-clustered standard errors, the thermal interaction coefficient is positive and statistically significant: β_{int} = +1.81 (t = 5.0). The thermal share main effect is also positive and significant, at γ = +10.0 (t = 11.9), while the base VRE coefficient is not statistically different from zero, at β_{base} = +2.66 (t = 0.58). The within-R^2 is 0.80.

Table 7. MF-MOM estimates and implied VRE coefficient.

Quantity	Estimate	(t, Clustered)	Implied ΔP ($\bar{R}$ = 5.83%)
β_{base} (R)	+2.66	0.58	—
β_{int} ($R \cdot C$)	+1.81	5.00	—
γ (C)	+10.00	11.90	—
β (THERM = 8.8%, low/hydro-dominated)	−9.92	—	−NZ$5.79/MWh
β (THERM = 15.8%, mean)	+2.65	—	+NZ$1.55/MWh
β (THERM = 23.1%, high/thermal-dominated)	+15.88	—	+NZ$9.26/MWh

Month-clustered standard errors, within-R^2 = 0.80. Coefficients are in NZ$/MWh per percentage point of VRE share. $\beta(C)$ is the implied VRE coefficient at each thermal share state. ΔP is the implied price change for a 10% relative VRE increase at $\overline{R}$ = 5.83%.

The insignificant base coefficient does not undermine the model. Since C is centred, β_{base} is the VRE effect at the mean thermal share state. The key hypothesis is not that this mean-state coefficient differs from zero, but that the VRE effect changes with the marginal fuel state. This is precisely what the significant β_{int} tests.

The implied VRE coefficient increases from −9.92 at low thermal share (8.8%) to +2.65 at the mean thermal share (15.8%) and to +15.88 at high thermal share (23.1%), as shown in Figure 10. In price terms, a 10% relative increase in VRE share lowers prices by approximately NZ$5.79/MWh in the hydro-dominated state. As thermal share rises, this suppressive effect weakens and becomes statistically indistinguishable from zero.

The positive high-THERM values are interpreted cautiously. They do not imply that VRE causally raises prices in high thermal share months. Rather, the robust finding is that the VRE coefficient changes across regimes. At high THERM, the merit-order effect attenuates as the system becomes more water value-constrained. The appropriate interpretation is therefore the weakening of price suppression, not a causal price-increasing effect of VRE.

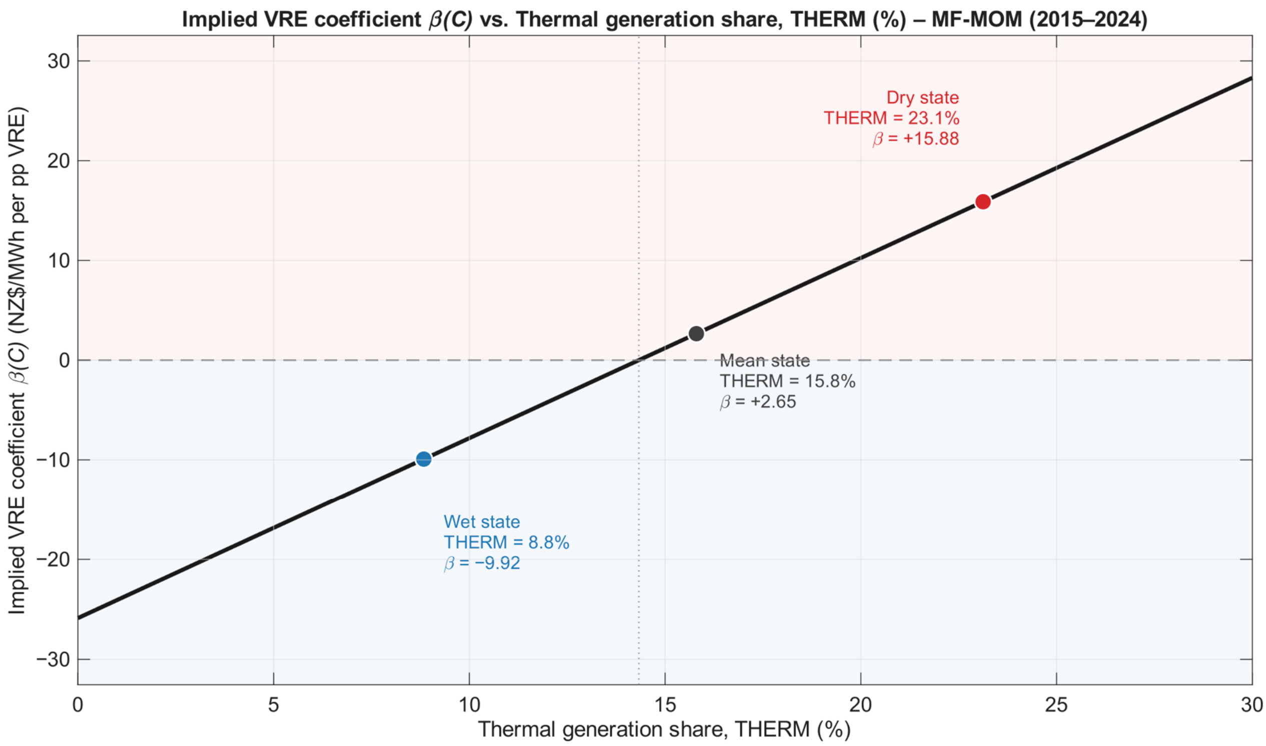

Figure 10. Implied VRE merit-order coefficient by marginal fuel state. Price suppression is strongest at a low thermal share and fades as the thermal share rises.

6.3. Robustness of the Interaction to Hydrological, Demand, and Simultaneity Controls

A natural objection to the marginal fuel interpretation is that the thermal share and the price may co-move because both respond to the same hydrological deficit. Dry months raise the thermal share and raise prices simultaneously, so a positive β_{int} might absorb the joint response of THERM and price to water scarcity rather than a genuine state dependence of the MOE. The concern is sharpened by the fact that the thermal share and reservoir storage are, as expected, negatively correlated over the sample (month-level correlation -0.60). Because the additive year and calendar-month effects absorb only the average annual and seasonal components, month-specific hydrological shocks common to all nodes are not removed and could in principle load on THERM and its interaction. We therefore re-estimate the full model under a sequence of controls that separate the marginal fuel channel from the hydrological, demand, and simultaneity channels. The results are reported in Table 8. All specifications include node fixed effects and additive year and calendar-month fixed effects. Standard errors are clustered by month and N (node month quantity) = 28,827 for specifications 1–5 and 28,587 in specification 6.

The interaction coefficient is stable in sign and magnitude across every contemporaneous specification. Adding reservoir storage as a main effect leaves β_{int} at +1.69 (t = 5.3), adding a quadratic storage term leaves it at +1.61 (t = 6.8), and controlling for national demand through the log of total generation leaves it essentially unchanged at +1.81 (t = 5.0). In each case, the storage main effect is large and precisely estimated (for example, -163, t = -6.1, in the linear specification of Table 8), confirming that reservoir level is a powerful driver of price level. What storage does not do is displace the interaction: the sensitivity of the VRE coefficient to the marginal fuel state is not an artefact of the water level itself.

