A Retrospective View of the Phillips Curve and Its Empirical Validity since the 1950s

Hoang-Phuong Do

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Economics

Aris Spanos, Chair
Kwok Ping Tsang
Melinda Miller
Shaowen Luo

April 16, 2021
Blacksburg, Virginia

Keywords: Phillips Curve, Empirical Evaluation, Statistical Adequacy, Macroeconomic History

A Retrospective View of the Phillips Curve and Its Empirical Validity since the 1950s

Hoang-Phuong Do

(ABSTRACT)

Since the 1960s, the Phillips curve has survived various significant changes (Kuhnian paradigm shifts) in macroeconomic theory and generated endless controversies. This dissertation revisits several important, representative papers throughout the curve's four historical, formative periods: Phillips' foundational paper in 1958, the wage determination literature in the 1960s, the expectations-augmented Phillips curve in the 1970s, and the latest New Keynesian iteration. The purpose is to provide a retrospective evaluation of the curve's empirical evidence. In each period, the preeminent role of the theoretical considerations over statistical learning from the data is first explored. To further appraise the trustworthiness of empirical evidence, a few key empirical models are then selected and evaluated for their statistical adequacy, which refers to the validity of the probabilistic assumptions comprising the statistical models. The evaluation results, using the historical (vintage) data in the first three periods and the modern data in the final one, show that nearly all of the models in the appraisal are misspecified - at least one probabilistic assumption is not valid. The statistically adequate models produced from the respecification with the same data suggest new understandings of the main variables' behaviors. The dissertations' findings from the representative papers cast doubt on the traditional narrative of the Phillips curve, which the representative papers play a crucial role in establishing.

A Retrospective View of the Phillips Curve and Its Empirical Validity since the 1950s

Hoang-Phuong Do

(GENERAL AUDIENCE ABSTRACT)

The empirical regularity of the Phillips curve, which captures the inverse relationship between the inflation and unemployment rates, has been widely debated in academic economic research and between policymakers in the last 60 years. To shed light on the debate, this dissertation examines a selected list of influential, representative studies from the Phillips curves' empirical history through its four formative periods. The examinations of these papers are conducted as a blend between a discussion on the methodology of econometrics (the primary quantitative method in economics), the role of theory vs. statistical learning from the observed data, and evaluations of the validity of the probabilistic assumptions assumed behind the empirical models. The main contention is that any departure of probabilistic assumptions produces unreliable statistical inference, rendering the empirical analysis untrustworthy. The evaluation results show that nearly all of the models in the appraisal are untrustworthy - at least one assumption is not valid. Then, an attempt to produce improved empirical models is made to produce new understandings. Overall, the dissertation's findings cast doubt on the traditional narrative of the Phillips curve, which the representative papers play a crucial role in establishing.

Acknowledgments

First and foremost, I wish to express my deepest gratitude to my advisor, Dr. Aris Spanos. His inspirational lectures, intellectual guidance, and persistent supports are the main reason for me to complete my graduate study.

I would like to thank the committee members, Dr. Kwok Ping Tsang, Dr. Melinda Miller, and Dr. Shaowen Luo. Their valuable and constructive advice enabled the completion of this work.

I am also very grateful to Dr. Hans Haller, Dr. Amoz Kats, Dr. Djavad Salehi-Isfahani, and Dr. Nicolaus Tideman. All of the discussions with them kept me motivated and helped me grow intellectually.

My sincere thanks also go to Dr. Steve Trost, who was very supportive when I performed my teaching duty. My appreciation also extends to the staff members of our department: Will Bebout and Amy Standford for their assistance during the last five years.

I am deeply thankful to Dr. Vu Thanh Tu Anh at Fulbright University. His mentoring during my first research assistant work is always the foundation for my intellectual development. Many thanks also go to Dr. Luciano Andreozzi. I learned my first proper lessons of economics from his lectures during my master's study.

My heartfelt appreciation goes to my Virginia Tech friends: Zichao Yang, ChangGeun Song, Trevin Glasgow, Brian D'Orazio, and Sunjin Kim. The discussions, coffee and dinner time, and, most importantly, our friendship were a constant source of encouragement and support. Similar gratitude also extends to my friends from the Vietnamese community in Blacksburg and many more around the world. The gatherings and their emotional support were always an essential aspect of my graduate life.

Last but not least, I am eternally grateful to my parents, Do Van Thanh and Nguyen Thi Hien, who have taught me the most valuable lessons of my life. My parents and my brother, Do Hoang Viet, never stops giving me their unconditional love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A Short History of the Phillips Curve

The history of the Phillips curve is full of twists and turns. The traditional story tells us that the New Zealand economist Alban Phillips first discovered the relationship between the unemployment and inflation rates in 1958. In his paper, Phillips collected annual United Kingdom data from 1861 to 1913 and showed a negative correlation between wage inflation and unemployment. The sample length lasting for more than a half of a century provides the relationship's stability proof. The curve got its second big moment in 1960 when the two economists Paul Samuelson and Robert Solow explored the US's same relationship. It was then widely accepted that Samuelson and Solow (1960) proposed a menu of choices for policymakers to exploit the perceived trade-off between unemployment and inflation. To lower the unemployment rate, the government can implement expansionary policies and incur higher inflation rates. These two studies were later thought to be favorably received by economists and unleashed a large follow-up literature that finer details of the trade-off were explored in the 1960s. On the policy-making level, the Phillips curve was also believed to become a theoretical vehicle to deliver arguments for the government's interventions in pushing up the economy to desired levels. It is the original and simple version of the Phillips curve that economists and the public recognize today.

By the end of the 1960s, the original version of the curve was vigorously challenged. The theoretical study by Phelps (1967) and the influential presidential speech by Friedman (1968) independently points out that the stable trade-off discovered by Phillips and accepted by the profession is actually unstable. What was missing in the previous analysis is the role of expectation. When workers observe their real purchasing power erode due to higher price levels, they subsequently raise their expectation of future inflation and lower their supply to the labor market. Newly updated expectations will drag the unemployment rate back to the starting (higher) level. The reverting threshold was proposed by Friedman as the "natural rate of unemployment" concept. Any attempts to push the unemployment rate below its natural level are prone to be unsuccessful. This theoretical analysis renders the original Phillips curve only as the short-run form; in the short run, the unemployment rate is responsive to changes in the inflation rate. However, in the long run, the unemployment rate is neutral to any inflationary attempts to lower it below its natural rate. The permanent trade-off is actually temporary. The short-run and the long-run relations can be integrated compactly as a single expectations-augmented Phillips curve (also called the 'accelerationist' curve). It is the second iteration of the curve.

The occurrence of stagflation during the 1970s, which featured both displeasingly rising inflation and an undesirably high level of unemployment, seemed to provide a fertile ground for empirical efforts to confirm Friedman's foresight idea. The empirical contest between the original and the expectations-augmented curves in that time came under the name of the Natural Rate Hypothesis (also known as 'the accelerationist hypothesis'). Then, by the end of the 1970s, empirical studies consensually confirmed the Natural Rate Hypothesis; the original Phillips curve was agreed as a naive and defunct understanding of the economy. During the 1970s, the initial wage Phillips curve equation, in which wage change is related to price change and unemployment, was also replaced by the price equation with price inflation and unemployment as the only two principal variables. Later on, the expectation-augmented Phillips curve has become the standard tool to investigate the inflation-unemployment-expectation relationship by economists.

Recently, the New Keynesian paradigm has become the workhorse, theoretical approach in macroeconomics. Under the rational expectation and the microfoundation apparatus, the New Keynesian Phillips curve (NKPC) was introduced as the third iteration. Compared to the 'old' curve, the expectation of the *current* inflation rate is replaced by the rational expectation of the *future* rate on the right-hand side (RHS) of the NKPC, and unemployment as a proxy of slackness in the economy is changed into the real marginal cost. A slightly different theoretical modification also incorporates past inflation into the NKPC to create the hybrid NKPC that both information from the past and anticipation toward the future inflation play a role in the determination of the current rate. On the empirical level, the NKPC can be estimated as a single independent equation or incorporated into the large dynamic stochastic general equilibrium modeling (DSGE) systems. The empirical merit of the NKPC is currently an unsettled debate in the literature.

This standard narrative of the Phillips curve is widespread recounted on many levels, including introductory textbooks to undergraduate students, historical papers by researchers making their names by their contributions to the topic (Fuhrer et al. 2009; Gordon 2011), and history books of the modern macroeconomics (Snowdon and Vane 2005; De Vroey 2016). However, Forder (2014a), as a comprehensive book about the history of the Phillips curve, impressively argues that the conventional narrative of the curve is deeply flawed.

Firstly, the widespread presumption, which says that Phillips' groundbreaking work in 1958 was initially well-received and inspiring, is ungrounded. Many criticisms questioned the paper's data reliability, the quantitative methodology, and the theoretical explanation. Later, Samuelson and Solow neither took a firm stand on the existence of a menu of choices between inflation and unemployment drawing nor supported inflationism as an appropriate measure to push down unemployment.

In the 1960s, the empirical literature, which studied the relationship between wage inflation and unemployment, was actually not induced by Phillips' work. Much of it belongs to the so-called wage-determination literature in which economists employed econometric techniques to explore various factors affecting wage change empirically. Moreover, what Friedman argued about the negligence of inflation expectation in his 1967 presidential speech is unfair. In many papers drawing from this literature in the 1960s, price change was intentionally taken into account in the regressions equation as an important explanatory factor.

Forder also shows that the expectation argument, which novelty is often attributed to Friedman and Phelp, was not a revolutionary idea. It is straightforward to describe the 1970s' stagflation as the shift of the Phillips curve in undergraduate textbooks on the empirical level.

However, the empirical literature during the 1970s did not produce a uniform verdict on the Natural Rate Hypothesis. In the first half of the turbulent decade, many studies rejected the hypothesis.

In the final chapter, Forder also attempts to explain why the Phillips-curve debate was beyond the point of agreement in the 1970s. Much of it comes from the multiplicity of theoretical arguments as well as countless empirical twists:

> *Three things can perhaps be suggested to explain the survival of the idea of the Phillips curve. One is that it has become very hard to show that it does not exist. One thing that happened in the 1970s is that introduction of the idea of 'expectations', along with any number of ways of measuring them, together with 'adjusted unemployment', with any number of ways of adjusting it, made for an explosion in the variety of Phillips curves that might be estimated - but did so without moving away from a rather basic idea of the curve as a relation between unanticipated inflation and unemployment. In this way, the inflation of the 1970s, by spurring these developments, created the epicycles of the theory that made the existence Phillips curve proof against refutation.* (Forder 2014a, page 296)

1.2 The Recent, Persistent Debate of the Phillips Curve

Forder's immense historical work only limits the majority of its investigation in the 1960s and the 1970s. However, the repeated theme can be found in the later Phillips curve literature. In particular, since the 1980s, various ad-hoc modifications have been added to the dominant expectations-augmented Phillips curve to explain away the behavior of inflation with responding changes in unemployment; see Summa and Braga (2019).

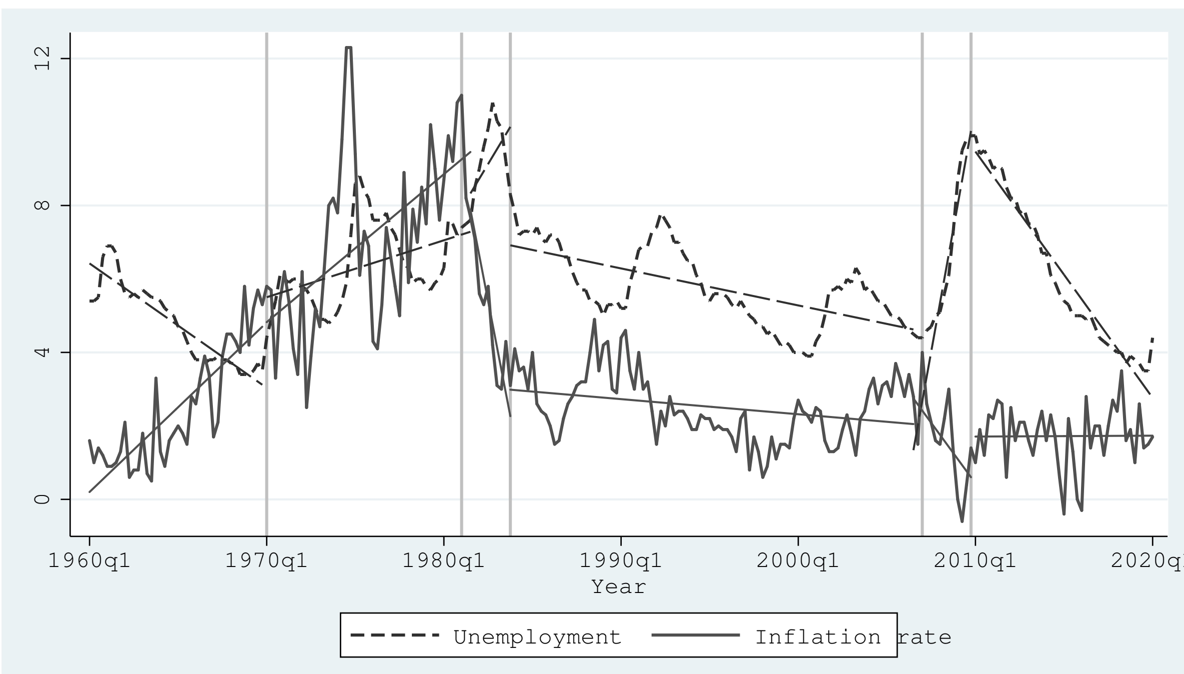
The standard expectations-augmented Phillips equation states that the price inflation rate is a function of the current expected rate, the difference between the actual rate of unemployment and the non-accelerating inflation rate of unemployment (NAIRU), and supply factors (Gordon 1977b, 1997's triangle model). The expected inflation is approximated by a sum of long lags of past inflation. The observed period of low and stable inflation and low level of unemployment in the late 1980s and early 1990s first challenged this equation's empirical robustness. Then, the time-varying (non-constant) NAIRU (Gordon 1997; Ball and Mankiw 2002) was proposed to show that the old framework is still empirically valid. Another explanation is that the slope of the Phillips curve is flattening over time (smaller effect of demand factors) due to a decrease in the frequency of firms' wage and price adjustment (Ball, Mankiw, and Romer 1991).

The next puzzling pattern appeared in the 2007-2010 period. During and after the Financial Crisis, the US economy had experienced a rising unemployment rate without getting under much inflationary pressure. Ball and Mazumder (2011) suggests that there is 'missing inflation' because the observed level of inflation is lower than the rate which the Phillips curve models predict. As a response, Gordon (2013) develops a modified model in which a short-run rate replaces the aggregate rate of unemployment, and the productivity measurement is introduced as another supply factor. Newly estimated results using this model reconfirms the empirical soundness of the old framework.

The post-Great Recession recovery of the US economy observed a steady reduction in the unemployment rate combined with a stable low inflation level. It raised an issue on why the inflation rate failed to surge up in response to the decreasing unemployment trend during the long decade. To answer this puzzle, some studies (Blanchard 2016, 2018; Gordon 2018; Ball and Mazumder 2019) propose that the recent US credible monetary policy has successfully anchored inflation expectation at low rates. Economic agents' expectations have become less sensitive to economic activities. In other words, the slope of the Phillips curve has flattened considerably. Others show that the Phillips curve's non-linear forms considerably fit well with the data (Nalewaik 2016; Hooper, Mishkin, and Sufi 2020).

