Asia & the Pacific Policy Studies

ORIGINAL ARTICLE OPEN ACCESS

Stage Segmentation of Rural Transformation and Comparisons Among Bangladesh, China, Indonesia, and Pakistan: Combining Machine Learning and New Structural Economics to Facilitate International Agricultural Development and Policy Design

Dong Wang[1] | Chunlai Chen[1] | Christopher Findlay[1] | Jikun Huang[2] | Justin Yifu Lin[3,4] | Abedullah[5] | Mohammad Jahangir Alam[6] | Abid Hussain[7] | Nunung Nuryartono[8] | Tahlim Sudaryanto[9] | David Shearer[10,11]

[1]Crawford School of Public Policy, Australian National University, Canberra, Australia | [2]China Centre for Agricultural Policy (CCAP), School of Advanced Agricultural Sciences, Peking University, Beijing, China | [3]Institute of New Structural Economics, Peking University, Beijing, China | [4]National School of Development, Peking University, Beijing, China | [5]Pakistan Institute of Development Economics (PIDE), Islamabad, Pakistan | [6]Department of Agribusiness and Marketing, Bangladesh Agricultural University, Mymensingh, Bangladesh | [7]Social Sciences Research Institute (SSRI), Pakistan Agricultural Research Council-National Agricultural Research Centre, Islamabad, Pakistan | [8]Faculty of Economics and Management, IPB University, West Java, Indonesia | [9]Centre for Behavioural and Circular Economy, National Research and Innovation Agency (BRIN), Jakarta, Indonesia | [10]Fenner School of Environment and Society, Australian National University, Canberra, Australia | [11]Agribusiness Program, Australian Centre for International Agricultural Research (ACIAR), Canberra, Australia

Correspondence: Dong Wang (D.Wang@anu.edu.au)

Received: 13 September 2024 | **Revised:** 19 January 2025 | **Accepted:** 25 January 2025

Funding: This research was funded by the Australian Centre for International Agricultural Research (ACIAR) project ADP-2017-024.

Keywords: Huang segmentation | machine learning | new structural economics | rural development | rural transformation

ABSTRACT

This paper contributes a new paradigm for international agricultural development research. It uses machine learning techniques to aid expert diagnosis of development problems in conjunction with New Structural Economics (NSE) to analyse and design policies to enable effective rural transformation. It conducts a multi-country, multi-regional, multi-level and multi-dimensional analysis in Bangladesh, China, Indonesia, and Pakistan to identify stage segmentations of rural transformation and examine stagewise associate policies and applicable learnings across each dimension. By presenting structured stages of rural transformation, we provide guidance on designing dynamic comparative-advantage-adapting policies that are able to adapt at each stage. This analytical procedure can serve other relevant agricultural development studies.

1 | Introduction

International agricultural research for development projects drives growth by facilitating changes in technology, institutions, governance, policies, and investment. It is problem-solving research that addresses complex challenges in developing countries' rural economies as well as drawing lessons from other nations. This usually involves conducting a multi-country, multi-level, multi-dimensional analysis, which requires researchers to possess a profound understanding of the 'nature' of development, identify 'real' issues, and propose 'feasible' solutions.

While conducting cross-country comparisons and policy analyses, researchers, policymakers, and policy practitioners often encounter a situation where their methodology limits their ability to effectively figure out specific problems in a particular stage and region, which results in some recommendations becoming too general to be practically useful. Addressing this dilemma calls for a new concept and methodology.

To address this dilemma, this paper aims to demonstrate a new paradigm by providing an innovative analytical toolkit that can assist in systematically reviewing historical development patterns through the lens of evolving comparative advantages and devising feasible future policy agendas and future strategies. We will present a cross-country, multi-level-and-regional, multi-dimensional comparative study on the extent of rural transformation within four countries: Bangladesh, China, Indonesia, and Pakistan. Our research involves segmenting rural transformation into analysable stages, diagnosing problems while comparing, and drawing policy recommendations. Conceptually, we incorporate two promising theories—the rural transformation stage segmentation proposed by J. Huang (2018) (hereafter, Huang Segmentation), which is based on summarised stylised facts from Chinese historical data, and the New Structural Economics (NSE) advanced by Lin (2011a). Methodologically, we integrate the conventional expert-driven approach and the newly data-driven approach—machine learning.

Rural transformation, defined by IFAD (2016, 23), is 'a process that involves rising agricultural productivity, increasing commercialisation and marketable surpluses, and diversification of production patterns and livelihoods. It also involves expanded decent off-farm employment and entrepreneurial opportunities, better rural coverage and access to services and infrastructure, and greater access to, and capacity to influence, relevant policy processes'. Recent research has documented that a successful rural transformation can significantly increase rural households' incomes and reduce poverty incidence in some Asian developing countries, such as China, Indonesia, and Pakistan (Abedullah, Farooq, and Naz 2023; Shi and Huang 2023; Sudaryanto et al. 2023). Accelerating rural transformation in third-world countries is particularly urgent at the moment in the post-Covid era (Huang et al. 2023). The impact of the Covid-19 pandemic, ongoing geopolitical conflicts, economic slowdown, inflation, and extreme weather events have led to a reversal in the global progress of poverty reduction, setting back achievements of global poverty reduction by approximately 3 years. As a result, achieving the United Nations' Sustainable Development Goal of poverty elimination by 2030 is increasingly unlikely (United Nations 2023). The challenge of pulling the world back on track with poverty reduction prompts researchers and policymakers to reflect on the policy framework of rural transformation, which has been serving as the main pathway out of poverty for all societies due to its rising productivity (Timmer 2017).

Rural transformation involves a series of structural changes in the rural economy, including structural transformation within the agricultural sector and the structural change between agricultural and non-agricultural sectors. According to J. K. Huang and Shi (2021) and D. Wang, Chen, and Findlay (2023), the within-sector transformation refers to agricultural transformation, which can be measured by the proportion of high-value agriculture production. The between-sector transformation can be measured by the proportion of non-farm employment.

The economics studies of rural transformation are based on the concepts of structural change economics, which have been extensively discussed since the 1960s. Important contributions include the seminal article '*The Role of Agriculture in Economic Development*' by Johnston and Mellor (1961) and the book '*Transforming Traditional Agriculture*' by Schultz (1964). Rural transformation enables the modernisation of the traditional agricultural sector in countries that rely on agriculture and natural resources, leading to economic prosperity by increasing productivity, commercialisation, and other factors. This transformation involves evolving conventional farm activities into modern agriculture (increasing productivity and efficiency of resources within the sector), and it also leads to a shift of resources and capital towards more productive sectors.

However, it is challenging to change traditional agriculture due to the state of the art and the state of the preference, that is what is technologically possible and what works best for farmers. Thus, rural transformation may not be initiated because of market or government failure. Effective policy facilitation requires understanding the characteristics of various rural transformation stages, as the comparative advantage can vary from stage to stage, so policy instruments need to be adjusted accordingly. For example, in the earlier stages, a rural economy is likely to be primarily focused on producing staple food and reliant much more on labour and natural resources that are relatively abundant. Conversely, as rural development progresses, it becomes increasingly dependent on more skilled human capital, investments, and technology due to the factor endowment component shifts and factors mobility between sectors. As a result, different regions may have a particular pace and issues to address in various stages of transforming their economic structure. It gives rise to a particular interest in recognising the phases of rural transformation and executing tailored strategies linked to the fundamental traits of various stages.

In practice, stage segmentation is useful because it can suggest one policy intervention's 'window period' and the degrees of tightness of one policy implementation and its supervision (Kingdon 2013). Wrongly identifying the stages or selecting inappropriate policy instruments may slow down the rural transformation process. The order in which interventions are implemented also affects the outcome. For instance, Pakistan initiated the Green Revolution in the 1960s before China's agricultural reform began in 1978. However, Pakistan's Green Revolution did not lead to the same subsequent success as China's agricultural reform. The Green Revolution heavily relied on technology and foreign aid in a subsistence food production agrarian economy in a country that lacked capital and an enabling environment. Such a policy is hard to sustain in many aspects. On the contrary, China's rural reform started from the Household-Responsibility-System (HRS), which is an institutional-led policy change (Lin 1992). It first created incentives for production and then enabled a market environment to encourage capitalisation and technological adoption. Eventually, it fostered endogenous growth in the Chinese rural economy and triggered widespread industrialisation. Regardless of other potential factors, this example shows that even though different policies may be effective in increasing production in the short term, the order in which they are implemented at various stages could result in significantly different outcomes in the long run.

Our research will start by testing Huang Segmentation of rural transformation stages in four developing Asian countries: Bangladesh, China, Indonesia, and Pakistan, using regional-level data for each country. We incorporate machine learning techniques into the conventional expert evaluation approach. Then, we adopt the New Structural Economics perspective to analyse the results and propose policy recommendations.

We aim to answer three questions. First, is the Huang Segmentation applicable to other countries as a general pattern? Second, what stages did those countries experience, and what caused the disparities between countries and regions? Third, how can we use New Structural Economics to understand the 'nature' of rural transformation and make appropriate policies according to stages?

The contributions are multifaceted. Overwhelmingly, this paper contributes a new paradigm that incorporates a purely data-driven approach and an expert-driven approach to diagnosing regional rural development distinctions. This methodology combines four pillars: expert evaluation, machine learning at the technical level; Huang Segmentation, and New Structural Economics at the conceptual level. This new analytical toolkit can equip academics and think tanks for future development studies in the relevant area. Secondly, this paper provides the 'justification' and 'applicability' of Huang Segmentation and New Structural Economics, as well as demonstrates the 'usefulness' of New Structural Economics in creating rural development strategies. Third, the discussion based on country-side stage segmentation will assist policymakers in implementing policy facilitation to accelerate rural transformation.

2 | Concepts and Related Literature

This paper is placed on four pillars. At the conceptual level, the two foundational theories are Huang Segmentation of rural transformation stages initially based on the Chinese experience with further expansion to other Asian developing countries (J. Huang 2018), and New Structural Economics, which is generated from the experiences of success and failure of structural transformation in their development and transition process in China, East Asian economies and other developing countries in Former Soviet Union, Latin America, South Asia and Africa.

Huang argues that Asian developing countries share a general path and trend of rural transformation, which includes four major stages. The first stage is primarily related to staple food production. The second stage is characterised by shifting agriculture to more diversification and commercialisation, which features labour-intensive and high-value agricultural production. The third stage features agricultural specialisation, mechanisation, and rural labour's non-farm employment. The last stage is the move to sustainable agriculture and urban-rural integration development.