The decisive test places the two interactions in direct competition. Specification 5 includes both the marginal fuel interaction $R{\cdot}C$ and a storage interaction $R{\cdot}S$, forcing them to compete for the same within-panel variation. The marginal fuel interaction retains its sign and significance (+1.26, t = 3.7), whereas the storage interaction is small and statistically indistinguishable from zero (−19.5, t = −1.4). The variance inflation factor for $R \cdot C$ remains well below conventional thresholds (1.98), so the result is not a collinearity artefact. When the drought level and the marginal fuel regime are allowed to compete, the data retain the marginal fuel regime and discard the storage interaction.

Table 8. Robustness of the marginal fuel interaction β_{int}.

Sr. #	Specification	β_{int}	SE	t	Storage Main (t)	Within-R^2
1	Baseline (Table 7)	+1.81	0.36	5.00	—	0.803
2	+storage level S	+1.69	0.32	5.26	−163.4 (−6.1)	0.847
3	+storage + S^2	+1.61	0.24	6.77	−173.4 (−6.7)	0.861
4	+demand, log (total generation)	+1.81	0.36	5.01	—	0.805
5	+S + $R \cdot S$	+1.26	0.34	3.67	−161.9 (−6.5)	0.852
6	Lagged conditioner, THERM (t − 1)	+1.46	0.75	1.95	—	0.604
7	+national inflow control	+1.82	0.37	4.95	—	0.803
8	Excluding top 5% price months	+1.12	0.23	4.85	—	0.774

Although thermal share and hydrological scarcity are correlated, their effects on the merit-order coefficient are empirically separable, and it is the marginal fuel regime, not the reservoir level, that conditions the VRE price suppression effect. The single exception to uniform significance is the specification replacing the contemporaneous thermal share with its 1-month lag (specification 6, +1.46, t = 1.95). A lagged conditioner is, by design, a weaker proxy. It discards the contemporaneous dispatch signal, so a lower within-sample fit is expected. This attenuation therefore does not undermine the interpretation. If anything, it reinforces the view that the contemporaneous marginal fuel state is what governs the VRE price suppression effect. We report the lagged specification for completeness, rather than as a preferred model.

Specification 3 additionally includes S^2 (+216, t = 2.6), specification 5 additionally includes the storage interaction $R \cdot S$ (−19.5, t = −1.4), and the demand specification's log-generation coefficient is −185 (t = −1.0). A storage anomaly control (the deviation of storage from its calendar-month normal) coincides with specification 2 because the calendar-month fixed effects absorb the seasonal normal. Specification 7 adds the national monthly inflow, which is reconstructed from the ten major storage lakes and hydro generation by monthly water balance and standardised. Its own coefficient is −3.73 (t = −1.15), so inflow carries no independent price signal once the node and time effects are present. Specification 8 excludes the 6 months whose mean price falls in the top 5%, retaining 114 of 120 months. The interaction remains significant but falls to +1.12, so extreme price months carry part of the estimated effect, while its sign and significance do not depend on them.

6.4. Stability Relative to Storage Conditioning

The marginal fuel interaction is substantially more stable over time than the storage interaction. Its sign is consistent across sub-periods (+1.09 in 2015–19, +1.57 in 2020–24), and its rolling-window dispersion is much smaller: a standard deviation of 2.46 against 57.4 for the storage interaction (see Table 8). Because the two interaction coefficients are measured on different scales, the comparison should be interpreted cautiously and primarily qualitatively in descriptive terms. The thermal interaction retains a consistent sign and remains tightly bounded across sub-periods, whereas the storage interaction changes sign. As shown in Table 9, the issues that undermined the storage model, which are sign reversal across sub-periods and substantial variation in rolling estimates, are not observed here.

Table 9. Stability: storage vs. marginal fuel conditioning.

Diagnostic	Storage (*S*)	Fuel (THERM)
Interaction sign	flips	consistent
2015–19/2020–24	+58/−51	+1.09/+1.57
Full-sample t	−4.22	+5.00
Rolling dispersion	57.4	2.46

6.5. Out-of-Sample Predictive Value of the Marginal-Fuel Interaction

The leakage caveat noted in Section 5.4 applies here too: because R and THERM are system-wide monthly variables, random-fold cross-validation overstates absolute accuracy. We retain it for continuity but treat the time-ordered designs as primary. Table 10 evaluates the models under the same strictly time-ordered protocols on the estimation panel of 245 nodes and 28,827 node-months. Adding *R*.THERM improves prediction by +14.2% under leave-one-month-out folds, by +17.0% under blocked five-fold time folds after level anchoring, and by +18.7% in the temporal hold-out, against +15.8% under the random folds originally reported (clustered $p = 0.048$). The gain is stable across protocols. Leakage inflates every model by a similar amount and so does not disturb the comparison between them. The BT error rises by about a quarter, from NZ$33.5/MWh to NZ$42.1/MWh, while its improvement over A2T is essentially unchanged. The model is therefore not merely a better in-sample fit, but a better predictor of held-out nodal prices under every design that respects the time ordering of the panel.

Table 10. Out-of-sample RMSE comparison for the marginal fuel interaction.

Design	A1	A2T	BT	Δ%
Five-fold CV (leakage-prone)	57.5	39.6	33.3	+15.8
Leave-one-month-out	69.8	49.0	42.1	+14.2
Blocked five-fold (raw)	101.3	74.9	72.5	+3.2
Blocked five-fold (level-anchored)	71.4	50.0	41.5	+17.0
Temporal	110.5	81.1	65.9	+18.7

Leave-one-month-out holds out each of the 120 months from January 2015 to December 2024, estimating on the remaining 119. The blocked folds hold out 2015–16, 2017–18, 2019–20, 2021–22, and 2023–24, in turn. The temporal hold-out estimates on January 2015 to December 2021 and tests on January 2022 to December 2024. For a held-out year, the year effect is not identified, so it is anchored to the training mean year effect, the convention already applied to the walk-forward and forward tests below. Table 10 therefore reports the blocked design in raw and level-anchored form; the mean absolute per-block bias is NZ$47.6/MWh for A2T and NZ$48.2/MWh for BT.

The blocked-raw column is the one design under which the interaction adds little, with the gain falling to +3.2%. That design requires each model to predict a 2-year era it has never observed, so its error is dominated by the shift in price level between eras rather than by within-era dynamics, as the bias figures indicate. No monthly conditioning variable claims to explain movements in fuel and carbon costs between eras, and we draw no conclusion from that comparison beyond what it measures.

Every design in Table 10 draws its test observations from within the 2015–24 sample. Two further designs respect the time-ordered structure of the panel directly. The first is an expanding origin walk-forward. Here, the model is estimated on all months from January 2015 through December of year $t - 1$ and used to predict every month of year t, for t = 2020–24, so each test year is predicted using only information available before it. As Table 11 shows, the interaction coefficient re-estimated on each training window is positive in four of the five windows (+0.71 to +1.64). The exception (−0.31 on training window 2015–20) is close to zero. Forward accuracy varies with how far the test year departs from the training distribution. The monthly tracking correlation between predicted and realised prices is 0.88–0.97 for the 2021, 2022, and 2024 test years, but 0.34 in 2020 and 0.52 in 2023, and the level-anchored RMSE ranges from NZ$24 to NZ$83/MWh. The pattern anticipates the regime evolution discussed in Section 6.7, and the state-dependent slope structure transfers forward more reliably than the price level. Tracking correlation is between predicted and realised monthly mean prices; anchored metrics remove each test window's mean level error and are therefore comparable across rows.