In fact, since the 1960s, the US economy has experienced almost all possible combined patterns between unemployment and inflation. On Figure 1.1, it can be seen that the two leading macroeconomic indicators: moved oppositely in the 1960s, rocketed synchronously in the 1970s, declined identically after the Volcker deflation in the 1980s, had a negative correlation during and in the aftermath of the 2007-2008 Financial Crisis and showed no relation from 2010 to 2019. A fair mind, who is not informed on various explanations in the literature, can easily ask himself why the Phillips curve can fit nearly all conceivable co-movements between unemployment and inflation. The possible answer for an empirics-related question can only be learned from the empirical aspect. Researching this aspect of the Phillips curve history is the primary objective for this dissertation.

Figure 1.1: The Unemployment and Inflation Rates, Quarterly Data, 1960-2019



1.3 Primary Objectives

From the above short account of the Phillips curve history, it can come to the conclusion that the unsettled state of the recurring debate around the Phillips curve stems from two things. Firstly, the Phillips curve as an economic concept is very flexible and has continuously evolved to accommodate endless theoretical twists from economists when they confronted economic problems in their times. The specific notion of the Phillips curve that was debated yesterday may not be exactly the same notion in today's conversations. Secondly, empirical research has continuously produced supporting evidence for each round of new theoretical adjustments. Furthermore, the history of the Phillips curve, as shown by Forder (2014a), has never been recounted to explain all of these historical evolutions and complications. One without advanced knowledge of the literature can easily get the impression that the Phillips curve was an empirical regularity at some points during the curve's history.

Forder (2014a) examines the traditional narrative and comprehensively exposes that the development of the Phillips curve is nothing but a sequence of stories to justify the theoretical turns in macroeconomics. However, his book still leaves the empiric side of the curve untouched. To shed light on the recurring debates and controversies relating to the nature of the Phillips curve since its formation in 1958, this dissertation revisits several influential, empirical papers during the curve's long history. The ultimate intention is to evaluate whether the conventional wisdom relating to the Phillips curve being an empirical regularity is well founded.

Since the 1950s, the traditional econometric textbook approach has been the main methodology to generate empirical credentials for the Phillips curve. This approach can be seen as the preeminence of theory perspective and has many crucial weaknesses (Spanos 2006, 2010; Hoover 2006). Its primary objective, to a certain degree, is not to learn from data about particular phenomena of interest, but to select fitted curves that enable the modeler to tell convincing enough stories: *"Both academic economists and policy makers use DSGE models to tell stories about how the economy responds to unexpected movements in the exogenous variables."* (Canova 2007, page 160). This methodology cannot be reliably applied to assess the empirical aspect of the Phillips curve.

This dissertation, therefore, chooses the Probability Reduction (PR) approach proposed in Spanos (2006, 2010) to forward statistical evaluations. Ultimately, the PR approach shows that behind any (empirical) structural models $\mathcal{M}_{\varphi}(x)$ there is an implicit statistical model $\mathcal{M}_{\theta}(x)$ which comprises the probabilistic assumptions imposed on data $\mathbf{x}_0$. Only when the validity of the probabilistic assumptions is secured, can one ensure the reliability of the statistical inferences and obtain trustworthy evidence. The main focus of this dissertation is to evaluate the trustworthiness of empirical evidence from a selected list of the Phillips curve's most influential papers.

To accomplish this goal, the history of the Phillips curve is divided into four major, monumental periods: (1) the original curve of Phillips, (2) the forgotten wage-determination literature in the 1960s, (3) the mainstream expectations-augmented Phillips curve, and (4) the latest New Keynesian Phillips Curve. There are a vast number of papers with endless variations in the literature that either confirm or reject the Phillips curve. It is impossible to evaluate all of these studies. However, it is possible to pick a few landmark papers from the Phillips curve's major development phases and appraise their statistical validity. Consequently, a couple of representative empirical works is picked for each period. Evaluation results of these papers can provide a concise and emblematic insight into the overall quality of the empirical Phillips curve literature. Except for Chapter 6, all vintage data will be used to ensure that the historical context is truly reflected.

1.4 A Brief Overview

Chapter 2 will bring out the serious weaknesses of the traditional textbook econometrics approach and introduce the PR approach to empirical modelling as the foundation for the subsequent statistical evaluations. The next four chapters - Chapter 3 to Chapter 6 - will focus on the four development periods of the Phillips curve: (1) the curve of Phillips, (2) the wage-determination literature, (3) the expectations-augmented Phillips curve, and (4) the New Keynesian Phillips Curve. In each of these four chapters, a short introduction and a discussion of the representative papers are first provided. Statistical evaluations for selected models are next performed. If misspecifications are found, an attempt to respecify with the same data will be carried out. The final chapter (Chapter 7) is the conclusion.

Chapter 2

Model-based Approach to Trustworthy Evidence

2.1 The Textbook Approach to Econometrics

Econometrics is a branch of economics giving the empirical content through statistical applications and mathematical models to economic theory. Hoover (2006) summarizes the four principal functions of econometrics. First, economists use econometric models to confirm or reject economic theories. Second, econometrics can be used to quantify unobserved theoretical parameters. Third, econometrics also helps to forecast future economic variables. Fourth, stylized economic facts (hidden relationships among economic variables) can be revealed from econometric research.

In serving these purposes, econometric modeling emphasizes economic theory playing the preeminent role over the learning from the data (Hoover 2006; Spanos 2006). However, economists disagree on the degree of commitment to economic theory over characterizations of observed data. The disagreements have led to the existence of more than one methodological approach in econometrics. In chronological order, the first well-known approach is the Cowles Commission. The subsequent developments - Vector Autoregression approach (Sims 1980), the London School of Economics methodology (Mizon 1995), calibration approach (Kydland and Prescott 1991), and the 'textbook' approach - are more or less the responses to criticisms received by as well as further development of the key concepts developed by the Cowles Commission; see Hoover (2006) and Spanos (2006) for a detailed discussion. Among these approaches, this section will focus exclusively on the textbook approach because it is the principal method of the single-equation Phillip-curve empirical research in the literature.

The textbook approach centers on how to provide estimation for single-equation regression models in econometric textbook (e.g. Greene 2012). The starting point is the quintessential Classical Linear Regression (CLR) model, specified in the form of:

$$Y_t = \beta_0 + \boldsymbol{\beta}_1' \mathbf{x}_t + \varepsilon_t,\ t \in \mathbb{N}, \tag{2.1}$$

where $\boldsymbol{\beta}$ and $\mathbf{x}_t$ are respectively $1 \times k$ and $T \times k$ matrix.

The statistical structure is introduced by imposing assumptions primarily on the *unobserved* error term ε_t. The statistical optimality of the CLR estimators is achieved through the classic Gauss-Markov (G-M) theorem with the assumptions $[1] - [7]$ on Table 2.1. However, in most practical applications, the non-experimental nature of economic data and the incomplete nature of the underlying economic theory render the error terms to be non white-noise. As a result, the ordinary least square (OLS) estimator $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$ loses its optimality. To keep the commitment to a prior theory, the textbook approach applies the 'error-fixing' strategies to achieve unbiasedness, consistency, and efficiency. For example, when the error term is presumed to be heteroskedastic, the heteroskedasticity-robust estimator is used to fix the perceived problem.

Being equipped with various estimation procedures, the methodological framework can be roughly described as follow. In the first step, a priori structural mathematical model capturing the economic relationship of the phenomena of interest is provided. Then, assumptions on the attached error term (e.g., uncorrelated, serial correlation, heteroskedasticity) are made, and potential violations of the G-M theorem are preliminarily tested (e.g., Durbin and Watson (D-W) test, White test). Synthesizing all this information, the appropriate estimation procedure(s), which is usually called the econometric strategy or identification strategy, are selected. Finally, the best available data is used to produce estimation results. It is important to note that the choice of the appropriate estimation procedure can come from both as a prior theoretical choices (e.g., nonlinear parameters or existence of omitted variables) and statistical considerations (e.g., discovering autocorrelation from the appropriate tests). At the most sophisticated level, the identification strategy is also an integral part of the development of the prior theory.

The textbook approach was further developed to add refinements to the causal analysis in microeconomics; see Heckman (2000). For example, the concepts of exogenous and endogenous variables were adopted from the Cowles Commission; and suitable techniques (e.g., instrumental variables, difference-in-difference) were proposed as statistical solutions for the assumption violations associated with the CR model in the context of causal analysis.

This next section will discuss the textbook approach's weaknesses and relate them to the case of the Phillips curve.

Table 2.1: Linear Regression Model: Traditional Specification

Substantive Generation Mechanism: $Y_t = \beta_0 + \boldsymbol{\beta}_1' \mathbf{x}_t + \varepsilon_t,\ t \in \mathbb{N}$	
[1] Normality:	$(\varepsilon_t \vert \mathbf{X}_t = \mathbf{x}_t) \backsim \mathsf{N}(.,.),\ t \in \mathbb{N}.$
[2] Zero mean:	$E\left(\varepsilon_t \vert \mathbf{X}_t{=}\mathbf{x}_t\right) = 0,\ t \in \mathbb{N}.$
[3] Homoskedasticity:	$Var\left(\varepsilon_t^2 \vert \mathbf{X}_t{=}\mathbf{x}_t\right) {=} \sigma_\varepsilon^2,\ t \in \mathbb{N}.$
[4] Zero correlation:	$\{(\varepsilon_t \vert \mathbf{X}_t{=}\mathbf{x}_t),\ t{\in}\mathbb{N}\}$ is uncorrelated.
[5] Fixed $\mathbf{X}_t$:	$\mathbf{X}_t$ is fixed at $\mathbf{x}_t$ in repeated samples, $t \in \mathbb{N}$.
[6] No omitted variables:	All relevant variables have been included.
[7] No collinearity:	$\text{rank}(\mathbf{X}'\mathbf{X}){=}p{+}1,\ \mathbf{X}{=}(\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p).$

2.2 Weaknesses of The Textbook Approach

2.2.1 Falsifiability

From the methodological perspective, the commitment to theory in econometric, especially in the textbook approach, poses a weakness that no single piece of evidence can decisively reject or confirm an economic question. Hoover (2006) summarily described:

> *...[In economics] neither verification nor falsification is typically the sharp result*

of single tests; rather empirical research is gradually redirected from lines less consonant to lines more consonant with the accumulated weight of evidence. Econometrics thus operates in a Lakatosian spirit, albeit with no commitment to the fine details of Lakatos's methodology.

In other words, economists are only on the verge of abandoning a theory or a theoretical framework when there is overwhelming multiple evidence rejecting it. This description perfectly fits Forder's account of the Phillips curve until the 1980s and the short discussion on the Phillips curve's persistence for the later periods. The "hardcore" unemployment-price-expectation relationship, from the Laktatosian point of view, has not been entirely rejected because the auxiliary hypotheses have been relentlessly introduced to provide empirical proof for the hardcore theory.

Part of this non-falsification weakness can be attributed to the inseparability between the economic and statistical assumptions of the textbook approach. When an empirical study provides evidence that is not consonant with the core theory, it is not straightforward to interpret the empirical result. In one way, it can be argued that the evidence rejects the core theory. However, on the other hand, it can be equally reasonable to say that the econometric (identification) strategy is not sufficiently sound.

2.2.2 Inference Reliability and Trustworthy Evidence

Apart from the problem of falsifiability, the textbook approach can potentially provide estimation results with unreliable statistical inferences. In the big picture, the current statistical inference of the textbook approach is closely related to the Fisher-Neyman-Person (F-N-P) approach to frequentist inference (Spanos 2006, 2010). In a nutshell, the primary tool of investigation under the F-N-P framework is the concept of a prespecified (parametric) statistical model $\mathcal{M}_\theta$ as:

$$\mathcal{M}_\theta(x) = \{f(x;\theta), \theta \in \Theta\}, x \in \mathbb{R}^n_X,$$

where $f(x;\theta)$ is the joint distribution of the sample $\mathbf{X} := (X_1, ..., X_n), n \in \mathbb{R}^n_X$ represents the sample space and Θ as the parameter space. $\mathcal{M}_\theta(x)$ contain all necessary statistical information to carry out statistical analysis. The observed data $\mathbf{x}_0 := (x_1, x_2, ..., x_n), n \in \mathbb{R}^n$ is then considered as a realization of the generating mechanism (GM) from the statistical model $\mathcal{M}_\theta(x)$. For example, the GM of the classic simple Normal model can be written as:

$$\mathcal{M}_\theta(x) : X_n \equiv \text{NIID}(\mu, \sigma^2), \theta := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+, \tag{2.2}$$

where NIID is the abbreviation of the Normal, Independent and Identically Distributed probability function with mean μ and standard deviation σ^2.

In the ideal world, there must exist a single unique value θ^* corresponding to every observed data $\mathbf{x}_0$ as:

$$\mathcal{M}^*_\theta(x) = \{f(x;\theta^*)\}, x \in \mathbb{R}^n_X\}, \tag{2.3}$$

where θ^* is the true value of θ. The stochastic nature of the statistical method prevents us from identifying the true value θ^* with the absolute level of certainty. Instead, the general purpose of statistical inference is to learn about θ^* by narrowing down its values as much as possible in the parameter space Θ.

Then, the learning of θ^* from the data $\mathbf{x}_0$ happens through the sampling distribution $f(y_n;\theta)$ of desired statistics $Y_n = g(X_1, X_2, ..., X_n)$, such as: estimator, test, predictor. The sampling distribution $f(y_n;\theta)$ derived from the joint distribution $f(x;\theta)$ of the statistical model $\mathcal{M}_\theta(x)$ as:

$$F(Y_n \leq y) = \underbrace{\int\int\cdots\int}_{\{\mathbf{x}:\ g(\mathbf{x})\leq y\}} f(\mathbf{x};\theta)d\mathbf{x},\ \forall y \in \mathbb{R}. \tag{2.4}$$

For example, under the simple Normal model the sampling distribution of the estimator $Y_n = \sum_{i=1}^{n} X_n$ of the mean μ is $N(\mu, \frac{\sigma}{\sqrt{n}})$.

Next, choosing a nominal error probability (e.g. 1% or 5%), the narrowing process of θ^* can be made under the following two forms of reasoning:

1. Factual (estimation and prediction): assume that $\theta = \theta^*$, whatever the true value θ^* is, we derive the Confidence Intervals (CIs) for the sampling distributions.

2. Hypothesis testing (hypothetical reasoning): investigating different hypothetical scenarios: how the sampling distribution becomes, if we assume $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$. The process is centered on the type I and type II error probabilities; see Spanos (2019).

The optimal properties (unbiasedness, consistency, and efficiency) of a statistic are desirable goals of frequentist inference. To ensure these optimal properties, one need to make sure the statistical assumptions imposed upon the statistical model $\mathcal{M}_\theta(x)$ being valid - $\mathcal{M}_\theta(x)$ is *statistically adequate*. If $f(y_n;\theta)$ from $\mathcal{M}_\theta(x)$ turns out to be wrong, the error probabilities will differ from the assumed values. In other words, the statistical inferences become *unreliable*. For example, using a test at 5% significance level will cause unreliable inference when the actual type I error is much higher, say 20%.

In the textbook econometric approach, the strong commitment to theory emphasizes the economic theory as a prior knowledge. The statistical model selection is conducted either by the strategy to fixing the error terms or by using the appropriate procedure to conform with variable characterizations of the economic theory. Furthermore, the tests for probabilistic assumptions (the statistical information) are not comprehensive and usually conflate with assertions from the economic theory (the substantive information). For example, the problem of omitted variable obtaining as non-statistical knowledge invalidates the statistical assumption [2] of the G-M theorem and leads to various fixing strategies (such as the instrumental variables (IV) approach). The ignorance of evaluating statistical adequacy - the validity of statistical assumptions imposed on the data - potentially renders empirical evidence applying the textbook econometric approach to be “untrustworthy”. The term “untrustworthy” can be understood in the sense that when the statistical inferences are unreliable, the empirical learning regarding supporting/rejecting economic theories is likely misleading.