Furthermore, J. K. Huang and Shi (2021), based on a Chinese historical review of policy response to rural transformation according to stages, highlight in each stage, specific institutional reforms, policies, and investments (IPIs) can stimulate and accelerate the rural transformation from one stage to the next stage. Although this fundamental contribution provides a framework for thinking about the process of rural transformation and associated policies, Huang Segmentation has not addressed several practical issues, especially for policy practitioners, when they apply this concept to local cases. Firstly, the identification of stages relies heavily on experts' evaluation, which is based on their knowledge and experience in China. This uses a historical induction method to assess complexity and specific content. Experts from different countries with different backgrounds might question the repeatability of this method or worry about over-generalisation issues due to the gaps of knowledge and cognitive abilities. Policymakers may also concern that such high-quality expertise requirements could induce bias and subjectivity in practice. This theoretical difficulty could hinder the wider dissemination and application of Huang Segmentation. Secondly, while the 'stylised' summary can be backed by pattern or trend analysis of historical data, Huang Segmentation does not thoroughly explain the methods of stage division or rigorously illustrate the principles by which stage-based policies should be implemented. This point for improvement might confuse researchers when they conduct cross-country comparisons. It might also be unclear for policymakers and practitioners when they deal with highly diverse development problems across regions within their countries.

In this regard, we seek insights from another area of development economics literature, specifically focusing on structural change and policy facilitation.

New Structural Economics advocates a strategic approach to creating effective industrial policies that can address market failures and drive economic development by accelerating structural change according to a country's comparative advantages determined by its factor endowments structure, which is given at any specific time and changeable over time. It reveals the underlying driver of structural change and provides a guide for making policy facilitations (Lin 2012). Lin argues that for each specific 'level'[1] of development, the economy's structure is inherently and endogenously determined by the structure of factor endowments, which may bring about factor-biased technological progress to enhance productivity, accordingly, the rural transformation upgrade. The comparative advantages evolve along the changing factor endowment structure as a result of capital accumulation in a market economy. The comparative advantages change directs the production towards commodities that encompass relatively profitability so that the structural upgrade can occur. Thus, an 'optimal' economic structure corresponds to the given factor endowment structure for each level of development, which can maximise the speed of development, capital accumulation and welfare of the entire economy. If the economic and endowment structures do not align, there is likely to be a loss of efficiency, indicating the presence of market failures, such as distortions, frictions or fragmentation. In such cases, the economy is unlikely to achieve its maximum potential productivity, and may reversely slow down the improvements of factor endowment structure in the following stages, thus the economy may fail to advance to the next level of development. This underscores the critical need for improved designed comparative-advantage-adapting industrial policies.

According to this idea, developing an effective policy necessitates policymakers to understand several crucial points. First, policymakers must accurately identify the level or stage they are currently experiencing. Second, policymakers need to clearly understand the economic and endowment structures during this particular period of development. This involves referencing other regions that have experienced similar pathways and stages of development. Third, policymakers need to carefully examine whether a market failure exists and what its causes are. Fourth, policy interventions should provide both 'hard' and 'soft' infrastructure based on the compatibility between economic and endowment structures. These interventions could focus on, sequentially or simultaneously, the factor market (D. Wang, Mugera, and White 2019), product market (Y. Wang 2019), and financial market (Lin et al. 2022). The term improvements in 'hard' infrastructure typically refers to capital investment in physical infrastructure, while improvements in 'soft' infrastructure pertain to institutional reforms. These interventions aim to achieve the goals such as efficient resource allocation and market mechanisms by eliminating distortions and conflicts. Additionally, they involve government participation in facilitating industrial upgrading and infrastructure improvements to prevent potential government failures.

Huang's stage segmentation and Lin's New Structural Economics both indicate that economic development involves a series of consecutive and dynamically adjustable structural changes, and tailored policy facilitation is needed. In this regard, they both transcend the oversimplified dichotomy of development studies, which classifies economies as either 'developed' or 'underdeveloped', or 'agrarian' or 'industrial'. They both oppose the one-size-fits-all policy. This justifies the need for continuous monitoring of the rural transformation process and consistent improvement of policy design. It also reminds policymakers that domestic capacity-building should be encouraged through participation in foreign aid programs to foster an endogenous growth model.

While in practice, policymakers may still be unsure about how to effectively diagnose the stages of development, which involves determining the extent to which the stages of rural transformation are clearly defined. To conduct this evaluation process effectively, it is important to incorporate the expertise and opinions of scholars and stakeholders. A conventional way is the Delphi method, which is a structured communication and systematic and interactive forecasting method that relies on the views of a panel of experts (Dalkey and Helmer 1963) because they process domain-specific expertise that is of greater 'true' knowledge than their peers and the public. This method, though, has been well-developed and widely used in policymaking (Antonelli et al. 2022; Frewer et al. 2011; Tiberius, Gojowy, and Dabić 2022), it has some drawbacks when applied to comparative studies of international agricultural development. First, multiple rounds of interactive communication and reflection can prolong the process and consume more resources, especially when experts are based in multiple countries. Second, in a multinational panel, the position of experts, the knowledge gap of local affairs and the value placed would hinder the achievement of the final consensus within the limited time (Fischer et al. 2013). Third, Individual experts with greater influence may pressure others, forcing a reluctant and compromising consensus.

The traditional approach of involving experts should be complemented by a more advanced and objective method to assess the results and improve the efficiency of the implementation. This is especially important for international agricultural development projects that involve multiple stakeholders at various levels.

This paper is the first attempt to adopt a machine-learning approach to facilitate development policy studies on the segmenting stages of rural transformation. Machine learning has been adopted by recent agricultural and applied economics research (Athey and Imbens 2019; Heckelei, Baylis, and Storm 2020; L. Lu, Tian, and Hatzenbuehler 2022; Mullainathan and Spiess 2017), and environmental and resource economics (Kvamsdal et al. 2021; Stetter, Huber, and Finger 2024). Machine learning is to turn complex information into valuable knowledge by 'letting the data speak'. Given the big data and computational power, the machine learning approach is helpful because it reduces dimensionality and the flexibility of functional form. It can also serve for feature-importance detection, signal-from-noise extraction, model-free clustering, model-free classification and correct specification via optimal model selection (Cerulli 2023). The supervised machine learning algorithm is mainly used for prediction (Hossain, Mullally, and Asadullah 2019; Shao et al. 2020; Traoré, Jimbira, and Sall 2021; Yu et al. 2022). Unsupervised machine learning is used to patterns of cognition using the clustering algorithm.

Such research could provide helpful, practical and specific policy recommendations for policymakers and practitioners. For example, Khalaf, Michaud, and Jolley (2021) use the clustering techniques to reveal the typology of the U.S. rural idiosyncrasies in resources, opportunities and challenges by county-level data. The findings conclude on what type of rurality requires what kind of specific economic development strategies. H. Wang and Yu (2023) use the clustering techniques to reveal the driving pattern of carbon dioxide emission in China's provinces, and they imply the energy transition pathway for each group of provinces that share the same driving forces. Liu et al. (2023) classify the pattern of egg price dynamics of Chinese provinces. By clustering consumers' expenditure and price elasticities on fruit and vegetables, Blumberg and Thompson (2021) identify what groups of consumers with similar behaviours to inform them of the marketing methods to boost consumption. In development studies, Zeng and Chen (2023) use the provincial-level panel data to categorise the types of urban-rural integration. They classified provinces into four types—low-level integration, high-level integration, integration in transition, and early integration in the backward stage. Furthermore, they present the evolution of spatial distribution over three decades. They further argue that different types of regions require different policy instruments. Apart from these, there are two papers focusing on classification regions based on development indicators. One is on clustering Russian regions based on the level of socio-economic development by using a set of selective indicators (Ketova, Kasatkina, and Vavilova 2021); the other international study classifies all countries into three groups based on the Human Development Index (H. Wang, Feil, and Yu 2023). However, these two papers only present the clustering method without analysing policies.

We argue here that machine learning does not diminish the value of expert involvement; rather, it serves as a supplementary tool by emulating human cognitive functions to learn and address problems. It provides extra support in diagnosing the diversity of developmental issues, ensuring that cross-country and cross-region comparisons are more precise and comprehensive.

3 | Data

The degree of rural transformation is measured in two dimensions—the share of high-value (or non-staple food) agriculture (RT1) and the share of rural labour's non-farm employment (RT2). These indicators were proposed by J. Huang (2018), J. Huang (2020) and J. K. Huang and Shi (2021). It has been widely used in recent research (Abedullah, Farooq, and Naz 2023; Rola-Rubzen et al. 2023; Shi and Huang 2023; Sudaryanto et al. 2023; D. Wang, Chen, and Findlay 2023). High-value agricultural production (e.g., vegetable, fruit, livestock and aquaculture) requires relatively intensive investment in capital and technology, along with diversifying crop and animal varieties and creating a more complex supply chain involving food processing, storage services etc. The higher return leads to capital accumulation and creates new job opportunities in agribusiness, thereby fostering the transformation of the rural economy. Non-farm employment of rural labour reflects the movement of the labour market, leading to the reallocation of resources, increased productivity and the growth of non-agricultural industries in rural areas. It involves a more profound structural change in the economy. Therefore, the higher values of RT1 and RT2 can indicate a higher level of rural transformation. In this paper, RT1 is defined by the value of high-value agricultural products divided by the gross agricultural output values, and RT2 is defined by the rural non-agricultural employment divided by total rural employed labourers.

We gathered raw data from different sources in four countries. Data from Bangladesh were obtained from the Bangladesh Bureau of Statistics (BBS), including the Yearbook of Agriculture, the Statistical Yearbook of Bangladesh and Labour Force Survey. Chinese data were collected from various editions of provincial Statistical Yearbooks. Indonesian data were sourced from Statistics Indonesia (Badan Pusat Statistik—BPS), while data from Pakistan were collected from the Pakistan Bureau of Statistics (PBS) and Various Annual Reports of the Pakistan Economic Survey (PES). In Bangladesh and Pakistan, based on availability, data were collected at the district level, with 32 districts in Bangladesh and 45 districts in Pakistan selected. Data were collected at the provincial level in China and Indonesia, with 24 provinces in China and 29 provinces in Indonesia selected. The selection criterion of regions is based on data availability and whether it is a grain-based region. The yearly data ranges are different in four countries due to the availability. Bangladesh's data span from 1991 to 2020; China's data span from 1978 to 2018; Indonesia's data span from 2001 to 2020; Pakistan's data span from 1998 to 2019.