Table 11. Time-ordered validation of MF-MOM: expanding origin walk-forward (2020–24) and the 2025–26 forward test.

Test Window	Train Window	β_{int} (Train)	Track Corr.	RMSE Anchored (NZS/MWh)	Level R^2 (Anchored)
2020	2015–19	+0.71	0.34	36.9	0.15
2021	2015–20	−0.31	0.97	61.4	0.72
2022	2015–21	+1.64	0.95	24.4	0.88
2023	2015–22	+1.33	0.52	40.9	−0.26
2024	2015–23	+0.73	0.88	83.2	0.74
2025–May 2026 (forward)	2015–24	+1.81	0.94	34.0	0.87

The second is a true forward test on data which fall entirely after the period used for model development. The pooled 2015–24 coefficients are frozen and used to predict monthly nodal prices from January 2025 to May 2026 (17 months, 3982 matched node-months), a window whose thermal share traverses nearly the full historical range (2.7% to 22.3%, against 2.8% to 27.2% in training). Reservoir storage records are published only through December 2024, so the storage-conditioned model cannot be evaluated on this window at all, whereas the marginal fuel model requires only generation and price records. The frozen model tracks the month-to-month movement of prices closely (monthly tracking correlation 0.94) with a modest upward level bias (+NZ$21/MWh, anchored RMSE NZ$34/MWh, anchored level R^2 = 0.87).

The decisive comparison again isolates the interaction. Both A2T and BT are estimated on the training sample and then held fixed for prediction in the forward validation window. Under this design, adding the *R*.THERM term improves forward RMSE by 33.0%. The improvement is statistically significant, with a month-clustered p = 0.031 and a wild cluster bootstrap p = 0.047 across 17 month clusters. This forward validation gain is larger than the corresponding contribution in the within-sample designs reported in Table 10.

Re-estimating the model on the forward validation window alone is necessarily weakly identified. The window contains only 17 monthly observations, while the specification includes 12 additive time dummies. The resulting interaction estimate is −2.9, but it is statistically indistinguishable from zero under wild cluster bootstrap inference (p = 0.45). We therefore do not interpret its sign.

Pooling January 2015 to May 2026 gives 137 months and 32,809 node-month observations. In this extended sample, β_{int} = +1.50, with t = 5.5. The coefficient has the same sign as in the development sample and is about 17% below the development sample estimate of +1.81. This result is consistent with the time-varying magnitude documented in Table 12 of Section 6.7.

Table 12. Cross-period MF-MOM coefficient comparison.

Quantity	2011–12 Panel	2015–24 Panel
N (node-months)	240	28,827
Nodes/months	10/24	245/120
Mean VRE share	4.5%	5.8%
Mean THERM share	24.6%	15.8%
β_{base} (R)	−5.44 (t = −0.99)	+2.66 (t = +0.58)
β_{int} $(R{\cdot}C)$	−1.87 (t = −2.35)	+1.81 (t = +5.00)
γ (C)	+12.85 (t = +3.53)	+10.00 (t = +11.90)
Within-R^2	0.65	0.80

6.6. Economic Mechanism

The estimated pattern is consistent with hydro-thermal price theory, but it differs from the simple thermal-system merit-order intuition. In a purely thermal system, VRE would be expected to suppress prices most strongly when expensive thermal units are marginal. In New Zealand, however, a high thermal share often signals a different condition, namely hydro scarcity and a high water value [11,13,40,42].

When THERM is low, the system is hydro-abundant. Stored water is relatively plentiful, the water value is low, and flexible hydro plants can be displaced by wind and solar without tightening the intertemporal energy constraint [37]; under this regime, additional VRE lowers the clearing price strongly. When THERM is high, the system is more likely to be energy-constrained. Thermal plants are running because hydro energy is scarce or being conserved, and the relevant price is then governed not only by the current dispatch stack, but also by the opportunity cost of water [39]. A marginal increase in VRE may reduce dispatch in the current month, but it does not necessarily relax the multi-period water constraint, so the estimated price suppression effect weakens and becomes statistically indistinguishable from zero.

This is why the thermal share is a better conditioning variable than storage. It directly reflects the realised dispatch regime, whereas reservoir level records the stock of water but not whether the system is drawing it down, conserving it, or replenishing it [2].

6.7. Cross-Period Re-Estimation

The purpose of this exercise is not to require identical coefficients across the two periods. The 2011–12 panel represents an earlier market regime, with lower VRE penetration, a higher average thermal share, and a different role for the thermal fleet. The aim is instead to test whether the form of the state dependence is stable across market regimes.

A natural test of whether MF-MOM captures a structural feature of the New Zealand market is whether the same specification, re-estimated on the 2011–12 panel used by Wen et al. [1], returns coefficients of similar sign and magnitude to the 2015–24 fit. We constructed the monthly thermal share for 2011–12 using the same procedure as in Section 3 (mean THERM = 24.6%, range 16.5–37.5%, against 15.8%/2.8–27.2%, respectively for 2015–24). We then re-fit Equation (5) on the 240 node-month observations (10 nodes $\times$ 24 months) with node and year fixed effects and month-clustered standard errors over 24 clusters.

The 2011–12 fit yields $\beta_{int} = -1.87$ (clustered SE 0.80, t = −2.35), $\gamma = +12.85$ (t = +3.53), $\beta_{base} = -5.44$ (t = −0.99), and within-R^2 = 0.65. The interaction estimate is opposite in sign to the 2015–24 result (see Table 12). The asymptotic clustered t-statistic suggests significance at the conventional 5% level (two-sided $p = 0.019$), but this panel contains only

24 monthly clusters, so conventional asymptotic inference is likely to overstate precision. To address this, we apply a wild cluster bootstrap-t procedure (Cameron–Gelbach–Miller, restricted under the null $\beta_{int} = 0$, 4999 Rademacher replications). The bootstrap yields a two-sided p-value of 0.053 (one-sided $p = 0.027$ against the directional alternative $\beta_{int} < 0$). With this more conservative test, the evidence for a negative interaction is best viewed as borderline rather than decisive. The point estimate therefore remains directionally opposite to the later-period fit, but sits at the boundary of significance under small-cluster inference. The baseline coefficient β_{base} is not statistically different from zero under either approach (bootstrap two-sided $p = 0.40$).

The two periods imply opposing $\beta(C)$ curves. In 2015–24, β rises with THERM. A strong negative MOE at low thermal share (hydro-dominated, wet conditions) attenuates toward zero, and is no longer distinguishable from zero, at high thermal share. In 2011–12, β falls with THERM. Weak (insignificantly positive) effects at low thermal share strengthen into substantial negative MOE at high thermal share, the textbook merit-order pattern in which VRE displaces thermal at the margin. The two specifications are consistent only at the qualitative level. Both report a point estimate of state dependence. The 2015–24 estimate is robust, well identified, and out-of-sample-validated, while the 2011–12 estimate is marginal under proper small cluster inference; the form of that point estimate dependence has changed.