A Monte Carlo simulation will be run next to demonstrate the unreliability of statistical inferences happening under misspecified statistical models and an exemplary error-fixing method.

2.2.3 Monte Carlo Simulation

In this part, two Monte Carlo experiments are performed to illustrate the unreliable inferences of the 'error-fixing' strategy in the textbook econometrics approach. These experiments are designed to closely resemble the misspecifications that are discovered from the evaluations of the Phillips curve's empirical evidence in the following chapters. The overall experiment design shares some similarities with the ones in Spanos and McGuirk (2001) and McGuirk and Spanos (2009), but it is reformulated to be relevant with the case of the Phillips curve.

Experiment 1

In the Experiment 1, the observed data are generated via a Normal, Markov(1) process with the following joint distribution:

$$\begin{pmatrix} Y_t \\ X_t \\ Y_{t-1} \\ X_{t-1} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 1.20 \\ 1.00 \\ 1.20 \\ 1.00 \end{bmatrix}, \begin{bmatrix} 1.40 & 0.90 & 0.75 & 0.60 \\ 0.90 & 1.80 & 0.80 & 0.50 \\ 0.75 & 0.80 & 1.40 & 0.90 \\ 0.60 & 0.50 & 0.90 & 1.80 \end{bmatrix} \right) \tag{2.5}$$

The choices of these numerical values imitate the typical behavior of macroeconomic time series. This joint distribution can be reparameterized to generate a Dynamic Linear Normal (DLR) model (Spanos 1994) as:

$$Y_t = 0.423 + 0.352X_t + 0.270Y_{t-1} + 0.101X_{t-1} + u_t \tag{2.6}$$

This distribution has temporal dependence for both X_t and Y_t. We will discover how statistical inferences under the misspecified models perform. The following five models are chosen to be estimated:

$$\begin{aligned}
&\text{(A)}: Y_t = c + \beta_0 X_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + u_t \\
&\text{(A1a)}: Y_t = c + \beta_0 X_t + u_t \\
&\text{(A1b)}: Y_t = c + \beta_0 X_t + u_t, \text{ with the Cochrane-Orcutt fix} \\
&\text{(A2a)}: Y_t = c + \beta_0 X_t + \beta_1 X_{t-1} + u_t \\
&\text{(A2b)}: Y_t = c + \beta_0 X_t + \beta_1 X_{t-1} + u_t, \text{ with the Cochrane-Orcutt fix}
\end{aligned}$$

Model A is the true model getting from the data's statistical GM. In models A1a and A1b, their specifications lack all of the temporal dependence (without X_{t-1} and Y_{t-1}). Models A2a and A2b, both without Y_{t-1}, do not have sufficient terms of lagged variables. Models A, A1a, and A2a are estimated by the OLS method to compare inferences from the true models and the misspecified models, which do not fully take into account the temporal dependence. Model A1b and model A2b respectively applied the iterative Cochrane–Orcutt correction procedure for model A1a and model A2a. This generalized least squares (GLS) procedure is a typical error-fixing technique to deal with the autocorrelation problem, which exists in models A1a and A2b due to their lack of temporal dependence. The results from the Cochrane–Orcutt estimators can well represent estimation results of other error-fixing strategies.

The estimation and the inference results for the five models with the same simulated data are reported on Tables 2.2 and 2.3. There are two simulations with the sample sizes at 50 and 100. The number of replications is 20000. For all five models, the arithmetic mean and standard deviations of the estimated coefficients are reported. The hypothesis testings (the t-test) at the true values of the correct specification are also performed. From the t-tests, the arithmetic mean of the t-statistic and the actual rejection rates (for the 0.05 significance level) are reported.

In Table 2.2 which presents the simulation results from the true DLR model, it can be seen that all of the means, the standard deviations, and actual rejection rates are very close to their nominal values. On the other hand, the simulation of the misspecified models A1a and A2a in Table 2.3 produces misleading results. The means of the estimated coefficients under the two models are highly biased. The actual rejection rates are not close to their nominal value at 0.05. Besides, the actual sizes of inconsistency (wider standard deviations) and statistical unreliability (worse actual rejection rates) are higher when the sample size n increases from 50 to 100.

The Cochrane-Orcutt fix for models A1a and A2a - the estimation results from models A1b and A2b - creates slightly better results. The extent of biasedness and inference unreliability in all estimated coefficients, especially for X_t, decreases. However, these values are far from the nominal levels.

Table 2.2: Simulation Results for Model A

True data GM: $Y_t = 0.423 + 0.352X_t + 0.270Y_{t-1} + 0.101X_{t-1} + u_t$				
N=20000	(A) True DLR model: $Y_t = c + \beta_0 X_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + u_t$			
	n=50		n=100	
Parameters	Mean	SD	Mean	SD
$[c = 0.423]\ \hat{c}$	0.478	0.215	0.450	0.142
$[\beta_0 = 0.352]\ \hat{\beta}_0$	0.352	0.117	0.352	0.081
$[\alpha_1 = 0.270]\ \hat{\alpha}_1$	0.227	0.151	0.248	0.105
$[\beta_1 = 0.101]\ \hat{\beta}_1$	0.100	0.121	0.101	0.084
t-statistics	Mean	%R(0.05)	Mean	%R(0.05)
$\tau_{\hat{c}} = \frac{\hat{c}-c}{\hat{\sigma}_c}$	0.163	0.056	0.114	0.053
$\tau_{\hat{\beta}_0} = \frac{\hat{\beta}_0-\beta_1}{\hat{\sigma}_{\beta_0}}$	-0.003	0.048	-0.001	0.051
$\tau_{\hat{\alpha}_1} = \frac{\hat{\alpha}_1-\alpha_1}{\hat{\sigma}_{\alpha_1}}$	-0.261	0.047	-0.190	0.051
$\tau_{\hat{\beta}_1} = \frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	-0.002	0.046	0.006	0.048

Table 2.3: Simulation Results for Models A1a, A1b, A2a and A2b

True data GM: $Y_t = 0.423 + 0.352X_t + 0.270Y_{t-1} + 0.101X_{t-1} + u_t$								
N=20000	(A1a) Misspecified model #1: $Y_t = c + \beta_0 X_t + u_t$,				(A1b) Cochrane-Orcutt fix for model A1a			
	n=50		n=100		n=50		n=100	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$[c = 0.423]$ $\hat{c}$	0.730	0.240	0.714	0.167	0.871	0.291	0.880	0.205
$[\beta_0 = 0.352]$ $\hat{\beta}_0$	0.471	0.126	0.485	0.089	0.330	0.164	0.319	0.117
t-statistics	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)
$\tau_{\hat{c}} = \frac{\hat{c}-c}{\hat{\sigma}_c}$	1.723	0.415	2.353	0.611	1.922	0.468	2.778	0.761
$\tau_{\hat{\beta}_0} = \frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	1.152	0.241	1.830	0.446	-0.221	0.223	-0.475	0.248

N=20000	(A2a) Misspecified model #2: $Y_t = c + \beta_0 X_t + \beta_1 X_{t-1} + u_t$,				(A2b) Cochrane-Orcutt fix for model A2a			
	n=50		n=100		n=50		n=100	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$[c = 0.423]$ $\hat{c}$	0.585	0.241	0.566	0.165	0.678	0.271	0.661	0.185
$[\beta_0 = 0.352]$ $\hat{\beta}_0$	0.426	0.108	0.434	0.075	0.352	0.122	0.355	0.084
$[\beta_1 = 0.101]$ $\hat{\beta}_1$	0.189	0.108	0.200	0.075	0.171	0.112	0.184	0.077
t-statistics	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)
$\tau_{\hat{c}} = \frac{\hat{c}-c}{\hat{\sigma}_c}$	0.809	0.174	1.054	0.228	1.023	0.184	1.410	0.295
$\tau_{\hat{\beta}_0} = \frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	0.699	0.103	1.106	0.192	-0.001	0.087	0.041	0.088
$\tau_{\hat{\beta}_1} = \frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	0.856	0.133	1.368	0.275	0.689	0.119	1.170	0.229

Experiment 2

Aside from the presence of temporal dependence, as discovered in Experiment 1, most macroeconomic data also exhibit time trends (increasing, decreasing, or complex patterns) in their means. The second set of simulated data via a linear-trend Normal, Markov(1) process is generated to incorporate the changing mean pattern from the following joint distribution:

$$\begin{pmatrix} Y_t \\ X_t \\ Y_{t-1} \\ X_{t-1} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 1.20+0.04t \\ 1.00+0.01t \\ 1.20+0.04(t-1) \\ 1.00+0.01(t-1) \end{bmatrix}, \begin{bmatrix} 1.40 & 0.90 & 0.75 & 0.60 \\ 0.90 & 1.80 & 0.80 & 0.50 \\ 0.75 & 0.80 & 1.40 & 0.90 \\ 0.60 & 0.50 & 0.90 & 1.80 \end{bmatrix} \right) \tag{2.7}$$

The true data GM of the mean-trending DLR model can be obtained as:

$$Y_t = 0.435 + 0.025t + 0.352X_t + 0.270Y_{t-1} + 0.101X_{t-1} + u_t \tag{2.8}$$

with $\sigma^2 = 0.820$.

The estimation and the inference results of the following six models are compared:

$$\begin{aligned}
\text{(B)}&: Y_t = c + \gamma t + \beta_0 X_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + u_t \\
\text{(B1)}&: Y_t = c + \beta_0 X_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + u_t \\
\text{(B2a)}&: Y_t = c + \gamma t + \beta_0 X_t + u_t \\
\text{(B2b)}&: Y_t = c + \gamma t + \beta_0 X_t + u_t, \text{ with the Cochrane-Orcutt fix} \\
\text{(B3a)}&: Y_t = c + \gamma t + \beta_0 X_t + \beta_1 X_{t-1} + u_t \\
\text{(B3b)}&: Y_t = c + \gamma t + \beta_0 X_t + \beta_1 X_{t-1} + u_t, \text{ with the Cochrane-Orcutt fix}
\end{aligned}$$

Among the six models, model B is the true DLR model; and model B1 excludes the time trend. In models B2a, B2b, B3a, and B3b, the linear time trend is included; but they lack sufficient temporal dependence. In models B2a and B2b, their specifications are without both Y_{t-1} and X_{t-1}, with B2b using the Cochrane-Orcutt procedure. Models B3a and B3b keep out Y_{t-1}, with B3b applying the Cochrane-Orcutt procedure.

The simulation results are reported in Tables 2.4 and 2.5. In the first misspecified model B1 in Table 2.4 which lacks the linear time trend, the estimated coefficients are misleading, and the actual rejections of the hypothesis testing at the true values are bigger than the nominal level at 0.05. These results suggest that the inclusion of the time trend is crucial to ensure estimation and inference precision.

Moving to the four misspecified models without sufficiently controlling for temporal dependence (B2a, B2b, B3a, B3b), the simulation results, as in the previous experiment, also reveal the same problems of bias and inference unreliability for both all OLS and the Cochrane-Orcutt estimations. The levels of inconsistency (wider standard deviations of the estimated coefficients) and inference unreliability also increase with the sample size n.

In conclusion, these simulation experiments provide an example of inference unreliability under the error-fixing strategy of the textbook econometrics approach. Spanos (2006, 2010) proposes the Probability Reduction (PR) approach to address these two key weaknesses of the textbook approach. The PR will be introduced in the next section.

Table 2.4: Simulation Results for Models B and B1

True data GM: $Y_t = 0.423 + 0.025t + 0.352X_t + 0.270Y_{t-1} + 0.101X_{t-1} + u_t$								
N=20000	(B) True DLR model: $Y_t = c + \gamma t + \beta_0 X_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + u_t$,				(B1) Misspecified #1: $Y_t = c + \beta_0 X_t + \alpha_1 Y_{t-1} + \beta_1 X_{t-1} + u_t$,			
	n=50		n=100		n=50		n=100	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$[c = 0.423]\ \hat{c}$	0.504	0.358	0.465	0.229	0.907	0.285	0.795	0.186
$[\gamma = 0.025]\ \hat{\gamma}$	0.028	0.012	0.026	0.005	-	-	-	-
$[\beta_0 = 0.352]\ \hat{\beta}_0$	0.350	0.119	0.352	0.081	0.315	0.123	0.263	0.089
$[\alpha_1 = 0.270]\ \hat{\alpha}_1$	0.198	0.152	0.234	0.106	0.394	0.150	0.648	0.074
$[\beta_1 = 0.101]\ \hat{\beta}_1$	0.101	0.122	0.102	0.084	0.049	0.128	-0.020	0.093
t-statistics	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)
$\tau_{\hat{c}} = \frac{\hat{c}-c}{\hat{\sigma}_c}$	0.133	0.067	0.078	0.061	1.609	0.305	1.506	0.224
$\tau_{\hat{\gamma}} = \frac{\hat{\gamma}-\gamma}{\hat{\sigma}_\gamma}$	0.189	0.067	0.224	0.056	-	-	-	-
$\tau_{\hat{\beta}_0} = \frac{\hat{\beta}_0-\beta_1}{\hat{\sigma}_{\beta_0}}$	-0.016	0.049	-0.007	0.050	-0.314	0.061	-1.015	0.171
$\tau_{\hat{\alpha}_1} = \frac{\hat{\alpha}_1-\alpha_1}{\hat{\sigma}_{\alpha_1}}$	-0.442	0.059	-0.318	0.057	0.908	0.157	4.568	0.989
$\tau_{\hat{\beta}_1} = \frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	0.011	0.047	0.016	0.049	-0.413	0.068	-1.330	0.260

2.3 The Probabilistic Reduction Approach

2.3.1 Statistical vs. Substantive Models

To avoid the textbook approach's conflation between the economic and statistical assumptions, the PR approach emphasizes the full separation between two types of models: substantive models and statistical models. Substantive models represent the empirically-feasible equations that are derived from economic theory to explain economic phenomena. Statistical models refer to a set of a complete, internally consistent, and testable (vis-a-vis observed data) set of statistical assumptions which aim to capture the true statistical generation mechanism (GM) of observed data. As a result, there is no longer incongruous inseparation between substantive assumptions (originating from economic theory) and statistical assumptions (components of the prespecified statistical models). The link between substantive models and statistical models comes from testing the restrictions of the substantive parameters on statistical models.

This separation allows one to put appropriate blames on the two types of misspecifications: statistical (invalid probabilistic assumptions) and substantive (failures of economic theories to explain data). In the context of the Phillips curve's unfalsifiability, the accurate descriptions of the inflation-unemployment relation potentially help to reject the inadequate protection-belt theories decisively.

