**Evolutionary Economics and Social Complexity Science 30**

Yuichi Ikeda

Data Science of Renewable Energy

Integration

The Nexus of Energy, Environment, and Economic Growth

**Evolutionary Economics and Social Complexity Science**

Volume 30

**Editors-in-Chief**

Takahiro Fujimoto, Waseda University, Tokyo, Japan

Yuji Aruka, Institute of Economic Research, Chuo University, Hachioji-shi, Japan

The Japan Association for Evolutionary Economics (JAFEE) always has adhered to its original aim of taking an explicit “integrated” approach. This path has been followed steadfastly since the Association’s establishment in 1997 and, as well, since the inauguration of our international journal in 2004. We have deployed an agenda encompassing a contemporary array of subjects including but not limited to: foundations of institutional and evolutionary economics, criticism of mainstream views in the social sciences, knowledge and learning in socio-economic life, development and innovation of technologies, transformation of industrial organizations and economic systems, experimental studies in economics, agent based modeling of socio-economic systems, evolution of the governance structure of firms and other organizations, comparison of dynamically changing institutions of the world, and policy proposals in the transformational process of economic life. In short, our starting point is an “integrative science” of evolutionary and institutional views. Furthermore, we always endeavor to stay abreast of newly established methods such as agent-based modeling, socio/econo-physics, and network analysis as part of our integrative links. More fundamentally, “evolution” in social science is interpreted as an essential key word, i.e., an integrative and/or communicative link to understand and re-domain various preceding dichotomies in the sciences: ontological or epistemological, subjective or objective, homogeneous or heteroge neous, natural or artificial, selfish or altruistic, individualistic or collective, rational or irrational, axiomatic or psychological-based, causal nexus or cyclic networked, optimal or adaptive, micro- or macroscopic, deterministic or stochastic, historical or theoretical, mathematical or computational, experimental or empirical, agent

based or socio/econo-physical, institutional or evolutionary, regional or global, and so on. The conventional meanings adhering to various traditional dichotomies may be more or less obsolete, to be replaced with more current ones vis-à-vis contemporary academic trends. Thus we are strongly encouraged to integrate some of the conventional dichotomies. These attempts are not limited to the field of economic sciences, including management sciences, but also include social science in general. In that way, understanding the social profiles of complex science may then be within our reach. In the meantime, contemporary society appears to be evolving into a newly emerging phase, chiefly characterized by an information and communication technology (ICT) mode of production and a service network system replacing the earlier established factory system with a new one that is suited to actual observations. In the face of these changes we are urgently compelled to explore a set of new properties for a new socio/economic system by implementing new ideas. We thus are keen to look for “integrated principles” common to the above mentioned dichotomies throughout our serial compilation of publications. We are also encouraged to create a new, broader spectrum for establishing a specific method positively integrated in our own original way.

**Editors in Chief**

Takahiro Fujimoto (Prof., Waseda Univ.)

Yuji Aruka (Prof. em., Chuo Univ.)

**Editorial Board**

Satoshi Sechiyama (Prof. em., Kyoto Univ.)

Yoshinori Shiozawa (Prof. em., Osaka City Univ.)

Kiichiro Yagi (Prof. em., Kyoto Univ., Prof., Setsunan Univ.) Kazuo Yoshida (Prof. em., Kyoto Univ., Prof., Kyoto Sangyo Univ.) Hideaki Aoyama (Prof., Kyoto Univ.)

Hiroshi Deguchi (Prof., Tokyo Institute of Technology)

Makoto Nishibe (Prof., Hokkaido Univ.)

Takashi Hashimoto (Prof., Japan Advanced Institute of Science and Technology) Masaaki Yoshida (Prof., Sensyu Univ.)

Tamotsu Onozaki (Prof., Rissho Univ.)

Shu-Heng Chen (Prof., National Chengchi Univ.)

Dirk Helbing (Prof., Eidgenössische Technische Hochschule Zürich)

Yuichi Ikeda

Data Science of Renewable Energy Integration

The Nexus of Energy, Environment, and Economic Growth

Yuichi Ikeda

Adv. Integ. Stu. in Human Survivability

Kyoto University

Kyoto, Kyoto, Japan

ISSN 2198-4204 ISSN 2198-4212 (electronic)

Evolutionary Economics and Social Complexity Science

ISBN 978-981-99-8778-8 ISBN 978-981-99-8779-5 (eBook) https://doi.org/10.1007/978-981-99-8779-5

© Springer Nature Singapore Pte Ltd. 2024

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Paper in this product is recyclable.

**Foreword**

Dr. Fatih Birol of the International Energy Agency (IEA) says that renewable energy becomes a mainstream and it is getting more important in the future as a main tool to achieve carbon neutral. As everyone notices, energy issue is very complex and it is difficult for any single discipline to understand the whole nature of the energy issue. For example, electrical engineers analyze power system design and study technological elements of various devices. Economists concentrate on the economics of power systems and the institutional design of electricity markets. Environmentalists point out the danger of rising temperatures and sea level rise due to increased greenhouse gases. Undoubtedly, these individual studies are essential, but the actual implementation of their recommendations may often contradict each other. When such contradictions occur, a methodology is needed to resolve the various contradictions and obtain a higher-level solution. The authors of this book argue that such a methodology can be possible by a trans-disciplinary research approach using data science, computational science, network science, and economics. The book attempts to explain the contents as simply as possible without assuming a wide range of knowledge in the natural sciences, economics, and informatics. Practitioners and industrialists may not have the time to read this book in its entirety from beginning to end. In that case, I recommend reading through the chapters that interest you. This book will give you an unprecedented insight. I recommend this book to all those who want to think about and understand the energy issues about renewable energy and its significant positive impact on society.

Tokyo, Japan Nobuo Tanaka 2023 Chair of the Steering Committee of Innovation for Cool Earth Forum,

Executive Director Emeritus of International

Energy Agency, and CEO, Tanaka Global, Inc.

vii

**Preface**

This book explains various data scientific approaches to analyze the grid inte gration of renewable energy, for which grid flexibility is the key to coping with its intermittency or fluctuation. It provides the readers with the scope to view renewable energy integration as establishing a distributed energy network instead of the traditional centralized energy system. Data science attempts to elucidate a system through data analysis and takes a bottom-up approach. This explanation is data science in the narrow sense. Computational science, on the other hand, attempts to depict the behavior of a system using mathematical models and takes a top-down approach. Both bottom-up and top-down approaches are needed to understand a system sufficiently. In this book, we use the term “data science” broadly, including computational science. This book covers a wide range of topics. Topics addressed in the book include fluctuation and correlation of renewable energy, unit commitment model to capture intermittency of renewable energy and to operate interconnected power grid, linear programming model of smart grid, causality inference for finance in renewable energy and carbon pricing, network model and coupled oscillator model of evolving micro-grids, hydrogen production using nuclear power. This book also explains the desired innovation to reduce the integration cost significantly using innovative technologies, e.g., energy storage with hydrogen production and vehicle-to-grid technology. Illustrated by analyzing carefully selected examples of renewable integration using different types of grid flexibility, this book is indispensable to readers who make policy recommendations to establish the distributed energy network integrated with large-scale renewable energy by disentangling the nexus of energy, environment, and economic growth.

Kyoto, Japan Yuichi Ikeda September 2023

ix

**Acknowledgments**

I would like to express my deep and sincere gratitude to the following individuals equally. This book would not have been possible without the collaborations, cooperations, advices, and other forms of support of the following individu als: Nobuo Tanaka (Chair of the Steering Committee of Innovation for Cool Earth Forum, Executive Director Emeritus of International Energy Agency, and CEO, Tanaka Global, Inc.), Prof. Peter Taylor (University of Leeds), Dr. Uwe Remme (International Energy Agency), David Elzinga (Asian Development Bank), Dr. Steve Heinen (Vector Ltd.), Prof. Kazuhiko Ogimoto (University of Tokyo), Kazuto Kataoka (University of Tokyo), Prof. Takashi Ikegami (Tokyo University of Agriculture and Technology), Akira Yabumoto (J-Power), Keisuke Inoue (Tokyo Electric Power Company), Hidetoshi Nishimura (Economic Research Institute for ASEAN and East Asia, ERIA), Shigeru Kimura (Economic Research Institute for ASEAN and East Asia, ERIA), Prof. Alexander Keeley (Kyushu University), Hitomi Tanaka (DMG Mori Co., Ltd), Emi Yoshiakwa (Ripple Labs Inc.), Lauren Weymouth (Ripple Labs Inc.), Ken Weber (Ripple Labs Inc.), Zelda Marquardt (Kyoto University), Yu Kimura (Kyoto University, Interoperability Labs Ltd.), Sena Omura (Interoperability Labs Ltd.), and Takeru Shimojima (Interoperability Labs Ltd.).

xi

**Contents**

**Part I Current World**

**1 Introduction: Today’s Our System** ....................................... 3 1.1 Man-Made Systems .................................................. 3 1.1.1 Sustainability ............................................... 3 1.1.2 Beyond Reductionism ...................................... 6 1.2 Centrally Managed Power System .................................. 7 1.2.1 Energy Balance Table ...................................... 7 1.3 Elementary Power System Engineering ............................. 12 1.3.1 Mechanism of Frequency Control ......................... 12 1.3.2 Power System Stability ..................................... 15 1.4 Intermittent Renewable Energy ...................................... 17 1.5 Summary.............................................................. 18 References.................................................................... 18

**2 Statistical Analysis** ......................................................... 21 2.1 Elementary Statistics................................................. 21 2.1.1 Statistical Quantities........................................ 21 2.1.2 Probability Distribution .................................... 24 2.1.3 Correlation Coefficients .................................... 27 2.2 Statistical Estimation ................................................. 28 2.2.1 Parent Population and Parameter .......................... 28 2.2.2 Point Estimation ............................................ 28 2.2.3 Maximum Likelihood Estimation.......................... 30 2.2.4 Interval Estimation ......................................... 31 2.2.5 Hypothesis and Statistical Test............................. 32 2.3 Regression Analysis.................................................. 34 2.3.1 Single Regression Analysis ................................ 34 2.3.2 Analysis of Variance........................................ 36 2.3.3 Interval Estimation of Parameters ......................... 38 2.3.4 Multiple Regression Analysis.............................. 38 2.3.5 Test of Parameter Estimation............................... 40

xiii

xiv Contents

2.4 Time Series Analysis................................................. 42 2.4.1 Origins of Temporal Variations ............................ 42 2.4.2 Stationary Process and Auto-correlation .................. 43 2.4.3 AR(p) Model................................................ 47 2.4.4 Model Selection Using Box–Jenkins Method ............. 48 2.4.5 MA(q) Model ............................................... 49 2.4.6 ARIMA(p,d,q) Model ...................................... 50 2.4.7 Model Selection Using Information Criterion............. 50

2.5 Summary.............................................................. 52 References.................................................................... 52

**3 Fluctuation and Correlation of Renewable Energy** ..................... 53 3.1 Principal Component Analysis ...................................... 53 3.1.1 Basic Concept of PCA...................................... 53 3.1.2 Maximization of *V* [*f* ] ..................................... 56 3.1.3 Important Indices ........................................... 59 3.2 Solar Photovoltaic Power ............................................ 59 3.2.1 Cross-correlation of PV Output Fluctuation............... 59 3.2.2 System-Wide Output Fluctuation .......................... 60 3.2.3 Random Matrix Theory .................................... 61 3.2.4 Data Analysis ............................................... 64 3.2.5 Estimation of Forecast Error ............................... 68 3.3 Wind Power........................................................... 69 3.3.1 Random Walk Model of Wind Speed Time Series........ 70 3.3.2 Correlation of Wind Speed in Europe ..................... 71 3.4 Summary.............................................................. 73 References.................................................................... 73

**Part II Toward Decarbonized World**

**4 Optimization and Dynamical System** .................................... 77 4.1 Optimization Techniques ............................................ 77 4.1.1 Linear Programming........................................ 77 4.1.2 Simplex Method ............................................ 79 4.1.3 Combinatorial Programming ............................... 82 4.1.4 Metaheuristic Optimization ................................ 86 4.2 Dynamical System ................................................... 88 4.2.1 Harmonic Oscillator ........................................ 88 4.2.2 Attractors.................................................... 90 4.3 Kuramoto Oscillator Model.......................................... 92 4.3.1 Collective Motions ......................................... 92 4.3.2 Synchronization in Physical and Biological Systems..... 93 4.3.3 Kuramoto Oscillator ........................................ 93 4.3.4 Concept of Phase Time Series ............................. 94 4.3.5 Measure of Synchronization ............................... 95

Contents xv

4.4 Kuramoto Model with Inertia ........................................ 96 4.4.1 Formulation ................................................. 96 4.4.2 Illustrative Example ........................................ 99

4.5 Analysis of Generator Synchronization ............................. 101 4.5.1 Synchronization............................................. 101 4.5.2 Fluctuating Electricity Market ............................. 101

4.6 Summary.............................................................. 103 References.................................................................... 104

**5 Grid Flexibility**.............................................................. 105 5.1 Technologies for Balancing Power .................................. 105 5.1.1 Gass Turbine Materials..................................... 105 5.1.2 Electricity Storage .......................................... 117 5.1.3 Demand Side Management................................. 121 5.2 Simple Model of Smart Grid ........................................ 123 5.2.1 Assumptions ................................................ 124 5.2.2 Formulation of Linear Programming ...................... 125 5.2.3 Parameters .................................................. 127 5.2.4 Load Curve and Load Shifting ............................. 129 5.2.5 Results and Discussions.................................... 130 5.3 Summary.............................................................. 136 References.................................................................... 137

**6 Grid Integration of Renewable Energy** .................................. 139 6.1 Unit Commitment Model ............................................ 139 6.1.1 Power Grid with Fluctuated Renewables .................. 139 6.1.2 Model: Objective Function ................................. 140 6.1.3 Model: Global Constraints ................................. 141 6.1.4 Model: Local Constraints for Thermal Power Plants ..... 144 6.2 Application 1: A Small Power Grid ................................. 145 6.2.1 Analyzed Power Grid....................................... 145 6.2.2 Results and Discussions.................................... 147 6.2.3 Summary 1 .................................................. 153 6.3 Application 2: Two Large Power Grids ............................. 154 6.3.1 Integrated Operation of Power Grids ...................... 154 6.3.2 Analysis of Tokyo and Tohoku Power Grids .............. 155 6.3.3 Wind Power Curtailment in Tohoku System .............. 160 6.3.4 Tokyo–Tohoku Grid Interconnection ...................... 162 6.3.5 Summary 2 .................................................. 164 References.................................................................... 165

**7 Evolving Microgrid Network and Power Market** ....................... 169 7.1 Complex Networks................................................... 169 7.1.1 Adjacency Matrix........................................... 170 7.1.2 Node Degree ................................................ 170 7.1.3 Path Length and Clustering Coefficient ................... 171

xvi Contents

7.1.4 Weight Matrix .............................................. 172 7.1.5 Node Strength............................................... 173 7.1.6 Small-World Network ...................................... 173 7.1.7 Strogatz *β* Model ........................................... 175 7.1.8 Scale-Free Network ........................................ 177 7.1.9 Barabasi–Albert Model..................................... 178

7.2 Microgrids ............................................................ 179 7.2.1 Small Island Developing States ............................ 179 7.2.2 Renewable Energy in Microgrid ........................... 180 7.2.3 Analysis Conditions ........................................ 184 7.2.4 Estimation of System-Wise LCOEs ....................... 186 7.2.5 Nexus of Energy, Environment, and Economic

Growth ...................................................... 187 7.2.6 Evolving Microgrid for SIDS Economies ................. 190 7.3 Summary.............................................................. 193 References.................................................................... 194

**8 Foreign Direct Investment in Renewable Energy** ....................... 197 8.1 Causal Inference ...................................................... 197 8.1.1 Exploratory Factor Analysis ............................... 197 8.1.2 Structural Equation Modeling.............................. 198 8.1.3 Illustrative Example ........................................ 199 8.2 Carbon Pricing and Investment to Developing Countries .......... 203 8.2.1 Development Assistance by Governments ................ 203 8.2.2 Foreign Direct Investment by the Private Sector .......... 204 8.2.3 Types and Objectives of Foreign Direct Investment ...... 204 8.2.4 Foreign Direct Investment and New Trade Theories...... 205 8.2.5 Carbon Pricing Mechanism ................................ 207 8.2.6 Carbon Pricing and FDI .................................... 207 8.3 Investment and Development in Renewable Energy................ 209 8.4 Precedent Studies on the Determinants of FDI ..................... 210 8.4.1 Traditional FDI Determinants.............................. 210 8.4.2 Determinants of FDI in the Renewable Energy Sector ... 212 8.5 Analysis of FDI in Wind Power Plants.............................. 213 8.5.1 Data of Wind Power Plants................................. 213 8.5.2 Analysis Results ............................................ 214 8.6 Summary.............................................................. 217 References.................................................................... 217