The sign reversal is consistent with structural evolution in the New Zealand market. In 2011–12, lower VRE penetration (about 4.5% wind only, against 5.83% wind plus solar in 2015–24) and a steadier thermal baseload made the textbook displacement mechanism more visible. When thermal generation was higher, wind more directly displaced fuel-cost plants. By 2015–24, with a more variable, peaking-oriented thermal fleet and more frequent hydro-storage stress, a high thermal share more often signals hydro scarcity and high water value. In that regime, VRE does not fully relax the binding intertemporal water constraint, so the price suppression effect weakens as THERM rises.

The pattern is consistent with the broader literature on time-varying merit-order effects [7,9]. The magnitude and the form of the MOE evolve with penetration and market structure. The implication for this paper is that MF-MOM, as specified, is a stable description of the current (2015–24) New Zealand market regime, not a panel-invariant structural model. The validation battery against storage-level conditioning in Sections 5 and 6.5 remains valid in its own terms. What it does not buy is unrestricted extrapolation backward in time, and we treat this limit as a feature to be reported (see Sections 9 and 10) rather than a finding to be hidden.

The marginal fuel specification therefore resolves the two failures of the storage-conditioned model. First, the interaction adds predictive value relative to a model that already includes the relevant main effect. Second, the interaction is stable across sub-periods within the 2015–24 regime. The improvement arises not from greater model complexity, but from replacing an indirect stock measure with a contemporaneous dispatch-based indicator of the hydro-thermal price-setting state.

7. Spatial Heterogeneity Across the Transmission Network

7.1. Motivation and Estimation Strategy

The pooled MF-MOM reported in Sections 6.2–6.7 imposes a single set of coefficients (β_{base}, β_{int}, γ) on all 245 nodes, identified separately from node-specific intercepts that the fixed effects absorb. This is the standard simplification, but it discards information. The MOE at the Auckland load centre may differ in strength from the same effect at a generation-rich southern node. Because the New Zealand market clears at nodal prices, inter-nodal price differences arise from transmission losses and congestion, so a coefficient estimated at one node need not hold at another [22,43]. To test whether the pooled estimate masks economically important spatial variation, we re-estimated MF-MOM separately at each node and mapped the results (see Figure 11). We then applied empirical-Bayes (EB) shrinkage and aggregated the results to the five Electricity Authority regional zones (Upper North Island (UNI), Central North Island (CNI), Lower North Island (LNI), Upper South Island (USI), and Lower South Island (LSI)) used in the Authority's wholesale market reporting. Empirical-Bayes shrinkage is a smoothing step. A node with only a few months of data, whose own estimate is therefore unreliable, is pulled toward the regional average. A node with many months of data keeps its own estimate. This prevents a few poorly measured nodes from producing extreme values.

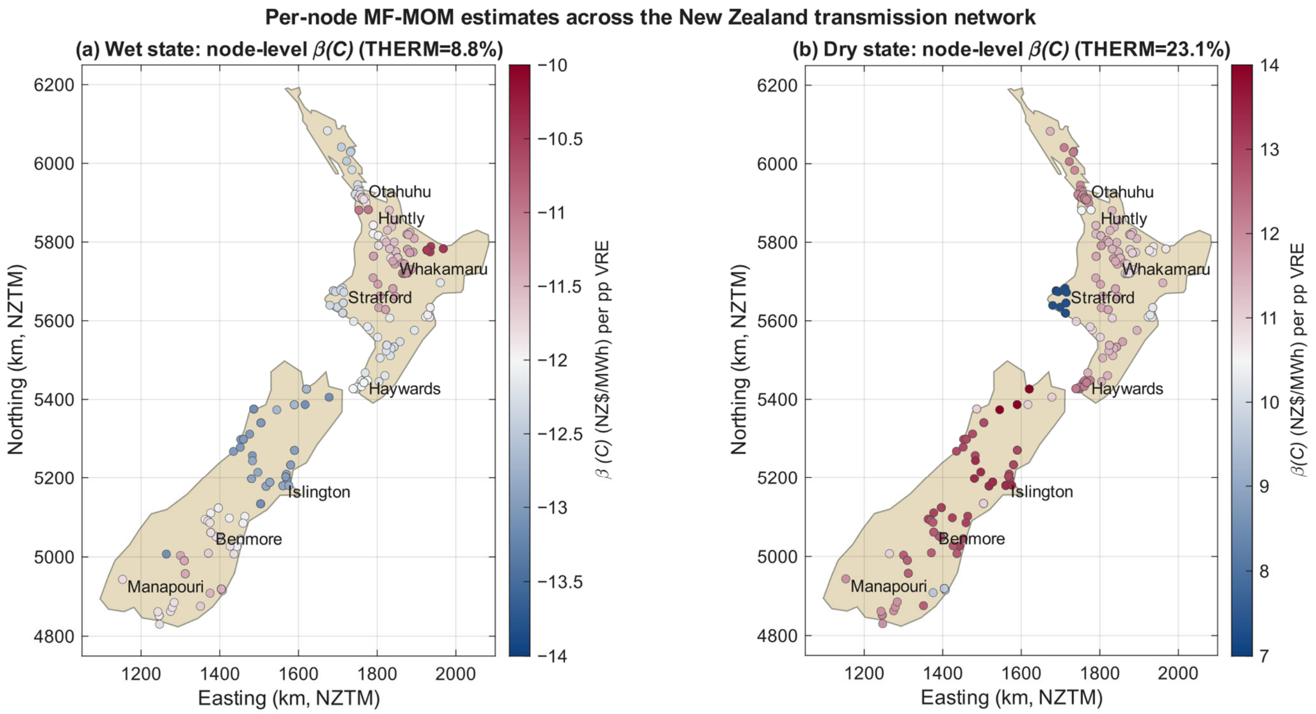

Figure 11. Per-node MF-MOM estimates across the New Zealand transmission network. (**a**) Implied VRE coefficient at low (8.8%) THERM. (**b**) Implied VRE coefficient at high (23.1%) THERM.

Node-level coefficients are stabilised using James–Stein shrinkage, with noisier estimates pulled more strongly toward the precision-weighted cross-node mean. The shrinkage weight depends on the estimated between-node variance relative to each node's sampling variance. Reported intervals are posterior intervals based on the shrunk estimates and posterior standard deviations and are therefore narrower than raw node-level intervals. The prior is exchangeable rather than spatial, because the purpose is descriptive mapping, not spatial inference. Spatial-lag or spatial-error modelling is left for future work. These node-level results complement the main analysis; the pooled findings in Sections 6.2–6.7 do not depend on them.

7.2. Distribution of Node-Level Estimates

Out of a total of 245 nodes, 243 converged with at least 60 months of price observations each. The per-node interaction coefficient, written β_{int}, is positive in every region (see Table 13 and Figure 12), with regional medians ranging from +1.60 (CNI) to +1.81 (USI). The state-dependent direction of MOE is therefore spatially universal across the New Zealand transmission network. Implied wet-state coefficients $\beta(C)$ at THERM = 8.8% range from a regional median of -11.34 in CNI to -12.93 in USI, and dry-state coefficients $\beta(C)$ at

THERM = 23.1% from +11.36 in CNI to +12.94 in USI. The implied wet-to-dry swing of price effect is approximately NZ$13-15/MWh per 10% relative VRE increase regardless of region.