To illustrate the differences between the PR and the textbook econometric approaches, a brief delineation of the reformulation of the CLR model in Table 2.1 as a statistical model will be presented; see Spanos (2019) for a full discussion. As provided in Section 2.2, the core of the

Table 2.5: Simulation Results for Models B2a, B2b, B3a and B4a

True data GM: $Y_t = 0.423 + 0.025t + 0.352X_t + 0.270Y_{t-1} + 0.101X_{t-1} + u_t$								
N=20000	(B2a) Misspecified model #2: $Y_t = c + \gamma t + \beta_0 X_t + u_t$,				(B2b) Cochrane-Orcutt fix for mispecified model B2a			
	n=50		n=100		n=50		n=100	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$[c = 0.423]$ $\hat{c}$	0.727	0.443	0.705	0.303	0.838	0.541	0.851	0.367
$[\gamma = 0.025]$ $\hat{\gamma}$	0.036	0.014	0.035	0.005	0.037	0.017	0.037	0.006
$[\beta_0 = 0.352]$ $\hat{\beta}_0$	0.450	0.127	0.475	0.089	0.332	0.161	0.321	0.116
t-statistics	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)
$\tau_{\hat{c}} = \frac{\hat{c}-c}{\hat{\sigma}_c}$	0.916	0.242	1.250	0.310	0.919	0.208	1.363	0.301
$\tau_{\hat{\gamma}} = \frac{\hat{\gamma}-\gamma}{\hat{\sigma}_\gamma}$	1.099	0.279	3.005	0.767	0.923	0.206	2.463	0.639
$\tau_{\hat{\beta}_0} = \frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	0.933	0.191	1.671	0.396	-0.192	0.207	-0.435	0.238

N=20000	(B3a) Misspecified model #3: $Y_t = c + \gamma t + \beta_0 X_t + \beta_1 X_{t-1} + u_t$,				(B3b) Cochrane-Orcutt fix for mispecified model B3a			
	n=50		n=100		n=50		n=100	
Parameters	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$[c = 0.423]$ $\hat{c}$	0.592	0.413	0.566	0.277	0.668	0.473	0.650	0.309
$[\gamma = 0.025]$ $\hat{\gamma}$	0.034	0.013	0.034	0.005	0.035	0.015	0.035	0.005
$[\beta_0 = 0.352]$ $\hat{\beta}_0$	0.414	0.110	0.428	0.075	0.350	0.122	0.354	0.084
$[\beta_1 = 0.101]$ $\hat{\beta}_1$	0.178	0.110	0.195	0.076	0.163	0.112	0.180	0.077
t-statistics	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)	Mean	%R(0.05)
$\tau_{\hat{c}} = \frac{\hat{c}-c}{\hat{\sigma}_c}$	0.455	0.148	0.580	0.160	0.521	0.118	0.738	0.135
$\tau_{\hat{\gamma}} = \frac{\hat{\gamma}-\gamma}{\hat{\sigma}_\gamma}$	0.964	0.224	2.622	0.689	0.830	0.163	2.229	0.578
$\tau_{\hat{\beta}_0} = \frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\beta_0}}$	0.579	0.090	1.023	0.173	-0.023	0.083	0.023	0.086
$\tau_{\hat{\beta}_1} = \frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$	0.742	0.114	1.286	0.249	0.604	0.104	1.110	0.210

statistical model is a joint distribution $f(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_t; \Theta)$. In the PR approach, the probabilistic assumptions can be divided into three general categories: dependence, heterogeneity, and distribution. For the case of the CLR model, one needs to invoke three probabilistic assumptions (corresponding to the three categories): independent, stationary and Normal distribution. They result in the joint multivariate Normal distribution as:

$$\mathbf{Z}_t := \begin{pmatrix} Y_t \\ \mathbf{X}_t \end{pmatrix} \sim \text{N}\left(\begin{bmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}'_{12} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \tag{2.9}$$

Then, the conditional distribution is obtained as:

$$(Y_t|\mathbf{X}_t{=}\mathbf{x}_t) \ \sim \ \mathsf{N}(\beta_0 + \boldsymbol{\beta}'_1\mathbf{x}_t, \sigma^2) \tag{2.10}$$

This conditional distribution creates the same regression function as in Equation 2.1. The exhaustive list of assumptions, which is comparable to the list of the CLR model in Table 2.1, is provided in Table 2.6.

Comparing between Table 2.1 and Table 2.6, the first four assumptions of Table 2.1 are equivalent to the the first four assumptions of Table 2.6. The key difference between them is at how each table makes use of substantive and statistical knowledge. For example, consider the assumption [2] $E(\varepsilon_t|\mathbf{X}_t{=}\mathbf{x}_t) = 0$ of Table 2.1, it is saying that there are no correlation between the *external variables* W_t and the unobserved error terms ε_t ($Cov(X_t, W_t) \neq 0$). The economic theory in the form of substantive models decides the appropriateness of this assumption by asserting whether there are omitted variables or not. In contrast, the assumption (2) $E(Y_t|\mathbf{X}_t{=}\mathbf{x}_t){=}\beta_0{+}\boldsymbol{\beta}_1\mathbf{x}_t$ only describe the characterization of the observed statistical errors u_t using only the observed data Y_t and $\mathbf{X}_t$.

Table 2.6: Normal, Linear Regression (LR) Model

Statistical Generation Mechanism (GM): $Y_t = \beta_0 + \boldsymbol{\beta}'_1\mathbf{x}_t + \varepsilon_t,\ t \in \mathbb{N}$	
(1) Normality:	$(Y_t|\mathbf{X}_t{=}\mathbf{x}_t) \backsim \mathsf{N}(.,.),\ t \in \mathbb{N}.$
(2) Linearity:	$E(Y_t|\mathbf{X}_t{=}\mathbf{x}_t){=}\beta_0 + \boldsymbol{\beta}'_1\mathbf{x}_t,\ t \in \mathbb{N}.$
(3) Homoskedasticity	$Var(Y_t|\mathbf{X}_t{=}\mathbf{x}_t){=}\sigma^2,\ t \in \mathbb{N}.$
(4) Independence	$\{(Y_t|\mathbf{X}_t{=}\mathbf{x}_t),\ t{\in}\mathbb{N}\}$ independent process, $t \in \mathbb{N}$.
(5) t-invariance:	$\theta{:=}\left(\beta_0, \boldsymbol{\beta}_1, \sigma^2\right)$ are *not* changing with t, $t \in \mathbb{N}$.
$\boldsymbol{\beta}{:=}(\beta_0, \boldsymbol{\beta}_1)',\ \beta_0{=}E(Y_t){-}\boldsymbol{\beta}'_1E(\mathbf{X}_t),\ \boldsymbol{\beta}_1{=}[Cov(\mathbf{X}_t)]^{-1}Cov(\mathbf{X}_t, Y_t),$ $\sigma^2{=}Var(Y_t){-}Cov(\mathbf{X}_t, Y_t)'[Cov(\mathbf{X}_t)]^{-1}Cov(\mathbf{X}_t, Y_t)$	

2.3.2 Testing the Validity of Probabilistic Assumptions

The formulation of the statistical model via the observed opens the gateway to address the second weakness of the textbook approach on making sure the statistical model's probabilistic assumptions are satisfied under a particular data. The PR approach evaluates the validity of the probabilistic assumptions via the Mis-Specification (M-S) testing; see Spanos (2018) for a complete discussion.

Intuitively, the M-S testing is built on the principle that residuals obtained from the estimation of a statistically adequate model are truly random. The test of randomness is performed by running auxiliary least-squares regressions on the residuals. The auxiliary regressions are designed to look for any significant departures from model assumptions. In general, model assumptions cannot be tested individually and must be evaluated altogether because their interrelation can cause misdiagnosis.

Considering a single explanatory variable (X_t is a scalar) for the LR model in Table 2.6, the (indicative) auxiliary regressions are performed through the following equations:

$$\hat{u}_t = \gamma_0 + \gamma_1 x_t + \overbrace{\gamma_2 x_t^2}^{\text{(2) Linearity}} + \overbrace{\gamma_3 t}^{\text{(5) t-invariance}} + \overbrace{\gamma_4 x_{t-1} + \gamma_5 Y_{t-1}}^{\text{(4) Independence}} + v_{1t} \tag{2.11}$$

$$\hat{u}_t^2 = \delta_0 + \overbrace{\delta_1 x_t^2}^{\text{(3) Homoskedasticity}} + \overbrace{\delta_2 t}^{\text{(5) t-invariance}} + \overbrace{\delta_3 x_{t-1}^2 + \delta_4 Y_{t-1}^2}^{\text{(4) Independence}} + v_{2t} \tag{2.12}$$

Equations 2.11 and 2.12 are called respectively as the regression and skedastic auxiliary regressions. The two auxiliary regressions are used to probe different assumption departures for the first and the second conditional moments in Table 2.6. The null hypothesis for assumptions (2)-(5) of Table 2.6 are presented on Table 2.7.

Table 2.7: M-S Testing's Null Hypotheses

Assumption	Null hypothesis	Auxiliary equation
(2) Linearity	$H_0 : \gamma_2 = 0$	Equation 2.11
(3) Homoskedasticity	$H_0 : \delta_1 = 0$	Equation 2.12
(4) Independence	$H_0 : \gamma_4 = \gamma_5 = 0$	Equation 2.11
	$H_0 : \delta_3 = \delta_4 = 0$	Equation 2.12
(5) t-invariance	$H_0 : \gamma_3 = 0$	Equation 2.11
	$H_0 : \delta_2 = 0$	Equation 2.12

When any null hypothesis H_0 in Table 2.7 are found not to be rejected, it indicates the departure of the corresponding assumptions. For example, if p-value of δ_1 is smaller than 0.05, it shows that the (3) homoskedasticity assumption is not supported. After securing the assumptions (2) - (5), the Normality assumption can be tested from the formal tests, such as the skewness and kurtosis test or the D'Agostino's K-squared test[1].

From the M-S testing, if any of the five assumptions are found not to be supported, the estimated model can be said to be misspecified. As illustrated in the Monte Carlo simulation, the actual error probabilities from statistical inferences of the misspecified model are different from the nominal probabilities chosen in the hypothesis testing. Also, the signs and values of estimated coefficients are highly misleading. Therefore, the empirical evidence inferred from the misspecified model becomes untrustworthy. On the other hand, when the M-S testings show the five assumptions cannot be rejected, it can be said that the estimated model is *statistically adequate*. It is the ultimate goal of statistical modeling in the PR approach. Achieving statistical adequacy guarantees reliable references - the trustworthiness of the empirical evidence.

1. When the evaluated statistical models are more complicated, or the sample size is too small, the two indicated auxiliary regressions in Equations 2.11 and 2.12 are modified accordingly to have appropriate degrees of freedom.

As discussed in the previous chapter, the Phillips curve's empirical regularity has been recurringly rediscovered in the literature. The preeminence of theory in the empirical modeling of the Phillips curve poses a big question on the validity of statistical assumptions behind the estimated models proving its empirical regularity. If the estimated models are statistically inadequate, the obtained evidence is untrustworthy, indicating Phillips curve empirical regularity was not truly established. The next four chapters will apply the M-S testing to learn about the trustworthiness of the Phillips curve's evidence in its four developmental periods.

Chapter 3

The Curve of Phillips

3.1 Introduction

3.1.1 Phillips' Correlation

The conventional history of the Phillips curve starts with a groundbreaking paper written by the New Zealand economist A. W. Phillips in 1958. In this paper, Phillips collected annual United Kingdom data from 1861 to 1913 and showed that there was a negative correlation between the wage inflation and unemployment rates. That the sample length lasts for more than a half of a century provides the stability proof of the relationship. This discovery marked the relation name as the Phillips curve.

However, Forder (2014a) convincingly argues that Phillips' study was not particularly influential immediately after its publication. Firstly, chapter 1 of the book shows that Phillips was not the first person who proposed the negative relationship between the two variables. The ideas were neither innovative nor there had not existed any exploration predating Phillips' 1958 paper. Secondly, the popular presumption, which says that Phillips' groundbreaking work in 1958 was initially well-received and inspiring, is ungrounded. There were many angles of criticisms questioning the paper's data reliability, the quantitative methodology, and the theoretical explanation. In fact, Phillips himself showed a due amount of reluctance in his conclusion section:

> *These conclusions [draw from the paper] are of course tentative. There is need for much more detailed research into the relations between unemployment, wage rates, prices and productivity.* (Phillips 1958, page 299)

Phillips' quantitative method is not the type of econometric study with which economists are familiar today. The first group of Phillips' data points includes 53 annual observations spanning from 1861 to 1913. Phillips' original scatter plot for these points is redrawn on Figure 3.1. To quantify the relationship between the unemployment and wage inflation rates, Phillips first divided these 53 points into six groups according to the values of unemployment rates: 0-2%, 2-3%, 3-4%, 4-5%, 5-7%, and 7-11%. He then averaged all points in each group to obtain six average points (the diamond symbols $\diamond$ on Figure 3.1). The 'statistical' function, which has a log-log function form with three unknown parameters, was constructed from these six points. Phillips ran a simple regression with the rightest four average points to obtain two parameters. The last one was selected by a trial and error procedure to ensure the best fit for his estimated curve. The end-product of his curve-fitting process is the famous equation that is commonly recognized as the Phillips' original correlation:

$$\ln(\Delta w_t + 0.9) = 0.984 - 1.394 \ln(U_t), \tag{3.1}$$

where $\Delta w_t = \frac{w_{t+1} - w_{t-1}}{2w_t}$ is the (central differences) rate of change of the nominal wage rate, and U_t is the unemployment rate. Because the resulting equation is one-half graphical fitting and one-half regression, there are no attaching t-statistics and p-values as in standard econometric studies.

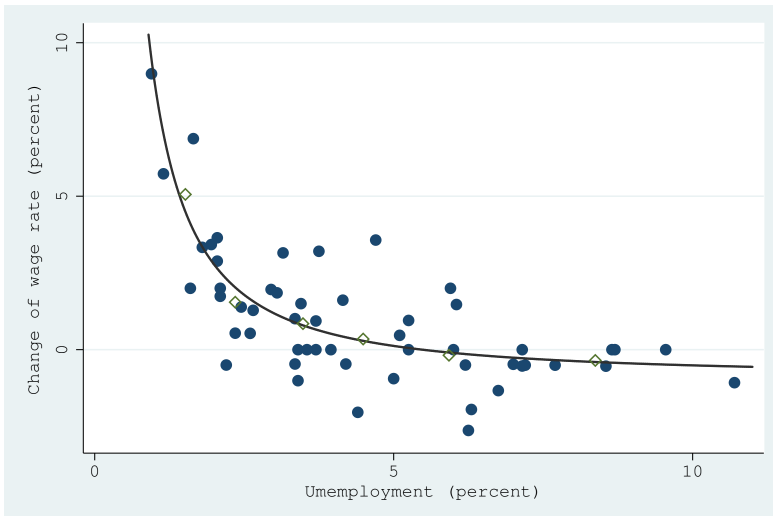


Figure 3.1: Phillips' Original Scatter Plot, Redrawn from Figure 1 of Phillips (1958).

Phillips devoted the rest of his paper to analyze the fitness of his newly estimated curve to data points for subsequent years. The conclusion is simple: the 1861-1913 curve can also explain the data from the 1913-1948 and 1948-1957 samples. Overall, it suggests that the inverse relationship between unemployment and wage inflation rates had been stable for nearly a century. It is important to stress that Phillips was well aware of the role of (domestic and import) price change as an additional essential factor in the formation of wage change and took account of the price levels into his analyses.

3.1.2 Lipsey's Regressions

As the only study directly inspired by Phillips, Lipsey (1960) attempted to improve Phillips' unconventional quantification method by employing the standard econometric technique of his time. To preserve Phillips' nonlinear function form, Lipsey firstly proposed the following equation:

$$\Delta w_t = \alpha_0 + \alpha_1 U_t^{-1} + \alpha_2 U_t^{-2} + \varepsilon_t \quad (3.2)$$

Because Phillips also considered the role of price change, Lipsey adds a "cost of living index" (p_t) as another explanatory variable as:

$$\Delta w_t = \alpha_0 + \alpha_1 U_t^{-1} + \alpha_2 U_t^{-2} + \alpha_2 \Delta p_t + \varepsilon_t \quad (3.3)$$

where Δp_t is also the central difference of p_t.