**Part III Our Future World**

**9 Hydrogen Energy Storage and Nuclear Energy** ........................ 223 9.1 Hydrogen as Energy Storage ........................................ 223 9.2 Analysis Conditions.................................................. 224

9.2.1 Load Duration Curve ....................................... 224 9.2.2 Time and Generation Technologies........................ 224

Contents xvii

9.2.3 Model Variables ............................................ 225 9.2.4 Model Parameters .......................................... 226 9.3 Analysis Model....................................................... 231 9.3.1 Setting of Analysis Conditions............................. 231 9.3.2 Linear Programming Model................................ 234 9.4 Analysis Results...................................................... 239 9.4.1 Combination of Power Generation Sources in 2009 ...... 239 9.4.2 Scenario 1: HTGR-SI Cycle, Without Emission Limit ... 240 9.4.3 Scenario 2: HTGR-SI Cycle, with Emission Limit ....... 241 9.4.4 Scenario 3: Solar-AWE, Without Emission Limit ........ 242 9.4.5 Scenario 4: Solar-AWE, with Emission Limit ............ 243 9.5 Discussion ............................................................ 244 9.6 Summary.............................................................. 245 References.................................................................... 246

**10 Blockchain Energy Trade System**......................................... 249 10.1 Crypto Asset .......................................................... 249 10.1.1 Blockchain Technology and the Crypto Asset Bitcoin ... 249 10.1.2 Ripple’s Crypto Asset XRP ................................ 256 10.1.3 Using Blockchain Technology to Solve Global Issues.... 257 10.2 Energy Trading System EDISON-X ................................ 258 10.2.1 System Architecture ........................................ 258 10.2.2 Web Application ............................................ 260 10.2.3 Monthly Process ............................................ 261 10.2.4 Demonstration Experiment................................. 262 10.3 Topological Characterization of Acquired Data .................... 264 10.3.1 Hypergraph of Transactions................................ 264 10.3.2 Cavity Detection Using Persistent Homology............. 265 10.4 Summary.............................................................. 268 References.................................................................... 269

**11 V2G: Vehicle to Grid** ....................................................... 271 11.1 System Dynamics .................................................... 271 11.1.1 Predator–Prey Model ....................................... 271 11.1.2 Stability of Dynamical System............................. 272 11.1.3 Feedback of Dynamical System ........................... 274 11.2 Electric Vehicle ....................................................... 275 11.2.1 G2V and V2G............................................... 275 11.2.2 Charging Power............................................. 277 11.2.3 Discharging Power ......................................... 278 11.2.4 Simple Grid-EV Model..................................... 279 11.3 Summary.............................................................. 284 References.................................................................... 284

xviii Contents

**12 Conclusion: Prospects for the Future** .................................... 287 12.1 Society with Distributed Autonomous Organizations .............. 287 12.2 Decarbonized Society ................................................ 288 12.3 Alternative to Economic Growth .................................... 290

**A Time Series of Wind Speed** ................................................ 291 **B Random Walk Model for Wind Speed** ................................... 299 **C Correlation of Wind Speed** ................................................ 307 **Index** ............................................................................... 321

**Part I**

**Current World**

**Chapter 1** 

**Introduction: Today’s Our System**

**1.1 Man-Made Systems**

***1.1.1 Sustainability***

On the Earth, there is an inflow and outflow of energy, such as the intake of sunlight (ultraviolet, visible, and infrared light) and the emission of thermal radiation into space. On the other hand, very little matter enters or leaves the Earth. Although it is not a perfect isolated system because of the energy flow in and out, the Earth can be regarded as an isolated system as a rough approximation. Thermodynamics tells us that, in the equilibrium state of an isolated system, a homogeneous state with no structure that maximizes entropy is realized.

In the past ten years of research on the global economy from a microscopic viewpoint by network analysis with many students and collaborators, we have accumulated examples that show that the global economy is not a homogeneous state but has a characteristic pattern structure [1]. For example, international trade is actively conducted among nearby firms in similar industries. Capital ownership ties among firms are also marked by investment among firms in similar industries. International trade in intermediate goods and investment for their production facility form the world production system. The same pattern is observed in migration, with migrants sending international remittances at the same level as the investment amount among international firms. This characteristic pattern structure is evidence that implies the global economy is not in an ideal equilibrium state.

Let us consider a more realistic system than the isolated one discussed above. The Earth has a geographically uneven distribution of various resources, such as raw materials and various factors of production. For example, crude oil is unevenly distributed in the Middle East, North America, and South Asia. The population is unevenly distributed among countries in the Northern Hemisphere’s mid-latitudes. By maximizing configuration entropy with the geographic maldistribution of these

© Springer Nature Singapore Pte Ltd. 2024

3

Y. Ikeda, *Data Science of Renewable Energy Integration*,

Evolutionary Economics and Social Complexity Science 30,

https://doi.org/10.1007/978-981-99-8779-5\_1

4 1 Introduction: Today’s Our System

resources as given constraints, some structure reflecting the maldistribution of resources will emerge. However, the uneven distribution of resources alone cannot fully explain the observed characteristic pattern structure of the global economy.

Let us recall that the global economy is embedded in the global environment. The global economy takes in various resources from the environment, referred to as an external system. It emits greenhouse gases and waste to the environment as the external system based on specific values and social norms. This indicates that the global economy is not an isolated system but an open nonequilibrium system in which energy and materials flow in and out of the external system. Furthermore, the global economy can be treated as a steady state since economic growth is small in the short term on an annual scale. Therefore, the global economy can be considered a nonequilibrium steady state driven by the exchange of energy and materials with the external system. In such a state, entropy need not be maximized, which is consistent with having a characteristic structure.

The concept of general equilibrium in economics states that an economic equilibrium state is realized by maximizing the utility of all economic agents with the supply–demand equilibrium as a given constraint. This concept corresponds to entropy maximization in an isolated system and does not explicitly consider the system is nonequilibrium. If the global economy is not in equilibrium as realized by utility maximization, what is the basic principle that explains the global economy? This question provides a new research topic for the statistical physics of nonequilibrium stationary systems. The research for a new fundamental principle to replace general equilibrium is the research in econophysics [2, 3], which I have been working on with my collaborators for the past 17 years, and in network science, where I have been working with many students for the past ten years to clarify examples of pattern structures in the global economy.

Earlier, we pointed out that the global production system takes in various resources from the external system and vents greenhouse gases and waste to the external system based on specific values and social norms. Humans constitute this economic system. Humans and animals take in oxygen and grain and emit CO2 and waste; plants take in CO2 and waste emitted by humans and animals and generate oxygen and grain. Thus, we can see that the human–animal and plant systems are complementary, as shown in Fig. 1.1a. In the past, the Earth has evolved as the

(a)

Human-animaI system

(b)

Human-animaI

system

PIant system

PIant system

New system

**Fig. 1.1 Complementarity**: (**a**) Two way and (**b**) Three way, “New system” takes in the global issues represented by global warming and generates what could be useful inputs to the global economic system

1.1 Man-Made Systems 5

totality of these human–animal and plant systems, achieving a state of equilibrium under the given constraints. This is the meaning of sustainability. In recent years, however, global warming caused by excessive CO2 emissions from the economic system and other global issues resulting from economic globalization have been pointed out as increasingly severe threats to sustainability.

It would be convenient if there were a “new system” to take in the global issues represented by global warming and generate what could be valuable inputs to the global economic system. A “new system” that is complementary to the economic system, like animal–plant systems, would make it possible to improve sustainability, as shown in Fig. 1.1b. There is a very useful new tool for presenting solutions to global issues. It is a distributed ledger based on blockchain technology. Some people may have a negative impression of blockchain technology, thinking of speculation on crypto assets with bursting prices, money laundering, and fraud related to such speculation. However, it is unwise not to take great advantage because of the drawbacks. By entirely using network science, topology, machine learning, quantum informatics, and computational science, it is possible to predict the price burst caused by anomalies such as money laundering fraud to some extent. By using the power of mathematics to suppress shortcomings and taking full advantage of distributed ledgers based on blockchain technology, we can create a “new system” that can solve global issues.

Renewable energy is an effective way to reduce CO2 emissions. However, solar power and wind energy have the disadvantage of fluctuating power generation depending on the weather. An energy storage device is needed to compensate for unintended fluctuations in power generation. However, having an energy storage device to compensate for fluctuations in renewable energy output is cost-prohibitive. Fortunately, electric vehicles are becoming increasingly popular. Most vehicles spend little time in motion and much of their time parked. Therefore, we consider connecting the energy storage devices of these parked electric vehicles to the power grid and using them as a countermeasure against output fluctuations of renewable energies. Suppose a market for buying and selling the right to use electricity generated by renewable energy or the right to use the energy storage devices of electric vehicles is popularized by a distributed ledger based on blockchain technology. In that case, it will be possible to integrate more renewable energy and electric vehicles into the power grid without significantly increasing costs. This would curb CO2 emissions and provide a solution to the global problem of global warming. At this time, new and nearly forgotten traditional values will arise, such as acting not only for one’s own benefit but also for the benefit of all, that is, large scale use of renewable energy and lending one’s unused capital, which is storage devices in parked electric vehicles, to be used for other people’s purposes. This “new system” incorporates the global warming problem and generates new values complementary to the global economic system.

As a prototype of such a “new system,” we are currently developing an energy trading system using blockchain technology and conducting a demonstration exper iment at a dormitory with the cooperation of many students at our graduate school. Furthermore, in addition to measures against global warming using the energy

6 1 Introduction: Today’s Our System

trading system, we are also developing a life monitoring system that regularly measures electricity consumption, temperature, humidity, and PM2.5 concentration in the living environment using the smart contract function of blockchain technology and is developing measures to prevent depopulation in rural areas by ensuring safe living conditions. We also consider developing a system to manage personal identification to medical care based on personal records for health care using blockchain technology to improve wellbeing for an aging society.

***1.1.2 Beyond Reductionism***

To further improve sustainability, we need a major change in our thinking. To achieve this goal, first, we consider the simplicity. The concept of simplicity has its origin in the reductionism. Reductionism is the idea that the object under consideration is composed of many lower elements and that if the elements are simple enough, we can understand the nature of the object under consideration from the properties of its elements. Simplicity means that the object under consideration can be divided into sufficiently small pieces and the divided elements can be understood.

From elementary particles to the universe, it constitutes a reductionism hier archy. The orders of length and energy of each hierarchy are “particle and nuclei” *(*10−15 m*, >*GeV*)*, “atom and molecule” *(*10−9 m*,* eV*)*, “macromolecule” *(*10−6 m*,* meV*)*, “society and economy” *(*m*,* 10−3eV*)*, “planet” *(*107 m*,* 10−2 eV*)*, “solar system” *(*1012 m*,* eV-MeV*)*, and “universe” *(*1028 m*,* eV − 1020 eV*)*. The

reductionism is based on a linear theory. The laws of the higher hierarchy with a larger scale are expressed differently from the laws of the lower hierarchy with a smaller scale. When the laws of the higher hierarchy can be explained by deduction based on the laws of the lower hierarchy, we consider that we have understood the phenomenon. Each level of theory is linear, and it is possible to predict another phenomenon by superposition of solutions that explain one phenomenon at the same level. Historically, the theories at each hierarchical level have constituted a separate body of academic discipline.

However, we note that the trend of academic research has grown since the middle of the twentieth century to overcome academic barriers beyond the boundaries of a specific academic discipline in expanding the research fields of various sciences. This trend began with a theory for discussing the formation of living organisms using “the negative entropy”, i.e., the emission of entropy outside a system [4], game theory to explain the rationale and deductive aspects of human decision making [5], and cybernetics to attempt to establish a general theory of control in nature, biological systems, and society [6].

In 1970, nonlinear phenomena called deterministic chaos, characterized by sensitive dependence on initial conditions, strange attractors, and self-similarity, were found in various hierarchical layers of the universe. The existence of chaos is considered to be the most substantial evidence of the breakdown of reductionism.

1.2 Centrally Managed Power System 7

At the end of the twentieth century, the study of nonlinear and nonequilibrium systems evolved into complex systems and complex adaptive systems [7, 8]. Today, a quiet revolution to overcome academic barriers beyond a specific academic discipline’s boundaries is in progress by adding a new strong perspective of complex networks [9, 10]. This was the first era the complexity was explicitly recognized.

Looking at the history of the development of complexity from a different angle, it is also the history of the development of systems science. Systems science is the study of natural, artificial, and social systems. Natural systems are mainly the subject of natural science. It includes systems in physical science. Artificial systems are the systems mainly targeted by engineering. These include power systems, information systems, and transportation systems. The artificial system is often called the man-made system. Social systems are systems that are mainly targeted by social sciences. They include economic systems, such as labor markets, financial markets, and political systems. Of the social system, international trade and foreign direct investment, in particular, constitute the global production system. In global issues, various social and natural sciences factors are intertwined in a complicated manner. However, we need a general methodology for integrating social and natural sciences. The difficulty in tackling global issues is distilled by the need for a general methodology for integrating social sciences and natural sciences.

In the modern history of humanity, those who establish the man-made system have dominated the world. For example, the following people are well known as such: Thomas Edison (the founder of General Electric Company), Alexander Graham Bell (the founder of the company later known as AT&T), Henry Ford (the founder of Ford Motor Company), Jeff Bezos (the founder of Amazon), Larry Page (one of the founders of Google), and Mark Zuckerberg (the founder of Facebook, now renamed Meta Platforms). Here, we exclude religious and political thinkers and practitioners who contributed to the construction of social systems. Engineers who do not aim to build systems are rather recognized as craftsmen.

This book discusses the electric power system as a man-made system classified as a complex system. It examines the nexus of energy, environment, and economic growth, which is a characteristic of this man-made system.

**1.2 Centrally Managed Power System**

***1.2.1 Energy Balance Table***

The energy flow from the primary energy supply, through the conversion process, to final consumption is shown in a tabular form and is called an energy balance table. It usually shows the energy flow over a year. Globally, the International Energy Agency (IEA) of the Organization for Economic Cooperation and Development (OECD) produces an annual table by country for OECD member countries and many non-member countries. Different countries use different units of energy

8 1 Introduction: Today’s Our System

**Table 1.1** Energy Conversion Factors

| TJ | Gcal | Mtoe | MBtu |
| --- | --- | --- | --- |
| 1 | 238.8 | 2*.*388 × 10−5 | 947.8 |
| 4*.*2868 × 10−3 | 1 | 10−7 | 3.968 |
| 4*.*1868 × 104 | 107 | 1 | 3*.*968 × 107 |
| 1*.*0551 × 10−3 | 0.252 | 2*.*52 × 10−8 | 1 |
| 3.6 | 860 | 8*.*6 × 10−5 | 3412 |

From \To GWh TJ 0.2778 Gcal 1*.*163 × 10−3 Mtoe 11630 MBtu 2*.*931 × 10−4 GWh 1

depending on their statistical data. The conversion factors of the units are summa rized in Table 1.1.

The energy balance table shows primary energy production, imports, exports, inventory changes, secondary energy conversion by the energy industry, and consumption by each final energy consumption sector by energy type [11]. For energy industries such as petroleum refining, power generation, and heat supply, energy used as raw materials is shown minus, and energy produced as products is shown plus. For power generation by nuclear power, hydroelectric power, primary energy production is calculated based on the average thermal efficiency of thermal power generation. A basic understanding of different energies [12, 13] facilitates a detailed understanding of energy balance tables. The energy balance tables for Japan, Germany, and the USA in 2020 are shown in Tables 1.2, 1.3, and 1.4. In

those tables, TPES and TFC are abbreviations of “Total Primary Energy Supply” and “Total Final Consumption”, respectively.

Each column corresponds to the production of each energy, and each row shows the energy flow. Production comes from coal, gas, and oil mines. Nuclear is the estimated value, which is the primary heat equivalent of the electricity produced by a nuclear power plant with an average thermal efficiency of 33%. The flow is negative for stock building and positive for stock use. Two rows under TPES are transformation: “Electricity, CHP and heat plants” and “Oil refineries, transformation”. Transformation means a change from the primary energy to the secondary energy. Transformation is negative for input and positive for output. In Tables 1.2, 1.3, and 1.4, there are only two items for transformation, but in reality, detailed data items are maintained. Five rows under TFC are end use: “Industry”, “Transport”, “Residential”, “Commercial and public services”, and “Other final consumption”. End use means energy consumption by sector for each energy.