Table 13. Per-zone median MF-MOM coefficients (243 nodes aggregated to five EA regions).

EA Zone	N	β_{base}	β_{int}	γ	$\beta_{(TH\_lo)}$	$\beta_{(TH\_hi)}$
Upper N.I. (UNI)	35	−0.39	+1.67	+9.91	−11.80	+12.02
Central N.I. (CNI)	73	−0.34	+1.60	+9.71	−11.34	+11.36
Lower N.I. (LNI)	51	−0.67	+1.65	+9.88	−12.18	+11.39
Upper S.I. (USI)	41	−0.27	+1.81	+11.42	−12.93	+12.94
Lower S.I. (LSI)	43	+0.17	+1.70	+10.85	−11.82	+12.60
Pooled (Section 6.2)	243	+2.66	+1.81	+10.00	−9.92	+15.88

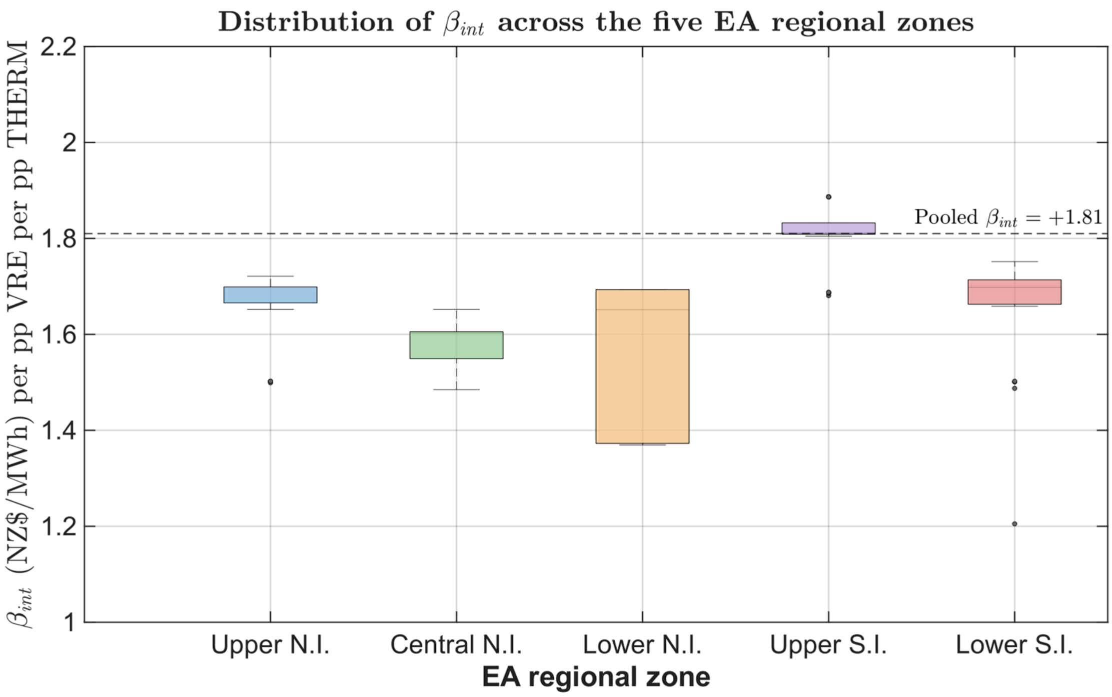

Figure 12. Distribution of the per-node MF-MOM interaction coefficient β_{int} across the five EA regional zones. Box widths represent the inter-quartile range (IQR), the centre line is the regional median, whiskers extend to 1.5× IQR, and outlier nodes are shown as grey dots.

7.3. Regional Patterns and Economic Interpretation

The relationship between wet-state and dry-state per-node coefficients is summarised in Figure 13, where the tight clustering around the pooled estimate confirms that cross-node heterogeneity is small once empirical-Bayes shrinkage is applied. The South Island zones exhibit marginally stronger MOE in both wet and dry conditions than the North Island zones, consistent with the South Island's greater hydro-dependence and lower local thermal capacity [42]. Within the North Island, CNI exhibits the weakest merit-order effect, consistent with its location near the central geothermal–hydro generation cluster (e.g., Whakamaru, Aratiatia, Kawerau), where the effective supply curve is relatively flat. A small number of nodes depart from this general pattern. Stratford, located close to the Taranaki gas-fired fleet, records a $\beta(C)$ at high THERM of around +7, compared with values of roughly +11 to +13 observed at most other nodes. This likely reflects the fact that additional VRE output at such locations primarily displaces nearby thermal generation, rather than interacting more broadly with the hydro-dominated system.

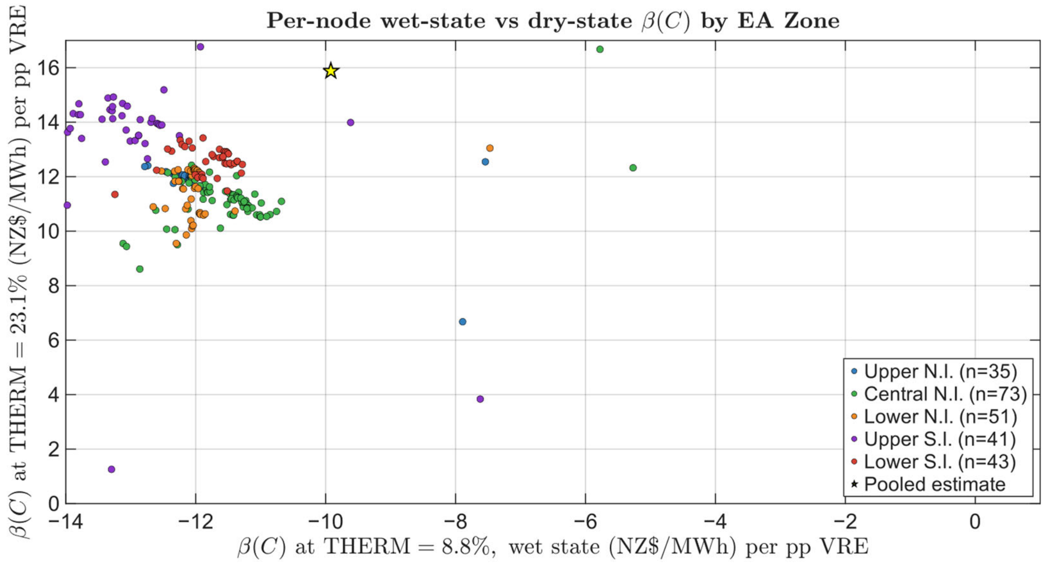

Figure 13. Per-node wet-state vs. dry-state $\beta(C)$, coloured by EA regional zone. The pooled estimate is from Section 6.2.