From the OLS estimation results with the same Phillips' data for Equations 3.2 and 3.3, Lipsey concluded that there is a "significant relation" between wage inflation and unemployment rates and a "weak" effect of the price change[1]. In his footnotes, Lipsey remarked that there is no evidence of autocorrelation without reporting the Durbin-Watson statistic that was widely used at the time.

3.2 Models in Evaluation

Essentially, Phillips (1958)'s original fitted curve is the first model to be evaluated under the PR approach. Despite the fact that Phillips' fitted curve was not based on traditional regression methods, one can evaluate its empirical validity since the two pieces of information needed for the M-S testing are the fitted values and residuals, which are available. In addition to Phillips' equation, Lipsey's two specifications in Equations 3.2 and 3.3 are also picked to be evaluated because his OLS standard treatment with Phillips' curve as well as his theoretical model in the paper played an important role in its development[2].

Phillips' vintage data and unorthodox quantitative method also attracted a couple of sporadic interests. In the 1970s, started by Desai (1975) and responded by Gilbert (1976), the duo studies revisited old estimated equations from Phillips (1958) and Lipsey (1960) with the updated technique of the pre-time series era. Recently, Granger and Jeon (2011) employed the modern time-series methodology to evaluate the merits of (revisioning) Phillips' non-linear form in forecasting. These all studies show that Phillips deserves credits in capturing statistical regularity in the data. These studies are not evaluated because they are minor studies and do not play an important role in the later literature.

As discussed in the introduction chapter, most statistical evaluations in this dissertation use vintage data to recreate the historical context. Hence, the replicated regressions for the three models will be run with the original dataset from Phillips (1958). The sample size includes 53 observations. In Chapter 2, the CLR model, which provides the foundation for the OLS estimation, is shown to equivalent to the LR model of the PR approach (Table 2.6). Thus, evaluation of evidence produced by the OLS estimation can be conducted by testing the statistical assumptions of the LR model. The M-S testings will next be performed to assess the validity of the five underlying statistical assumptions.

3.3 M-S Testing Results

Lipsey's two model in Equations 3.2 and 3.3 are called "Lipsey (1960)a" and "Lipsey (1960)b", respectively. The OLS replications of the two yield:

$$\Delta w_t = -\underset{(0.588)}{1.134} + \underset{(3.033)}{5.406} U_t^{-1} + \underset{(3.075)}{3.811} U_t^{-2} + \varepsilon_t \tag{3.4}$$

$$\Delta w_t = -\underset{(0.522)}{0.700} + \underset{(2.681)}{3.960} U_t^{-1} + \underset{(2.683)}{4.437} U_t^{-2} + \underset{(0.077)^{**}}{0.337} \Delta p_t + \varepsilon_t \tag{3.5}$$

where the standard errors are reported in parentheses, and (*) and (**) represent significance levels at 5% and 1% respectively.

1. In addition to the two specifications in Equations 3.2 and 3.3, Lipsey also included the growth rate of the unemployment rate as another explanatory variable due to the qualitative analysis with the unemployment growth rate by Phillips. However, the OLS regressions reject this variable as a significant factor.

2. These two equations in the evaluation are equations (7) and (11) in Lipsey (1960).

The M-S testing results for the three models are summarized on Table 3.1. From the test's *p*-values of Phillips' 'model', it is found that the first six *p*-values (of eight) are well below the 5% threshold. These numbers indicate a failure to reject the first four null hypotheses that are employed to detect assumption departures. Therefore, they indicates that the first four assumptions - (1) Normality, (2) linearity, (3) homoskedasticity, and (4) independence - are not supported. On the other hand, the two bottom t-values representing the (5) t-invariant assumption are large; their corresponding *p*-values exceed the 5% threshold. It means that the fifth assumption is met. In short, the M-S testing result provides that four of the five model assumptions from Phillips (1958)'s fitted curve are not valid. It is hardly a surprise due to Phillips' ad hoc method.

The M-S testing results for Lipsey's two models reveal fewer assumption departures. In both models, only the (4) independence and (5) t-invariant assumptions are found not supported. These departures reflect a lack of dynamic understanding at the time that the concept of time-heterogeneity had not been properly studied yet. In short, both empirical evidence provided by Phillips and Lipsey is built on statistical inadequate statistical models. These misspecifications render the two authors' evidence untrustworthy.

Table 3.1: M-S Testing Results: the Curve of Phillips

		Phillips (1958)	Lipsey (1960)a	Lipsey (1960)b
(1) Normality		**χ(2)=7.60 [0.022]**	χ(2)=3.90 [0.143]	χ(2)=2.71 [0.257]
(2) Linearity		**t(47)=2.61 [0.012]**	t(46)=-0.45 [0.658]	t(44)=-0.94 [0.351]
(3) Homoskedasticity		**t(48)=10.31 [0.000]**	**t(48)=-1.99 [0.052]**	t(47)=-1.44 [0.156]
(4) Independence	R	**t(47)=5.96 [0.000]**	**t(46)=7.15 [0.000]**	**t(44)=4.99 [0.000]**
	S	**t(48)=5.43 [0.000]**	**t(48)=2.03 [0.048]**	t(47)=1.30 [0.199]
(5) T-invariant	R	t(47)=0.02 [0.981]	t(46)=-1.24 [0.222]	t(44)=-0.95 [0.346]
	S	t(48)=0.53 [0.602]	**t(48)=-2.85 [0.006]**	**t(47)=-2.93 [0.005]**

Note: R and S denote the auxiliary regression and skedastic functions, respectively.

3.4 Respecification

To provide trustworthy evidence, empirical modeling in the PR approach consists of three essential stages: specification, M-S testing, and respecification; see Spanos (2010) for a step-by-step example. Essentially, when a model is found to be misspecified, the modeler will proceed to perform respecification. The direction of respecification is based on the statistical information (invalid assumptions) learning from the M-S testing. The process of M-S testing and respecification is repeated until an ultimate statistically adequate model is achieved.

In this part, the respecification for Phillips' and Lipsey's models is performed. Firstly, the rate of changes for wage and price in the three models were computed as the central growth rate ($\Delta w_t = \frac{w_{t+1}-w_{t-1}}{2w_t}$, $\Delta p_t = \frac{p_{t+1}-p_{t-1}}{2p_t}$). From the modern perspective, taking the central average seems to be arbitrary. It is essential to respecify with the original indices.

In Phillips' original correlation, only two variables - unemployment and wage growth - are included. Therefore, the respecification for the model with these two variables first is performed first. Then, a model with the three wage, unemployment, and price variables is attempted.

From the M-S testings for the two Lipsey's models, only the (4) independence assumption is found invalid. It suggests that it is necessary to employ a model with lags, and no further attempt to go beyond a Normality-based model featuring non-linearity and heteroskedasticity/heterogeneity is needed. Thus, the single-equation Dynamic Linear Regression (DLR) model of the PR approach is a well-qualified initial respecified candidate. Appendix A summarizes the set of the five testable statistical assumptions that constitute the DLR model.

The first respecification attempt is performed with the wage annual growth rate, the original unemployment rate, and the price annual growth rate. This choice of variables resembles the modern form of the Phillips curve. However, despite the best effort, no statistically adequate models are achieved. There still exist the departures from the (2) linearity assumption and the (4) second-order t-invariant assumption.

Next, the log form for the three variables is used. The fine-tuned repetition of M-S testing and respecification revolving around the DLR model succeeds in achieving statistical adequacy. The respecified statistical adequate models are DLR(2) with time trends. The M-S testing results for the statistically adequate DLR models are summarized on Table 3.2, showing that none of the assumption departures is found.

Estimation of the first statistically adequate model, which includes only logs of wage and unemployment, yields:

$$\begin{aligned}\ln w_t = &\underset{(0.157)^{**}}{0.565} + \underset{(0.134)^{**}}{1.128}\ln w_{t-1} - \underset{(0.122)^{*}}{0.244}\ln w_{t-2} - \underset{(0.006)^{*}}{0.021}\ln U_t - \underset{(0.008)}{0.001}\ln U_{t-1} + \underset{(0.006)}{0.001}\ln U_{t-2} \\ &+ \underset{(0.007)^{**}}{0.022}\, t + \underset{(0.014)^{**}}{0.059}\, D_{1872} + \varepsilon_t, \qquad (3.6)\end{aligned}$$

where D_{1872} is a dummy variable for the year of 1872. The inclusion of the dummy variable captures an idiosyncratic positive change in the behavior of the wage series.

From the statistically adequate model, what can be learned about the economic (substantive) relation between unemployment and wage? The estimation results show the estimator of U_t and two lags of $\ln w_t$ are significant at 5%. The significance of U_t and its negative sign are sufficient evidence to support Phillips' conclusion that there is an inverse relationship between wage change and unemployment. However, the estimated coefficient of U_t at 0.021 is relatively small to the sum of lags of w_t's at 0.088. It means that 1% increase in the contemporaneous unemployment U_t leads to 2.1% decrease in the wage index (due to log-log form). When the three unemployment variables' estimated coefficients are summed up, the cumulative effect is nearly the same, at -0.021. On the other hand, the two lags of wages (w_{t-1} and w_{t-2}) with the estimated values respectively at 1.173 and -0.287 have much more explanatory power. In the modern macroeconomic language, it may suggest the presence of (wage) price stickiness.

The estimation of the second statistically adequate model, which have all the three variables, produces:

$$\begin{aligned}\ln w_t = &\underset{(0.223)^{*}}{0.525} + \underset{(0.143)^{**}}{1.173}\ln w_{t-1} - \underset{(0.130)^{**}}{0.287}\ln w_{t-2} - \underset{(0.007)^{**}}{0.019}\ln U_t - \underset{(0.008)}{0.003}\ln U_{t-1} + \underset{(0.006)}{0.002}\ln U_{t-2} \\ &+ \underset{(0.085)}{0.066}\ln p_t - \underset{(0.116)}{0.132}\ln p_{t-1} + \underset{(0.084)}{0.072}\ln p_{t-2} + \underset{(0.010)^{*}}{0.023}\, t + \underset{(0.014)^{**}}{0.057}\, D_{1872} + \varepsilon_t \qquad (3.7)\end{aligned}$$

This estimation shows that the effect of price is insignificant; all three p-values of the price index's variables are lower than the 5% threshold. For the wage and the unemployment variables, their estimated coefficients are nearly identical to the previous estimation results that the lags of wage play a significant role, and the effect of U_t is also nearly 2%.

Table 3.2: M-S Testing Results for the Respecified Models: the Curve of Phillips

		Model 1 with $\ln w_t$ and $\ln U_t$	Model 2 with $\ln w_t$, $\ln U_t$ and $\ln p_t$
(1) Normality		$\chi(2)$=0.72 [0.699]	$\chi(2)$=1.26 [0.533]
(2) Linearity		t(39)=-0.91 [0.845]	t(36)=-0.11 [0.911]
(3) Homoskedasticity		t(46)=-0.27 [0.788]	t(46)=-0.26 [0.799]
(4) Markov(p) dependence	R	t(39)=-1.81 [0.077]	t(36)=-1.60 [0.119]
	S	t(46)=0.11 [0.914]	t(46)=0.07 [0.941]
(5) T-invariant	R	t(39)=0.20 [0.845]	t(36)=-1.43 [0.162]
	S	t(46)=-0.81 [0.421]	t(46)=-0.98 [0.334]

Note1: R and S denote the auxiliary regression and skedastic functions, respectively.
Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

3.5 Conclusions

This chapter replicates estimates of the three specifications using Phillips' original 53 data points and evaluates their statistical adequacy. From the M-S testing results, it is found that Phillips' famous equation and Lipsey's follow-up models are misspecified. Then, the respecification successfully produces statistically adequate models with the same data. From the new estimation results, there also exists a negative relationship between unemployment and wage inflation. However, the presence of the temporal dependence, especially the significance of lags of wage change, suggests that the role of lagged variables needs to be taken into account to explain the behavior of wage change. The static form of the original version of the curve provides an insufficient understanding of the phenomena.

Chapter 4

The 1960s Wage-determination Literature

4.1 Introduction

The conventional story tells us that the Phillips curve had its second big moment in 1960 when the two famous economists, Samuelson and Solow, explored the same relationship using US data and used the apparent trade-off between unemployment and inflation to propose a menu of choices for policymakers. To lower the unemployment rate, the government can implement expansionary policies and incur costs at higher rates of inflation. These two studies are supposed to be favorably received by economists and unleashed a large follow-up literature that finer details of the trade-off were explored in the 1960s.

As pointed out by Forder (2014a), however, with the exception of Lipsey (1960), very few papers in the 1960s were directly inspired by Phillips (1958). Moreover, Samuelson and Solow did not take a firm stand on the potential choices between inflation and unemployment. Their paper considers Phillips' very long sample as the most interesting point. However, they neither saw it as a stable relationship nor support inflationism as an appropriate measure to push down unemployment.

On the contrary to the later presumptions, such as Friedman (1977) and Santomero and Seater (1978), that there were many follow-up studies to Phillips' groundbreaking paper, the literature during the 1960s focused primarily on the determinants of wage change, which is only loosely related to the Phillips curve trade-off. Phillips (1958) argues that the behavior of wage - the price of labor in the market - operates according to the supply and demand framework. A high demand of labor in the market, which leads to a lower unemployment rate, causes a higher wage growth rate. Similarly, changes in wages get smaller when the demand for labor is low, and the unemployment rate is accordingly high. However, as Forder (2014a, 2014b) explained, the existing econometric literature approaches the question of the determinants of wage change in a different way. The dominant view of the beginning of the 1960s considers wage to be set via collective bargaining between employees and employers. Other factors are contributing to the bargaining process. It results in a long list of potential explanatory variables in regression equations which range from the (currently) familiar variables (e.g., unemployment or price change) to the specific factors of bargaining considerations (such as profit rates or productivity growth). This related econometric literature to Phillips' original estimated equation can be called the wage-determination literature.

Because the wage-determination literature in the 1960s, despite its non-genuine tie, is still an important chapter of the Phillip curve history, empirical evidence learning from this literature must also be evaluated. The choices of important studies are guided by a survey by

Qin (2011). Qin documented the development of the modern econometric through a case study of the Phillips curve. The three papers Klein and Ball (1959), Dicks-Mireaux and Dow (1959), and Sargan (1964) were chosen by Qin to represent important advances in the econometric technique of the Phillips curve literature in the 1960s.

The three papers obtained from Qin (2011) only use the UK data. Perry (1964), which is an important study at the time with the US data, is also considered to balance with the literature on the other side of the Atlantic.

In the next section, these papers will be briefly discussed. It is noted that the authors of the four studies experimented with a considerable number of specifications to ensure the robustness of their estimation results. In order to keep the evaluation results concise and representative, one exemplary model in each paper will be picked and introduced. Then, Section 4.3 will replicate the chosen models and perform the M-S testing.

4.2 Brief Evaluation of Different Models

4.2.1 Klein and Ball (1959)

In 1961, the future Nobel winner Lawrence Klein published a large-scale macroeconometric model including more than 30 estimation equations for the UK. This study is a typical empirical study under the Cowles Commission methodology. As a part of a large project, a preliminary small-scale result with four simultaneous equations was published in Klein and Ball (1959). In one of these four equations, unemployment, domestic price change, and quarter dummy variables are put together to explain the behavior of nominal wage change. The system of equations with endogenous and exogenous variables is estimated using the Limited Information Maximum Likelihood (LIMI) method. For comparison, Klein and Ball also made a single OLS estimation for wage change as:

$$\Delta w_t = \alpha_0 + \alpha_1 U_t^a + \alpha_2 \Delta p_t + \alpha_3 F_{1952} + \sum_{i=1}^{3} \beta_i Q_i + u_t, \tag{4.1}$$

where $\Delta w_t = w_t - w_{t-4}$, $U_t^a = \frac{U_t + U_{t-1} + U_{t-2} + U_{t-3}}{4}$, $\Delta p_t = \frac{p_t - p_{t-4} + p_{t-1} - p_{t-5} + p_{t-2} - p_{t-6} + p_{t-4} - p_{t-7}}{4}$, F_{1952} as a before/after-1952 dummy variable and Q_i as three quarterly dummies.