Using the energy balance tables, important figures are readily calculated. For instance, oil dependence of Japan in 2020 is calculated as *(*5076 + 1118*)/*16*,*109 × 100 = 38*.*45%. Further, the import dependence of Japan in 2020 is calculated as 14*,*941*/(*14*,*941 + 1816*)* × 100 = 89*.*16%. The oil dependence of Germany in 2020 is calculated as *(*3026 + 323*)/*11*,*654 × 100 = 28*.*74%. Further, the import dependence of Germany in 2020 is calculated as 8981*/(*8981 + 4046*)* × 100 = 68*.*94%. Oil dependence of the USA in 2020 is calculated as *(*36*,*261 − 6860*)/*85*,*324 × 100 = 34*.*46%. Further, the import dependence of the USA in 2020 is calculated as 19*,*852*/(*19*,*852 + 90*,*437*)* × 100 = 18*.*00%.

Total

1816

14*,*941

−497

16*,*109

−3667

119

11*,*009

3146

2615

1844

1947

1456

| Heat | . 0 0 | 0 | 150 | 23 | 0 0 1 22 0 |
| --- | --- | --- | --- | --- | --- |
| Electricity | 0 0 0 | 0 | 3633 0 | 3267 | 1145 62 952 1095 12 |
| Renewables and waste | 1277 84 0 | 1360 | −1087 0 | 265 | 140 18 9 75 22 |
| Nuclear | 423 0 0 | 423 | −423 0 | 0 | 0 0 0 0 0 |
| Natural gas | 3761 0  81 | 3858 | −2738 0 | 1123 | 424 1 390 298 10 |
| Oil products | 0  1841 −403 | 1118 | −351 5270 | 5548 | 689 2534 491 452 1382 |
| Crude, NGL, and feedstocks | 4904 0  18 | 5076 | −11  −5150 | 0 | 0 0 0 0 0 |
| Coal, peat, and oil shale | 18  4351 −94 | 4274 | −2705 0 | 784 | 748 0 0 5 30 |

**Table 1.2** Japan’s Energy Balance Table in 2020 (PJ)

Flow \Product

Production

Imports

Exports

TPES

Electricity, CHP,

and heat plants

Oil refineries,

transformation

TFC

Industry

Transport

Residential

Commercial and

public services

Other final

consumption

Total

4046

8981

−1356

11*,*654

−1789

−59

8955

2269

2150

2374

1116

1045

| Heat | 0 0 0 | 0 | 4270 | 376 | 168 0 160 48 0 |
| --- | --- | --- | --- | --- | --- |
| Electricity | 0  172 −241 | −69 | 2038 0 | 1727 | 760 40 457 451 19 |
| Renewables and waste | 2065 145 −127 | 2083 | −1339 0 | 722 | 173 140 288 86 34 |
| Nuclear | 702 0 0 | 702 | −702 0 | 0 | 0 0 0 0 0 |
| Natural gas | 169 2783 0 | 3123 | −804 0 | 2275 | 812 20 910 416 116 |
| Oil products | 0  1481 −929 | 323 | −48 3989 | 3614 | 145 1950 544 115 860 |
| Crude, NGL, and feedstocks | 130 3516 0 | 3626 | 0  −4048 | 0 | 0 0 0 0 0 |
| Coal, peat, and oil shale | 979 883 −59 | 1865 | −1361 0 | 240 | 210 0 13 0 17 |

**Table 1.3** Germany’s Energy Balance Table in 2020 (PJ)

Flow \Product

Production

Imports

Exports

TPES

Electricity, CHP,

and heat plants

Oil refineries,

transformation

TFC

Industry

Transport

Residential

Commercial and

public services

Other final

consumption

Total

90*,*437

19*,*852

−23*,*244

85*,*324

−17*,*947

−903

11*,*090

22*,*980

11*,*049

8407

7653

| Heat | 0 0 0 | 0 | 4570 | 61,178 | 197 0 .. 52 .. |
| --- | --- | --- | --- | --- | --- |
| Electricity | 0  221−51 | 170 | 15*,*2600 | 13*,*600250 | 255737533145741101 |
| Renewables and waste | 7477 82 −146 | 7387 | −3881 0 | 3503 | 1336 1408 529 166 64 |
| Nuclear | 8980 0 0 | 8980 | −89800 | 0 | 0 0 0 0 0 |
| Natural gas | 32*,*984 2471 −5203 | 30*,*105 | −12*,*1050 | 14*,*925 | 5697 1040 4603 3122 464 |
| Oil products | 0  3021 −8569 | −6860 | −358 30*,*692 | 27*,*963 | 20*,*495 587 478 5643  761 |
| Crude, NGL, and feedstocks | 30*,*208 13*,*947 −7555 | 36*,*261 | 0  −31*,*595 | 381 | 0 0 0 0  381 |
| Coal, peat, and oil shale | 10*,*788 111 −1720 | 9280 | −8341 0 | 556 | 542 0 0 14 0 |

**Table 1.4** The USA’s Energy Balance Table in 2020 (PJ)

Flow \Product

Production

Imports

Exports

TPES

Electricity, CHP,

and heat plants

Oil refineries,

transformation

TFC

Industry

Transport

Residential

Commercial and

public services

Other final

consumption

12 1 Introduction: Today’s Our System

**Table 1.5** Carbon Emission

Factors Fuel Carbon Emission Factor (*tC/T J* ) Crude Oil 20.0

Natural Gas 17.2

Gasoline 18.9

Other important figures are CO2 emission derived from energy use (CO2). This quantity is calculated as “Energy consumption (*T J* ) ×CO2 emission factors (*t*CO2*/T J* )”. This is identical to “Energy consumption (*T J* ) × C emission factors (*t*C*/T J* ) × 44/12”. Carbon emission factors are shown in Table 1.5 for fossil fuels. CO2 emission from oil consumption of Japan in 2020 is 5076 × 20*.*0 = 101*,*520*tC*. CO2 emission from natural gas consumption of Japan in 2020 is 3858 × 17*.*2 = 66*,*358*tC*. CO2 emission from oil consumption of Germany in 2020 is 3026 × 20*.*0 = 60*,*520*tC*. CO2 emission from natural gas consumption of Germany in 2020 is 3123 × 17*.*2 = 53*,*716*tC*. CO2 emission from oil consumption of the USA in 2020 is 36*,*261 × 20*.*0 = 725*,*220*tC*. CO2 emission from natural gas consumption of the USA in 2020 is 30*,*105 × 17*.*2 = 517*,*806*tC*.

**1.3 Elementary Power System Engineering**

***1.3.1 Mechanism of Frequency Control***

Since the demand for electric power changes from time to time, the output from the generators must be adjusted to these changes. To do so, the amount of change in demand must be detected in some way. In a power system based on AC transmission, changes in power demand can be detected from the rotational speed, or frequency of rotation, of the generator. Consider a simple power system consisting of one generator and one load.

*M dfdt* = *Pm* − *Pe,* (1.1)

where *Pm* is, for example, the rotational power of a steam turbine in a thermal power generator, *Pe* is the generator’s electrical output, *M* is the inertia constant, and *f* is the AC frequency. An imbalance between the supply and demand of electricity causes a difference between *Pm* and *Pe*, which appears as a change in *f* .

Table 1.6 summarizes the demand–supply relationship and frequency change. Based on the relationship between the supply–demand relationship and the fre quency change, the feedback control to keep the supply–demand balance of the power system is called frequency control. Frequency control is performed by two types of control loops: local control loops and global control loops.

1.3 Elementary Power System Engineering 13

**Table 1.6** Supply–Demand Balance and Frequency

| Rotor Acceleration |
| --- |
| Positive (Acceleration) |
| 0 (No acceleration/deceleration) |
| Negative (Deceleration) |

Supply–Demand Balance Frequency Change *Pm > Pe* Increase *f >f*0 *Pm* = *Pe* Constant *f* = *f*0 *Pm < Pe* Decrease *f <f*0

In the local control loop, the supply–demand balance *Pm* = *Pe* can be maintained by measuring the rotor frequency *f* (equal to the AC frequency) at individual generators and adjusting the amount of steam entering the thermal generator using a regulator. However, the deviation from the reference frequency *F*0 (frequency deviation) cannot be returned to 0. To eliminate the frequency deviation, a global control loop is required.

A load dispatching center can issue commands to change the rotating power to multiple generators and adjust the rotating power by speed changers at each generator to keep the AC frequency *f* of the system equal to the reference frequency *f*0.

The deviation of the AC frequency *f* to the reference frequency *f*0 is written as *Δf* = *f* − *f*0*.* (1.2)

When the governor detects a frequency deviation, the rotational force is changed accordingly.

*ΔPm* = −1*rΔf,* (1.3)

where *r* is a parameter that describes the characteristics of the governor and is called the speed regulation ratio. If the AC frequency falls below the reference value, *Δf* will be negative, meaning insufficient rotational power. The local control loop using the governor and the global control loop using the speed changer are shown in Fig. 1.2.

We consider a power system with two generators rotating synchronously and supplying power at load *PL*.

*M*1*dfdt* = *Pm*1 − *Pe*1 (1.4)

*M*2*dfdt* = *Pm*2 − *Pe*2 (1.5)

*Pe*1 + *Pe*2 = *PL.* (1.6)

Here, for generators 1 and 2, the generator inertia constants are *M*1 and *M*2, respectively, the outputs are *Pe*1 and *Pe*2, respectively, and the rotational forces are *Pm*1 and *Pm*2, respectively.

14 1 Introduction: Today’s Our System

Rotation speed

down

Speed changer

(fuIcrum at point B)

up

A

B C Up Down

To

turbine

Outward by

centrifugaI

force

Rotation speed change

Governor (fuIcrum at point A)

VaIve opening VaIve cIosing

**Fig. 1.2 Control Loops**: The Governor provides local control, and the speed changer provides global control

Suppose the load changes stepwise from *PL*0 to *ΔPL*.

*PL* → *PL*0 + *ΔPL.* (1.7)

The AC frequency of the supplied power is accordingly increased from *f*0 by *Δf* , and the rotational power is increased from *Pm*10 to *Pm*10. The rotational force changes from *Pm*10 by *ΔPm*1 and from *Pm*20 by *ΔPm*2.

*f* → *f*0 + *Δf* (1.8)

*Pm*1 → *Pm*10 + *ΔPm*1 (1.9)

*Pm*2 → *Pm*20 + *ΔPm*2*.* (1.10)

From equations (1.4), (1.5), (1.6), the following equations describe the entire power system:

*(M*1 + *M*2*)dfdt* = *Pm*1 + *Pm*2 − *PL.* (1.11)

Also, substituting equations (1.7), (1.9), and (1.10) into the above equation, we obtain the following equation for the frequency deviation:

*r*1+1*r*2)*Δf* − *ΔPL.* (1.12)

*(M*1 + *M*2*)dΔf*

*dt* = −

( 1

1.3 Elementary Power System Engineering 15

This equation can be solved immediately and yields the following solution: *Δf (t)* = *e*−*At* + *B* (1.13)

*A* =

1*~~r~~*1 + 1*~~r~~*2

*M*1 + *M*2(1.14)

*B* = *ΔPL*

*M*1 + *M*2*.* (1.15)

This result shows that the frequency deviation *Δf* does not become 0 after an infinite time and that the frequency cannot be returned to the reference frequency only by adjusting the rotational power by the governor. From the viewpoint of power quality, it is necessary to keep the frequency at the standard reference value. For this purpose, global frequency control using a speed changer called Load Frequency Control (LFC) is used.

***1.3.2 Power System Stability***

Power system stability is one of the most critical issues for integrating fluctuating renewable energy into a power grid. Consider a mass point in potential energy shown in Fig. 1.3, and explain its dynamical stability. Figure 1.3a shows that a mass point placed at point S is stationary. Even if its position is slightly changed, it will return to point S. On the other hand, a quality point placed at point U can be made to remain stationary if sufficient care is taken. However, if its position is changed even

**Fig. 1.3 System stability**:A mass point placed in potential energy is a simplification of a single machine connected to

an infinite bus. (**a**) Stable equilibrium. (**b**) Unstable equilibrium. (**c**) Transiently stable. (**d**) Transiently

unstable

(a) (b) (c) (d)

16 1 Introduction: Today’s Our System

slightly, as in Fig. 1.3b, it will fall off the potential energy pile. Point S is stable equilibrium, and point U is unstable equilibrium.

Consider the motion of masses placed at different positions as in Fig. 1.3c. The mass placed at point A oscillates near point S, dissipates kinetic energy due to friction, and eventually comes to rest at point S. The mass placed at point A’ oscillates near point S, dissipates kinetic energy due to friction, and eventually rests at point S. The same is true for the mass point placed at point A’. Such a system is transiently stable. However, Fig. 1.3d shows that the mass point placed at point B passes through point S, overcomes point U, and tumbles down to the proper valley. Such a system is transiently unstable.

In the following, we will discuss the stability of the power system based on this discussion of dynamical stability. Power systems made up of many synchronous machines and transmission lines exhibit synchronization. The stability and control of this synchronization of the machines have been studied in power system engineering [14].

The dynamical property of a single synchronous machine is often studied in the single machine connected to the infinite bus shown in Fig. 1.4a, where *Vg*, / *δ*, *I* , *xl*, *rl*, *V*∞, and / 0 are the generator voltage, generator phase, current, transmission line reactance, transmission line resistance, infinite bus voltage, and infinite bus phase, respectively. Here, the infinite bus is characterized by constant voltage and constant frequency. The active power *Pe* and reactive power *Qe* are calculated by

*Pe* + *iQe* = *Vg* exp*(iδ)I* ∗*,* (1.16)

where ∗ indicates the complex conjugate, and the current *I* is given by *I* = *Vg* exp*(iδ)* − *V*∞

*rl* + *ixl.* (1.17)

By substituting Eq. (1.17) into Eq. (1.16), the active power *Pe* is obtained as the real part of the equation

*Pe* = *VgV*∞

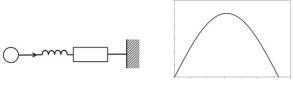
*xl*sin *δ* = *P* max sin *δ,* (1.18)

by considering relation *rl* < *xl*. *Pe* is plotted as a function of the generator phase *δ* in Fig. 1.4b. It should be noted that *Pe* is the electric power transmitted through the transmission line and works as a synchronizing force in power systems.

1.4 Intermittent Renewable Energy 17

(a) 1.20(b)

1.00

PePmax 

*Vg*/& *x V*∞/0 *l rl* G

0.80

0.60

0.40

0.20

0.00

0.00 0.50 1.00 1.50 2.00 &**(rad)**

2.50 3.00 3.50

**Fig. 1.4 Power system stability** (**a**) single machine connected to infinite bus and (**b**) synchroniz ing force are shown

wind speed log diff of wind speed

20

15

10

5

0

0 200 400 600 Time(hour)

1.0

0.0

−1.0

0 200 400 600 Time(hour)

**Fig. 1.5 NASA The POWER Project Data:** Wind speed (m/s) and its change rate at 50 m from the surface in London: Jan. 2020

**1.4 Intermittent Renewable Energy**

Hourly changes of wind speed at 50 m from the surface in London for January 2020 and July 2020 are shown in the following figures: Figs. 1.5 and 1.6, respectively. The top panel shows wind speed *st* in m/s, and the bottom panel shows its change rate defined as the log-difference of the wind speed: *rt* = log *st* − log *st*−1. As seen in the upper panel, the wind continues to blow throughout the day. This contrasts sharply with the sunshine, which is always zero at night. However, wind speed constantly fluctuates and behaves erratically. This trend is consistent throughout the year. The change rate in wind speed shown in the lower panel shows long periods with the same level of magnitude but occasionally substantial changes. The change rate is greater in winter than in summer in London.

For a time series of each month in 2020, see Appendix A.

18 1 Introduction: Today’s Our System

**Fig. 1.6 NASA The**

15

**POWER Project Data:**

wind speed

10

Wind speed (m/s) and its

change rate at 50 m from the

5

surface in London: Jul. 2020

0 200 400 600

Time(hour)

log diff of wind speed

0.2 0.4

−0.2

**1.5 Summary**

0 200 400 600 Time(hour)

In this chapter, I have attempted to give an overview of the characteristics of the socio-technical systems we use daily. In particular, the electric power system is a typical example of a system based on central control, and advanced control technologies have been developed to ensure stable operation. Today, however, we are required to integrate renewable energies such as solar and wind power into the power grid to reduce greenhouse gas emissions. Renewable energies are known to have output fluctuations, and to understand this in practice, a time series of wind speeds in London is presented.