7.4. Implications for Pooled vs. Node-Level Modelling

Two observations are worth emphasising. First, the strong central tendency around the pooled estimate, combined with EB shrinkage weights below 0.01 for all parameters, indicates that the pooled MF-MOM reported in Section 6.2 is a faithful summary of the underlying spatial structure. Per-node deviations exist but are dominated by within-node sampling noise rather than genuine cross-node heterogeneity. Second, the spatial heterogeneity that does exist is concentrated at nodes with idiosyncratic local supply features (thermal fleet adjacent or geothermal adjacent), exactly the locations where node-specific siting analysis would matter most. From a policy standpoint, the pooled coefficient suffices for system-level analysis, while node-level coefficients are needed only for site-specific cannibalisation or hedging calculations at the small number of locally distinctive nodes.

8. Discussion

8.1. Marginal Fuel vs. Storage Conditioning

The contrast is not about model flexibility; both conditioners enter through the same linear interaction. It is about choosing a variable that indexes the economic regime contemporaneously. Thermal share is a flow that reflects the actual dispatch outcome in each month, whereas storage level is a stock, whose mapping to the price-setting regime shifts with inflows, demand, expectations, and hydro release strategy. The published finding by Peña [2] that reservoir levels do not explain marginal prices in hydro systems is the same phenomenon seen from the price side, and this is why a storage-level coefficient is non-stationary, while a marginal fuel coefficient is stable.

8.2. Comparison with New Zealand Studies

Models by Wen et al. [1,16] report a negative, season-varying wind effect over 2011–12. As shown in Section 6.7, the same MF-MOM specification re-estimated in this study on their work returns a directionally opposite point estimate whose statistical significance is marginal (two-sided $p = 0.053$). The strong, robust, positive interaction observed in 2015–2024 does not extend backward. The 2011–12 panel instead yields a weakly identified estimate of the opposite sign. This pattern is consistent with evolution of the New Zealand market over the decade, rather than with a single time-invariant relationship. The present 2015–24 results therefore extend rather than reproduce the earlier NZ literature. They identify a state dependence whose mechanism (water value) and direction (suppression strongest in hydro-dominated conditions) are appropriate to the current market regime but should not be projected backward without re-estimation. The agent-based studies [18,19] reproduce New Zealand prices only with explicit hydro and water value tracking, which corroborates the central premise here. Relative to all prior New Zealand work, the contribution is to estimate the merit-order coefficient itself as a function of a direct marginal technology state on observed nodal data and to validate that function out of sample within the 2015–24 regime.

8.3. Comparison with International Studies

The international MOE literature [3–10] reports single average coefficients in thermal-dominated markets and, where examined, finds the effect to be state-dependent [7,9]. The present work is consistent with that broad picture but differs in two ways. It conditions on a hydro-thermal marginal state, and it is set in a market where that state varies enough to identify the dependence.

Direct comparison is constrained by structural differences across markets. In particular, most studies do not consider systems with a hydro-storage regime comparable to that of New Zealand, and reported effect sizes are defined using different metrics. As a result, the magnitudes should be interpreted as contextual, rather than directly comparable benchmarks. The closest hydro-based studies [2,12,13,35,37,44] are informative for understanding the underlying mechanism, but their reported quantities are not directly comparable with the implied-price measure used here.

8.4. Implications

The implications extend beyond estimation. The value of additional VRE [45], the cannibalisation risk to existing VRE [8,9,46], and retailers' hedging exposure [39,47] all depend on the prevailing marginal fuel regime rather than on a single average MOE. A framework that distinguishes between regimes can therefore better inform siting decisions, market design [48], and price-risk forecasting [49] than a model based on a constant coefficient.

More broadly, these results highlight a methodological point. Strong in-sample fit, conventional robustness checks, and a successful placebo test provide important internal validation, but they do not by themselves establish a structural relationship. Identifying such relationships requires evidence that is robust to the appropriate out-of-sample comparisons and stable across sub-periods.

9. Limitations

Five limitations are stated plainly. First, within the 2015–24 sample, the out-of-sample gain is significant in cross-validation ($p = 0.048$) but marginal in the temporal hold-out ($p = 0.066$). The 2025–26 forward test in Section 6.5 strengthens the extrapolation evidence. The frozen interaction improves forward RMSE by 33% (wild cluster bootstrap $p = 0.047$), but the forward window remains short, at 17 months, and re-estimation on that window alone is too weakly identified to be informative.

Second, the VRE coefficient is positive at high thermal share, but the base coefficient is imprecisely estimated, so the defensible claim is that the suppression effect vanishes when the thermal share is high, not that VRE raises prices.

Third, the analysis uses the monthly thermal generation share as a proxy for the price-setting regime rather than unit-level marginal plant data. The half-hourly diagnostics in Section 6.1 indicate that the regime is persistent within months, so monthly conditioning captures the relevant state reasonably well. Nevertheless, the proxy does not identify the marginal unit in each trading interval. Some residual window-level variation therefore remains in the interaction.

Fourth, MF-MOM, as specified, is a stable description of the current (2015–24) New Zealand market regime. Direct re-fit on the 2011–12 panel (see Section 6.7) returns a directionally opposite point estimate, consistent with structural evolution of the market over the decade. The model is therefore current regime-specific rather than panel-invariant. We report this limit honestly because it is itself informative about how the New Zealand market has evolved.

Fifth, several price-level drivers are not directly controlled, including gas availability, generator outages, and transmission constraints, because consistent monthly data are unavailable for the full period. We treat this as an omitted-variable risk. However, node and additive time effects absorb common spatial and temporal components, and the interaction remains stable across multiple robustness checks, including storage, nonlinear storage, demand, inflows, lagged THERM, storage interactions, and exclusion of extreme price months. An omitted factor would therefore need to be correlated specifically with the VRE-by-THERM interaction, not merely with the price level. We cannot rule this out, and we do not claim to.

10. Future Work

Four extensions follow. One natural direction would be to incorporate unit-level marginal plant (price-setting) data, which would allow the generation-share proxy to be replaced with a direct measure of the marginal technology. In practice, however, such data are not consistently available for the New Zealand market, and assembling a reliable series may require substantial reconstruction. The per-node analysis of Section 7 could be deepened with a formal spatial-econometric specification (for example, a spatial-lag or spatial-error model) to test for cross-nodal spillovers directly, rather than relying on cross-sectionally robust standard errors. Newly published hourly and regional energy-demand projections for New Zealand could support such spatially explicit and per-region extensions [50]. Richer functional forms (for example, a quadratic in THERM or interactions involving the squared VRE share) could test whether a single specification can reconcile the 2011–12 and 2015–24 regimes via a non-monotonic $\beta(C)$ profile. It can also confirm whether the observed sign reversal reflects a structural change that cannot be captured within a single parametric form or not. Finally, external validation on another hydro-dominated market with reservoir data, such as the Nordic system (Nord Pool, with generation and reservoir data via ENTSO-E and the Nordic TSOs etc.), would provide a test of whether the marginal fuel conditioning framework and its water value interpretation extend beyond New Zealand.

Forward out-of-sample validation of the marginal fuel model against the January 2025 to May 2026 market data is reported in Section 6.5. Extending that comparison to the storage-conditioned model remains future work, because reservoir storage records from the national hydrology archive are published only through December 2024, so the storage conditioner cannot yet be evaluated on the forward window. The asymmetry is itself informative. A dispatch-based conditioner is testable as soon as generation and price records appear, whereas a hydrological stock conditioner is not. Once post-2024 storage records are released, the three-way forward comparison (baseline, storage-conditioned, marginal fuel) will complete the sequence.