This model resembles the choices of variables from the original Phillips curve. However, Klein and Ball only mentioned Phillips' paper on a footnote and emphasized union aggressions in their discussion. In presenting the estimation results, Klein and Ball provided the Durbin-Watson test and the von Neumann ratio test to check for the problem of residual autocorrelation.

4.2.2 Dicks-Mireaux and Dow (1959)

In the same year of Klein and Ball's paper, Dicks-Mireaux and Dow (1959) attempted to explain the annual wage growth by the annual price inflation rate and their index of excess labor demand. This excess labor index was proposed to be equivalent to unemployment as a variable capturing the labor market's slack. Dicks-Mireaux and Dow's basic model with quarterly UK data is:

$$\ln(w_t/w_{t-4}) = \alpha_0 + \alpha_1 \ln(p_t/p_{t-4}) + \alpha_2 \log(d_t) + u_t, \tag{4.2}$$

where d_t is the index of excess labor demand which the authors previous estimated in Dow and Dicks-Mireaux (1958).

Detecting the presence of autocorrelation from the Durbin-Watson test, Dicks-Mireaux and Dow put much effort into dealing with the problem. They first tried to transform their variables by lag operators $(1-\frac{3}{4}L(.))$ and $(1-\frac{7}{8}L(.))$ and experimented with various lag lengths of their explanatory variables. Their adoption of the log-linear form was innovative at their time. The basic specification in Equation 4.2 is selected to be evaluated.

4.2.3 Sargan (1964)

Significant econometric developments culminated with Sargan (1964), an London School of Economics (LSE) econometrician, who revolutionized time series modeling by proposing a different perspective on empirical modeling that become known as the "LSE tradition"; see Hendry (2003) and Spanos (2014). In addition to proposing a different perspective on 'residual autocorrelation' as 'unmodeled dynamics', he put forward several novel techniques and procedures for testing the validity of the probabilistic assumptions, as well as comparing different specifications for model selection purposes. In modeling the wage-change equation, Sargan performed a specification search which consists of trying different lag lengths, adding time trends, and experimenting with the log-linear functional form. His search concluded that the "most satisfactory" specification is:

$$\Delta \ln w_t = \alpha_1 + \alpha_2 \ln U_{t-4} + \alpha_3 \ln(w_{t-1}/p_{t-1}) + \alpha_4 \Delta \ln w_{t-1} + \alpha_5 \Delta \ln w_{t-2} + \alpha_6 t + u_t \tag{4.3}$$

This specification has been very influential in the LSE tradition that has influenced the time series literature in the 1980s and 1990s greatly due to the introduction of several novel features, in addition to modeling temporal dependence using lags and the heterogeneity using trend terms. For instance, the term $\ln(w_{t-1}/p_{t-1})$ became known as the error-correction term that played an important role in relating the estimated equation to the long-run theoretical relationship; see Hendry (2003). Because of its impact on the subsequent literature, the statistical adequacy of Equation 4.3 is evaluated.

4.2.4 Perry (1964)

In addition to the empirical studies with UK data, Perry (1964) used US data to test his wage-determination model. Citing all keys papers including Dicks-Mireaux and Dow (1959), Lipsey (1960), and Klein and Ball (1959), he chose a nonlinear equation with the inverse of unemployment U_t on the RHS and added cost of living and profit rate (R_t) as explanatory variables. To refine the estimation results, he experimented with various specifications in which contemporaneous, lagged and different terms of the RHS variables were mixed. Among them, the following specification with one-period lag for Δp_t and R_t is chosen to evaluate:

$$\Delta w_t = \alpha_1 + \alpha_2 U_t^{-1} + \alpha_3 \Delta p_{t-1} + \alpha_4 R_{t-1} + u_t, \tag{4.4}$$

where $\Delta w_t = \frac{w_t - w_{t-4}}{w_{t-4}}$, $\Delta p_t = \frac{p_t - p_{t-1}}{p_{t-1}} + \frac{p_{t-1} - p_{t-2}}{p_{t-2}} + \frac{p_{t-2} - p_{t-3}}{p_{t-3}} + \frac{p_{t-3} - p_{t-4}}{p_{t-4}}$, and R_t is the profit rate measured by "net profits after taxes as a percentage of stockholder's equality".

4.3 M-S Testing Results

The original dataset is used to replicate the four chosen models in the four papers. Table 4.1 shows the summary of the numbers of observations, the exact positions of each model in the original papers, and the included variables. Since these models are replicated by the OLS estimation, the same M-S testing as in the previous chapter will be performed.

Table 4.1: Summary of the Models in Evaluation

Paper	Replicated model	Data	Variables
Klein and Ball (1959)	Equation 1e'	Quarterly UK, 1947 - 1958 with 36 observations.	Wage index, unemployment index, consumer price index and dummies for policy change (after/before 1952) and three quarters.
Dicks-Mireaux and Dow (1959)	Second model in Table 4	Quarterly UK, 1947 - 1958 with 43 observations.	Price index and excess labor demand index.
Sargan (1964)	Equation 18	Quarterly UK, 1947 - 1961 with 51 observations. An extension of Klein and Ball (1959)'s sample.	Wage index, unemployment index and consumer price index.
Perry (1964)	Equation 3.6	Quarterly US, 1948 - 1960 with 50 observations.	Hourly earnings, unemployment rate, consumer price index and profit rate.

The summary of the M-S testing results is provided on Table 4.2. All four models are found statistically inadequate, with at least one assumption departure existing in each model. Among the four, Sargan's model is nearly adequate when only the second assumption relating to the functional form is invalid. This result exemplifies Sargan's diligent effort to account for the temporal dependence and heterogeneity in the data directly and not model the error term, as the traditional econometric literature often does to this day. For the three other models, there is a similar departure of the (4) Markov(p) dependence assumption. It indicates that Dicks-Mireaux and Dow and Perry's ad hoc specifications of the lag structure could not fully account for the temporal dependence existing in the data.

4.4 Respecification

Because all of the models in the evaluation are statistically inadequate, this part proceeds to respecify them. The first three papers use the UK data, and only Perry (1964) is a study with American time series. Therefore, two respecifications are performed: one using the UK data and another with the US data. The UK dataset is a copy of Sargan (1964)'s, which has the most extended sample.

Table 4.2: M-S Testing Results: the Wage-determination Literature

		Klein and Ball (1959)	Dicks-Mireaux and Dow (1959)
(1) Normality		$\chi(2)$=0.56 [0.756]	$\chi(2)$= 4.01 [0.1349]
(2) Linearity		**t(25)=-2.76 [0.011]**	t(40)=1.92 [0.063]
(3) Homoskedasticity		t(31)= 0.81 [0.424]	t(42)=-0.87 [0.389]
(4) Markov(p) dependence	R	**t(25)=2.68 [0.013]**	**t(40)=7.45 [0.000]**
	S	t(31)=0.80 [0.429]	**t(42)=5.21 [0.000]**
(5) T-invariant	R	**t(25)=3.12 [0.005]**	t(40)= -0.28 [0.788]
	S	**t(31)=-2.48 [0.019]**	t(42)=0.57 [0.570]
		Sargan (1964)	Perry (1964)
(1) Normality		$\chi(2)$= 0.77 [0.680]	$\chi(2)$=0.82 [0.664]
(2) Linearity		**t(45)=2.35 [0.023]**	t(42)=-0.94 [0.352]
(3) Homoskedasticity		t(50)=1.82 [0.076]	t(45)=-1.50 [0.141]
(4) Markov(p) dependence	R	t(45)=1.04 [0.306]	**t(42)=3.69 [0.001]**
	S	t(50)=-0.43 [0.667]	t(45)=1.65 [0.106]
(5) T-invariant	R	t(45)=-0.82 [0.415]	t(42)=-0.24 [0.811]
	S	t(50)=-0.60 [0.549]	t(45)=-1.70 [0.096]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.
Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

4.4.1 UK Dataset

In respecification with the UK data, only the three essential variables of the Phillips curve are used: wage, unemployment, and price. These variables are then put under the same log form. It is due to two reasons. Firstly, the log form resembles Sargan's log-linearization that nearly achieves statistical adequacy. Secondly, the unemployment variable is provided as an index of unemployment (out of 100). It is easier to interpret the estimation in the log-log form. As in the previous chapter, two models are respecified: the first one with only wage and unemployment, and the second including the price index. The targeted respecified model is also the DLR model. The respecification uses the most extended sample from Sargan's paper.

The respecification successfully produces two statistically adequate models. The M-S testing results are summarized in Table 4.3. Estimation of the first model (without the price variable) yields:

$$\begin{aligned}\ln w_t = &\underset{(0.246)^{**}}{0.737} + \underset{(0.134)^{**}}{1.489}\ln w_{t-1} - \underset{(0.212)^{**}}{1.136}\ln w_{t-2} + \underset{(0.209)^{**}}{0.968}\ln w_{t-3} - \underset{(0.130)^{**}}{0.455}\ln w_{t-4}\\ &- \underset{(0.008)^{*}}{0.018}\ln U_t + \underset{(0.013)^{**}}{0.035}\ln U_{t-1} - \underset{(0.013)^{**}}{0.047}\ln U_{t-2} + \underset{(0.013)^{*}}{0.030}\ln U_{t-3} - \underset{(0.008)^{*}}{0.017}\ln U_{t-4} + \underset{(0.018)^{*}}{0.045}\,t + \varepsilon_t\end{aligned} \tag{4.5}$$

This model is a DLR(4) model with a linear trend. It is not a surprise to use up to the fourth lag because Sargan's specification already involves the third lag of w_t. The estimation results show all explanatory variables (level, lags, and trend) are significant at the 5% level. The original Phillips' hypothesis about the inverse relationship between unemployment and wage is supported. The sum of estimated coefficients of wage and unemployment are respectively 0.865 and -0.017. These numbers suggest the dominant role of the wage 'stickiness' that the original hypothesis does not take into account.

The estimation of the second statistically adequate model including the price index is:

$$\begin{aligned}\ln w_t =& \underset{(0.438)}{0.244} + \underset{(0.150)^{**}}{0.691} \ln w_{t-1} - \underset{(0.161)^{**}}{0.711} \ln w_{t-2} + \underset{(0.128)^{**}}{0.365} \ln w_{t-3} - \underset{(0.008)^{*}}{0.019} \ln U_t + \underset{(0.011)}{0.011} \ln U_{t-1} \\ & - \underset{(0.011)^{*}}{0.023} \ln U_{t-2} + \underset{(0.008)}{0.002} \ln U_{t-3} + \underset{(0.099)^{*}}{0.240} \ln p_t + \underset{(0.132)}{0.122} \ln p_{t-1} + \underset{(0.130)}{0.240} \ln p_{t-2} + \underset{(0.125)}{0.146} \ln p_{t-3} \\ & + \underset{(0.028)}{0.046} t + \underset{(0.005)^{**}}{0.018} t^2 - \underset{(0.003)^{*}}{0.007} t^3 + \varepsilon_t \qquad (4.6)\end{aligned}$$

This model is a DLR(3) model with third-order trends. The inclusion of the price variables slightly changes the significance of the variables of wage and unemployment. Compared to the first respecified model without price, the two lags of U_t (U_{t-1} and U_{t-2}) in the second are no longer significant at 5% level. Only the current rate is significant for the price variables, and the three lags are not. However, the inclusion of the price variables changes the relative contribution of the three factors to the formation of the current wage change. The sum of estimated coefficients for the wage, unemployment, and price variables are respectively 0.344, -0.029, and 0.748. The current price change and its lags become the most significant factor; lags of wage change play a much less role (0.344 against 0.865 in the first model). However, the insignificances of the three lags of the price variable ($\ln p_{t-1}, \ln p_{t-2}, \ln p_{t-3}$) make the interpretation of the effect of price change not clear-cut.

In short, the statistically adequate models obtaining from the UK dataset confirm the original Phillips' hypothesis that there is an inverse statistical relationship between wage growth and unemployment. However, it also reveals a considerable effect of the lags of wage growth and smaller influence from the price factor.

Table 4.3: M-S Testing Results for the Respecified Models with the UK Dataset: the Wage-determination Literature

		DLR(4) model with w_t and U_t	DLR(3) model with w_t, U_t and p_t
(1) Normality		$\chi(2)$=2.81 [0.245]	$\chi(2)$=1.01 [0.604]
(2) Linearity		t(36)=-0.91 [0.371]	t(35)=1.54 [0.132]
(3) Homoskedasticity		t(46)=0.29 [0.776]	t(47)=0.01 [0.990]
(4) Markov(p) dependence	R	t(36)=-1.18 [0.245]	t(35)=1.64 [0.327]
	S	t(46)=-1.23 [0.225]	t(47)=-0.30 [0.764]
(5) T-invariant	R	t(36)=0.43 [0.667]	t(35)=-0.99 [0.327]
	S	t(36)=-0.34 [0.738]	t(47)= -0.06 [0.953]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.

Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

4.4.2 US Postwar Dataset

In respecification with Perry's US dataset, the three wage, unemployment, and price variables are also kept. Because Perry did not provide the raw wage and price indices, the provided growth rates for these two variables are for respecify. The wage variable is the annual wage growth rate, and the price variable the sum of the one-quarter changes in the price index (approximately equal to the annual inflation rate). The inverse of the unemployment rate U_t^{-1} originally used by Perry is switched to the level rate U_t. Fine-tuned respecification also achieves statistical adequacy, with Table 4.4 summarizing the M-S testing results.

Estimation of the first model with the wage and the unemployment variables yields:

$$\begin{aligned}\Delta w_t = &\underset{(1.071)^{**}}{3.490} + \underset{(0.135)^{**}}{0.832}\Delta w_{t-1} - \underset{(0.167)}{0.105}\Delta w_{t-2} + \underset{(0.166)}{0.184}\Delta w_{t-3} - \underset{(0.158)^{**}}{0.799}\Delta w_{t-4} + \underset{(0.112)^{**}}{0.414}\Delta w_{t-4} \\ &- \underset{(0.695)^{**}}{1.982}U_t + \underset{(1.618)}{2.363}U_{t-1} + \underset{(1.691)}{0.163}U_{t-2} - \underset{(1.697)}{1.650}U_{t-3} + \underset{(1.653)}{0.490}U_{t-4} + \underset{(0.722)}{0.334}U_{t-5} + \varepsilon_t\end{aligned} \tag{4.7}$$

The first statistically adequate model is a DLR(5) model. Of the RHS variables, only three lags of wage growth (Δw_{t-1}, Δw_{t-2} and Δw_{t-3}) and the current unemployment rate U_t are significant at the 5% level. The sum of estimated coefficients of wage growth and unemployment are respectively 0.525 and -0.283. It means that the current wage growth rate is strongly influenced by its lags and changes in unemployment.