**References**

1. Yuichi Ikeda, Hiroshi Iyetomi, Takayuki Mizuno, “Big Data Analysis on Global Community Formation and Isolation -Sustainability and Flow of Commodities, Money, and Humans”, Springer Nature, 2021, ISBN: 9789811549434.

2. Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda, Hiroshi Iyetomi, Wataru Souma, “Econo physics and Companies –Statistical Life and Death in Complex Business Networks–”, Cambridge University Press, 2010, ISBN: 9780521191494.

3. Hideaki Aoyama, Yoshi Fujiwara, Yuichi Ikeda, Hiroshi Iyetomi, Wataru Souma, Hiroshi Yoshikawa, “Macro-Econophysics –New Studies on Economic Networks and Synchronization–”, Cambridge University Press, 2017, ISBN: 9781107198951.

4. Erwin Schrëdinger, “What Is Life? The Physical Aspect of the Living Cell”, Cambridge University Press, 1944, ISBN: 0-521-42708-8

5. John von Neumann, Oskar Morgenstern, “Theory of Games and Economic Behavior”, Princeton University Press, 1944, ISBN: 978-0691130613

6. Norbert Wiener, “Cybernetics Or Control and Communication in the Animal and the Machine”, MIT Press, 1961, ISBN: 026273009X, 9780262730099

References 19

7. John H. Holland, “Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence”, MIT Press, 1992, ISBN: 9780262581110, 0262581116

8. Murray Gell-Mann, “The Quark and the Jaguar: Adventures in the Simple and the Complex”, Little, Brown & Company, 1994, ISBN-13: 978-0316907415

9. D. J. Watts , S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, Nature. 1998 Jun 4;393(6684):440-2. https://doi.org/10.1038/30918.

10. Albert-László Barabási, “Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life”, Basic Books, 2014, ISBN-13: 978- 0465085736

11. IEA world energy balances highlights (2022 edition) https://www.iea.org/data-and-statistics/ data-product/world-energy-balances

12. David J. c. Mackay, “Sustainable Energy – Without the Hot Air”, UIT Cambridge Ltd; Illustrated edition, May 2009, ISBN-13: 978-0954452933

13. Robert L. Jaffe and Washington Taylor, “The Physics of Energy”, Cambridge University Press, January 2018, ISBN-13: 978-1107016651

14. P. Kunder, “Power System Stability and Control” (McGraw-Hill, New York, 1993), p. 221.

**Chapter 2** 

**Statistical Analysis**

This chapter briefly explains the basics of statistics, regression analysis, and time series analysis. There are excellent textbooks published on these basics [1, 2].

**2.1 Elementary Statistics**

This section reviews elementary statistics as a first step in the data science needed to study various man-made systems.

***2.1.1 Statistical Quantities***

Consider data *x* with *n* components.

*x* = {*x*1*, x*2*,* ··· *, xn*}*.* (2.1)

Data *x* can be divided into the following *m* bins:

*X*0 = min[*x*]*,* (2.2)

*Xm* = max[*x*]*,* (2.3)

*ΔX* = *Xm* − *X*0

*m ,* (2.4)

© Springer Nature Singapore Pte Ltd. 2024

21

Y. Ikeda, *Data Science of Renewable Energy Integration*,

Evolutionary Economics and Social Complexity Science 30,

https://doi.org/10.1007/978-981-99-8779-5\_2

22 2 Statistical Analysis

**Fig. 2.1 Frequency**

**distribution:** Each bin

contains milk up to the red

line

where *X*0 is the lower limit of the smallest bin and *Xm* is the upper limit of the largest bin. *ΔX* is a bin size. The smallest bin is *i* = 1, and the largest bin is *i* = *m*. The frequency *Fi* is the number of data samples in each bin *Xi* − *Xi*−1. The frequency distribution

{*(X*1*, F*1*), (X*2*, F*2*),* ··· *, (Xm, Fm)*} (2.5)

is shown in Fig. 2.1.

The mean is the value obtained by dividing the sum of each element by the number of elements belonging to a population,

*μ* = 1*N*Σ*N i*=1

*xi.* (2.6)

The deviation is the difference between the value of each element belonging to the mean of the population,

*δi* = *xi* − *μ.* (2.7)

The variance is the mean square of the deviations,

*σ*2 = 1*N*Σ*N*

*i*=1

where *σ* is called the standard deviation.

*δ*2*i ,* (2.8)

We apply the transformation that makes the mean zero and the standard deviation one.

*zi* = *xi* − *μ*

*σ (i* = 1*,* ··· *, n).* (2.9)

Skewness is a measure of the symmetry of the distribution and is 0 for a normal distribution. Compared to the normal distribution, it takes a positive value when the distribution is shifted to the left and the tail extends to the right, and a negative value

2.1 Elementary Statistics 23

**Fig. 2.2 Skewness and kurtosis:** The red dotted curve represents the normal distribution being compared. (**a**) Negative skewness. (**b**) Positive skewness. (**c**) Positive kurtosis

when the distribution is shifted to the right and the tail extends to the left, as shown in Fig. 2.2a and b:

*skewness* = E[z3] =1nΣni=1z3i *.* (2.10)

The kurtosis is a measure of the sharpness of a distribution. Compared to a normal distribution, kurtosis is positive for distributions with sharp peak and long tail, as shown in Fig. 2.2c, and negative for distributions with more rounded peak and short tail:

*kurtosis* = E[z4]

E[z2]2 − 3*.* (2.11)

Next, we discuss the inequality among the elements of the population. We obtain *y* = {*y*1 *<* ··· *< yk <* ··· *< yn*} by *y* = *Sort*[*x*] in ascending order. A two dimensional plot with

rank[yk]

*n* (2.12)

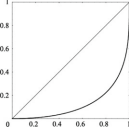
on the horizontal axis and

Σ*ki*=1 *yi*

~~Σ~~*~~n~~i*=1 *yi*(2.13)

on the vertical axis yields a Lorenz curve. The Gini coefficient is defined as the area between the Lorenz curve and a diagonal line multiplied by two. The Gini coefficient is an indicator of inequality, and its range is from 0 to 1. The larger the Gini coefficient means, the more inequality (Fig. 2.3).

24 2 Statistical Analysis

**Fig. 2.3 Lorenz curve and** 

**Gini coefficient:** Gini

coefficient is an area between

the Lorenz curve and a

diagonal line multiplied by

two

**Fig. 2.4 Probability density function of the normal distribution:**

*μ* = 0*, σ* = {0*.*75*,* 1*.*0*,* 2*.*0}

|  |  |
| --- | --- |

0.5

0.4

0.3

0.2

0.1

0.0

-6 -4 -2 0 2 4 6

***2.1.2 Probability Distribution***

In many statistical analyses, the reference probability density is the normal distribu tion or Gaussian distribution:

[

*p(x)* = *N (μ, σ*2*)* = 1

~~√~~2*πσ*exp

− *(x* − *μ)*2 2*σ*2

]

*,* (2.14)

where *μ* and *σ* are a mean and a standard deviation, respectively. We plot the probability density function of the normal distribution *N (μ, σ*2*)* for *μ* = 0, *σ* = 0*.*75*,* 1*.*0*,* 2*.*0 in Fig. 2.4.

The binomial distribution *fx* is a discrete distribution of the number of success *x* in *n*-times trials, when the probability of success is equal to *p*:

*fx* =

(*n x*

)

*px (*1 − *p)n*−*x* (2.15)

(*n x*

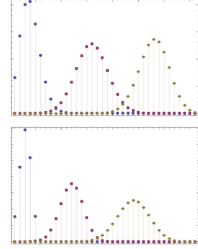
)

= *n*!

*(n* − *x)*!*x*! (2.16)

2.1 Elementary Statistics 25

**Fig. 2.5 The binomial distribution:** In the upper panel, *n* is fixed and plotted for different *p*. In the lower panel, *p* is fixed and plotted for different *n*

0.20 

*n* = 40, *p* = {0.1,0.4,0.7}

0.15

0.10

0.05

0.00

5 10 15 20 25 30 35 0.35

0.30 0.25 0.20 0.15 0.10 0.05 0.00

*n* = {5,20,40}, *p* = 0.6

5 10 15 20 25 30 35

*μ* = *np* (2.17)

*σ*2 = *np(*1 − *p).* (2.18)

Here, both *x* and *n* are integers. Figure 2.5 shows the parameter dependences of the binomial distributions. The distribution shifts toward the right as *p* or *n* increases. When *p* is small, the binomial distribution can be approximated by the Poisson distribution, a continuous probability density function.

*p(x)* = *λx*

*x*!*e*−*λ,* (2.19)

where *μ* = *λ* and *σ*2 = *λ*. Note that the mean is equal to the variance. Figure 2.6 shows the parameter dependences of the Poisson distributions for *λ* = 5*.*0*,* 10*.*0*,* 20*.*0.

Using the probability density function *p(x)*, the probability *P* can be written as *P* = *p(x)dx*. The height of the elongated rectangular area in Fig. 2.7 is the probability density *p(x)*, and the width is *dx*, so the area equals the probability *P*. The cumulative probability *P<(x)* is given by

*P<(x)* =

f *x*

−∞

*p(x*'*)dx*'*,* (2.20)

26 2 Statistical Analysis

**Fig. 2.6 Probability density function of the Poisson distribution**:

*λ* = {5*.*0*,* 10*.*0*,* 20*.*0}

**Fig. 2.7 Probability density** *p(x)* **and probability**

*P* = *p(x)dx*: The width of the rectangle in the vicinity of

|  |
| --- |

0.15

0.10

0.05

0.00

0 5 10 15 20 25 30 0.50

0.40

*x* = 1 is *dx*, the height is *p(x)*, and the area of the rectangle is equal to the probability *P*

**Fig. 2.8 Complementary cumulative probability** *P>(x)*: See Eq. (2.21) for definition

**)x(p )x(>P**

0.30 0.20 0.10 0.00

1.20 1.00 0.80 0.60 0.40 0.20 0.00

-6.0 -4.0 -2.0 0.0 2.0 4.0 6.0 **x**

-6.0 -4.0 -2.0 0.0 2.0 4.0 6.0 **x**

and the complementary cumulative probability *P>(x)* is given by

*P>(x)* =

f ∞ *x*

*p(x*'*)dx*'*.* (2.21)

The complementary cumulative probability *P>(x)* calculated from the probabil ity density *p(x)* shown in Fig. 2.7 using Eq. (2.21) is shown in Fig. 2.8.

2.1 Elementary Statistics 27

Therefore, we obtain the probability density *p(x)* by calculating the derivative of *P>(x)*:

*p(x)* = −*P>(x)*

*x .* (2.22)

The probability of event *x* and event *y* occurring simultaneously is described by a joint probability density *pxy* :

*pxy* = *p(x, y).* (2.23)

In this case, the probability of event *x* occurring, regardless of event *y*, is given by the marginal probability *px* :

f ∞

*px* =

−∞

*pxy dy.* (2.24)

Similarly, the probability of event *y* occurring, regardless of event *x*, is given by the marginal probability *py* :

f ∞

*py* =

***2.1.3 Correlation Coefficients***

−∞

*pxy dx.* (2.25)

The correlation coefficient is an indicator that relates the change in the first random variable with the change in the second random variable.

Figure 2.9 shows a scatter plot for the relationship between the US dollar (USD) and Aussie dollar (AUD) change rates. When a straight line is applied to the relationship between these variables, the square of the correlation coefficient shows

**Fig. 2.9 Scatter plot of USD and AUD:** The data are distributed near the regression line, showing a moderate positive correlation

0.03

0.02

0.01

AUD

0.00

0.01

0.02

0.03

0.02 0.01 0.00 0.01 0.02 USD

28 2 Statistical Analysis

how well the line represents the relationship. Thus, we can tell that the correlation coefficient indicates linear correlation.

Even if there is a correlation between two variables, we should not assume that one variable caused the other. Sometimes, there is no causal link when we obtain a definite correlation coefficient.

When there is a relationship between USD and AUD change rates, we can think of the following three cases:

1. USD changes cause AUD changes.

2. AUD changes cause USD changes.

3. Changes in a third variable X cause USD and AUD changes.

The first and second cases show a causal link between the two variable changes. The correlation coefficient, however, cannot be used to specify the cause of the causal link. In the third case, a finite correlation coefficient might be obtained even if there is no causal link between the two variables, called spurious correlation.

**2.2 Statistical Estimation**

***2.2.1 Parent Population and Parameter***

The parent population is a set consisting of all individuals. Parameter *θ* is a generic name for statistical quantity in a parent population. For instance, population mean *μ* is the mean value in a parent population, and population variance *σ*2 is the variance in a parent population.

On the other hand, the sample is a set consisting of sampled individuals. Sampled parameter *Θ* is a generic name of statistical quantity in a sample set. For instance, sample average *X*¯ is the mean value in a sample set, and sample variance *S*2 is the variance in a sample set.

*X*¯ = 1*n*Σ*ni*=1*xi* (2.26)

*S*2 = 1*n*Σ*ni*=1*(xi* − *X)*¯ 2*.* (2.27)

***2.2.2 Point Estimation***

Statistical estimation is an estimation of parameters for a parent population using a sample set. Point estimation is an estimation of a parameter as a scalar value. An

2.2 Statistical Estimation 29

unbiased estimator is an estimation for a parameter without bias. Sampled parameter *Θ* is called unbiased estimator of parameter *θ*, if *θ* = E[*Θ*].

For instance, sampled parameter *X*¯ is called unbiased estimator of mean *μ*, if *μ* = E[X¯ ]:

E[X¯ ] = E

[

1

n

Σn i=1

]

xi

= 1*n*Σ*ni*=1E[xi] = 1*n*Σ*ni*=1*μ* = *μ.*

(2.28)

Sampled parameter *S*2 is called unbiased estimator of variance *σ*2, if *σ*2 = *n*

*~~(n~~*~~−1~~E[S2]:

*S*2 = 1*n*Σ*ni*=1*(xi* − *X)*¯ 2

= 1*n*Σ*ni*=1{*(xi* − *μ)* − *(X*¯ − *μ)*}2

= 1*n*Σ*ni*=1*(xi* − *μ)*2 − *(X*¯ − *μ)*2*,*

E[S2] =1*n*Σ*ni*=1E[*(*xi − *μ)*2] − E[*(*X¯ − *μ)*2] = 1*n*Σ*ni*=1E[*(*xi − *μ)*2] − E⎡⎣(1nΣni=1xi − *μ*)2⎤⎦ = 1*n*Σ*ni*=1E[*(*xi − *μ)*2] −1n2Σni=1E[*(*xi − *μ)*2] = *n* − 1

*n σ*2*,*

(2.29) (2.30)

30 2 Statistical Analysis

where the independent relation E[*(*xi − *μ)(*xj − *μ)*] = E[xi − *μ*]E[xj − *μ*] = 0 is used in the second term of the L.H.S.

***2.2.3 Maximum Likelihood Estimation***

Maximum likelihood estimation is a method to estimate parameters for a parent population using a sample set when a probability distribution for the parent population is known. The estimated parameter is called a maximum likelihood estimator. The likelihood means probability and is often written using *L*.

We show some examples of the maximum likelihood estimation as follows. The first example is the Poisson distribution. Suppose *n* samples were randomly sampled from a parent population with the Poisson distribution, {*x*1*, x*2*,* ··· *, xn*}. The likelihood *L* is written as

*L* = *p(x*1*, x*2*,* ··· *, xn)* = *p(x*1*)* · *p(x*2*)*··· *p(xn)*

= *e*−*μ μx*1

*x*1! · *e*−*μ μx*2

*x*2!··· *e*−*μ μxn*

(2.31)

= *e*−*nμ μ*Σ*ni*=1 *xi* ~~|~~*~~n~~i*=1 *xi*! *.*

*xn*!

The likelihood *L* is maximized when the condition,

(

*dμ* = *Lμ dL*

−*nμ* +Σ*n i*=1

*xi*

)

= 0*,* (2.32)

is satisfied. Therefore, we have *μ* = Σ*ni*=1 *xi/n*. The concept of the maximum likelihood estimation for the Poisson distribution is shown in the upper panel of Fig. 2.10.