11. Conclusions

A merit-order effect conditioned on reservoir level is non-stationary and does not generalise well. The storage-conditioned model is highly significant in sample and passes placebo and outlier checks. On internal diagnostics alone, it would appear defensible. However, it adds no significant out-of-sample value once a storage main effect is already included. Its interaction coefficient also reverses sign across sub-periods, and this instability cannot be explained by collinearity.

The weakness of the storage model is therefore conceptual rather than purely statistical. Reservoir level is an indirect proxy for the price-setting regime, consistent with the Spanish evidence that reservoir levels alone do not explain hydro-market prices [2]. Conditioning the merit-order effect on monthly thermal generation share performs better. Conditioning on the realised dispatch regime produces a stable, significant, and out-of-sample-validated interaction. The thermal share is a proxy for that regime and not a measurement of the marginal unit, so the finding is that the merit-order effect varies systematically with the realised hydro-thermal dispatch state, not that a particular technology sets the price in a given interval. Its implied VRE coefficient is strongly negative in hydro-dominated conditions and fades as thermal generation becomes more important at the margin. This pattern is consistent with the water value logic of hydro-thermal dispatch [13].

The result provides a stable and interpretable description of the merit-order effect in the current New Zealand market regime. It also provides a template for testing state-dependent merit-order effects before reporting them as structural relationships. The cross-period test adds an important qualification. When the model is re-estimated on the 2011–12 panel, the interaction has the opposite sign. The estimate is marginal under wild cluster bootstrap-t inference, with a two-sided p-value of 0.053. The strong, out-of-sample-validated state dependence found for 2015–24 therefore characterises the current market regime rather than a structurally invariant law. It is reported as a property of the present, more variable, renewable-intensive system, not projected backward onto the earlier one.

The evidence suggests that the nature of state dependence in Aotearoa New Zealand has evolved alongside changes in the electricity system. In a hydro-dominated market, the merit-order effect is real, but it is also time-varying. Rigorous specification testing is therefore essential before drawing structural conclusions.

Author Contributions: Conceptualization, A.H. and K.P.; Methodology, A.H. and K.P.; Software, A.H.; Validation, A.H.; Investigation, A.H. and K.P.; Resources, K.P.; Data curation, A.H.; Writing—original draft, A.H.; Writing—review & editing, K.P. and J.K.; Supervision, K.P. and J.K.; Project administration, K.P. and J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wen, L.; Suomalainen, K.; Sharp, B.M.H.; Yi, M.; Sheng, M. Impact of wind-hydro dynamics on electricity price: A seasonal spatial econometric analysis. *Energy* **2022**, *238*, 122076. [CrossRef]
2. Peña, J.I. A note on hydropower as a marginal price setter for Spain's electricity market in 2021–2022. *Util. Policy* **2024**, *87*, 101726. [CrossRef]
3. Cludius, J.; Hermann, H.; Matthes, F.C.; Graichen, V. The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications. *Energy Econ.* **2014**, *44*, 302–313. [CrossRef]
4. Zarnikau, J.; Woo, C.K.; Zhu, S.; Tsai, C.H. Market price behavior of wholesale electricity products: Texas. *Energy Policy* **2019**, *125*, 418–428. [CrossRef]

5. Csereklyei, Z.; Qu, S.; Ancev, T. The effect of wind and solar power generation on wholesale electricity prices in Australia. *Energy Policy* **2019**, *131*, 358–369. [CrossRef]
6. Mwampashi, M.M.; Nikitopoulos, C.S.; Konstandatos, O.; Rai, A. Wind generation and the dynamics of electricity prices in Australia. *Energy Econ.* **2021**, *103*, 105547. [CrossRef]
7. Tselika, K. The impact of variable renewables on the distribution of hourly electricity prices and their variability: A panel approach. *Energy Econ.* **2022**, *113*, 106194. [CrossRef]
8. Halttunen, K.; Staffell, I.; Slade, R.; Green, R.; Saint-Drenan, Y.-M.; Jansen, M. Global assessment of the merit-order effect and revenue cannibalisation for variable renewable energy. *SSRN Electron. J.* **2020**. [CrossRef]
9. López Prol, J.; Steininger, K.W.; Zilberman, D. The cannibalization effect of wind and solar in the California wholesale electricity market. *Energy Econ.* **2020**, *85*, 104552. [CrossRef]
10. Ajanaku, B.A.; Collins, A.R. Comparing merit order effects of wind penetration across wholesale electricity markets. *Renew. Energy* **2024**, *226*, 120372. [CrossRef]
11. Read, E.G.; Ring, B.J.; Starkey, S.R.; Pepper, W. An LP based market design for natural gas. In *Handbook of Networks in Power Systems II*; Sorokin, A., Rebennack, S., Pardalos, P.M., Iliadis, N.A., Pereira, M.V.F., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 77–113. [CrossRef]
12. Kleiven, A.; Risanger, S.; Fleten, S.-E. Co-movements between forward prices and resource availability in hydro-dominated electricity markets. *Energy Syst.* **2023**, *14*, 859–889. [CrossRef]
13. Pavičević, M.; Huallpara, A.; Yu, A.; Balderrama, S.; Ploussard, Q. Water value and market response functions in hydropower-dominated electricity systems: A systematic review and methodological comparison. *Renew. Sustain. Energy Rev.* **2026**, *235*, 116952. [CrossRef]
14. Ministry of Business, Innovation and Employment. *Electricity Demand and Generation Scenarios 2024*; MBIE: Wellington, New Zealand, 2024. Available online: https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-modelling/electricity-demand-and-generation-scenarios (accessed on 18 June 2026).
15. Transpower New Zealand. *Whakamana i Te Mauri Hiko—Empowering Our Energy Future*; Transpower: Wellington, New Zealand, 2020. Available online: https://www.transpower.co.nz/about-us/our-strategy/whakamana-i-te-mauri-hiko-empowering-our-energy-future (accessed on 18 June 2026).
16. Wen, L.; Sharp, B.; Sbai, E. Spatial effects of wind generation and its implication for wind farm investment decisions in New Zealand. *Energy J.* **2020**, *41*, 47–72. [CrossRef]
17. Suomalainen, K.; Pritchard, G.; Sharp, B.; Yuan, Z.; Zakeri, G. Correlation analysis on wind and hydro resources with electricity demand and prices in New Zealand. *Appl. Energy* **2015**, *137*, 445–462. [CrossRef]
18. Browne, O.; Poletti, S.; Young, D. How does market power affect the impact of large scale wind investment in "energy only" wholesale electricity markets? *Energy Policy* **2015**, *87*, 17–27. [CrossRef]
19. Young, D.; Poletti, S.; Browne, O. Can agent-based models forecast spot prices in electricity markets? Evidence from the New Zealand electricity market. *Energy Econ.* **2014**, *45*, 419–434. [CrossRef]
20. Poletti, S. Market power in the New Zealand wholesale electricity market 2010–2016. *Energy Econ.* **2021**, *94*, 105078. [CrossRef]
21. Poletti, S.; Staffell, I. Understanding New Zealand's wind resources as a route to 100% renewable electricity. *Renew. Energy* **2021**, *170*, 449–461. [CrossRef]
22. Macedo, D.P.; Marques, A.C.; Damette, O. Challenges in assessing the behaviour of nodal electricity prices in insular electricity markets: The case of New Zealand. *Economies* **2023**, *11*, 159. [CrossRef]
23. Wang, G.; Wan, H.; Sbai, E. Nonlinear impacts of distributed generation on extreme electricity prices under policy shocks: Evidence from New Zealand's oil and gas exploration ban. *Energy* **2025**, *335*, 137754. [CrossRef]
24. Kapoor, G.; Wichitaksorn, N.; Zhang, W. Analyzing and forecasting electricity price using regime-switching models: The case of New Zealand market. *J. Forecast.* **2023**, *42*, 2011–2026. [CrossRef]
25. Kapoor, G.; Wichitaksorn, N. Electricity price forecasting in New Zealand: A comparative analysis of statistical and machine learning models with feature selection. *Appl. Energy* **2023**, *347*, 121446. [CrossRef]
26. Sensfuß, F.; Ragwitz, M.; Genoese, M. The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Energy Policy* **2008**, *36*, 3086–3094. [CrossRef]
27. Antweiler, W.; Muesgens, F. On the long-term merit order effect of renewable energies. *Energy Econ.* **2021**, *99*, 105275. [CrossRef]
28. Antweiler, W.; Muesgens, F. The new merit order: The viability of energy-only electricity markets with only intermittent renewable energy sources and grid-scale storage. *Energy Econ.* **2025**, *145*, 108439. [CrossRef]
29. Hirth, L. The market value of variable renewables: The effect of solar and wind power variability on their relative price. *Energy Econ.* **2013**, *38*, 218–236. [CrossRef]
30. Acar, B.; Selcuk, O.; Dastan, S.A. The merit order effect of wind and river type hydroelectricity generation on Turkish electricity prices. *Energy Policy* **2019**, *132*, 1298–1319. [CrossRef]