When the price variables are included, estimation of the respecified statistically adequate model yields:

$$\begin{aligned}\Delta w_t = &\underset{(1.980)^{**}}{13.408} + \underset{(0.147)}{0.284}\Delta w_{t-1} - \underset{(0.163)}{0.111}\Delta w_{t-2} + \underset{(0.152)}{0.071}\Delta w_{t-3} - \underset{(0.128)^{**}}{0.608}\Delta w_{t-4} \\ &- \underset{(0.642)^{**}}{2.296}U_t + \underset{(1.311)}{1.715}U_{t-1} + \underset{(1.431)}{0.320}U_{t-2} - \underset{(1.329)}{1.339}U_{t-3} + \underset{(0.633)}{0.018}U_{t-4} \\ &+ \underset{(0.111)}{0.190}\Delta p_t + \underset{(0.156)}{0.068}\Delta p_{t-1} + \underset{(0.144)}{0.175}\Delta p_{t-2} - \underset{(0.144)}{0.129}\Delta p_{t-3} + \underset{(0.100)}{0.125}\Delta p_{t-4} \\ &- \underset{(0.239)^{**}}{0.910}\,t + \underset{(0.308)^{**}}{1.356}\,t^2 - \underset{(0.185)^{**}}{0.613}\,t^3 + \varepsilon_t\end{aligned} \tag{4.8}$$

This is a DLR(4) model with the third-order of trends. The estimation results show that all price variables are not significant at the 5% threshold.

4.5 Conclusions

This chapter evaluates the statistical adequacy of a few selective models representing the wage-determination literature in the 1960s and finds them statistically inadequate, indicating the evidence obtained from the four papers is all untrustworthy. Except for Sargan's model, the inchoate attempts from the researchers at that time to incorporate the dynamic nature of the data do not provide a sufficient characterization. Then, the respecification reveals that the lags of the regressor (wage change) must be included to achieve statistical adequacy.

Table 4.4: M-S Testing Results for the Respecified Models with the US dataset: the Wage-determination Literature

		DLR(5) model with w_t and U_t	DLR(4) model with w_t, U_t and p_t
(1) Normality		$\chi(2)$=2.12 [0.346]	$\chi(2)$=0.68 [0.711]
(2) Linearity		t(30)=0.52 [0.609]	t(25)=-1.11 [0.278]
(3) Homoskedasticity		t(41)=-0.70 [0.490]	t(42)=0.05 [0.959]
(4) Markov(p) dependence	R	t(30)=-0.05 [0.961]	t(25)=0.32 [0.754]
	S	t(41)=-0.67 [0.506]	t(42)=0.75 [0.455]
(5) T-invariant	R	t(30)=-0.41 [0.684]	t(25)=-1.52 [0.141]
	S	t(41)=0.42 [0.676]	t(42)=0.64 [0.526]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.
Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

Chapter 5

The Expectations-augmented Phillips Curve

5.1 The Phillips Curve in the 1970s

5.1.1 Introduction

The late 1960s and the 1970s were a period of great events in the conventional Phillips curve history. The history of this period often starts with the new insights provided by Milton Friedman's famous presidential address (Friedman 1968) and a theoretical paper by Edmund Phelps (Phelps 1967). Different from the original curve, when the effect of inflation expectation is taken into account, the unemployment rate is revealed to be only responsive to changes in the inflation rate only in the short run. In the long run, the unemployment rate becomes neutral to any inflationary pressures from fiscal or monetary policies. The newly discovered short-run and the long-run relations can be combined to the expectations-augmented Phillips curve (also called the accelerationist Phillips curve). There was a debate on the 'new' curve versus the 'old' curve. This debate was called the Natural Rate Hypothesis (NRH) in the literature. The stagflation of the early 1970s is widely considered as a confirmation of the NRH, rendering the original Phillips curve as a naive and defunct empirical relationship.

Forder (2014a, 2014b) provide an in-depth account of the literature on the Phillips curve in the 1970s, making a strong case that (a) the label 'Phillips curve' carried a multiplicity of meanings at that time, (b) there was no unified understanding of the 'naive', the 'short-run', and 'long-run' versions of the curve, and (c) different authors utilized the same label for various claims. Phelps (1969) presumed the existence of the short-run trade-off in need of further explanation. Later, the curve became the point of departure for other theoretical developments, such as Kydland and Prescott (1977)'s time inconsistency. In the econometric literature, the wage-unemployment-price Phillips curve equation provided the basis for several empirical studies, including evaluating the NRH. Also, shifts of the Phillip curve were used as an oversimplified description for the 1970s' stagflation that lingers on to this day in macroeconomics textbooks.

Because the Phillips curve literature in the 1970s is huge and diverse, it is reasonable to limit the evaluation to a smaller group of empirical studies which revolves around the wage-unemployment-price equation. This focus retains the continuation to the wage-determination literature that the same wage-unemployment-price equation was modified and reinterpreted to answer newly emerging questions. Moreover, the modern mainstream expectations-augmented Phillips curve framework is a direct descendant of the developments in the literature in the 1970s that will be discussed later.

At the end of the 1960s, the older wage-unemployment-price equation of the wage-determination period was modified to incorporate the expectation concept. In particular, the current price change was replaced by the inflation expectation, and the basic structural equation is under the form as:

$$\pi_t^w = \beta_0 + \boldsymbol{\beta}_1' \mathbf{U}_t + \gamma E(\pi_t^p) + \boldsymbol{\alpha}' \mathbf{z}_t + \varepsilon_t, \tag{5.1}$$

where π_t^w is the wage growth rate, $\mathbf{U}_t$ is a vector of various measurements of unemployment, $E(\pi_t^p)$ is a single-variable approximation of the expected inflation, and $\mathbf{z}_t$ include other explanatory variables.

The NRH hypothesis was tested by looking at the estimated value of the coefficient γ of the expected inflation $E(\pi_t^p)$. When γ equals one, any changes in the expected inflation rate fully affect the wage growth rate. It causes an endless spiral cycle of the price-wage hike. In other words, no permanent trade-off between inflation and unemployment is proven empirically. The NRH is confirmed. On the other hand, in the case of γ being less than one, inflation expectations partially reflect back to wage change. After a while, the feedback of price change on wage change dies out; and stable growth rates of wage and price are established. It means that a certain level of permanent trade-off exists. The NRH is then rejected.

Although the 1970s is a defining historical period of the Phillips curve, there lack of modern surveys devoted to providing an extensive and focused examination of econometric studies using the wage-unemployment-price equation and how the NRH was empirically rejected. Theoretical developments are the main focus of the historical explorations. For example, Snowdon and Vane (2005), one of the most scrupulous accounts of macroeconomics history, only cite an old survey paper (Santomero and Seater 1978) as a review of the empirical literature without further elaboration. Recently, Robert Gordon, one of the main protagonists of the Phillips-curve empirical literature, in a paper on the history of the Phillips curve (Gordon 2011) only writes a short paragraph and a footnote on the empirical rejection of the NRH. Fuhrer et al. (2009) as a collection of papers that aim to "provides an intellectual history of the Phillips curve" completely ignore the literature on empirical testing of the Phillips curve in the 1970s.

For the purpose of a statistical evaluation, a short survey of important papers in the 1970s using the wage-unemployment-price equation will be performed in the next section. The aim of the survey is not to provide a comprehensive review of the literature but learn key features of econometric techniques. The review can help to bridge the gap in the history of the Phillips curve literature that was mentioned above. The sole criterion of picking papers is the number of citations.

5.1.2 Important Empirical Phillips-curve Studies in the 1970s

In the 1970s, there were two comprehensive surveys (Goldstein 1972; Santomero and Seater 1978) reviewing the econometric Phillips curve literature in the 1970s[1]. However, these two papers only cover the literature until 1977. The literature review in this part was extended to the beginning of the 1980s to make sure the final list is exhausted. It was found that starting from the beginning of the 1980s, the debate on the NRH was no longer the primary concern of economists. The adverse impacts of the Volcker disinflation policy drove researchers

1. Forder (2014a) also explored a long list of econometric studies during that period. However, as a book on the history of economic thoughts, Forder focuses on challenging the conventional story rather than reviewing the literature.

to use the expectations-augmented Phillips curve equation to understand the newly emerging phenomena. Since 1983, the literature has stopped producing any NRH-testing paper.

Table 5.1 provides the list of the important papers. The list includes eight papers that use US data and have at least 100 citations chronologically. The estimated values of the expected inflation's coefficients and the conclusion on the NRH of each paper are also presented in Table 5.1.

Table 5.1: List of Important Papers

Paper	Citations	Expected inflation coeff.	Conclusion on the NRH
Solow (1969)	224	0.37-0.55	Reject
Gordon (1970)	154	0.45	Reject
Perry (1970)	390	0.33-0.44	Reject
Gordon (1971)	245	0.35-0.60	Reject
Eckstein and Brinner (1972)	108	0.42-0.77	The coefficient of the expected inflation is unity when inflation rate exceeds 2.5 percent.
Gordon (1972)	183	All less than 1	The coefficient of the expected inflation is unity when inflation rate exceeds 7.0 percent.
Turnovsky and Wachter (1972)	159	0.43-0.49	Reject
Gordon (1977a)	170	Approximately 1	Confirm

Before exploring each paper in the list, a summary of a few common features among the eight papers is introduced first. Firstly, the single-equation expectations-augmented Phillips curve in Equation 5.1 is the principal structural equation of the regression analyses. Different authors provided their own specification modifications, such as different measurements of the unemployment rate or a distinct method to better approximate the expected inflation rate. Secondly, OLS is the predominant estimation method. In a few cases, a price equation is included to form a system of equations that is estimated by the two-stage least square method. Also, there was no consideration of the statistical features of the involved variables, specifically time series complications. The approaches were typically theory-driven.

In exploring each individual paper, it is needed to note that half of the eight papers in the list were contributions from the economist Robert Gordon. It highlights the predominant role of Gordon's studies in this period; see Goutsmedt and Rubin (2018) for an elaboration of Gordon's works in the 1970s. Thus, this section's discussion will explore the four papers from other authors first and then move to Gordon's studies. Table 5.2 provides a summary of the specifications in these papers.

In the list, the well-known Solow (1969) is the earliest direct response to Friedman's assertions in 1967. In this paper, Solow uses an equation with the price index on the left-hand side (LHS) to test the NRH. On the RHS, nominal wage change, output per manhour, and farm price change are used to capture the cost factors. An index for capability utilization is employed to capture the demand pressure. Solow approximates inflation expectation as a weighted average between the closest lagged expected rate and the actual rate. Using the US quarterly data from 1948 to 1966, the estimated coefficients of the expected inflation rate range from 0.37 to 0.55, which is well below unity. Solow (1969) rejects the NRH.

The next paper, Perry (1970), inherits much influence of the preceding wage-determination literature. Perry argues that the official unemployment rate is not a satisfying measurement of the labor market's tightness. To better track its structural changes, he develops two derivative indices: weighted unemployment rate and unemployment dispersion (these new two measurements were widely adopted in later important papers). Apart from the unemployment adjustments, social insurance contribution and a proxy for secondary employment changes were also included as controlling variables. To approximate the expected inflation, Perry only uses a simple one-period lag of the current rate. Perry's models can be generally written down as:

$$\pi_t^w = \beta + \boldsymbol{\beta}_1' \mathbf{U}_t + \gamma \pi_{t-1}^p + \boldsymbol{\alpha}' \mathbf{z}_t + \varepsilon_t, \tag{5.2}$$

where $\mathbf{U}_t$ is a vector of the unemployment indices, and $\mathbf{z}_t$ is a vector of controlling variables. Using the US data, Perry's estimation results reject the NRH because the estimated coefficients of the expected inflation fall on the 0.3-0.4 range.

Eckstein and Brinner (1972) is a study on inflation with the purpose to advise a congressional committee. In this paper, wage and price growth are put in the LHS variables to understand their behaviors. The wage equations, which include various measurements of labor market tightness in the RHS, are almost identical to the ones of Perry (1970). However, Ekstein and Brinner incorporate a new small but crucial twist in their specification. An index of inflationary severity, which equals the difference between the eight-quarter inflation rate and the 5 percent threshold, is added as an explanatory variable. From the estimation results, the significance of this new variable allows the authors to learn that the coefficient of the expected inflation reaches one when the annual inflation rate goes beyond a 2.5 percent threshold. It is evidence supporting the NRH.

In Turnovsky and Wachter (1972), the authors use various price expectation formations (including extrapolative, adaptive, and general lags) to test the NRH. The paper's most important contribution comes from the usage of a direct measurement of inflation expectation. This measurement was obtained from the Livingston survey, which is a biannual publication of predictions of important economic indicators sent by Joseph A. Livingston, the financial editor of the Philadelphia Bulletin, to many economists. The estimation results from this paper reject the NRH.

The remaining four papers in the list all belong to Robert Gordon. The first paper (Gordon 1970) estimates price and wage equations with the purpose of forecasting future inflation. Comparing the other studies, Gordon, however, made a couple of small twists in his paper. On the LHS, the simple wage growth rate is adjusted to become the standard unit labor cost which is computed by dividing the compensation per hour w_t by an estimation of productivity growth. On the RHS, unemployment U_t is measured by the level of the total employment rate of manhours which is claimed to capture tightness in the labor market better. Inflation expectation $E(p_{t-1})$ is calculated by a sophisticated method. In other important papers, the simplest way (as in Perry 1970) is to use one lag of the observed inflation rate to approximate the expectation. A more complicated way, such as in Solow (1969), captures the expected rate as a weighted average of many lags. Gordon proposes that the weights of the lags can be computed more precisely by an auxiliary regression. It is to regress the Treasury bond's interest rates on lagged inflation rates and the controlling variables. The obtained estimated coefficients of the auxiliary regression are then utilized as the new weights. The expected inflation rate is finally computed as:

$$E_t(\pi_t^p) = \sum_{i=1}^{8} w_i \pi_{t-i}^p, \tag{5.3}$$

where w_i is the newly obtained weights from the auxiliary regression.

In the main regression, the set of the controlling variables $\mathbf{X}_t$ includes growth rate in output, the growth rate of employer's tax rate, a dummy variable for the 1962-1966 period in which the Council of Economic Advisers recommended a voluntary wage-price control (called as the Guidepost dummy). Using the 1951 - 1969 sample, the estimated coefficient of the expected inflation was 0.453. This number, which is well below unity, indisputably rejects the NRH.

Gordon (1971) is an extended version of the paper in the previous year. Instead of using one variable to capture the slackness in the labor market, Gordon incorporates all measurements of unemployment from Perry (1970) and tries with various combinations among them. The same method of computing improved weights is kept to approximate the expected inflation. Two other simpler methods of expectation approximation - one-period lag and the Livingston survey - are also experimented with.

The regressor (wage inflation) is further modified to match Gordon's ideal thinking. It is an adjustment to neutralize changes in the industry mix at the disaggregate level and compensate for earnings outside the official salaries. Besides, the difference between nonfarm deflator and personal consumption expenditure deflator is included in the controlling set. This variable is supposed to capture firms' behaviors in the scenario that the union's bargaining process puts pressure on firms when there is excess demand on the market. Control for tax change is also adjusted by adding both measurements of employee and employer's tax rates. From the 1954-1970 sample, the estimated coefficients of expected inflation range from 0.35 to 0.60. The NRH is rejected by Gordon (1971) as well.