The second example is the normal distribution. Suppose *n* samples were randomly sampled from a parent population with the normal distribution, {*x*1*, x*2*,* ··· *, xn*}. The likelihood *L* is written as

*L* = *p(x*1*, x*2*,* ··· *, xn)* = *p(x*1*)* · *p(x*2*)*··· *p(xn)*

= 1

~~√~~2*πσ*exp

[

− *(x*1 − *μ)*2 2*σ*2

]

· 1

~~√~~2*πσ*exp

[

− *(x*2 − *μ)*2 2*σ*2

]

·

··· 1

~~√~~2*πσ*exp

[

− *(xn* − *μ)*2 2*σ*2

]

(2.33)

=

( 1

~~√~~2*πσ*

)*n*exp [−Σ*ni*=1*(xi* − *μ)*2 2*σ*2

]

*.*

2.2 Statistical Estimation 31

**Fig. 2.10 Maximum**

**likelihood estimation:** The

upper panel is the Poisson

distribution, and the lower is

the normal distribution

The likelihood *L* is maximized when the condition,

*dμ* = −*nμ* − Σ*ni*=1 *xi*

*dL*

*σ*2 *L* = 0*,* (2.34)

is satisfied. Therefore, we have *μ* = Σ*ni*=1 *xi/n*. The concept of the maximum likelihood estimation for the normal distribution is shown in the lower panel of Fig. 2.10.

***2.2.4 Interval Estimation***

Next, we explain the interval estimation. The interval estimation estimates a range of a parameter *θ* with a given confidence level *γ* .

Estimate *Θ*1 and *Θ*2 for the sampled values in order that the probability where the unknown parameter *θ* is in the range *Θ*1 *<θ<Θ*2 is equal to the given confidence level *γ* . Here [*Θ*1*, Θ*2] is the confidence interval. The confidence level *γ* is defined

32 2 Statistical Analysis

as the probability in the confidence interval and is often quite close to 1, such as *γ* = 95%, or 98%.

An example of the interval estimation is explained. *n* samples were chosen from a parent population with *N (μ, σ*2*)*. The sample average *X*¯ is distributed according to *N (μ, σ*2*/n)*, where the mean and the variance of *X*¯ are calculated as *nμ/n* and

*(*√*~~nσ~~/n)*2, respectively. When the population variance *σ*2 is known, we estimate a confidence interval with a given confidence level *γ* . The sample average *X*¯ is standardized by

*Z* = *X*¯ − *μ*

*σ/*~~√~~*~~n~~ ,* (2.35)

and the variable *Z* is distributed with the standardized normal distribution *N (*0*,* 1*)*, where the confidence intervals are [−*Z*1*, Z*1] = [−1*.*282*,* 1*.*282] and [−2*.*576*,* 2*.*576] for the confidence levels *γ* = 98% and 99*.*9%, respectively. For a given confidence interval, we have

− *Z*1 *<Z<Z*1*.* (2.36)

Thus we have the confidence interval of the parent population with *N (μ, σ*2*)*, − *X*¯ − *σ*~~√~~*~~n~~Z*1 *<μ< X*¯ + *σ*~~√~~*~~n~~Z*1*.* (2.37)

***2.2.5 Hypothesis and Statistical Test***

A statistical hypothesis is a hypothesis about a distribution in a parent population. Statistical test means true–false decision of the hypotheses. Significance level, or risk rate, *α* is the criterion for the decision and is often chosen as 1%, or 5%.

We explain an example of the statistical test below. Assume that a distribution of the length of radish harvested in village A is known, as shown in Fig. 2.11. When

10 30 50 70 90 **Length (cm)**

10 30 50 70 90 **Length (cm)**

**Fig. 2.11 Statistical Test:** The left panel is with significance level *α* = 5% and the right *α* = 1%

2.2 Statistical Estimation 33

a radish is 71 cm, we will test a hypothesis “This radish is harvested in village A” with significance level *α* = 5%. The 71 cm radish is located in the tail region with *α* = 5%. Therefore, the hypothesis *H*0 is rejected. Then, we will test the hypothesis with significance level *α* = 1%. The 71 cm radish is not located in the tail region with *α* = 1%. Therefore, the hypothesis *H*0 is adapted.

Adoption of the hypothesis does not mean that the hypothesis is proven but means that it is not specific hypothesis is incorrect. That is double negation. Therefore, null hypothesis *H*0 should be designed to be rejected in the statistical test. If the null hypothesis is rejected, alternative hypothesis *H*1 is given to be adopted. On the other hand, if the null hypothesis is adapted, it comes to nothing. We have two kinds of errors. One is the type I error or false positive. This means rejection of the null hypothesis *H*0, despite the correctness of *H*0. The other is the type II error or false negative. This means the adoption of the null hypothesis *H*0, despite the correctness of *H*1.

The t-statistic was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Ireland [3]. “Student” was his pen name. Using sample average *X*¯ and sample variance *s*2, we define variable *t*:

*X*¯ = *x*1 +···+ *xn*

*n* (2.38)

*s*2 = 1 *n* − 1

Σ*n i*=1

(*Xi* − *X*¯)2 (2.39)

*t* = *X*¯ − *μ*

*s/*~~√~~*~~n~~ .* (2.40)

The variable *t* follows a probability density *p(t)*, called Student’s t-distribution: *p(t)* = *Γ ((m* + 1*)/*2*)*

~~√~~*~~mπ~~Γ (m/*2*)(*1 + *t*2*/m)*−*(m*+1*)/*2*,* (2.41)

where *m* = *n*‘1 is degrees of freedom and *Γ (*·*)* is the gamma distribution. The t distribution with a large degree of freedom *m* can be approximated by the standard normal distribution *N (*0*,* 1*)*.

The hypothesis “parameter *β* is equal to 0” is tested with a given significance level using the following *t*,

*t* = E[*β*] − 0

SE[*β*] *.* (2.42)

Here E[*β*] is estimation and SE[*β*] is standard error. If *t* = 1*.*8, the hypothesis is adopted at significance level 1%. We cannot say anything. If *t* = 2*.*9, the hypothesis is rejected. The estimated parameter is statistically significant at a significant level 1%. The probability of taking a value greater than t in a t-distribution *P r(>* |*t*|*)* is

34 2 Statistical Analysis -1.96 0 2.58

0 1.96 **t**

-2.58

**t**

**Fig. 2.12 Two-sided test with t-distribution:** The significance level *α* is allocated half to each side

called the p-value. The hypothesis is rejected at a significant level *α*, if *α >* p-value. Therefore, the p-value has to be small.

Consider the following simple example. You rolled a dice 400 times and obtained even numbers 224 times. Is this a good dice? *μ* = *np* = 400 × 0*.*5 = 200 *σ*2 = *np(*1−*p)* = 400×0*.*5×0*.*5 = 102 *t* = *(x*−*μ)/σ* = *(*224−200*)/*10 = 2*.*4. Assume that a distribution of the even number is t-distribution, as shown in Fig. 2.12. We test the null hypothesis *H*0 “Population mean is *t* = 2*.*4”.

In the one-sided or one-tailed test, alternative hypothesis *H*1 “Population mean *t* ≥ 2*.*4” or *H*1 “Population mean *t* ≤ 2*.*4”. In the two-sided or two-tailed test, the alternative hypothesis *H*1 “Population mean *t* /= 2*.*4”. Figure 2.12 depicts the significance level *α* of two-sided test; the left panel is for *α* = 5% and the right panel is for *α* = 1%.

*t* = 2*.*4 is located in the tail region with *α* = 5%. Therefore, the hypothesis *H*0 is rejected. Then, we will test the hypothesis with significance level *α* = 1%. *t* = 2*.*4 is not located in the tail region with *α* = 1%. Therefore, the hypothesis *H*0 is adapted.

**2.3 Regression Analysis**

***2.3.1 Single Regression Analysis***

An explained variable, or an objective variable *Yi(i* = 1*,*··· *,N)*, is modeled using a linear relation

*Yi* = *α* + *βXi* + *εi,* (2.43)

2.3 Regression Analysis 35

where *Xi(i* = 1*,* ··· *,N)* is an explanatory variable or an independent variable, and *εi(i* = 1*,* ··· *,N)* is residual or error. In this model, the squared sum of error, or the sum of squared error

*SSE* = Σ*N i*=1

*ε*2*i* = Σ*N i*=1

*(Yi* − *α* − *βXi)*2*,* (2.44)

is minimized to estimate regression coefficients, or regression parameters *α* and *β*. This is called the single regression analysis. The conditions to minimize the sum of square *SSE* are

*∂α* = −2Σ*N*

*∂SSE*

*i*=1

*(Yi* − *α* − *βXi)* = 0*,* (2.45)

*∂β* = −2Σ*N*

*∂SSE*

*i*=1

*(Yi* − *α* − *βXi)Xi* = 0*.* (2.46)

These conditions are rewritten as the normal equation,

*αN* + *β*Σ*N i*=1

*Xi* = Σ*N i*=1

*Yi,* (2.47)

*α*Σ*N i*=1

*Xi* + *β*Σ*N i*=1

*X*2*i* = Σ*N i*=1

*XiYi.* (2.48)

The regression parameters *α* and *β* are obtained by solving Eqs. (2.47) and (2.48),

*α* = 1*N*Σ*N i*=1

*Yi* − *β*1*N*Σ*N i*=1

*Xi* = *Y*¯ − *βX,* ¯ (2.49)

*β* =

Σ*Ni*=1*(Xi* − *X)(Y* ¯ *i* − *Y )*¯

~~Σ~~*Ni*=1*(Xi* − *X)*¯ 2 = *SXY*

*SSX.* (2.50)

This method of parameter estimation is often called the least square method.

36 2 Statistical Analysis ***2.3.2 Analysis of Variance***

The analysis of variance (ANOVA) is a methodology to decompose the sum of square *SSY* into the sum of the square of each factor and to compare the magnitude of each square of sum using the ANOVA table shown in Table 2.1 (Fig. 2.13). The residual *εi* is given by rewriting Eq. (2.43) using the predicted value *Y*ˆ*i*,

*εi* = *Yi* − *Y*ˆ*i* = *Yi* − *(α* + *βXi).* (2.51)

**Table 2.1** ANOVA table

| Squared Sum | Degree of Freedom |
| --- | --- |
| *SSR* | *φ*1 = 1 |
| *SSE* | *φ*2 = *N* − 2 |
| *SSY* | *Φ* = *N* − 1 |

Factor Mean Square

Regression *MSR* = *SSR* Error *MSE* = *SSE*

*~~N~~*~~−2~~

– –

**Fig. 2.13 ANOVA:**

Deviation (the upper panel)

equals the sum of regression

(the middle panel) and

residual (the lower panel). See Eq. (2.56)

Y

X

Y

X

Y

X

2.3 Regression Analysis 37

Residuals *εi(i* = 1*,* ··· *,N)* are not independent of each other because of the relation

Σ*N i*=1

*εi* = Σ*N i*=1

*(Yi* − *Y*ˆ*i)* = 0*.* (2.52)

Then we rewrite residual *εi* as a sum of deviation and regression as *Yi* − *Y*ˆ*i* = *(Yi* − *Y )*¯ − *(Y*ˆ*i* − *Y ).* ¯ (2.53)

By taking squared sum of both sides of Eq. (2.53), we have the relation

Σ*N i*=1

*(Yi* −*Y*ˆ*i)*2 = Σ*N i*=1

*(Yi* −*Y )*¯ 2 +Σ*N i*=1

*(Y*ˆ*i* −*Y )*¯ 2 −2Σ*N i*=1

2*(Yi* −*Y )(* ¯ *Y*ˆ*i* −*Y ).* ¯ (2.54)

Using the relation

Σ*N*

*i*=1

*(Yi* − *Y )(* ¯ *Y*ˆ*i* − *Y )*¯ = Σ*N i*=1

*(Y*ˆ*i* − *Y )*¯ 2 (2.55)

with Eq. (2.54), we obtain the decomposition for the sum of squares,

Σ*N i*=1

*(Yi* − *Y )*¯ 2 = Σ*N i*=1

*(Y*ˆ*i* − *Y )*¯ 2 +Σ*N i*=1

*(Yi* − *Y*ˆ*i)*2*,* (2.56)

where the left-hand side of the equation, the first term of the right-hand side of the equation, and the second term of the right-hand side of the equation are called the whole sum of squares *SSY* , the squared sum of regression *SSR*, and the squared sum of error *SSE*, respectively. *SSR* corresponds to the component explained by the regression, and *SSE* corresponds to the component that is not explained by the regression. We note that *SSR* is rewritten as

*SSR* = Σ*N i*=1

*(Y*ˆ*i* − *Y )*¯ 2 = *β*Σ*N i*=1

*(Xi* − *X)(Y* ¯ *i* − *Y ).* ¯ (2.57)

We have the relation of the sum of squares *SSY* . Therefore, the whole degree of freedom *Φ* is equal to *N* −1. On the other hand, the squared sum of regression *SSR* is determined by a single parameter *α*. Therefore, we have *φ*1 = 1. Consequently, we have *φ*2 = *Φ* − *φ*1 = *N* − 2. The mean square of regression *MSR* and mean square of error *MSE* shown in the most right column of Table 2.1 are defined as the squared sum of regression *SSR* divided by the degree of freedom *φ*1 and the squared sum of error *SSE* divided by the degree of freedom *φ*2.

38 2 Statistical Analysis ***2.3.3 Interval Estimation of Parameters***

From Eq. (2.57), we have

Σ*Ni*=1*(Xi* − *X)(Y* ¯ *i* − *Y )*¯

*β* =

*SSX*= Σ*N i*=1

*SSXYi* −Σ*N*

*Xi* − *X*¯

*i*=1

*SSXY*¯ = Σ*N*

*Xi* − *X*¯

*i*=1

*Xi* − *X*¯

*SSXYi.* (2.58)

As *Yi(i* = 1*,* ··· *,N)* are independent variables and *Xi*−*X*¯

numbers, the variance of *β* is calculated as follows:

*~~SS~~X* are regarded as constant

*V* [*β*] = Σ*N i*=1

(*Xi* − *X*¯ *SSX*

)2

*V* [*Yi*] = Σ*N i*=1

(*Xi* − *X*¯ *SSX*

)2

*σ*2 = *σ*2

~~Σ~~*Ni*=1*(Xi* − *X)*¯ 2*.*

(2.59)

Similarly, the variance of *α* is calculated from Eq. (2.49),

*V* [*α*] =

(*∂α ∂β*

)2

*V* [*β*] =*X*¯ 2

~~Σ~~*Ni*=1*(Xi* − *X)*¯ 2*σ*2*.* (2.60)

***2.3.4 Multiple Regression Analysis***

An explained variable or an objective variable *yi(i* = 1*,* ··· *,N)* is modeled using a linear relation,

*yi* = *θ*0 + *θ*1*xi*1 + *θ*2*xi*2 +···+ *θpxip* + *εi,* (2.61)

where *xip(i* = 1*,* ··· *,N)* are explanatory variables or independent variables. The number of model parameters is not equal to two but equal to *p* + 1. This model is called multiple regression analysis.

We introduce a vector of explained variable

*y* =

⎛ ⎜⎜⎜⎝

*y*1

*y*2*... yN*

⎞

⎟⎟⎟⎠ *,* (2.62)

2.3 Regression Analysis 39 a vector of residual

⎞

⎛

*ε*1

⎟⎟⎟⎠ *,* (2.63)

⎜⎜⎜⎝

*ε*2*...*

=

*εN*

a vector of model parameters

⎛

*θ*1

⎜⎜⎜⎝

*θ*2*...*

*θ* =

*θp*

and a matrix of explanatory variables

⎞

⎟⎟⎟⎠ *,* (2.64)

*X* =

⎛ ⎜⎜⎜⎝

1 *x*11 ··· *x*1*p* 1 *x*21 ··· *x*2*p ... ... ... ...* 1 *xN*1 ··· *xNp.*

⎞

⎟⎟⎟⎠ *.* (2.65)

Using these vectors and matrix, Eq. (2.61) is briefly rewritten as *y* = *Xθ* +  *.* (2.66)

We write parameter *θ* = ˆ*θ* when = **0**. In order to obtain a formal solution for parameters ˆ*θ*, we multiply *XT* from the left-hand side to Eq. (2.66),

*XT y* = *XT X*ˆ*θ,* (2.67)

where a matrix with suffix *T* indicates transposed matrix. Then a formal solution for parameters ˆ*θ* is obtained by multiplying *(XT X)*−1 from the left-hand side to Eq. (2.67),

ˆ*θ* = *(XT X)*−1*XT y,* (2.68)

where a matrix with suffix −1 indicates inverse matrix.