4 | Methodology

4.1 | Expert Panel Formation

The research started by establishing an expert panel representing Bangladesh, China, Indonesia, Pakistan and Australia who had both domestic and international experience. The panel was assembled through a selection process based on merit, with individuals invited to contribute by the Australian Centre for International Agricultural Research (ACIAR), which has established trustworthy relationships with agricultural experts from the four countries over the past 40 years. The initial stage involved identifying the necessary skills and participation in the investigation. The experts were chosen based on three criteria. Firstly, they had to have cognitive competency, which means they needed to possess true knowledge of their countries and have extensive and intensive domain-specific expertise. Secondly, they were required to have significant policy influence in their home country, demonstrated by their long-term connections with policymakers and stakeholders and their participation in policy decision-making. Thirdly, all experts involved needed to have foresight, meaning they had to understand how today's policy decisions impact or shape the future of the rural economy in their countries.

The expert panel was then organised into two layers. The core research group comprises 44 scholars from universities or research institutes who consistently worked together within each country. The core research group was supported by an advisory group consisting of 20 experts from government and industry who occasionally contributed insights and comments to the core research group. Researchers from the Australian National University (ANU) co-designed the research with both groups and led the research.

4.2 | Research Design

The research combines conventional expert evaluation methods with machine learning techniques to assist policy design. The whole process is illustrated below (Figure 1).

In the first step, based on literature and data availability, the expert panel decided to measure rural transformation using two indicators: the share of high-value agriculture (RT1) and the share of non-farm employment (RT2). The expert panel then considered whether Huang Segmentation, which was established and confirmed in China, is logical and applicable to the other countries involved in the study: Bangladesh, Indonesia, and Pakistan. From this consideration, the expert panel agreed that, conceptually, the rural transformation segmentation across multiple countries could be considered through the following stages.

- *Stage I*: Staple food production
- *Stage II*: Agricultural diversification and commercialisation

- *Stage III*: Specialisation/mechanisation and non-farm employment
- *Stage IV*: High-value and sustainable agriculture and integrated urban-rural development

Before applying the Huang Segmentation to other countries, it was empirically tested for validation. Since Huang Segmentation is backed by a historical review of China and widely adopted in literature, we first input Chinese data RT1 and RT2 into a machine learning model to see whether the machine can return the same result as Huang Segmentation asserted. We use K-mean clustering techniques to classify unit-free data points preprocessed by the principal component analysis (PCA) into a number of clusters generated by a computing algorithm. The clusters of data points were then demonstrated along with the

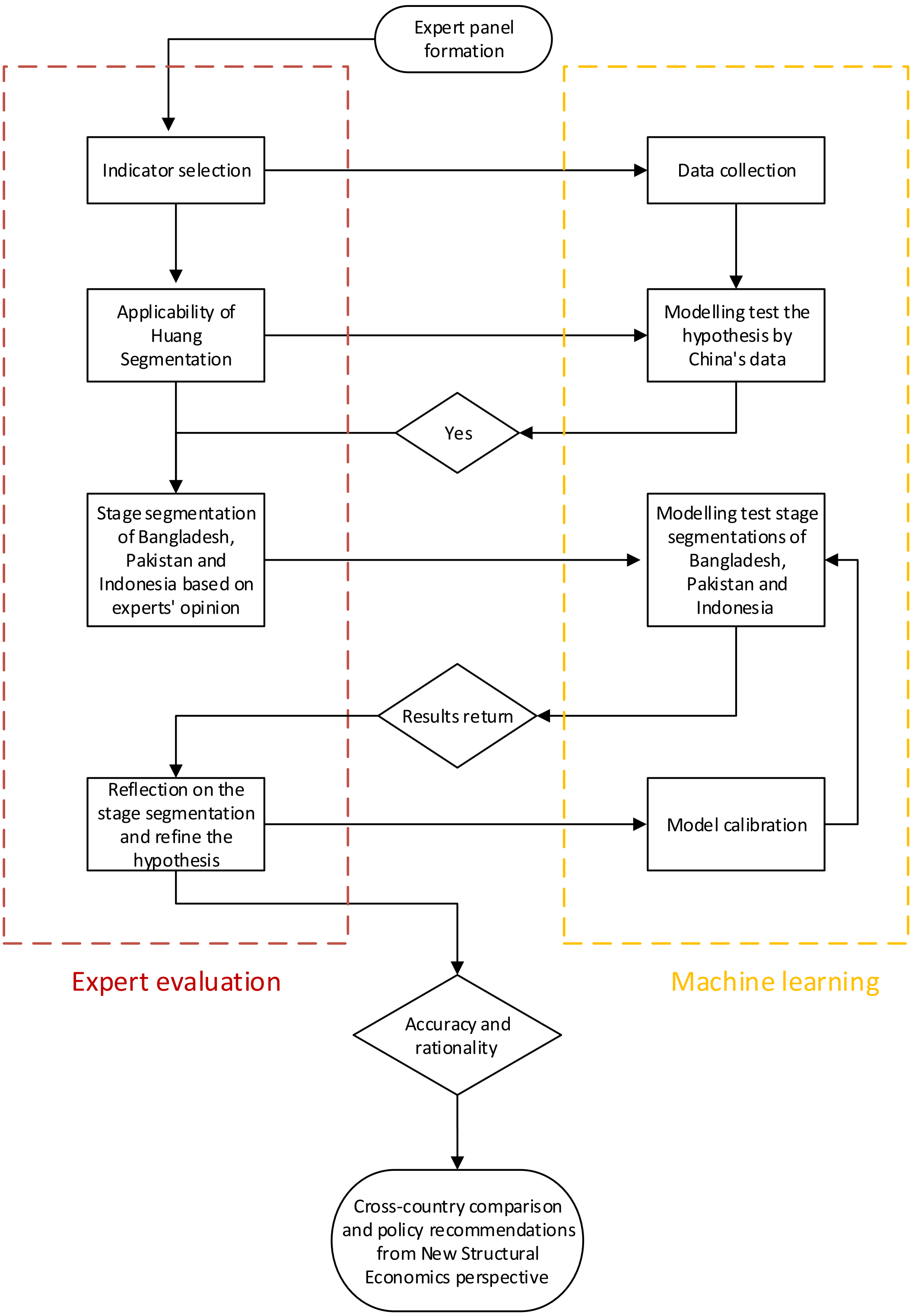

FIGURE 1 | The research process.

progress of the rural transformation over time. We can identify the sequence of stages based on the clusters' appearance along the time axis. We also observe how the results align with landmark-associated policies in history to verify the robustness and validity of the method. Then, we assess whether the clustering results produced by the machine align with Huang's assertion by examining the periods of time and the years of transition from stage to stage. If the machine learning model yields results consistent with Huang Segmentation, along with notable government policies from history across different stages, it validates and reinforces the effectiveness of Huang Segmentation for application in other countries. This also instils greater confidence in the expert panel regarding using Huang Segmentation for their respective nations.

Having tested the validity of Huang Segmentation with Chinese data, it was then used to evaluate stage segmentations for Bangladesh, Pakistan, and Indonesia with relevant country data. Research on the rural transformation stage in these countries is relatively rare. The relevant stage divisions are vague and unclear. Therefore, we need to combine expert content-specific knowledge with computing aid techniques to make decisions. For each country, experts first assemble to evaluate at the national level which stage they are experiencing. Second, we review the available data. Based on the time frame of the data coverage, the experts, based on their expertise, predict how many stages might be able to be observed within that period covered by available data. These hypotheses then turn to some pre-determined parameters that are tested in the model. Third, the model returns some preliminary results on showing the clusters of data as well as the stage numbers, which enable experts to review their hypotheses on the RT process they encountered and refine their segmentation while adjusting the model to enhance reliability. Since we have already determined the current stage one country is experiencing (the last stage), we can arrange the previous stages in reverse order. In this step, experts need to examine the rationale behind the clusters by comparing the stages with the landmark policies and utilising their expertise. We also need to reflect the selected stage numbers by combining the tests from machine learning and expert opinions. Some outliers also need special consideration to determine if they are reasonable and what policy implications can be drawn behind them. This iterative process was conducted in several rounds until we obtained reasonable and reliable results, which is called model calibration. This calibration loop also involves external stakeholders who were consulted through the country advisory groups to reflect on the modelling and placement of hypotheses. Figure 1 presents the research workflow. This process is iterative and interactive, involving expert opinion and data-driven modelling to ensure interaction between machine learning and expert evaluation activities across the workflow of two components. We repeat this workflow from country to country. The machine learning effectively assists with pattern recognition, aiding experts in segmenting stages and diagnosing associated development problems.

Finally, we compared the stages and associated policies of the four countries at both the national and regional levels and examined the distinctions. Using the principles of New Structural Economics, we assess the driving forces of transformation and the applicability of policies at each stage from this perspective. This holistic approach focuses on changes in comparative advantages and structural shifts in factor endowments, while considering social, natural, historical, and geographical conditions. Subsequently, the expert panel was able to provide a comprehensive policy recommendation template based on the research process and shared experiences of these countries. In order to identify suitable policy interventions corresponding to each phase of rural transformation, we undertook comprehensive historical analyses and examined a range of policies informed by the principles of New Structural Economics. Firstly, the expert panel assembles to conclude the shared characteristics of each stage across the four countries. We summarise these characteristics from the perspective of changing comparative advantages, considering natural conditions and resources, as well as labour and capital while comparing the situation with those of other countries and the home country's other stages. Secondly, the expert panel reflects on the policies implemented at each stage, drawing on shared experiences and lessons. It involves the investigation of common and particular policy interventions. The common policy intervention is that most countries often adopt similar policy interventions at specific stages. The particular policy intervention refers to the policies adopted only by specific countries at certain stages to address unique problems they face. Thirdly, by bringing together those with shared experiences, the expert panel produces a policy design matrix to guide policymakers. It highlights the policy interventions that should be considered at each stage; see Table A1.

4.3 | Principal Component Analysis (PCA)

We performed a tandem analysis that constructs a composite index of rural transformation followed by a machine-learning K-mean clustering algorithm to identify stages of the rural transformation. The composite index combines RT1 and RT2 to measure the extent of rural transformation. We use principal component analysis (PCA) to determine the loading factors (weights) of RT1 and RT2 in a linear combination, which follows the methodology illustrated by Greenacre et al. (2022) and the OECD *Handbook on Constructing Composite Indicators* (OECD, European Union, and Joint Research Centre of the European Commission 2008). Firstly, we group RT1 and RT2 by region[2] within a country and standardise the data to reduce dimensionality. Secondly, we carry out a PCA algorithm to identify the first principal component, which represents the direction in the feature space that exhibits the most variation with the data. We then use this principal component to create a linear combination specific to each region, which effectively explains the variance of RT1 and RT2 to the maximum extent. This linear combination represents a composite index that contains the most information on rural transformation derived from RT1 and RT2 because it captures the maximum variance from the original variables based on the first principal component scores. The weights of each original variable, RT1 and RT2, are determined by the eigenvectors of the covariance matrix of RT1 and RT2.