31. Benhmad, F.; Percebois, J. Photovoltaic and wind power feed-in impact on electricity prices: The case of Germany. *Energy Policy* **2018**, *119*, 317–326. [CrossRef]
32. de Lagarde, C.M.; Lantz, F. How renewable production depresses electricity prices: Evidence from the German market. *Energy Policy* **2018**, *117*, 263–277. [CrossRef]
33. Newbery, D. Shifting demand and supply over time and space to manage intermittent generation: The economics of electrical storage. *Energy Policy* **2018**, *113*, 711–720. [CrossRef]
34. Electricity Authority. Electricity Market Information (EMI), Wholesale Market Datasets. Available online: https://www.emi.ea.govt.nz/ (accessed on 18 June 2026).
35. Hirth, L.; Mühlenpfordt, J.; Bulkeley, M. The ENTSO-E Transparency Platform: A review of Europe's most ambitious electricity data platform. *Appl. Energy* **2018**, *225*, 1054–1067. [CrossRef]
36. Ministry of Business, Innovation and Employment. *Energy in New Zealand 2024*; MBIE: Wellington, New Zealand, 2024. Available online: https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-publications-and-technical-papers/energy-in-new-zealand/energy-in-new-zealand-2024 (accessed on 18 June 2026).
37. Stelling, P.; Brent, A.C.; Burmester, D. The impact of 100% renewable electricity on hydropower generation in Aotearoa New Zealand. *Renew. Energy Focus* **2026**, *56*, 100769. [CrossRef]
38. Zhang, Z.; Liu, X.; Zhao, D.; Post, S.; Chen, J. Overview of the development and application of wind energy in New Zealand. *Energy Built Environ.* **2023**, *4*, 725–742. [CrossRef]
39. Frontier Economics. *Review of Electricity Market Performance: Final Report to Ministers and MBIE*; Frontier Economics: Melbourne, Australia, 2025.
40. Bahrami Gholami, M.; Poletti, S.; Staffell, I. Wind, rain, fire and sun: Towards zero carbon electricity for New Zealand. *Energy Policy* **2021**, *150*, 112109. [CrossRef]
41. Ghelasi, P.; Ziel, F. A data-driven merit order: Learning a fundamental electricity price model. *Energy Econ.* **2026**, *154*, 109114. [CrossRef]
42. Philpott, A.; Read, G.; Batstone, S.; Miller, A. The New Zealand electricity market: Challenges of a renewable energy system. *IEEE Power Energy Mag.* **2019**, *17*, 43–52. [CrossRef]
43. Nagpure, M.; Wakte, G.; Kumar, M. Advanced nodal pricing and optimization techniques for integrating renewable energy in electricity markets. In *Proceedings of the 2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*; IEEE: Piscataway, NJ, USA, 2025; pp. 922–929. [CrossRef]
44. Pereira, J.P.; Pesquita, V.; Rodrigues, P.M.M. The effect of hydro and wind generation on the mean and volatility of electricity prices in Spain. In *Proceedings of the 2017 14th International Conference on the European Energy Market (EEM), Dresden, Germany, 6–9 June 2017*; IEEE: Piscataway, NJ, USA, 2017; p. 7981915. [CrossRef]
45. Mallapragada, D.S.; Junge, C.; Wang, C.; Pfeifenberger, H.; Joskow, P.L.; Schmalensee, R. Electricity pricing challenges in future renewables-dominant power systems. *Energy Econ.* **2023**, *126*, 106981. [CrossRef]
46. Market Development Advisory Group. *Price Discovery in a 100% Renewables Electricity Market: Literature Review*; Electricity Authority: Wellington, New Zealand, 2022.
47. Flottmann, J.; Simshauser, P.; Wild, P.; Todorova, N. The forward market dilemma in energy-only electricity markets. *Energy Econ.* **2025**, *148*, 108676. [CrossRef]
48. Zhu, Z.; Bu, S.; Chan, K.W.; Li, F.; Ye, Y.; Chung, C.Y.; Strbac, G. Designing the future electricity spot market with high renewables via reliable simulations. *Nat. Rev. Electr. Eng.* **2025**, *2*, 320–337. [CrossRef]
49. Sun, Y.; Gomes, H.M.; Pfahringer, B.; Bifet, A. Real-time energy pricing in New Zealand: An evolving stream analysis. In *PRICAI 2024: Trends in Artificial Intelligence, Proceedings of the 21st Pacific Rim International Conference on Artificial Intelligence, Kyoto, Japan, 18–24 November 2024*; Hadfi, R., Anthony, P., Sharma, A., Ito, T., Bai, Q., Eds.; Lecture Notes in Computer Science; Springer: Singapore, 2025; Volume 15285, pp. 91–97. [CrossRef]
50. Canessa, R.; Osorio-Aravena, J.C.; Peer, R.; Gils, H.C.; Wetzel, M.; Gulagi, A.; Breyer, C.; Haas, J. Projected hourly and regional energy demand for power, heat and transport in New Zealand to 2050. *Sci. Data* **2026**, *13*, 193. [CrossRef] [PubMed]