The squared sum of errors of parameter *θ* is given by

*SSE(θ)* = 2 = *(y* − *Xθ)T (y* − *Xθ)* (2.69)

40 2 Statistical Analysis

and rewritten as

*SSE(θ)* = *(y* − *Xθ)T (y* − *Xθ)*

= {*y* − *X*ˆ*θ* + *X(*ˆ*θ* − *θ)*}*T* {*y* − *X*ˆ*θ* + *X(*ˆ*θ* − *θ)*} = *(y* − *X*ˆ*θ)T (y* − *X*ˆ*θ)* + *(y* − *X*ˆ*θ)T X(*ˆ*θ* − *θ)* + {*X(*ˆ*θ* − *θ)*}*T (y* − *X*ˆ*θ)* + {*X(*ˆ*θ* − *θ)*}*T X(*ˆ*θ* − *θ).*

(2.70)

By substituting the relations that the second and third terms of the r.h.s. of Eq. (2.70) are equal to zero,

*(y* − *X*ˆ*θ)T X(*ˆ*θ* − *θ)* = {*XT (y* − *X*ˆ*θ)*}*T (*ˆ*θ* − *θ)* = 0*,* (2.71) and

{*X(*ˆ*θ* − *θ)*}*T (y* − *X*ˆ*θ)* = *(*ˆ*θ* − *θ)T XT (y* − *X*ˆ*θ)* = 0*.* (2.72) We have the relation

*SSE(θ)* = *(y* − *X*ˆ*θ)T (y* − *X*ˆ*θ)* + {*X(*ˆ*θ* − *θ)*}*T X(*ˆ*θ* − *θ).* (2.73) Then we obtain the squared sum of errors of the estimated parameter ˆ*θ SSE(*ˆ*θ)* = *(y* − *X*ˆ*θ)T (y* − *X*ˆ*θ),* (2.74)

by applying

{*X(*ˆ*θ* − *θ)*}*T X(*ˆ*θ* − *θ)* ≥ 0*,* (2.75)

for Eq. (2.73). Therefore, we have inequality:

*SSE(θ)* ≥ *SSE(*ˆ*θ).* (2.76)

We can note that the estimated parameters ˆ*θ* minimize the squared sum of errors *SSE(θ)*. This means that the estimated parameters ˆ*θ* are equal to the least square solution.

***2.3.5 Test of Parameter Estimation***

First we describe *Y*ˆ*i* − *Y*¯ or *Yi* − *Y*ˆ*i* as *xi*. The variables *x*1*, x*2*,* ··· *, xN* are independent of each other and obey to normal distribution *N (*0*,* 1*)*. Then *MSR* or

2.3 Regression Analysis 41

*MSE* is described by *z* = *x*21 + *x*22 +···+ *x*2*N* , and it is known that variable *z* obeys to *χ*2 distribution,

{ 1

2*~~n/~~*~~2~~*Γ (n/*2*)zn*−2

*Tn(z)* =

*~~n~~ e*− *z*~~2~~ *(z >* 0*)*

0 *(z* ≤ 0*),* (2.77)

where *n* and *Γ (x)* are the degree of freedom and the gamma function, respectively. *Γ (x)* is defined by

*Γ (x)* =

f ∞ 0

*yx*−1*e*−*y dy.* (2.78)

In the case that *x* is an integer or half-integer, it is written respectively as *Γ (n* + 1*)* = *n*!*,* (2.79)

(

*Γ*

*n* +12)= *(*2*n)*! 22*nn*!

√*~~π.~~* (2.80)

Next, we describe F-value *F*, which is defined by a ratio,

*F* = *MSR*

*MSE,* (2.81)

and discuss the distribution to which F-value *F* obeys.

When *z*1 and *z*2 are independent of each other and obey to *χ*2 distributions with degree of freedom *φ*1 and *φ*2, respectively, variable *r* = *φ*2*z*1

*~~φ~~*1*~~z~~*2, defined as the ratio of *z*1 *~~φ~~*1 and *z*2

*~~φ~~*2, obeys the distribution called F-distribution,

⎧⎪⎪⎨

*φ*

*φ*1~~2~~1 *nφ*2~~2~~ *rφ*1−2 ~~2~~

*fφ*1*,φ*2 *(r)* =

*B( φ*1~~2~~ *,φ*2~~2~~ *)(φ*1*r*+*φ*2*)~~φ~~*1~~+~~*~~φ~~*2 ~~2~~*(r >* 0*)* ⎪⎪⎩

0 *(r* ≤ 0*)*

*,* (2.82)

where *B(s, t)* is the beta function defined using *Γ (*·*)*,

*B(s, t)* = *Γ (s)Γ (t)*

*Γ (s* + *t) .* (2.83)

Then, we test the null hypothesis *H*0 that all parameters are equal to zero. If *F >t*, *H*0 is rejected at a given significant level, e.g., *α* = 0*.*5%. Thus, we say that all parameters are not equal to zero. On the other hand, if *F <t*, *H*0 is accepted at a given significant level, e.g., *α* = 0*.*5%. Thus, we do not say that all parameters are not equal to zero. F-value *F* is the most important statistical value to show the validity of the regression analysis (Fig. 2.14).

42 2 Statistical Analysis

**Fig. 2.14 F-distribution:**

*fφ*1*,φ*2 *(x)* with *φ*1 = 10 and

*φ*2 = 10. See Eq. (2.82)

0.6

density

0.4

0.2

0.0

F(10,10)

0123 4 x

Finally, we can write the coefficient of determination *R*2 from Eq. (2.56) as follows:

*R*2 = *SSR*

*SSY*= 1 − *SSE*

*SSY.* (2.84)

Here, the smaller error *SSE* means larger *R*2.

**2.4 Time Series Analysis**

***2.4.1 Origins of Temporal Variations***

Time series of economic indicator *x(t)*, such as GDP or index of industrial production, shows temporal variations. Some variation shows a cyclical change, which is often referred to as a business cycle. Business cycles are categorized into four types: (1) the Kitchin inventory cycle (3–5 years), (2) the Juglar fixed investment cycle (7–11 years), (3) the Kuznets infrastructural investment cycle (15– 25 years), and (4) the Kondratiev wave or long technological cycle (45–60 years).

The following variables are used to characterize the temporal variation of a time series. The index, change rate, and growth rate of *x(t)* are defined by

*x*'*(t)* = *x(t)*

*x(t*0*),* (2.85)

*s(t)* = *x(t)* − *x(t*0*)*

*x(t*0*) ,* (2.86)

*r(t)* = *x(t)* − *x(t* − 1*)*

*x(t* − 1*) ,* (2.87)

respectively. Here *t*0 is a base year.

Suppose the case where the time series *x(t)* is monthly. Consider decomposing a time series into multiple components. Most monthly time series include seasonal

2.4 Time Series Analysis 43

variations. Therefore, it is necessary to decompose the time series into its compo nents. The original time series *x(t)* is decomposed into trend variation *T (t)*, cyclical variation (cycle due to inventory adjustment) *C(t)*, seasonal variation *S(t)*, and random fluctuation *I (t)*. Two different models of decomposition are used; one is the additive model *x(t)* = *T (t)* + *C(t)* + *S(t)* + *I (t)* and the other is the multiplicative model *x(t)* = *T (t)*·*C(t)*·*S(t)*·*I (t)*. Note that the log transform of the multiplicative model is regarded as the log-additive model.

Moving average is a smoothing method by replacing data with an average value. The central, backward, and the forward-moving averages are defined by

*x(t)* ¯ = *x(t* − *k)* +···+ *x(t)* +···+ *x(t* + *k)*

2*k* + 1 *,* (2.88)

*x(t)* ¯ = *x(t* − 2*k)* +···+ *x(t)*

2*k* + 1 *,* (2.89)

*x(t)* ¯ = *x(t)* +···+ *x(t* + 2*k)*

2*k* + 1 *,* (2.90)

respectively.

A simple time series *x(t)* decomposition is possible using the additive model and the central moving average. Here, we assume that the ranges of cycles of variations are known as follows. The cycle of *T* is longer than 39 months. The cycle of *C* is between 39 months and 13 months. The cycle of *S* is between 12 months and 3 months. The cycle of *I* is shorter than 3 months. If we calculate moving average of *x(t)* with period *p* = 2*k* + 1, the variation with cycle shorter than *p* will be removed. Thus the moving averages with periods 39, 13, and 3 obtain time series *T (t)*, *a(t)* ≡ *T (t)* + *C(t)*, and *b(t)* ≡ *T (t)* + *C(t)* + *S(t)*, respectively. Then we obtain cyclical variation *C(t)* by subtracting *T (t)* from *a(t)*, and seasonal variation *S(t)* by subtracting *a(t)* from *b(t)*, and random fluctuation *I (t)* by subtracting *b(t)* from *x(t)*. However, we note that the moving average time series decomposition is explained only for educational purposes, but actual time series decomposition or seasonal adjustment is based on the X-12-ARIMA model.

***2.4.2 Stationary Process and Auto-correlation***

The log-growth rate *r(t)* is defined by

*r(t)* = log *x(t)* − log *x(t* − 1*)*

(

= log

1 + *x(t)* − *x(t* − 1*) x(t* − 1*)*

)

≈ *x(t)* − *x(t* − 1*)*

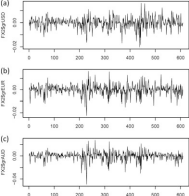
*x(t* − 1*) .* (2.91)

44 2 Statistical Analysis

**Fig. 2.15 Exchange rate:** The original time series for (**a**) JPY/USD, (**b**) JPY/EUR, (**c**) JPY/AUD are plotted

**Fig. 2.16 The log-growth rate of exchange rate:** The time series of the log-growth rate for (**a**) JPY/USD, (**b**) JPY/EUR, (**c**) JPY/AUD are plotted

(a) (b) (c)



The log-growth rate *r(t)* is a good approximation to the growth rate defined by Eq. (2.87). In order to calculate the growth rate, we take a time difference, which allows us to remove the linear trend. Similarly, if we calculate the second-order finite difference, we can remove the quadratic trend (Figs. 2.15 and 2.16).

Next, we consider the conditions of the stationary process. We divide the time series of interest into multiple time series of interval *T* . We calculate mean value

2.4 Time Series Analysis 45

*μx* , variance *σ*2*x* , and covariance Cov[*x(t), x(t* + *τ )*] for each time series of interval *T* , as follows:

*μx* = 1*T*Σ*T t*=1

*x(t)* = const (2.92)

*σ*2*x* = 1 *T* − 1

Σ*T t*=1

*(x(t)* − *μx )*2 = const (2.93)

Cov[*x(t), x(t* + *τ )*] =1 *T* − 1

Σ*T t*=1

*(x(t)* − *μx )(x(t* + *τ )* − *μx )* ≡ *γ (τ )* (2.94) *γ (*0*)* = *σ*2*x .* (2.95)

If mean value *μx* is the same independent of the time series, and variance *σ*2*x* is the same independent of the time series, covariance Cov[*x(t), x(t* + *τ )*] is the same independent of the time series and depends only on lag *τ* , time series *x(t)* is called stationary process.

Here, we defined the auto-correlation function *ρ(τ )* using *γ (τ )*: *γ (*0*)* = Cov[*x(t), x(t* + *τ )*]

*ρ(τ )* = *γ (τ )*

*σ*2*x.* (2.96)

Specifically, if the auto-correlation function *ρ(τ )* satisfies the following relation:

Σ∞

*τ*=−∞

*ρ(τ )* = const*,* (2.97)

time series is called the short-term memory process. Contrary, if *ρ(τ )* satisfies the following relation:

Σ∞

*τ*=−∞

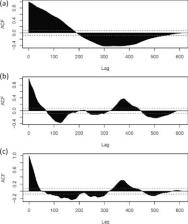
*ρ(τ )* = ∞*,* (2.98)

time series is called the long-term memory process.

The unit root test is another known method to check whether a process is stationary. Assume that the time series *xt* evolves according to the following equation:

*xt* = *axt*−1 + *ut .* (2.99)

46 2 Statistical Analysis

**Fig. 2.17 Auto-correlation** 

**function of exchange rate:**

(**a**) JPY/USD, (**b**) JPY/EUR,

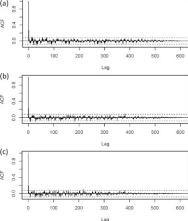
(**c**) JPY/AUD

Here *a* is a model parameter, and *ut* is the residual. In this case, it is known that the time series exhibits the following properties. When parameter *a <* 1, time series *xt* is stationary. When *a* = 1, time series *xt* is non-stationary. *a* = 1 is called a unit root. Our null hypothesis H0 is “a unit root is present”, which means that the time series is non-stationary. Using t value *t* = *(*E[*a*] − 1*)/(*SE[*a*]*)*, p-value *p* = *P r(>* |*t*|*)*, and significance level (risk rate) *α*, we perform the t test. If *p<α*, H0 is rejected. This means that “a unit root is not present”. Therefore, time series is a stationary process. On the other hand, if *p>α*, H0 is adopted. This means that “a unit root is present”. Therefore, time series is a non-stationary process.

Auto-correlation functions of exchange rate time series for (a) JPY/USD, (b) JPY/EUR, (c) JPY/AUD are shown in Fig. 2.17. Auto-correlation functions do not converge to a specific value; therefore, time series is the long-term memory process. The unit root test obtains p-values 0.5827, 0.1933, and 0.317 for (a) JPY/USD, (b) JPY/EUR, (c) JPY/AUD, respectively. We obtained a large p-value, so H0 was adopted. Thus, the time series is a non-stationary process. The non-stationary process is modeled by a random walk (standard Brownian motion with H = 1/2).

Auto-correlation functions of the log-growth rate of exchange rate time series for (a) JPY/USD, (b) JPY/EUR, (c) JPY/AUD are shown in Fig. 2.18. Auto-correlation functions converge to a specific value; therefore, time series is the short-term memory process. The unit root test obtains a p-value less than 2*.*2 × 10−16 for all cases (a) JPY/USD, (b) JPY/EUR, (c) JPY/AUD. We obtained a small p-value, so H0 was rejected. Thus, the time series is a stationary process.

2.4 Time Series Analysis 47

**Fig. 2.18 Auto-correlation** 

**function of the log-growth**

**rate of exchange rate:** (**a**)

JPY/USD, (**b**) JPY/EUR, (**c**)

JPY/AUD

***2.4.3 AR(p) Model***

The pth order auto-regression model of time series *xt* is given by

*xt* = Σ*p m*=1

*amxt*−*m* + *ut,* (2.100)

where *am* is the model parameter and *ut* is the residual. If *ut* satisfies the following equations:

E[*ut*] = 0 (2.101)

{*σ*2 *(t* = *s)*

E[*utus*] =

0 *(t* /= *s).* (2.102)

*ut* is called the white noise. *am* satisfies the characteristic equation:

1 = Σ*p m*=1

*amzm.* (2.103)

|*am*| *<* 1 for all *m* is the condition of stationary process.

48 2 Statistical Analysis We write the auto-covariance function for the AR model as follows:

*γ (s)* = *γs* = E[*xt*+*sxt*] *(s* ≥ 0*)*

= E[*(ut*+*s* + *aut*+*s*−1 +···+ *akut*+*s*−*k* +···*)*

× *(ut* + *aut*−1 +···+ *akut*−*k* +···*)*]

= *as*[1 + *a*2 +···]*σ*2*.* (2.104) Using *γ (s)*, we define the auto-correlation function for the AR model as follows: *ρs* := *ρ(s)* = *γ (s)*

*γ (*0*)* = *as (s* = 0*,* 1*,* 2*,* ···*).* (2.105)

For the AR(1) model, the auto-correlation function *ρ(s)* decreases at a geometric rate when |*a*| *<* 1 for the stationary process. Therefore, time series with the stationary process exhibits the properties of a short-term process.

***2.4.4 Model Selection Using Box–Jenkins Method***

Auto-correlation function *pk*, defined as follows, is the indicator to show the strength of correlation at different lag *k*:

*pk* =

⎛ ⎜⎜⎜⎝

1 *ρ*1 *ρ*2 *... ρk*−1 *ρ*1 1 *ρ*1 *... ρk*−2 *... ... ... ... ... ρk*−1 *ρk*−2 *ρk*−3 *...* 1

⎞

⎟⎟⎟⎠ *.* (2.106)

We remove the effect from lag *k* by replacing the *k*th column by (*ρ*1*, ρ*2*,* ··· *, ρk*):

*p*∗*k* =

⎛ ⎜⎜⎜⎝

1 *ρ*1 *ρ*2 *... ρ*1 *ρ*1 1 *ρ*1 *... ρ*2 *... ... ... ... ... ρk*−1 *ρk*−2 *ρk*−3 *... ρk*

⎞

⎟⎟⎟⎠ *.* (2.107)

Using *pk* and *p*∗*k* , we define the partial auto-correlation function: *Φk* = |*p*∗*k* |

|*pk*|*(k* = 1*,* 2*,* ···*).* (2.108)

2.4 Time Series Analysis 49

*Φk* is the indicator to show the strength of correlation after removing the effect from lag *k*.