We allow the weights of RT1 and RT2 to vary by region. That is, the relative importance of the structural change within the agricultural sector and that between agriculture and other sectors is heterogeneous and determined by some inherent idiosyncrasies of a region, say, factor endowment structure. The rationale behind this is based on the theory of New Structural Economics, which asserts that one region's economic structure is endogenously determined by its factor endowment structure. For example, in some natural resource-rich areas, the transformation of rural areas may be primarily driven by high-value agricultural production, which heavily relies on natural resources. Conversely, in regions with abundant labour, rural transformation may be more linked to the growth of non-farm employment.

Further, we normalise the composite index using the maximum and minimum values of the whole sample of each country. That is,

$$\text{index}_{\text{normalised}\_i} = \frac{\text{index}_i - \text{index}_{\min}}{\text{index}_{\max} - \text{index}_{\min}}$$

This conversion translates the numbers into a percentage form ranging from 0 to 100. It marks the ranking position of each data point within a country, making it interpretable in terms of economics. Here, the index is used to represent the process of rural transformation in one region not only compared to its own previous and future states but also in comparison to other regions during the observed period, rather than representing a magnitude of combined shares of RT1 and RT2.

4.4 | K-Means Clustering

After creating a composite index, we then apply the K-mean clustering technique to segment the rural transformation stages of each country. We establish a two-dimensional coordinate system to depict the rural transformation's 'status' using both the index and the year as the time reference. That is, each data point has two attributes: the index represents the relative ordering of rural transformation degree compared to the other data points, and the year represents the time effect of the associated data point. We consider the time effect as it can be interpreted as the technological progress and the 'environment' of the rural transformation process at that time, which is provided by the New Structural Economics—at every specific time, the factor endowment structure is exogenously determined. Economically speaking, we distinguish stages not just based on the magnitude of the index but also consider the prevailing conditions of the era and associated endowment conditions.

In the algorithm, K-mean clustering is an unsupervised iterative learning procedure that partitions the data points into K clusters, in which each data point is assigned to one of the K stages based on the feature of similarity—the similarity is measured by the nearest mean of data points. Thus, this algorithm aims to find stages in data and the number of stages, K. Intuitively, we categorise the 'status' of rural transformation similar to the others into one cluster to represent one stage.

In this study, we choose the Manhattan distance to compute the similarity of every two points. Manhattan distance measures the absolute differences between two points along axes in each dimention. That is, for two points $P_i(\text{index}_i, \text{year}_i)$ and $P_j(\text{index}_j, \text{year}_j)$, the distance ($D$) is

$$D_{ij} = \left|\text{index}_i - \text{index}_j\right| + \left|\text{year}_i - \text{year}_j\right|$$

Economically speaking, Manhattan distance depicts the shortest rural transformation pathway as the sum of the rural economy's and factor endowments' structural change. In other words, we regard two data points as belonging to the same stage if they are simultaneously closer in terms of economic structure and endowment structure. Its merits include its use in situations where the two dimensions of data points are not comparable and its reduced sensitivity to the scale effect of data transformation, making it superior to Euclidean distance. Notably, the number of optimal clusters (stages), K, is a priori and needs to be chosen by the researcher. To deal with the choice of the number K, we should take into the context of development study as well as with computational tests. 'The clustering should not be treated as an application-independent mathematical problem but should always be studied in the context of its end-use' (Lesmeister 2019, 187). Therefore, we incorporate the expert's intimate knowledge of their country and the elbow method algorithm (Makles 2012) to determine how many numbers we should choose for the data analysed.

In the first step, the expert panel comprehensively reviews the historical process and characteristics of rural transformation for all participating countries. The experts evaluated what stage each country is experiencing at the aggregate level. Second, the country experts also qualitatively summarised how many stages each country has experienced and the stage segmentation of rural transformation in that country, following the framework of Huang Segmentation. Third, the expert panel reviewed the data coverage to propose the possible number of stages (K) that the data covers. Fourth, we test the experts' opinions on K selection and the numbers generated by the elbow method. Fifth, the results are returned to the expert panel for reflection and revision on the K selection. We go through the process of testing, reflecting, and revising multiple times until we arrive at a reasonable number that can be verified by the machine and aligns with the expert's knowledge.

5 | Results and Comparison

This section presents the Huang Segmentation results for the four countries at the regional level clustering. This result can provide valuable insights for policymakers, allowing them to understand the 'ease' and 'pace' of transformation at the regional level and identify the forerunners and laggards in every stage. These insights can inform targeted acceleration policies specific to various regions to improve rural transformation.

In policy discussions and comparisons, we will adhere to the analytical procedures outlined in New Structural Economics (Lin 2012). First, results are discussed, with particular interests in some outlier regions, from the perspective of their factor endowment structure and then assess the latent comparative advantages based on the region's characteristics. Afterwards, overall patterns are compared across countries and regions. During this comparison, we may identify another region as a reference that had a similar situation regarding natural conditions, endowment structure and comparable advantages in the past. As this reference region has further progressed through the stages, it allows learning for the observed region. Generally, successful rural transformation occurs gradually due to the relatively full utilisation of comparative advantages. Therefore, if the examined region aims to emulate this successful model in the reference region and expedite its rural transformation, policymakers must identify any barriers and bottlenecks specific to the region. For instance, if institutional or transaction costs are hindering the region from progressing to the next stage.

5.1 | China

Figure 2 shows China's stage segmentation of rural transformation by 24 provinces data from 1978 to 2018. China has progressed through the four stages according to the Huang Segmentation model. The model states that the first stage was before 1980, the second stage started in the 1980s, the third stage began in the 1990s, and the fourth stage began in 2010. Figure 2 provides a more detailed pattern of the stage transitioning years and regional heterogeneity as the thickness of the data point distribution represents the cross-regional disparities of rural transformation.

China's rural transformation would have been influenced by a number of landmark policies, which we can use to justify these results. The foundation of China's rural development was based on the Household Responsibility System (HRS), a well-known institutional reform in the late 1970s. In the HRS, land or specific tasks were contracted to individual households for a set period of time. Once farmers have met their state quota obligations, they are free to keep their surplus for personal use or sell it on the market. The HRS allowed farmers to accumulate and commercialise surplus from farming activities, which incentivised them to increase staple food production. The HRS enabled China to declare in 1984 that it had ensured the people's basic needs for food and clothing. Following the agricultural diversification and commercialisation process started in the mid-1980s. This period's policies was inclined to be market-oriented and focused on establishing the agri-products transaction and circulation system (F. Lu 1999). The market orientation stimulated rural enterprises to rapidly emerge in China in the 1990s. This period also saw part-time off-farm employment emerging in rural areas. These rural enterprises accumulated capital and brought about rural structural change. The policy approach in the 1990s focused on partial reform through a 'step-by-step' approach rather than a 'big-bang' change (Chen et al. 1994).

The increase in capital intensity of the rural economy cultivated new comparative advantages that resulted in the expansion of specialisation and mechanisation and associated full-time off-farm employment after the mid-1990s. During this period, the Chinese government implemented a wide range of policies to support the rapid rural transformation, including increased investment in infrastructure, stabilising and improving contracted management of rural land, and promoting extension services. For example, during this period, the Chinese government committed to a national modern agriculture development plan and provided subsidies for agricultural machinery,[3] consolidated farmland to facilitate larger-scale farming operations, promoted rural financial services and supply chain integration, and provided farmers with more capacity-building and extension services.[4] As a result, there was a significant increase in agricultural specialisation and mechanisation from the mid-1990s to the mid-2000s. Off-farm rural employment increased during this period, making the data distribution of stage three highly dense but less heterogeneous across regions.

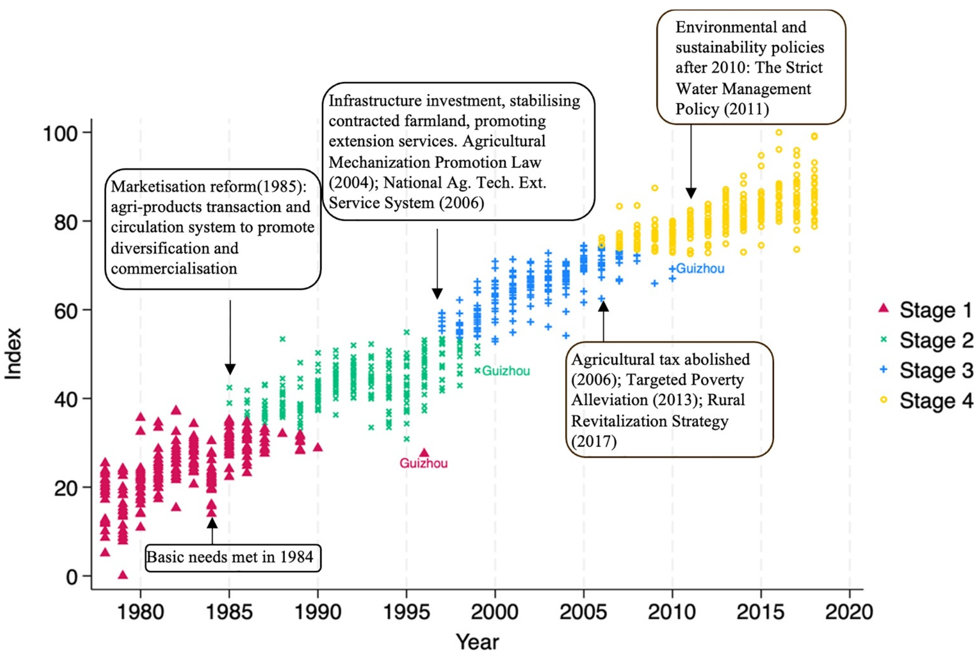

FIGURE 2 | The rural transformation stage segmentation in China.

In 2006, the agricultural tax was officially abolished nationwide and the country started moving from stage three to stage four. A key impact of the rural transformation was the reduction in the income gap between urban and rural areas. By improving the financial situation of farmers, the government aimed to promote social stability and reduce rural-urban migration pressures. After 2011, all provinces entered stage four (final stage), which is to achieve rural-urban sustainable development. More environmental and sustainability policy instruments were introduced in the final stage; for example, China launched a strict water management policy in 2011 (J. Wang et al. 2025). The final stage was also supported by the Targeted Poverty Alleviation (2013), the Rural Revitalisation Strategy (2017), and, more recently, a set of 'green' agriculture development policies.

The success of China's rural transformation was due to its pragmatic approach, starting by removing institutional barriers and aiming to cultivate endogenous driving forces of structural change and the public policy alignment to adjust with changing comparative advantages as the transformation progressed from one stage to the next.

Initially, the focus was on incentivising smallholders through market-oriented reform with the provision of required social protection. This was followed by a shift towards providing public goods, such as investing in infrastructure and extension services, while also incentivising private engagement to drive capital accumulation. Subsequently, the emphasis shifted to boosting specialisation and mechanisation through public and private financial assistance and consolidating associated land tenure reform. Finally, the policies shifted to ensuring inclusive and balanced growth between rural and urban areas.

In Figure 2, we can see that not all regions transition from one stage to another simultaneously. It takes approximately three to 5 years for all provinces to move from one stage to the next.

Lessons can be gained through consideration of the constrained rural transformation in Guizhou province, which lagged behind other provinces in its transition. This ongoing lag suggests that Guizhou province has some underlying challenges in transitioning. One limitation could be its mountainous terrain (Long et al. 2024), which requires terrace farming with high costs and labour demand. The mountainous terrain also increases transportation costs and hinders information spillover and market integration. In addition, it limits the utilisation of machines, which specifically affects stage three transformation. The underlying natural factors indicate that Guizhou has less comparative advantage in moving through the rural transformation process, with particular constraints in agricultural commercialisation and mechanisation during stages two and three.

Guizhou province also faced challenges in rural transformation in the 1990s due to misguided policies. Transitioning to a diversified and commercialised rural economy required the government to address information asymmetry, improve market access, and invest in infrastructure to reduce transaction costs. While, the policies remained focused on staple food production from 1988 to 1996 without transitioning from focusing on RT1 to RT2. This lack of policy transition meant labour was retained for low-value crops, hindering the release of excess labour and the accumulation of capital. This example highlights how policies that defy comparative advantage can hinder structural change.

5.2 | Indonesia

Figure 3 shows the Indonesian stage segmentation of rural transformation from the clustering of provincial data from 2000 to 2020.

Upon reviewing Figure 3, it becomes apparent that there are two discernible distinctions from China's pattern of stages. Firstly, the data dispersion in Indonesia seems greater than in China. This pattern suggests that during the observed years, greater regional disparities in the degree of rural transformation in Indonesia exist, which can be attributed to its status as an archipelago economy with diverse economic activities. This is likely to lead to different rural transformation processes across

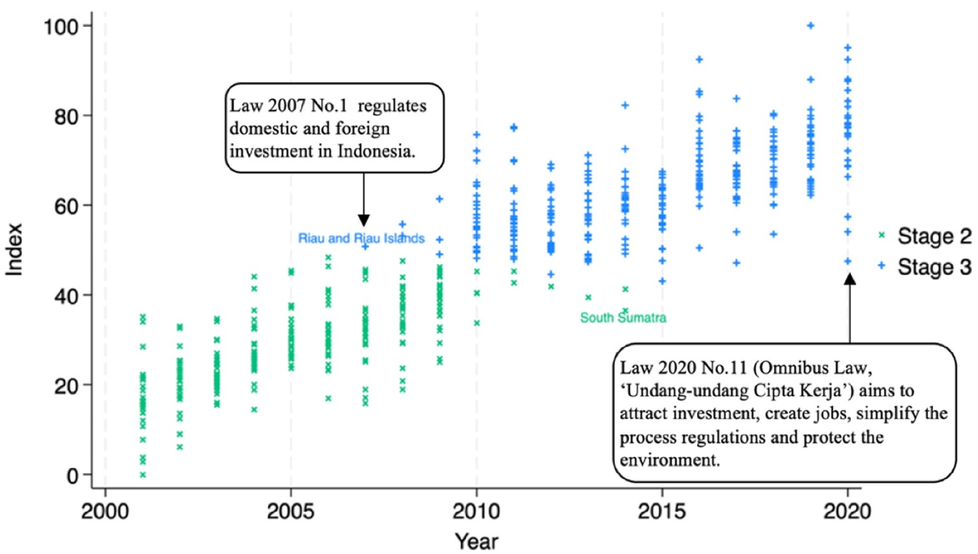

FIGURE 3 | The rural transformation stage segmentation in Indonesia.

the islands due to higher transaction and logistic costs, which may hinder the spillover effect of technology and market integration. Secondly, the two stages' data points overlap more, suggesting that the period of stage transition is notably more extensive than in China. This implies a prolonged transition period and an overall slower pace of transition than in China. Indeed, the decline of agriculture as a share of total production and employment has been slower in Indonesia compared to China over the past three decades. This could be partially due to certain agriculture sectors in Indonesia benefiting from the commodities boom of the past decade, such as palm oil, rubber, coffee, and tea. As a result, the increase in services relative to agriculture in Indonesia is also smaller compared to China. This reflects that lower value-added services have grown much larger than modern services in Indonesia.

In light of these observations, it is crucial for the Indonesian government to comprehensively understand and tactfully accelerate the transformation, taking into account the distinctive characteristics of various islands within the country. This may include governing the advancement of transformation at the regional level and effectively tackling barriers related to institutions, infrastructure, and policies.

We can observe that the Riau and Riau Islands region were pioneers in entering stage three. The Riau and Riau Islands region is adjacent to Singapore and Malaysia. In 1994, the three countries launched the Indonesia–Malaysia–Singapore Growth Triangle zone to establish free trade zones and industrial parks, attracting foreign investment and manufacturing industries. The region has been one of Indonesia's most rapid growth centres for years. International trade and foreign direct investment can significantly promote rural transformation (Hu et al. 2024). These institutional arrangements brought about rapid capital accumulation and capital deepening, thus boosting the specialisation of agricultural production and mechanisation in the rural economy. The two provinces are also endowed with rich natural resources, particularly petroleum and natural gas, which promote rapid economic development, including rural non-farm employment.

South Sumatra lagged behind in entering stage three. Agriculture in this province is dominated by staple food (rice) in the swampy environment, which is labour intensive, low productivity, high risk, and hinders mechanisation and commercialisation. Technology innovation developed by research centres up until now is more suited to irrigated land, located mostly on Java island.

The development of high-value commodities so far has been confronted with government policies biased towards staple food, particularly rice, to achieve self-sufficiency. First and foremost, the government allocated fertiliser subsidies, estimated at IDR 25.3 trillion (USD 1.66 billion) in 2023, which increased by 100.4% to IDR 50.7 trillion (USD 3.19 billion) in 2024. Second, roughly 60% of agriculture R&D has been allocated towards this strategic commodity (Sudaryanto et al. 2024). Between 2005 and 2021, the Ministry of Agriculture created 240 food crop varieties, 147 horticulture, 130 estate crops and 10 livestock breeds. A similar direction is also true in terms of infrastructure development, such as irrigation, which is more suited for rice farms. These policy frameworks have been instrumental in stage one, which brought Indonesia to achieve self-sufficiency in rice for the first time in 1984.

Diversification towards high-value commodities (stage two) was driven mainly by exploiting the international market (particularly in the case of palm oil). It is also supported by massive government support towards the development of estate crops (such as rubber, coffee, and cocoa), some of this support utilising donor funds, particularly from the World Bank and Asian Development Bank (ADB). Infrastructure development, particularly rural road and market infrastructure, contributes to this structural change.

5.3 | Bangladesh

Figure 4 presents the Bangladesh rural transformation stage segmentation. The data covers 32 districts from 1991 to 2020. The data distribution is very neat and thin, which suggests the regional heterogeneity is smaller than that of other countries. The transition period between every two stages is only 3 years, which is also shorter than that of other countries. This is mainly because Bangladesh has a smaller, flatter land area and a less complex economic structure compared to other countries.

Bangladesh's agricultural production is dominated by rice and smallholder farmers. It is rooted in its endowment structure of plenty of labour, a favourable climate for rice cultivation, and the vast plains of fertile soil. Since the Green Revolution in the 1960s and 1970s, the country has focused on achieving staple food self-sufficiency, emphasising the significant enhancement of productivity in the 1990s and early 2000s. We observe that stage two started in 2002. The early 2000s saw a gradual expansion in agricultural diversification with a continuing increase in productivity enhancements. For instance, the Northwest Crop Diversification Project (NCDP), implemented in the 2000s, aimed to promote high-value crops such as vegetables and fruits. This was followed by a subsequent phase of crop diversification projects that ran from 2010 to 2016. Another policy in 2003, the Comprehensive Village Development Programme (CVDP), highlighted infrastructure improvements such as building roads, bridges, and healthcare facilities. This programme also focused on improving education and vocational training activities. The CVDP enhanced market services and diversified farmers' income so as to help cultivate comparative advantages during this period, paving the way for transitioning to the next stage of agricultural specialisation and mechanisms accompanied by full-time non-farm employment. According to the endowment's structural change, the country wholly entered stage three after 2014, which featured agricultural specialisation, mechanisation and full-time non-farm employment. The Bangladesh government started Promoting Agricultural Commercialisation and Enterprises (PACE) projects in 2015 to support microenterprises and agribusiness services and create more non-farm employment opportunities. Apart from providing subsidies for purchasing modern farming machinery, the farm machinery manufacturing subsector had started to grow alongside the import and after-sales service provisions by private sector companies (Alam 2019). The country is also utilising excess labour to develop the garment industry and is even taking over some production capacities from China.

5.4 | Pakistan

Figure 5 presents the segmentation of the Pakistan rural transformation stage. The data covers 36 districts from 1981 to 2019. Overall, we can see the data's dispersion is thicker than that of Bangladesh, suggesting Pakistan has greater regional heterogeneity than Bangladesh. Accordingly, the transitional period between every two stages is three to 6 years, which is longer than Bangladesh's. This is due to Pakistan's larger land area and more diverse natural conditions among regions. The policy target of rural transformation from staple food production to agricultural diversification and commercialisation started in the early 1990s as a result of the Integrated Rural Development Programme (IRDP) (Gill et al. 1999). During the 1980s–1990s, Pakistani rural policy focused on market-oriented reforms, land redistribution, increasing access to basic facilities and infrastructure development. However, during this period, agriculture grew extensively rather than intensively, with growth primarily relying on area expansion while technological progress remained slow. Natural disasters further delayed the progress. Therefore, the agricultural diversification and commercialisation actually started in the mid-2000s and developed until the mid-2010s. During this period, the government launched the Khushali Bank to provide microfinance and the Khushal Pakistan Programme (KPP) to improve infrastructure (World Bank 2002, 2005). The stage one and stage two overlapped from 2003 to 2008 at the regional level. This progress lagged behind that in Bangladesh for about 5 years.