If *Φk* is non-zero, the past point (lag *k*) affects the present point. This implies that if *Φk* is 0, then the *k*th parameter *ak* of the AR model is 0. In AR(p) model,

*Φk* = 0 for *k > p.* (2.109)

Box–Jenkins method is the method to determine order p and parameter *ai* of the AR (p) model using the partial auto-correlation function *φs*. For AR(1) model *xt* = *a*1*xt*−1 + *ut* , the auto-correlation function is *ρs* = *as*1. The partial auto-correlation functions for different k are obtained as follows:

*Φ*1 = *ρ*11 = *a*1 (2.110)

*Φ*2 =

||||1 *ρ*1 *ρ*1 *ρ*2 ~~|~~|||1 *ρ*1 *ρ*1 1

||||

~~|~~|||= *ρ*2 − *ρ*21

1 − *ρ*21= *a*21 − *a*21

1 − *a*21= 0 (2.111)

*Φ*3 = *Φ*4 =···= 0*.* (2.112)

We confirmed *Φk* = 0 for *k >* 1 in AR(1) model.

***2.4.5 MA(q) Model***

AR(1) model is rewritten using only the residual terms *ut*−*k*:

*xt* = *ut* + *a*1*xt*−1 (2.113)

= *ut* + *a*1 *(ut*−1 + *a*1*xt*−2*)* (2.114)

= *ut* + *a*1*ut*−1 +···+ *ak*1*ut*−*k* +··· *.* (2.115)

This is obtained by repeatedly substituting the time-shifted AR(1) model for *xt*−1 on the right-hand side of the AR(1) model. By generalizing this expression, we define the following model:

*xt* = *ut* +Σ*q k*=1

*bkut*−*k,* (2.116)

which is called the *q*th order moving average model, MA(q) model.

50 2 Statistical Analysis ***2.4.6 ARIMA(p,d,q) Model***

Auto-regressive moving average model ARMA(p,q) is defined by combining the AR (p) model and the MA (q) model, as follows:

*xt* = *ut* + Σ*p m*=1

*amxt*−*m* +Σ*q k*=1

*bkut*−*k.* (2.117)

Applying ARMA(p,q) to the *d*th difference time series, we obtain the auto regressive integrated moving average model ARIMA(p,d,q). This is the general form of the primary time series model. The Box–Jenkins method cannot be used to determine the order of this model. In the following, we explain how to determine the order of the model.

***2.4.7 Model Selection Using Information Criterion*** For a model of time series *xt* ,

*xt* = *εt* + Σ*p*

*m*=1

we define the likelihood *L(θ )* as follows:

*αmxt*−*m* + *β,* (2.118)

*L(θ )* = *Πtp(xt, θ ),* (2.119)

where *θ* is the model parameter:

*θ* = {*αm, β*} *(m* = 1*,* 2*,* ··· *, p).* (2.120)

Specifically, the likelihood is written down as

*L(θ )* =

( 1

~~√~~2*πσ*

)*T*

exp

[

−

Σ*Tt*=1*(xt* − Σ*pm*=1 *αmxt*−*m* + *β)*2 2*σ*2

]

*.* (2.121)

Furthermore, we define the log-likelihood *l(θ )* by taking the logarithm of *L(θ )*:

*l(θ )* := log*L(θ )* = *T* log

( 1

~~√~~2*πσ*

)

−

Σ*Tt*=1*(xt* − Σ*pm*=1 *αmxt*−*m* + *β)*2 2*σ*2 *.*

(2.122)

2.4 Time Series Analysis 51

**Fig. 2.19 Akaike**

**information criterion:** The

log-likelihood dominates for

small *m* and the number of

model parameters *m*

dominates for large *m*

|  |
| --- |

Model order *p* is selected using the Akaike information criterion: AIC*(p)* = −2 log*L(θ )* + 2*p* = −2*l(θ )* + 2*p,* (2.123)

where *p* is the number of model parameters. For a small number of *p*, AIC*(p)* is dominated by the log-likelihood:

AIC*(p)* ≈ −2*l.* (2.124)

On the other hand, for a large number of *p*, AIC*(p)* is dominated by the number of model parameters *p*:

AIC*(p)* ≈ 2*p.* (2.125)

Both the model order and model parameters are determined to minimize the AIC. In general, the larger the model order, the larger the number of parameters, and thus the larger the likelihood. Minimizing the AIC means trying to obtain as large a likelihood as possible with as few parameters as possible. This is called the principle of parsimony (Fig. 2.19).

For a given number of *p*, the model parameters are estimated using the maximum likelihood estimation:

*∂l*

*∂θm*= 0 *(m* = 1*,* 2*,* ··· *, p).* (2.126)

Specifically, the likelihood equations are written down as

*∂l*

*∂σ* = 0*,* (2.127)

*∂l*

*∂αi*:= *S(αi)* = 0*, (i* = 1*,* 2*,* ··· *, p)* (2.128)

*∂l*

*∂β*:= *S(β)* = 0*.* (2.129)

52 2 Statistical Analysis

We use the Newton–Raphson algorithm to solve the likelihood equations in an iterative way:

*i* − *S(α(k)*

*α(k*+1*)*

*i* = *α(k)*

*i )*

*i ), (i* = 1*,* 2*,* ··· *, p)* (2.130) *H (α(k)*

*β(k*+1*)* = *β(k)* − *S(β(k))*

*H (β(k)),* (2.131)

where function *H (*·*)* is defined as follows:

*H (αi)* := −*∂S(αi)*

*∂αi, (i* = 1*,* 2*,* ··· *, p)* (2.132)

*H (β)* := −*∂S(β)*

*∂β .* (2.133)

**2.5 Summary**

This chapter explained the basics of statistics needed to deal with the nature of renewable energy with fluctuating output. After reviewing basic statistical concepts and statistical tests, regression analysis and time series analysis were explained. In regression analysis, the data reproducibility for the model as a whole and the statistical significance of each explanatory variable are important. In time series analysis, the fundamental analysis was the subject of stationary processes, and the importance of parameter estimation by the maximum likelihood method and model selection by the amount of information was explained.

**References**

1. David A. Freedman, “Statistical Models: Theory and Practice”, Cambridge University Press; 2nd edition, June, 2009, ISBN 978-0521743853

2. Charles Wheelan, “Naked Statistics: Stripping the Dread from the Data”, W W Norton & Co Inc; Reprint edition, January, 2014, ISBN 978-0393347777

3. J. F. Box, “Guinness, Gosset, Fisher, and Small Samples”, Statistical Science. 2 (1), pp.45–52, (1987). https://doi.org/10.1214/ss/1177013437.

**Chapter 3** 

**Fluctuation and Correlation**

**of Renewable Energy**

**3.1 Principal Component Analysis**

***3.1.1 Basic Concept of PCA***

We are interested in multi-dimensional data for various practical purposes. However, it is often seen that the multi-dimensional data are hard to understand intuitively due to correlations between some of the dimensions of the data. The purpose of the principal component analysis (PCA) is information contraction by converting a set of the original correlated data into a small set of uncorrelated variables. The uncorrelated variables are called the principal components. Suppose the data you are interested in are multi-dimensional, and the dimension of the data is higher than the data’s number. In that case, multi-regression analysis cannot be applied.

A multi-dimensional data *xi(i* = 1*,* ··· *, n)* is given by a column vector,

⎞

⎟⎟⎟⎠ *,* (3.1)

or a row vector,

*xi* =

⎛ ⎜⎜⎜⎝

*xi*1 *xi*2*... xip*

*xTi* = *(xi*1*, xi*2*,* ··· *, xip),* (3.2)

© Springer Nature Singapore Pte Ltd. 2024

53

Y. Ikeda, *Data Science of Renewable Energy Integration*,

Evolutionary Economics and Social Complexity Science 30,

https://doi.org/10.1007/978-981-99-8779-5\_3

54 3 Fluctuation and Correlation of Renewable Energy

where *p* is the dimension of data. Similarly, a *p*-dimensional weight vector *w* is given by a column vector,

⎞

⎛

*w*1

⎟⎟⎟⎠ *,* (3.3)

⎜⎜⎜⎝

*w*2*...*

*w* =

*wp*

or a row vector,

*wT* = *(w*1*, w*2*,* ··· *, wp).* (3.4)

Here, we define the principal component by

*fi* = Σ*p j*=1

*wj xij* = *wT xi* = *xTi w.* (3.5)

In the PCA, we determine the weight *w* to maximize the variance of the prinipal component *V* [*f* ] with the constraint |*w*|2 = *wT w* = 1.

*E*[*f* ] = *f*¯ = 1*n*Σ*ni*=1*fi* = 1*n*Σ*ni*=1Σ*p j*=1

*wj xij* = Σ*p j*=1

*wj x*¯*j* = *wT x*¯ (3.6)

*V* [*f* ] =1 *n* − 1

Σ*n*

*i*=1 Σ*n*

*(fi* − *f )*¯ 2 = 1 *n* − 1

Σ*n i*=1

{Σ*p j*=1

Σ*n*

*wj (xij* − ¯*xj )*}2

(3.7)

= 1

*n* − 1

*i*=1

{*wT (xi* − *x*¯*)*}2 = 1 *n* − 1

|*wT (xi* − *x*¯*)*|2*. i*=1

We obtain the following relation using the Pythagorean theorem (see Fig. 3.1),

|*xi* − *x*¯|2 = |*xi* − *x*ˆ*i*|2 + |*x*ˆ*i* − *x*¯|2

= |*xi* − *x*ˆ*i*|2 + |*wT (xi* − *x*¯*)*|2*.*(3.8)

The variance of the principal component *V* [*f* ] is rewritten using the above relation,

*V* [*f* ] =1 *n* − 1

Σ*n i*=1

|*xi* − *x*¯|2 − |*xi* − *x*ˆ*i*|2*.* (3.9)

Maximizing variance *V* [*f* ] means minimizing the sum of foot of perpendicular |*xi* −*x*ˆ*i*|2 because each term in the r.h.s. of Eq. (3.9) is positive. Minimizing the sum

3.1 Principal Component Analysis 55

**Fig. 3.1 The variance of the**

**principal component** *V* [*f* ]

is rewritten using the relation

obtained by the Pythagorean

theorem

=

**Fig. 3.2 The principal**

**component analysis**

minimizes the sum of foot of

perpendicular

**Fig. 3.3 The least square**

**method** minimizes the sum

of the square root of residuals

of the foot of perpendicular |*xi* − *x*ˆ*i*|2 is depicted in Fig. 3.2 for the case of *n* = 6 and *p* = 2. It is noted that maximizing the variance of the principal component *V* [*f* ] is different from minimizing the sum of the square root of residuals in the least square method (see Fig. 3.3).

56 3 Fluctuation and Correlation of Renewable Energy ***3.1.2 Maximization of*** *V* [*f* ]

For two-dimensional data, we write the expectation value and variance of the principal component *f*

*E*[*f* ] = *f*¯ = *w*1*x*¯1 + *w*2*x*¯2*,* (3.10)

*V* [*f* ] =1 *n* − 1

Σ*n i* = 1

*(fi* − *f )*¯ 2 = 1 *n* − 1

Σ*n i* = 1

{*(w*1*xi*1 + *w*2*xi*2*)* − *(w*1*x*¯1 + *w*2*x*¯2*)*}2

= 1

*n* − 1

= 1

*n* − 1

Σ*n*

*i* = 1 Σ*n*

*i* = 1

{*(w*1*(xi*1 − ¯*x*1*)* + *w*2*(xi*2 − ¯*x*2*)*}2

{*(w*21*(xi*1 − ¯*x*1*)*2 + 2*w*1*w*2*(xi*1 − ¯*x*1*)(xi*2 − ¯*x*2*)* + *w*22*(xi*2 − ¯*x*2*)*2}

= *w*21*S*11 + 2*w*1*w*2*S*12 + *w*22*S*22

= *(w*1*, w*2*)*

(*S*11 *S*12 *S*12 *S*22

) (*w*1 *w*2

)

= *wT Sw,*

(3.11)

where *S*11 = 1 *~~n~~*~~−1~~

Σ*ni*=1*(xi*1− ¯*x*1*)*2*, S*12 = 1 *~~n~~*~~−1~~

Σ*ni*=1*(xi*1− ¯*x*1*)(xi*2− ¯*x*2*), andS*22 =

1

*~~n~~*~~−1~~

Σ*ni*=1*(xi*2 − ¯*x*2*)*2*.*

We explain the method of Lagrange multiplier. We find *x*0 and *y*0 to maximize *f (x, y)* = 1 − *x*2 − *y*2 with constraint *g(x, y)* = *x* + *y* − 1 = 0 (see Fig. 3.4). By introducing Lagrange multiplier *λ*, we define function *Q(x, y)* by

*Q(x, y, λ)* = *f (x, y)* − *λg(x, y).* (3.12)

Point *(x, y)* = *(x*0*, y*0*)* at which function *f (x, y)* has maximum value satisfies equations,

*Q*

*x* = −2*x* − *λ* = 0*,* (3.13)

*Q*

*y* = −2*y* − *λ* = 0*, and* (3.14)

*Q*

*λ* = *x* + *y* − 1 = 0*.* (3.15)

3.1 Principal Component Analysis 57

**Fig. 3.4 The method of**

**Lagrange multiplier**

maximizes

*f (x, y)* = 1 − *x*2 − *y*2 with

constraint

*g(x, y)* = *x* + *y* − 1 = 0

By solving the above equations, we obtain solution *(x*0*, y*0*)* = *(* 1~~2~~ *,* 1~~2~~ *)*. Multiplier *λ* = −1 is equal to ratio of *f~~x~~* to *g~~x~~* , or to ratio of *f~~y~~*to *g~~y~~* .

We maximize *V* [*f* ] with constraint *w*21 + *w*22 = 1 using the method of Lagrange multiplier. Function *Q(x, y)* and conditions satisfied by solutions are

*Q(w*1*, w*2*, λ)* = *V* [*f* ] − *λ(w*21 + *w*22 − 1*).* (3.16)

*Q*

*w*1= 2*w*1*S*11 + 2*w*2*S*12 − 2*λw*1 = 0*,* (3.17)

*Q*

*w*2= 2*w*2*S*22 + 2*w*1*S*12 − 2*λw*2 = 0*, and* (3.18)

*Q*

*λ* = *w*21 + *w*22 − 1 = 0*.* (3.19)

Equations (3.17) and (3.18) are rewritten as a matrix form by

(*S*11 *S*12 *S*12 *S*22

) (*w*1 *w*2

)

= *λ*

(*w*1 *w*2

)

*, or* (3.20)

*Sw* = *λw, or* (3.21)

*(S* − *λI )w* = 0*,* (3.22)

which is referred as the eigenvalue problem.

58 3 Fluctuation and Correlation of Renewable Energy

We explain the eigenvalue problem. We can analyze the eigenvalue of the correlation matrix. Multiplying a vector by a matrix generally produces a vector with a different magnitude and direction. A vector with only a different magnitude without any change in direction is called an eigenvector. A multiplying factor that can only change the magnitude is called the eigenvalue of the matrix.

(*x*'1 *x*2

)

=

(cos *θ* − sin *θ* sin *θ* cos *θ*

) (*x*1 *x*2

)

*.* (3.23)

PCA is formulated as the eigenvalue problem. It focuses on the covariance or correlation between variables in terms of N-dimensional data and represents the overall characteristics of several indicators. In other words, PCA is another typical multivariate analysis method for information contraction.

From Eq. (3.22), we have the characteristic equation

|*S* − *λI* | =

||||*S*11 − *λ S*12

||||

*S*12 *S*22 − *λ*

(3.24)

= 0*, or*

*λ*2 − *(S*11 + *S*22*)λ* + *S*11*S*22 − *S*212 = 0*.* (3.25)

We note that two roots *λ*1 and *λ*2 are real number because

*D* = *(S*11 + *S*22*)*2 − 4*(S*11*S*22 − *S*212*)* = *(S*11 − *S*22*)*2 + 4*S*212 ≥ 0*.* (3.26) Thus, we have two principal components. The first component is *f* 1 = *wT*1 *x* (3.27)

and the second component is

*f* 2 = *wT*2 *x.* (3.28)

Variance of the principal components is calculated as follows:

*V* [*f*1] = *wT*1 *Sw*1 = *λ*1*wT*1 *w*1 = *λ*1*,* (3.29)

*V* [*f*2] = *wT*2 *Sw*2 = *λ*2*wT*2 *w*2 = *λ*2 (3.30)

and are equal to the eigenvalues (see Fig. 3.5).