The delayed transition from stage one to stage two is likely due to the fact that Pakistani agriculture was severely influenced by El Niño extreme climate events three times during the 1990s. These events delayed agricultural diversification by not

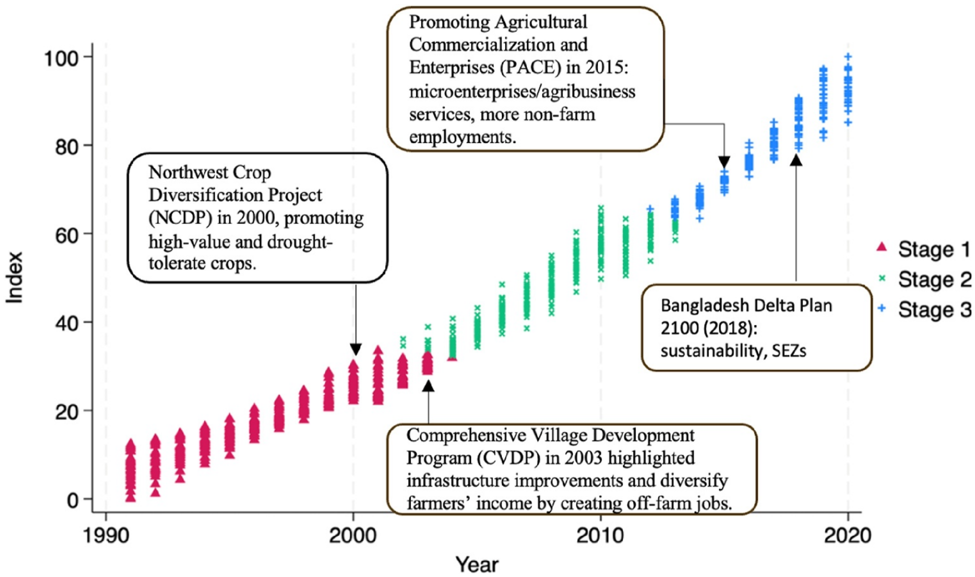

FIGURE 4 | The rural transformation stage segmentation in Bangladesh.

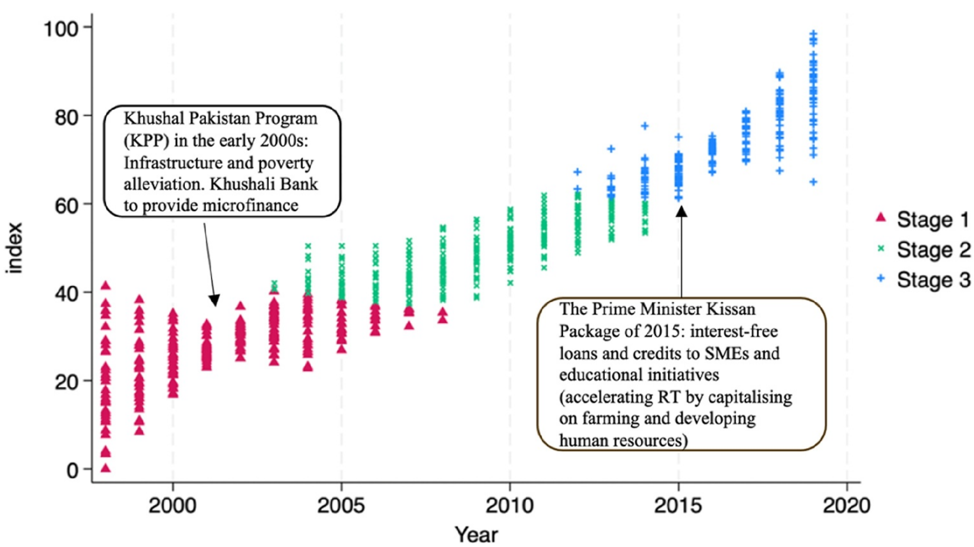

FIGURE 5 | The rural transformation stage segmentation in Pakistan.

significantly increasing the cultivation of high-value crops that demand more water supplies (Iqbal and Arif 2010). Due to high temperatures, the growing cycle of crops was disrupted and shortened, leading to early maturation and lower yields (Alam et al. 2007). Similarly, the transition to stage three was also delayed due to the devastating floods in Pakistan in 2010, which severely damaged the infrastructure, agricultural crops, livestock, and the irrigation system. The floods nearly wiped out the kharif crops, including cotton crops in south Punjab and north Sindh. In KPK (Khyber Pakhtunkhwa) and Barani areas of Punjab, the flood also caused substantial damage. Not only were the kharif crops damaged, but the cultivation of Rabi crops was also delayed, exacerbating the problem (Dorosh, Malik, and Krausova 2010). Overall, the crop output and livestock loss due to floods amounted to $2.2 billion and $870 million, respectively. Losses further translated into a reduction in agricultural income and low demand for rural non-farm goods and services, subsequently lowering the non-farm income as well (Dorosh, Malik, and Krausova 2010). This phenomenon is referred to as the multiplier effect of flood (Dorosh, Niazi, and Nazli 2003). Following the recovery from losses and in pursuit of rural transformation, agricultural specialisation and mechanisation were initiated after 2015. In this stage, the policies aimed to promote tractorization, land laser levellers and power tillers by providing financial support for agricultural machinery. For instance, the Prime Minister Kissan Package of 2015 was designed to provide interest-free loans and credit facilities to small and medium-sized enterprises (SMEs) and educational initiatives. These policies aim to capitalise on farming activities and develop human resources, which can be seen as accelerating rural transformation by enhancing the changing comparative advantages of the factors. The package specifically targeted resource-poor farmers affected by high production costs, declining prices for cotton and rice, and damage to cotton and sugarcane fields due to 2015 floods (Hussain, Akhtar, and Jabbar 2022).

6 | Conclusion and Policy Design

In this paper, we present how to incorporate the concept of Huang Segmentation and New Structural Economics to analyse the rural transformation process experienced in four grain-based economies of Asia—Bangladesh, China, Indonesia and Pakistan. At a methodological level, we demonstrated the benefit of integrating machine learning techniques into the conventional expert-driven approach to development diagnosis. We can draw conclusions below.

Firstly, we verified that Huang Segmentation truly happened in China and that the analytical framework is applicable to Bangladesh, Indonesia, and Pakistan. This would imply the potential applicability of the methodology in other developing economies to support rural transformation policy design. During the observed period, China has entered the last stage—rural-urban integrated and sustainable development. Indonesia is in the late stage three, where agricultural specialisation and mechanisation are well developed, and full-time off-farm employment is commonly widespread. Pakistan and Bangladesh are in the early-to-middle stage three of rural transformation.

Secondly, we see that Indonesia has the largest regional heterogeneity among these countries, which means it is likely that Indonesia will require more time to transition from stage to stage and that policy formulations need to take into account this regional heterogeneity (one size will not fit all). The regional heterogeneity is likely due to Indonesia's archipelago economy, which has high transaction costs and difficulties in technology spillover between islands. The important thing to achieve convergence and speed in rural transformation is to build an integrated logistics system that characterises an archipelago country. Integration of logistics systems between land and sea, and air can reduce transaction costs. In Indonesia, infrastructure development should be accelerated in the eastern part of Indonesia. The implementation of agricultural development and rural transformation that uses a regionalisation approach according to regional characteristics will be able to accelerate the convergence process. On the contrary, Bangladesh has the lowest regional heterogeneity among the four countries, partially due to its small territory on the alluvial plain formed by the Ganges and Brahmaputra (in Bangladesh, Brahmaputra is called Jamuna) rivers, noting that the exclusion of some district data may have affected the results. Indonesia could consider lessons on the integration of the domestic market through central government policy coordination to ensure cohesive development across the heterogeneous provinces, remove institutional and legal barriers between provinces, invest in transportation infrastructures like roads, railways, ports and airports, particularly to improve connectivity between remote areas and cities, to reduce logistic costs and speed up factors mobility.

Third, the special economic zone is a useful institutional arrangement. It can speed up rural transformation, particularly in the period towards stage three, observed from Indonesia's Riau and Riau Islands region. This is because international trade and foreign direct investment can accelerate capital accumulation and induce directed technology change to boost agricultural specialisation and mechanisation and create more off-farm employment opportunities. Within Indonesia, another example is the development of the Batam-Bintan-Karimun (BBK) Special Economic Zone, which other countries could consider.

Fourth, the rural transformation of mountainous terrains, such as those identified by China's Guizhou province case, can be challenging. In stage two, diversification and commercialisation may be hindered by highly fragmented areas, information gaps, and high transportation costs. In stage three, mechanisation may be impeded by the steep and hilly landscapes. To overcome some of the natural factors, in stage two, the policy should focus on increasing market accessibility, where the government needs to provide more infrastructure and technology information and services. In stage three, priority should be given to the extension of small-scale agricultural machinery and microfinance services that are suitable for mountainous agriculture. Moreover, policymakers need to understand that rural transformation in such regions would need more particular support as the transformation cost in those areas can be much higher than in other, less mountainous regions.

Fifth, policymakers can adopt the principle of New Structural Economics to examine stagewise policies and design development strategies. The comparative advantages vary from stage to stage, which calls for shifting policy focus and instruments accordingly. To guide policymakers, we provide a stage-based policy design matrix table in Appendix A; see Table A1. In the following, we will illustrate this design stage by stage.

Stage one is commonly labour and raw materials abundant but lacks capital stock, technology and managerial capacity. The focus should be on ensuring food production meets basic demand. Increasing productivity and protecting smallholders' interests and incentives should be prioritised.

In stage two, labour is relatively abundant but lacks capital. Farmers seek to diversify production to pursue high profits. This stage's policy priority should be market-oriented to help farmers generate income through commercialising farming activities. Marketisation includes removing institutional barriers to stimulate the diversification of production and commercialisation of rural products. It also includes investment in roads, bridges, agricultural schools, rural marketplace, service and extension stations. In this stage, policies supporting technology promotion should focus on simple but practical technology that smallholder farmers can easily adopt rather than promoting too sophisticated technologies. Subsidies, some protection, and microfinance support need to be considered at this stage to support smallholder farmers in the adoption of new technologies to improve productivity. The government may also enable the development of local labour-intensive manufacturing to generate non-farm jobs in industrial clusters through the Growth Identification and Facilitation Framework as proposed by New Structural Economics (Lin 2011b, 2017). The development of non-farm jobs will not only raise household incomes and increase demand for cash crops but also generate job opportunities for rural outmigration and modern agricultural inputs. The government's efforts to facilitate the development of manufacturing of the country's comparative advantages should be followed in stages three and stage four as well.

In stage three, capital becomes relatively more available than in stage two, making mechanisation feasible to release more labour resources from farming. This also enables agribusiness to emerge throughout the value chain. In this stage, policy should prioritise supporting farmers in accumulating wealth through capitalising on their farming and providing social protection. In this stage, the main policies include extending the application of machines and relatively more complex technology, promoting micro-to-medium rural financial services, more subsidies on equipment like engines, and more training and extension that enhance capacity. During this stage, the government must consider special policies and arrangements that support the unique transition for some mountainous areas. Another important thing is providing added value by downstream agricultural products, either by improving agricultural industrialisation or market downstream, which is able to absorb the workforce. The development of agricultural products downstream should be considered in relation to the development of the domestic market and integrated with the global chain.

In stage four, the rural economy has been well developed, underpinned by the modernisation of agriculture with urban areas industrialised. There are three key themes to ensure success in this stage: balancing rural and urban development, balancing rural industrialisation and the natural environment, and ensuring equal opportunity for wealthy and poor people. Addressing these three questions is crucial because it can help countries avoid the 'resource curse', 'pollution heaven' or 'middle-income trap'. Anderson (2022) argues Australia avoided the resource curse like other resource-rich country due to its inclusive political and economic institutions that can maintain and protect property rights, law and order, well-operating markets, openness, and equal distribution of income and wealth. China has shown some good practices in abolishing agricultural taxes and promoting the social welfare of rural residents (Shen, Li, and Wang 2021). To protect the natural heritage and the environment, policies need to consider 'natural capital' and how it is valued in rural environments. Policies could also consider ecosystem services and other economic activities such as rural tourism.

Finally, this paper introduces a new approach to conducting international agricultural research for development. It involves integrating machine learning to aid conventional expert-driven development diagnosis and then using New Structural Economics to analyse policies stage by stage and design stage-based development policies and strategies. This analytical methodology can be applied to similar research in the future.

Acknowledgements

This research is funded by the Australian Centre for International Agricultural Research (ACIAR) project ADP-2017-024. We thank (in no particular order) Pengfei Shi, Syamsul Hidayat Passaribu, Farah Naz, Subrata Saha, Al Amin Al Abbasi, Ismat Ara Begum, Shahid Shabir, Shujjat, Farooq, Yang Xu, Helena Purba for their great assistance. Open access publishing facilitated by Australian National University, as part of the Wiley - Australian National University agreement via the Council of Australian University Librarians.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] Here, Lin's 'level' and Huang's 'stage' can be synonymous replacements.

[2] For China and Indonesia, they are provinces; for Bangladesh and Pakistan, they are districts.

[3] The 'Agricultural Mechanization Promotion Law' was enacted in 2004.

[4] The 'National Agricultural Technology Extension Service System' was enhanced in 2006.

References

Abedullah, Farooq, S., and F. Naz. 2023. "Developing Strategy for Rural Transformation to Alleviate Poverty in Pakistan: Stylized Facts From Panel Analysis." *Journal of Integrative Agriculture* 22, no. 12: 3610–3623. https://doi.org/10.1016/j.jia.2023.11.028.

Alam, M. M. 2019. *Study to Find Out Extent and Constraints of Use of Farm Machineries Developed by National Institutes of Bangladesh to Accelerate Mechanization-Final Report*. Dhaka, Bangladesh: Krishi Gobeshona Foundation (KGF).

Alam, M., A. Rahman, M. Rashid, et al. 2007. "Background Paper Impacts, Vulnerability and Adaptation to Climate Change in Asia." https://

unfccc.int/sites/default/files/unfccc_asian_workshop_background_paper.pdf.

Anderson, K. 2022. *Structural Transformation in Growing Open Economies: Australia's Experience* [Working Paper]. Australian National University. https://acde.crawford.anu.edu.au/publication/working-papers-trade-and-development/20681/structural-transformation-growing-open.

Antonelli, M., L. Basile, F. Gagliardi, and P. Isernia. 2022. "The Future of the Mediterranean Agri-Food Systems: Trends and Perspectives From a Delphi Survey." *Land Use Policy* 120: 106263. https://doi.org/10.1016/j.landusepol.2022.106263.

Athey, S., and G. W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11, no. 1: 685–725. https://doi.org/10.1146/annurev-economics-080217-053433.

Blumberg, J., and G. Thompson. 2021. "Nonparametric Segmentation Methods: Applications of Unsupervised Machine Learning and Revealed Preference." *American Journal of Agricultural Economics* 104, no. 3: 976–998. https://doi.org/10.1111/ajae.12257.

Cerulli, G. 2023. "The Basics of Machine Learning." In *Fundamentals of Supervised Machine Learning: With Applications in Python, R, and Stata*, 1–17. Springer International Publishing. https://doi.org/10.1007/978-3-031-41337-7_1.

Chen, C., C. Findlay, A. Watson, and X. Zhang. 1994. "Rural Enterprise Growth in a Partially Reformed Chinese Economy." In *Rural Enterprises in China*, edited by C. Findlay, A. Watson, and H. X. Wu, 4–23. Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-23609-1_2.

Dalkey, N., and O. Helmer. 1963. "An Experimental Application of the Delphi Method to the Use of Experts." *Management Science* 9, no. 3: 458–467. https://doi.org/10.1287/mnsc.9.3.458.

Dorosh, P., S. J. Malik, and M. Krausova. 2010. "Rehabilitating Agriculture and Promoting Food Security After the 2010 Pakistan Floods: Insights From the South Asian Experience." *Pakistan Development Review* 49, no. 3: 167–192. https://doi.org/10.30541/v49i3pp.167-192.

Dorosh, P., M. K. Niazi, and H. Nazli. 2003. "Distributional Impacts of Agricultural Growth in Pakistan: A Multiplier Analysis." *Pakistan Development Review* 42, no. 3: 249–275. https://doi.org/10.30541/v42i3pp.249-275.

Fischer, A. R. H., M. T. A. Wentholt, G. Rowe, and L. J. Frewer. 2013. "Expert Involvement in Policy Development: A Systematic Review of Current Practice." *Science and Public Policy* 41, no. 3: 332–343. https://doi.org/10.1093/scipol/sct062.

Frewer, L. J., A. R. H. Fischer, M. T. A. Wentholt, et al. 2011. "The Use of Delphi Methodology in Agrifood Policy Development: Some Lessons Learned." *Technological Forecasting and Social Change* 78, no. 9: 1514–1525. https://doi.org/10.1016/j.techfore.2011.05.005.

Gill, Z. A., K. Mustafa, W. A. Jehangir, and M. A. Chaudhary. 1999. "Rural Development in the 21st Century: Some Issues [With Comments]." *Pakistan Development Review* 38, no. 4: 1177–1190. https://doi.org/10.30541/v38i4iipp.1177-1190.

Greenacre, M., P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina. 2022. "Principal Component Analysis." *Nature Reviews Methods Primers* 2, no. 1: 100. https://doi.org/10.1038/s43586-022-00184-w.

Heckelei, T., K. Baylis, and H. Storm. 2020. "Machine Learning in Agricultural and Applied Economics." *European Review of Agricultural Economics* 47, no. 3: 849–892. https://doi.org/10.1093/erae/jbz033.

Hossain, M., C. Mullally, and M. N. Asadullah. 2019. "Alternatives to Calorie-Based Indicators of Food Security: An Application of Machine Learning Methods." *Food Policy* 84: 77–91. https://doi.org/10.1016/j.foodpol.2019.03.001.

Hu, C., C. Chen, C. Findlay, and Y. Wu. 2024. "Foreign Direct Investment, Off-Farm Employment and Rural Labour Income: Evidence From China." *Applied Economics* 56, no. 41: 4974–4988. https://doi.org/10.1080/00036846.2023.2276086.

Huang, J. 2018. "Facilitating Inclusive Rural Transformation in the Asian Developing Countries." *World Food Policy* 4, no. 2: 31–55. https://doi.org/10.18278/wfp.4.2.4.

Huang, J. 2020. "Rural Revitalization: Rural Transformation, Structural Transformation and Government's Functions." *Issues in Agricultural Economy* 1: 4–16. https://doi.org/10.13246/j.cnki.iae.2020.01.001.

Huang, J., D. Wang, C. Findlay, and C. Chen. (2023). "Accelerating Rural Transformation in Asian Developing Countries." *Journal of Integrative Agriculture* 22, no. 12: 3565–3567. https://doi.org/10.1016/j.jia.2023.11.027.

Huang, J. K., and P. F. Shi. 2021. "Regional Rural and Structural Transformations and Farmer's Income in the Past Four Decades in China." *China Agricultural Economic Review* 24, no. 2: 278–301. https://doi.org/10.1108/caer-11-2020-0262.

Hussain, A., W. Akhtar, and A. Jabbar. 2022. "Risk Management for Small Farmers in Pakistan: A Review." *Pakistan Journal of Agricultural Sciences* 59, no. 2: 247–259. https://doi.org/10.21162/pakjas/22.334.

IFAD. 2016. Rural Development Report 2016: Fostering Inclusive Rural Transformation. https://www.ifad.org/documents/38714170/39155702/Rural+development+report+2016.pdf/347402dd-a37f-41b7-9990-aa745dc113b9.

Iqbal, M. M., and M. A. Arif. 2010. "Climate-Change Aspersions on Food Security of Pakistan." *A Journal of Science for Development* 15.

Johnston, B. F., and J. W. Mellor. 1961. "The Role of Agriculture in Economic Development." *American Economic Review* 51, no. 4: 566–593. http://www.jstor.org.virtual.anu.edu.au/stable/1812786.

Ketova, K. V., E. V. Kasatkina, and D. D. Vavilova. 2021. "Clustering Russian Federation Regions According to the Level of Socio-Economic Development With the Use of Machine Learning Methods." *Economic and Social Changes-Facts Trends Forecast* 14, no. 6: 70–85. https://doi.org/10.15838/esc.2021.6.78.4.

Khalaf, C., G. Michaud, and G. J. Jolley. 2021. "Toward a New Rural Typology: Mapping Resources, Opportunities, and Challenges." *Economic Development Quarterly* 36, no. 3: 276–293. https://doi.org/10.1177/08912424211069122.

Kingdon, J. 2013. *Agendas, Alternatives, and Public Policies (With an Epilogue on Health Care), Updated Edition: Pearson New International Edition*. Pearson Education, Limited. http://ebookcentral.proquest.com/lib/anu/detail.action?docID=5176150.