3.2 Solar Photovoltaic Power 59

**Fig. 3.5 Eigenvalue:**

Variances of the principal

components are equal to the

eigenvalues

***3.1.3 Important Indices***

The following are the important indices of PCA:

Contribution ratio

*rj* = *λj*

*tr(S)* (3.31)

Cumulative contribution ratio

*Cj* = Σ*j j* '=1

*r*'*j* = *λ*1 + *λ*2 +···+ *λj*

*tr(S)* (3.32)

The jth principal component score of the ith element

*fij* = *wj*1*xi*1 + *wj*2*xi*2 +···+ *wjpxip* = Σ*p k*=1

**3.2 Solar Photovoltaic Power**

*wjkxik* (3.33)

***3.2.1 Cross-correlation of PV Output Fluctuation***

Restructuring of the electric utility industry and large-scale grid integration of PV systems were intensively discussed after the East Japan Earthquake of 2011. The former includes separating electrical power generation from power distribution and transmission, establishing the retail power market, and revitalizing the wholesale

60 3 Fluctuation and Correlation of Renewable Energy

power markets. Although the institutional design of the power markets is still an open question in Japan, the market has to be designed to have an optimal operation schedule, which is obtained using a unit commitment calculation through competitions between generation companies.

The large-scale grid integration of PV systems brings another kind of problem, namely, the PV output fluctuation, into the power system operation. The planned installation capacity of PV systems will be 100 GW in 2030 in the whole of Japan [1], and about one-third of the capacity will be in the Tokyo area. A significant fraction of the PV system will be installed on the rooftop of the consumer’s residential houses and office buildings, which are widely distributed in Tokyo. Therefore, the forecast of PV output with high spatial resolution is a crucial problem to be considered, and the cross-correlations of the PV outputs will be essential quantities to estimate the forecast error of PV output.

Concerning the above discussion, the concept of “local production for local consumption of renewable energy” has been proposed in Japan. Because electric power is in great demand in the Tokyo area, the area price could be high enough to be close to the feed-in tariff price for PV power. For this reason, the concept of “local production for local consumption of renewable energy” of PV power is considered to be economically feasible [2]. This concept is also advantageous because of the mitigation of transmission loss. However, it is to be noted that this concept needs careful consideration for PV and wind power because of the inherent nature of output fluctuation, even though it is suitable for geothermal and biomass energies [3].

In this chapter, we analyzed the cross-correlation of PV output fluctuation for the actual PV output time series data [4] in both the Tokyo area and the whole of Japan using the principal component analysis with the random matrix theory. Based on the obtained cross-correlation coefficients, the forecast error for PV output was estimated for some extreme cases. Then, the operation schedule of thermal plants was calculated to integrate PV output using our unit commitment model [5, 6] with the estimated forecast error. The system-balancing cost of the PV system was also estimated with or without demand response. Finally, the validity of the concept of “local production for local consumption of renewable energy” and alternative policy implications were also discussed.

***3.2.2 System-Wide Output Fluctuation***

The forecast of system-wide PV output is decomposed as

*t* ≡ *X(t)* = Σ*N*

*pv(f )*

*i*=1

*xi(t)* = Σ*N i*=1

*ciyi(t),* (3.34)

3.2 Solar Photovoltaic Power 61

where *yi(t)* = *xi(t)/ci* and *ci* are the forecast of PV output per installed capacity (load factor) and the installed capacity in the ith site, respectively. Our unit commitment model [5, 6] requires the PV output forecast time series and the forecast error to estimate the optimal operation schedule considering the PV output fluctuation. Suppose both the accuracy and spatial resolution of the PV forecasting are high. In that case, the forecasted time series is similar to a moving average of actual PV output for each PV site. Consequently, the cross-correlation of residual time series, equal to subtracting the actual output from forecast output at each time point, is expected to be a white noise. Thus, the forecast error of system-wide PV output *σX* is

*σ*2*p* = Σ*N i*=1

(*∂X ∂yi*

)2*σ*2*i* = Σ*N i*=1

*c*2*i σ*2*i ,* (3.35)

where *σi* is the forecast error of PV output per installed capacity in the ith site. On the other hand, if the spatial resolution of the forecast is low and, for example, we have just a few forecasted sites in the Tokyo area, the residual time series includes the cross-correlation between the various PV sites located in different places. In this case, we have a larger forecasting error due to the cross-correlations. The forecast error of system-wide PV output *σX* is written as

*σ*2*p* = Σ*N i*=1

(*∂X ∂yi*

)2*σ*2*i* + 2Σ*N i*=2

Σ *j<i*

(*∂X ∂yi*

)( *∂X ∂yj*

)

*σij*

(3.36)

= Σ*N i*=1

*c*2*i σ*2*i* + 2Σ*N i*=2

Σ *j<i*

*cicj σij*

*σij* = *σiσjρij* (3.37)

by including covariance among different sites *σij* . Here, *ρij* is the cross-correlation coefficient among different sites. Generally, it is expected that the number of forecasted sites is smaller than that of the installed sites *N*. For instance, we cannot forecast PV output for each rooftop PV of all the residential houses and office buildings with high accuracy in the Tokyo area due to both technological and economic reasons. Therefore, it is required to consider the cross-correlation *σij* to estimate the forecast error of system-wide PV output *σX*.

***3.2.3 Random Matrix Theory***

We analyzed the de-trended PV output *zi(t)* obtained by filtering the actual PV output time series per installed capacity *yi(t)* using the Fourier series expansion.

62 3 Fluctuation and Correlation of Renewable Energy

In general, correlation coefficients are expected to be associated with random noise for a fluctuating time series such as PV output. The correlation coefficient between points *i* and *j* is calculated by

*Cij* = <*(zi(t)* − <*zi*>*)(zj (t)* − <*zj* >*)*>

~~/~~

*(*<*z*2*i* >−<*zi*>2*)(*<*z*2*j* >−<*zj* >2*),* (3.38)

where *zi(t)* is the de-trended PV output at the site *i(*= 1*,* ··· *,N)* and time *t (*= 1*,* ··· *, L)* and <·> indicates the time average for the time series. Now we consider the eigenvalue problem

*C*|*α*> = *λα*|*α*> (3.39)

for the correlation matrix *C*. *λα* and |*α*> are the eigenvalues and the corresponding eigenvector, respectively. We assume that the eigenvalues are arranged in decreasing order *(α* = 0*,* ··· *, N* − 1*)*. Once the eigenvalues are calculated using Eqs. (3.38) and (3.39), the distribution of eigenvalue *ρ(λ)E* is obtained.

According to the random matrix theory [7–10], distribution of the eigenvalue for the matrix 1*~~T~~ H HT* where all elements of the matrix *H* are given as a random number *N (*0*, σ*2*)* is given by

*ρ(λ)T* = *Q*2*π*√*~~(λ~~*max ~~−~~ *~~λ)(λ~~* ~~−~~ *~~λ~~*min*~~)~~*

*λ ,* (3.40)

where

*Q* = *LN ,* (3.41)

*λ* = [*λ*min*, λ*max]*,* (3.42)

*λ*min = *(*1 − 1~~√~~*~~Q~~)*2*,* and (3.43)

*λ*max = *(*1 +1~~√~~*~~Q~~)*2*.* (3.44)

Equation (3.40) is exact at the limit *N,L* → ∞. For a randomly fluctuating time series such as PV output, the distribution *ρ(λ)E* obtained by data analysis is expected to agree to the distribution *ρ(λ)T* calculated using Eqs. (3.40) to (3.44) for *λ* ≤ *λ*max. Therefore, only the small number of eigenvalues for *λ>λ*max has the information of genuine correlation.

In order to extract the genuine correlation, we rewrite the correlation matrix *C* using eigenvalue *λα* and the corresponding eigenvector |*α*> [11]. First, we define

3.2 Solar Photovoltaic Power 63

the complex conjugate vector of the eigenvector |*α*> by

<*α*|=|*α*∗>*t.* (3.45)

For the real symmetric matrix, such as the correlation matrix *C*, all elements of the eigenvector |*α*> are real. Thus, the complex conjugate denotes the transpose *t*. Then, the correlation matrix *C* is rewritten as

*C* = *N*Σ−1 *α*=0

*λα*|*α*><*α*| (3.46)

by multiplying Eq. (3.39) with the transposed vector <*α*| from the left-hand side and taking summation over *α*. Here, the property of the projection operator |*α*><*α*|

*N*Σ−1 *α*=0

|*α*><*α*| = 1 (3.47)

was used. As a result, the correlation matrix *C* of Eq. (3.46) is divided in the following components:

*C* = *Ct* + *Cr* = Σ*Nt α*=0

*λα*|*α*><*α*| +*N*Σ−1 *α*=*Nt*+1

*λα*|*α*><*α*|*.* (3.48)

The first term *Ct* corresponds to the genuine correlation component (*λ>λ*max). The second term *Cr* corresponds to the random component (*λ* ≤ *λ*max). The term *λ*0|0><0| is interpreted as the change as a whole system, such as the weather change.

We introduce the vector |*z(t)*>, which consists of the time series of PV output *zi(t)(i* = 1*,* ··· *,N)*. Then the vector |*z(t)*> is expanded on the basis of the eigenvectors |*α*> [11]:

|*z(t)*> =*N*Σ−1 *α*=0

*aα(t)*|*α*>*.* (3.49)

The expansion coefficient *aα(t)* is obtained using the orthogonality of the eigenvec tors:

*aα(t)* = <*α*|*z(t)*>*.* (3.50)

The time series corresponding to the genuine correlation *Ct* is extracted by truncating the summation up to *Nt* in Eq. (3.49):

|*z(t)*> = Σ*Nt α*=0

*aα(t)*|*α*>*.* (3.51)

64 3 Fluctuation and Correlation of Renewable Energy ***3.2.4 Data Analysis***

The genuine components of cross-correlation of the de-trended PV output per installed capacity were studied using the random matrix theory. The analyzed data are the output time series acquired every hour for each prefecture [4]. Before analyzing the data, two preprocessing were made. First, the data during nighttime were removed. Then, the trend was removed from the time series by filtering out the components with a period longer than six hours using the Fourier series expansion. Therefore, only the short-term fluctuation is the component that remained in the time series. The auto-correlation function and fluctuation distribution for Tokyo in May are shown in Fig. 3.6. The memory in the auto-correlation function gets lost within a few hours. This means that the trend component is well removed. The kurtosis of the fluctuation distribution is 5.0849, which is significantly larger than the value expected for the normal distribution, i.e., 3.0. This means the actual fluctuation distribution has a longer tail than the normal distribution. The two different types of functional forms of fluctuation distribution are shown in Fig. 3.7. If the fluctuation is distributed according to the normal distribution, the probability density function is

[

*p(x)* = 1

~~√~~2*πσ*2 exp

− *(x* − *μ)*2 2*σ*2

]

*,* (3.52)

and the cumulative distribution function is written using the error function erf[·] as

*φ(x)* = 12(1 + erf[*x* − *μ*

~~√~~2*σ*2

1.0

70

60

50

0.5

])

*,* (3.53)

ACF

0.0−0.5

Density

40 30 20 100

0 20 40 60 80 100 Lag (month)

−0.06 −0.04 −0.02 0.00 0.02 0.04 0.06 z

**Fig. 3.6 Auto-correlation function and fluctuation distribution** for Tokyo in May are shown in the left and right panels, respectively. The red curve in the right panel is the normal distribution. The kurtosis of the actual distribution is positive

3.2 Solar Photovoltaic Power 65

**Fig. 3.7 Functional Form**

**of Fluctuation Distribution:**

The positive kurtosis is

reproduced for the Laplace

distribution

where *μ* and *σ* are the mean and standard deviation, respectively. However, if the probability density function *p(x)* is a Laplace distribution

*p(x)* = 12*b* exp [− |*x* − *μ*|

*b*

then the cumulative distribution function *φ(x)* is

]

*,* (3.54)

*φ(x)* = 12(1 + sgn*(x* − *μ)*(1 − exp [− |*x* − *μ*| *b*

]))

*.* (3.55)

Here, a standard deviation is given by *σ* = √2*b* and sgn*(x* − *μ)* = +*(x* ≥ *μ),* −*(x < μ)*. The functional forms for these distributions are depicted for *μ* = 0 and *σ* = 1 in Fig. 3.7. It is to be noted here that the Laplace distribution shows a distribution tail longer than the normal distribution.

Eigenvalue distribution for the Tokyo area and Japan in May is shown in Fig. 3.8. For the Tokyo area, we calculate *λ*max = 1*.*35 using Eq. (3.44) with *N* = 9 and *L* = 420. The upper panel of Fig. 3.8 depicts that only the largest eigenvalue is larger than *λ*max. On the other hand, for the whole of Japan, we calculate *λ*max = 1*.*88 with *N* = 47 and *L* = 420. The lower panel of Fig. 3.8 depicts that the five largest eigenvalues are larger than *λ*max.

We show the distribution of genuine correlation coefficients calculated for the de trended PV output time series in both the Tokyo area and the whole of Japan. The cross-correlation coefficients for the Tokyo area are shown in Fig. 3.9. Panels (a) and (c) are genuine correlation *Ct*, and panels (b) and (d) are the random components

66 3 Fluctuation and Correlation of Renewable Energy

1.0

0.8

0.6

Frequency

0.4

0.2

0.0

151050 Eigen value

2.0

1.5

Frequency

1.0

0.5

0.0

151050 Eigen value

**Fig. 3.8 Eigenvalue distribution** for the Tokyo area (left) and the entire Japan (right) in May

(a) Genuine correlation in January

8

(c) Genuine correlation in July

8

Frequency

6

4

2

0

−0.5 0.0 0.5 1.0 Correlation coefficient

Frequency

6

4

2

0

−0.5 0.0 0.5 1.0 Correlation coefficient

(b) Random component in January

12

(d) Random component in July 10

Frequency

8

6

4

2

0

−0.5 0.0 0.5 1.0 Correlation coefficient

8

Frequency

6

4

2

0

−0.5 0.0 0.5 1.0 Correlation coefficient

**Fig. 3.9 Cross-correlation coefficients for the Tokyo area in January and July:** Panels (**a**) and (**c**) are correlations, and panels (**b**) and (**d**) are random noise

(a) The 1st eigen vector in January

(b) The 1st eigen vector in July

Eigen vector

0.0  −0.2−0.4

Eigen vector

−0.1 −0.3

**Fig. 3.10 The 1st eigenvector for the Tokyo area:** panel (**a**) is January, and panel (**b**) is July

*Cr*. The genuine correlation *Ct* was calculated using only the largest eigenvalue and the corresponding eigenvector. Figure 3.9 depicts that the genuine correlation *Ct* has a positive correlation, and on the other hand, the random components *Cr* distribute around 0*.*0. The 1st eigenvector for the Tokyo area is shown in Fig. 3.10. The nine components correspond to eight prefectures, and Tokyo was included in the Tokyo

3.2 Solar Photovoltaic Power 67

(a) Genuine correlation in January  100

(c) Genuine correlation in July  100

Frequency

60

0 20

−0.5 0.0 0.5 1.0 Correlation coefficient

Frequency

60

20

0

−0.5 0.0 0.5 1.0 Correlation coefficient

(b) Random component in January  200

(d) Random component in July

Frequency

100

50

0

−0.5 0.0 0.5 1.0 Correlation coefficient

150

Frequency

50

0

−0.5 0.0 0.5 1.0 Correlation coefficient

**Fig. 3.11 Cross-correlation coefficients for the whole of Japan in January and July:** Panels (**a**) and (**c**) are correlations, and panels (**b**) and (**d**) are random noise

(a) The 1st eigen vector in January

(d) The 1st eigen vector in July

Eigen vector

−0.05  −0.15−0.25

Eigen vector

0.00  −0.10−0.20

(b) The 2nd eigen vector in January

(e) The 2nd eigen vector in July

Eigen vector

0.1 −0.1−0.3

Eigen vector

0.0 0.1 −0.2

(c) The 3rd eigen vector in January

(f) The 3rd eigen vector in July

Eigen vector

0.2 0.0 −0.2

Eigen vector

0.3  0.1−0.1−0.3

**Fig. 3.12 The 1st to 3rd eigenvector for the whole of Japan:** Panels (**a**), (**b**), and (**c**) are in January, and Panels (**d**), (**e**), and (**f**) are in July

area. It was noted that all vector components had the same sign. This means that the PV output fluctuates simultaneously in the same direction for all prefectures in the Tokyo area.

The cross-correlation coefficients for Japan are shown in Fig. 3.11. Panels (a) and (c) are genuine correlation *Ct*, and panels (b) and (d) are the random components *Cr*. The genuine correlation *Ct* was calculated using only the five largest eigenvalues and corresponding eigenvectors. Figure 3.11 depicts that the genuine correlation *Ct* has a positive correlation, and on the other hand, the random components *Cr* distribute around 0*.*0. The cross-correlation of PV output fluctuation in the Tokyo area was larger than the cross-correlation in Japan throughout the year. The 1st to 3rd eigenvectors for Japan are shown in Fig. 3.12. Forty-seven