Kvamsdal, S. F., I. Belik, A. O. Hopland, and Y. Li. 2021. "A Machine Learning Analysis of the Recent Environmental and Resource Economics Literature." *Environmental and Resource Economics* 79, no. 1: 93–115. https://doi.org/10.1007/s10640-021-00554-0.

Lesmeister, C. 2019. "Cluster Analysis." In *Mastering Machine Learning With R: Advanced Machine Learning Techniques for Building Smart Applications With R 3.5*, 3rd ed., 186–211. Limited: Packt Publishing, Limited. http://ebookcentral.proquest.com/lib/anu/detail.action?docID=5667630.

Lin, J. Y. 1992. "Rural Reforms and Agricultural Growth in China." *American Economic Review* 82, no. 1: 34–51. http://www.jstor.org/stable/2117601.

Lin, J. Y. 2011a. "New Structural Economics: A Framework for Rethinking Development." *World Bank Research Observer* 26, no. 2: 193–221. https://doi.org/10.1093/wbro/lkr007.

Lin, J. Y. 2011b. "Growth Identification and Facilitation: The Role of State in the Process of Dynamic Growth." *Development Policy Review* 29, no. 3 (May): 264–290.

Lin, J. Y. 2012. *New Structural Economics: A Framework for Rethinking Development and Policy*. World Bank. https://hdl.handle.net/10986/2232.

Lin, J. Y. 2017. “Industrial Policies for Avoiding the Middle-Income Trap: A New Structural Economics Perspective.” *Journal of Chinese Economics and Business Studies* 15, no. 1: 5–18. https://doi.org/10.1080/14765284.2017.1287539.

Lin, J. Y., Z. Yang, Y. Li, and Y. Zhang. 2022. “Development Strategy and the MSMEs Finance Gap.” *Journal of Government and Economics*, 5: 100034. https://doi.org/10.1016/j.jge.2022.100034.

Liu, C., L. Zhou, L. Höschle, and X. H. Yu. 2023. “Food Price Dynamics and Regional Clusters: Machine Learning Analysis of Egg Prices in China.” *China Agricultural Economic Review* 15, no. 2: 416–432. https://doi.org/10.1108/CAER-01-2022-0003.

Long, M. S., Y. L. Zhao, C. F. Zhou, X. B. Li, L. L. Su, and Y. Zou. 2024. “Analysis of Factors Influencing Terrace Abandonment Based on Unmanned Aerial Photography and Farmer Surveys: A Case Study in Jianhe, Guizhou.” *Land Degradation & Development* 35, no. 2: 757–771. https://doi.org/10.1002/ldr.4950.

Lu, F. 1999. *Three Grain Surpluses (1984–1998) – An Alternative Explanation for the Evolution of My Country’s Grain Circulation Policy*. Peking University. https://www.nsd.pku.edu.cn/cbw/tlg1/19962015a/250405.htm.

Lu, L., G. Tian, and P. Hatzenbuehler. 2022. “How Agricultural Economists Are Using Big Data: A Review.” *China Agricultural Economic Review* 14, no. 3: 494–508. https://doi.org/10.1108/caer-09-2021-0167.

Makles, A. 2012. “Stata Tip 110: How to Get the Optimal K-Means Cluster Solution.” *STATA Journal* 12, no. 2: 347–351. https://doi.org/10.1177/1536867x1201200213.

Mullainathan, S., and J. Spiess. 2017. “Machine Learning: An Applied Econometric Approach.” *Journal of Economic Perspectives* 31, no. 2: 87–106. https://doi.org/10.1257/jep.31.2.87.

OECD, European Union, and Joint Research Centre (JRC) of the European Commission. 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. https://doi.org/10.1787/9789264043466-en

Rola-Rubzen, M. F., H. T. Vuong, C. Doll, et al. 2023. “Gender and Rural Transformation: A Systematic Literature Review.” *Journal of Integrative Agriculture* 22, no. 12: 3624–3637. https://doi.org/10.1016/j.jia.2023.10.035.

Schultz, T. W. 1964. *Transforming Traditional Agriculture*. Yale University Press.

Shao, Y., T. Xiong, M. Li, D. Hayes, W. Zhang, and W. Xie. 2020. “China’s Missing Pigs: Correcting China’s Hog Inventory Data Using a Machine Learning Approach.” *American Journal of Agricultural Economics* 103, no. 3: 1082–1098. https://doi.org/10.1111/ajae.12137.

Shen, Y., S. Li, and X. Wang. 2021. “Impacts of Two Tax Reforms on Inequality and Welfare in China.” *China and World Economy* 29, no. 3: 104–134. https://doi.org/10.1111/cwe.12377.

Shi, P.-F., and J.-K. Huang. 2023. “Rural Transformation, Income Growth, and Poverty Reduction by Region in China in the Past Four Decades.” *Journal of Integrative Agriculture* 22, no. 12: 3582–3595. https://doi.org/10.1016/j.jia.2023.10.037.

Stetter, C., R. Huber, and R. Finger. 2024. “Agricultural Land Use Modeling and Climate Change Adaptation: A Reinforcement Learning Approach.” *Applied Economic Perspectives and Policy* 46, no. 4: 1379–1405. https://doi.org/10.1002/aepp.13448.

Sudaryanto, T., Erwidodo, S. K. Dermoredjo, H. J. Purba, R. R. Rachmawati, and A. R. Irawan. 2023. “Regional Rural Transformation and Its Association With Household Income and Poverty Incidence in Indonesia in the Last Two Decades.” *Journal of Integrative Agriculture* 22, no. 12: 3596–3609. https://doi.org/10.1016/j.jia.2023.11.029.

Sudaryanto, T., A. Suryana, H. J. Purba, et al. 2024. “Reorienting Public Agriculture R&D to Achieve Resilient and Sustainable Food and Agriculture Systems in Indonesia.” *Journal of STI Policy and Management* 9 no. 2: 1–13. https://doi.org/10.14203/STIPM.2024.406.

Tiberius, V., R. Gojowy, and M. Dabić. 2022. “Forecasting the Future of Robo Advisory: A Three-Stage Delphi Study on Economic, Technological, and Societal Implications.” *Technological Forecasting and Social Change* 182: 121824. https://doi.org/10.1016/j.techfore.2022.121824.

Timmer, C. P. 2017. “Food Security, Structural Transformation, Markets and Government Policy.” *Asia & the Pacific Policy Studies* 4, no. 1: 4–19. https://doi.org/10.1002/app5.161.

Traoré, F., S. S. Jimbira, and L. M. Sall. 2021. “Nonlinear Price Transmission in the Rice Market in Senegal: A Model-Based Recursive Partitioning Approach.” *Applied Economics* 54, no. 20: 2343–2355. https://doi.org/10.1080/00036846.2021.1989369.

United Nations. 2023. The Sustainable Development Goals Report. https://unstats.un.org/sdgs/report/2023/.

Wang, D., C. Chen, and C. Findlay. 2023. “A Review of Rural Transformation Studies: Definition, Measurement, and Indicators.” *Journal of Integrative Agriculture* 22, no. 12: 3568–3581. https://doi.org/10.1016/j.jia.2023.10.038.

Wang, D., A. Mugera, and B. White. 2019. “Directed Technical Change, Capital Intensity Increase and Energy Transition: Evidence From China.” Supplement, *Energy Journal* 40, no. S1: 277–296. https://doi.org/10.5547/01956574.40.SI1.dwan.

Wang, H., J.-H. Feil, and X. Yu. 2023. “Let the Data Speak About the Cut-Off Values for Multidimensional Index: Classification of Human Development Index With Machine Learning.” *Socio-Economic Planning Sciences* 87: 101523. https://doi.org/10.1016/j.seps.2023.101523.

Wang, H., and X. Yu. 2023. “Carbon Dioxide Emission Typology and Policy Implications: Evidence From Machine Learning.” *China Economic Review* 78: 101941. https://doi.org/10.1016/j.chieco.2023.101941.

Wang, J., Z. Guo, D. Wang, et al. 2025. “Rural Transformation, Water Scarcity and Policy Responses in China.” *Asia & the Pacific Policy Studies* 12, no. 1: e70004. https://doi.org/10.1002/app5.70004.

Wang, Y. 2019. “A Model of Industrialization and Rural Income Distribution.” *China Agricultural Economic Review* 11, no. 3: 507–535. https://doi.org/10.1108/CAER-02-2019-0030.

World Bank. 2002. Document of the World Bank. https://documents1.worldbank.org/curated/en/868671468758720856/pdf/multi0page.pdf.

World Bank. 2005. Document of the World Bank. https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ppar_33612.pdf.

Yu, L., S. Liang, R. Chen, and K. K. Lai. 2022. “Predicting Monthly Biofuel Production Using a Hybrid Ensemble Forecasting Methodology.” *International Journal of Forecasting* 38, no. 1: 3–20. https://doi.org/10.1016/j.ijforecast.2019.08.014.

Zeng, Q. Y., and X. F. Chen. 2023. “Identification of Urban-Rural Integration Types in China – An Unsupervised Machine Learning Approach.” *China Agricultural Economic Review* 15, no. 2: 400–415. https://doi.org/10.1108/CAER-03-2022-0045.

Appendix A

TABLE A1 | The policy design matrix.

Stage	Path of RT	Characteristics	Policy focus
1	Staple food production	Labour and raw materials are abundant but lack capital stock, technology and managerial capacity	• Ensuring food production meets basic demand • Increasing productivity and protecting smallholders' interests and incentives should be prioritised
2	Agricultural diversification and commercialisation	Labour is relatively abundant but lacks capital	• Market-oriented reform • Investment in roads, bridges, agricultural schools, rural marketplaces • Service and extension stations • Promoting simple but practical technology • Subsidies, some protection, and micro-finance support need to be considered • The government may also enable the development of local labour-intensive manufacturing at this stage
3	Specialisation/mechanisation and non-farm employment	Capital becomes relatively more available than in stage two, making mechanisation feasible to release more labour resources from farming; and agribusiness to emerge	• Supporting farmers in accumulating wealth through capitalising on their farming and providing social protection • Extending the application of machines and relatively more complex technology • Promoting micro-to-medium rural financial services • More subsidies on equipment like engines and more training and extension that enhance capacity • Special arrangements to support some mountainous areas • Providing added value by downstream agricultural products • Further developing the domestic market and integrating with the global chain
4	High-value and sustainable agriculture and integrated urban-rural development	Rural economy has been well developed, underpinned by the modernisation of agriculture with urban areas industrialised	• Balancing rural and urban development • Balancing rural industrialisation and the natural environment (water, carbon, pollution, etc.) • Ensuring equal opportunity for wealthy and poor people and different genders • Avoiding the 'resource curse', 'pollution heaven' or 'middle-income trap'