As a follow-up effort, Gordon (1972) replicates and compares the estimation results between his own study (Gordon 1971) and the other two important papers (Perry 1970; Eckstein and Brinner 1972). The comparison criteria consist of the good of fitness, stability of estimated coefficients, and in-sample simulations. In the conclusion of his paper, Gordon partially accepts the NRH:

> *The only major conclusion of the 1971 paper that appears questionable is the assumption of a fixed coefficient on expected inflation in the wage equation. An alternative equation is specified in which this coefficient is estimated to be a linear function of expected inflation and eventually to reach unity when the inflation rate reaches 7 percent.* (Gordon 1972, page 416)

The final important paper - Gordon (1977a) - marks a transition to the modern form of the expectations-augmented Phillips curve. Firstly, this paper abandons the computation of the expected inflation as a weighted sum of lags and switches to a set of direct lags. Gordon also experiments with constraining these lags to unity as modern studies do. Secondly, the supply (cost-push) factors are brought forward to be accounted for. To consider the effect of the oil price hike, Gordon employs his own "non-food net of energy" price index (known today as the core price indices) as the replacement of the general index in his regressions. Thirdly, the adjusted measurements of the unemployment rate from Perry (1970) are also disposed of; and Gordon experiments with a single-variable output gap as the proxy for market pressure. In the same year, Gordon (1977b) also introduces the triangle model - in which price inertia (inflation expectation), a demand-pull pressure (e.g., unemployment, output gap), and cost-push factors (e.g., oil or import price) are the main determinants of the current inflation rate. Since the 1980s, these variables have become the standard set of explanatory variables in the Phillips curve literature. The transition was completed in Gordon (1981) in which he contends that the wage equation performs inferiorly to the price equation. Since then, price change has become the principal dependent variable in the Phillips curve literature as:

$$\pi_t^p = \beta(U_t - U_t^N) + \gamma E(\pi_t^p) + \alpha^\top \mathbf{z}_t^s + \varepsilon_t, \tag{5.4}$$

where π_t^p is the price inflation rate, U_t^N is the natural rate of unemployment (mostly used as the non-accelerating inflation rate of unemployment (NAIRU)), and $\mathbf{z}_t^s$ consists of supply factors.[2]

Compared to Gordon's previous studies, this paper made the opposite conclusion to the NRH. Firstly, Gordon adjusts his data and rerun one specification in Gordon (1971) with the same 1954-1970 period. Surprisingly, the newly estimated coefficient of the expected inflation is approximately one, leading to a remark:

> *Ironically, the 'natural rate hypothesis', in the form of a coefficient of unity on price inflation, is vindicated by the revisions in the official data.* (Gordon 1977a, page 265)

Secondly, the estimation results with extended sample size to 1976 and the mentioned changes in the estimated specifications all produce that the estimators of the expected inflation is approximately one. These results show support for the NRH.

2. Equation 5.4 has also been the main equation to estimate NAIRU; see the symposium on the natural rate of unemployment in the Journal of Economic Perspectives, Winter 1997, and Ball and Mankiw (2002).

Table 5.2: Main Variables and Estimated Specifications of the Important Papers

Paper	Estimated equation	Specification
Solow (1969)	$\pi_t^p = \beta_0 + \boldsymbol{\beta}_1'\mathbf{X}_t + \gamma E(\pi_t^p) + u_t$	$E(\pi_t^p)$ is a weighted average between one lag of the expected rate and the actual rate. $\mathbf{X}_t$ includes: unit labor cost (two variables), change in farm price, index for capability utilization (demand side) and two dummies for the Korean war and the Guide Post period.
Perry (1970)	$\pi_t^w = \beta_0 + \boldsymbol{\beta}_1'\mathbf{U}_t + \gamma\pi_{t-1}^p + \boldsymbol{\alpha}'\mathbf{z}_t + u_t$	$\mathbf{U}_t$ includes a weighted unemployment rate, a unemployment dispersion measurement and a hidden unemployment rate. The controlling set $\mathbf{z}_t$ consists of the guideposts dummy, one social insurance contribution, and a secondary unemployment variable.
Gordon (1970)	$\pi_t^w = \beta_0 + \beta_1 U_t + \gamma E(\pi_t^p) + \alpha z_t + u_t$	π_t^w is adjusted to become the standard unit labor cost (wage growth divided by productivity growth). U_t is the total employment rate of manhours. $E(\pi_t^p)$ is a weighted average of recent price changes. z_t is a change in social security tax rate.
Gordon (1971)	$\pi_t^w = \beta_0 + \boldsymbol{\beta}_1'\mathbf{U}_t + \gamma E(\pi_t^p) + \boldsymbol{\alpha}'\mathbf{z}_t + u_t$	The set $\mathbf{U}_t$ is adopted from Perry (1970). w_t is further adjusted. Large set of controlling variables $\mathbf{z}_t$ is added. $E(\pi_t^p)$ is the same as in Gordon (1970). Various measurements of the price index are used as a robustness check.
Eckstein and Brinner (1972)	$\pi_t^w = \beta_0 + \boldsymbol{\beta}_1'\mathbf{U}_t + \gamma_0 E(\pi_t^p) + \gamma_1 JP_t + \boldsymbol{\alpha}'\mathbf{z}_t + u_t$	The set $\mathbf{U}_t$ is adopted from Perry (1970). $E(\pi_t^p)$ is a weighted average of the four lags. An index of inflationary severity JP_t is included. $\mathbf{z}_t$ consists of the Guide Post dummy and a tax variable.
Gordon (1972)	$\pi_t^w = \beta_0 + \boldsymbol{\beta}_1'\mathbf{U}_t + \gamma E(\pi_t^p) + \boldsymbol{\alpha}'\mathbf{z}_t + u_t$	This study replicates and compares estimations from Perry (1970), Gordon (1970), and Eckstein and Brinner (1972).
Turnovsky and Wachter (1972)	$\pi_t^w = \beta_0 + \beta_1 U_t + \gamma E(\pi_t^p) + \varepsilon_t$	Different methods of approximating $E(\pi_t^p)$ are compared, including the direct biannual Livingstone survey.
Gordon (1977a)	$\pi_t^w = \beta_0 + \boldsymbol{\beta}_1'\mathbf{U}_t + \gamma E(\pi_t^p) + \boldsymbol{\alpha}'\mathbf{z}_t + \varepsilon_t$	A core price index is introduced. Output gap is also an alternative to the unemployment measurements. $E(\pi_t^p)$ is approximated by long lags of the inflation rate as in the modern approach.

5.1.3 Brief Evaluation of Different Models

In light of the discussion of the eight most-cited papers, it was showed that Gordon's efforts encompass important features of the other papers. In Gordon (1970), he adapts the weighting method from Solow (1969). Gordon (1971) incorporates the adjusted unemployment rates from Perry (1970) and experiments with the direct Livingston measurement of the expected inflation; and the estimation results from Perry (1970) and Eckstein and Brinner (1972) are reproduced and compared with his own results in Gordon (1972). It is reasonable to argue that the evaluation results of Gordon's studies can well represent the works from other authors.

Of the four Gordon's papers, Gordon (1977a) is chosen as the ultimate, representative work of this period because it manifests his radical stance in explaining inflation and the econometric specifications to estimate it. Gordon was contacted for the vintage data. Unfortunately, these data series were completely lost, and thus, the only possibility to replicate his results was to reconstruct them as closely as possible. However, since the dataset has many variables and some variables were compiled using complicated methods, it was not possible to faithfully reconstruct the complete list. It is reasonable to keep the three essential variables (wages, price, and unemployment). Specifically, in the reproduced models, all other explanatory variables are omitted, and a single official rate of unemployment is chosen (instead of a set of the adjusted measurements). A complete description of data reconstruction and specification choices is provided in Appendix D.

In light of the reconstructed dataset, two specifications in Gordon (1977a) - column (1) and (7) on Table 3 - are picked to be appraised. They are called G1977a and G1977b, respectively. The first - the G1977a - is a reproduced regression of Gordon (1971)'s best specification in which the NRH is rejected. The reproduced version of G1977a with only three essential variables is:

$$\pi_t^p = \beta_0 + \beta_1 U_t + \gamma E(\pi_t^p) + \varepsilon_t \tag{5.5}$$

In this model, $E(\pi_t^p)$ is computed as a weighted average of twelve past rates. The weighting scheme is the same as Gordon used (Gordon 1971, Appendix A).

The second reproduced model - G1997b - is a modified version of the G1977a that provides supporting evidence for the NRH (with an extended sample period to 1976). Most of the modern features of the expectations-augmented Phillips curve were incorporated into the G1977b: the newly developed core price index replaced the general index, and the direct lags superseded the weighted average. The reproduced G1977b is:

$$\pi_t^p = \beta_0 + \beta_1 U_t + \gamma_i \sum_{i=1}^{8} \pi_{core,t-i}^p + \varepsilon_t, \tag{5.6}$$

There are two data-reconstruction complications for reproducing G1977b. Firstly, the sample size (1960-1976) is shorter than the original one (1954-1976). Gordon uses his core price index in G1997b. It is almost impossible to reconstruct this core price index because part of the data sources was not published or completely lost, and a few paragraphs of instructions do not provide sufficient technical details. When contacted, Gordon recommended using the modern core Personal Consumption Expenditures (PCE) as the replacement for his vintage index. However, the PCE is only available from 1959. It results in the reconstructed sample size for G1977b to be smaller than in the original estimations (62 against 92). Secondly, a portion of the reconstructed wage series is not compiled in an exact way as that done in the Gordon series; see Appendix D for more details.

Apart from the reconstructed dataset, the modern data is also utilized to perform the reconstructed regressions for G1977a and G1977b. The intention is to discover how the results change under the revision of the old data. The choices of the modern series are also provided in Appendix D. The next section will present the reproduced regressions and compare them with the original ones. Then, the M-S testing results for these reproduced regressions will be discussed.

5.2 Reproducing Inference Results and M-S Testings Results

5.2.1 Reproducing Inference Results

Table 5.3 reports the reproduced OLS estimations for G1977a and 1977b with the reconstructed and modern data. For comparison, Table 5.3 also includes Gordon's estimated coefficients for the expected inflation $E(\pi_t^p)$. Gordon's estimators for the set of adjusted unemployment variables are not presented because only the standard unemployment rate is used on the RHS of the reproduced models.

Under the reconstructed data, the estimated coefficients of all variables for G1977a are significant, and their signs (negative for unemployment and positive for expected inflation) are as expected. The estimated coefficient of the expected inflation $E_t(\pi_t^p)$, at 0.594, is close to Gordon original estimate, at 0.6. For G1977b, the estimated coefficients of unemployment are not significant; however, their signs are negative. The estimated sum of lagged inflation rates is nearly the same as Gordon's original estimate (0.948 against 0.939) and is significantly different from zero (the t-ratio is computed using the Delta method). These essential similarities can make up for the insignificance of unemployment. In short, the reproduced estimation results suggest that the reconstructed data sufficiently emulates the vintage data.

With the modern dataset, the reproduced estimation results for G1977a do not resemble Gordon's original ones. Although the estimated coefficient for unemployment is negative and significant, the estimated coefficient of the expected inflation (at 0.246) is much lower than the original value (at 0.600) and not significant. For G1977b, the estimated outcome is closer to the original. The estimated value of the expected inflation is close to one, but the unemployment is not significant. Assuming that the modern revised data is close to the old vintage unmodified model, these results suggest that Gordon's intensive modifications of his data series likely produce considerable effects on the final estimated results.

5.2.2 M-S Testing Results

Table 5.4 summarizes the M-S testing results for the four reproduced models. For the G1977a model using the reconstructed data, only the (4) Markov(p) dependence assumption is invalid, failing to account for the temporal dependence in the data. On the other hand, the testing results for G1977b using reconstructed data show no apparent departure from the model assumptions. However, the Markov(p) dependence appears to be spurious, potentially resulting from the small sample size and the inclusion of 8 lags of the inflation rate. A similar departure of the (4) Markov(p) dependence assumption presents on both models using the

Table 5.3: Reproduced Estimations: the Expectations-Augmented Phillips Curve

	Gordon's original estimations		Estimations with reproduced data		Estimations with modern data	
	G1977a	G1977b	G1977a	G1977b	G1977a	G1977b
	n=68	n=92	n=68	n=62	n=68	n=62
$E_t(\pi_t^p)$	0.600**	0.939**	0.581**	0.948**	0.246	1.035**
	(0.150)	(0.055)	(0.191)	(0.114)	(0.204)	(0.111)
U_t	-	-	-0.004	-0.013	-0.005**	-0.0002
	-	-	(0.001)	(0.001)	(0.001)	(0.001)

Note: The standard errors are reported in parentheses. (*) and (**) represent significance levels at 5% and 1% respectively

modern data. There are also additional assumptions departures on the two models: the (1) Normality assumption on the G1977a and the (5) t-invariant assumption on the G1977b. In short, the M-S testing results show the four models using two sets of data are misspecified, mostly due to a lack of attention to the temporal dependence in the wage change series.

Considering the M-S testing results of models of the previous two periods and this period as a whole, the departure of the (4) Markov assumption is persistent by all models in evaluation (except for Sargan's special case). It indicates that the structural wage Phillips curve, which does take into account lags of the wage series, is an 'inappropriate' statistical characterization of the behavior of the data, and the empirical evidence learning from Gordon (1977a) is not trustworthy. Also, it can be further argued that the evidence from the other important papers, which share a similar estimation specification, is equally not trustworthy. Finally, because the reproduced regressions of Gordon's models are not statistically adequate, respecification forwards next.

5.2.3 Respecification

Due to the incomplete nature of the data reconstruction and Gordon's ad-hoc, intensive modifications of the original indices, the reconstructed dataset is not utilized, and respecification only uses the modern dataset. The intention of the respecification in this part to complement the previous evaluation and provide a preliminary look into the relationship between the wage and price growth rate. The wage change and the price inflation rates as the one-year growth rates are used to respecify.[3]

The initial respecifying effort with the Normal DLR models does not produce any statistically adequate model. The M-S testing results for the best-effort models - a DLR(1) model without any trend term - are reported in table 5.5. It is found that the three assumptions - (2) linearity, (4) Markov(2) dependence, and (5) t-invariant - are not met. The invalid (5) t-invariant assumption, which is detected in the skedastic auxiliary regression, suggests the existence of changing variances over time (variance heterogeneity). Among the three, the departure of the (2) linearity assumption is the most severe when the F-value is very high at

3. Respecifications with one-quarter and two-quarter growth rates also generate similar results.

Table 5.4: M-S Testing Results: the Expectations-Augmented Phillips Curve

Models using reconstructed data			
		G1977a	G1977b
(1) Normality		$\chi(2)$=0.68 [0.651]	$\chi(2)$=0.71 [0.700]
(2) Linearity		t(59)=0.91 [0.366]	t(46)=-0.11 [0.913]
(3) Homoskedasticity		t(63)=-1.61 [0.112]	t(61)=-0.78 [0.439]
(4) Markov(p) dependence	R	**F(3,59)=9.11 [0.001]**	F(3,46)=0.99 [0.404]
	S	t(63)=0.22 [0.827]	t(61)=0.60 [0.552]
(5) T-invariant	R	t(59)=-1.35 [0.181]	t(46)=0.86 [0.393]
	S	t(63)=0.30 [0.769]	t(61)=-0.21 [0.835]

Models using modern data			
		G1977a	G1977b
(1) Normality		$\chi(2)$=**24.24 [0.000]**	$\chi(2)$=5.13 [0.077]
(2) Linearity		t(59)=0.82 [0.418]	t(46)=1.70 [0.096]
(3) Homoskedasticity		t(63)=-0.27 [0.789]	t(61)=-0.82 [0.416]
(4) Markov(p) dependence	R	**F(3,59)=8.31 [0.000]**	**F(3,46)=5.35 [0.003]**
	S	t(63)=0.37 [0.714]	t(61)=1.39 [0.170]
(5) T-invariant	R	t(59)=-1.69 [0.095]	**t(46)= 2.14 [0.038]**
	S	t(63)=-1.76 [0.083]	t(61)=-0.53 [0.601]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.
Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

232.96.

Because the M-S testing detects variance heterogeneity, the next respecification round moves to employ the Student's t Dynamic Linear Model (StDLR) model. The StDLR model accommodates both variance heterogeneity and heteroskedasticity; see Spanos (1994) and Poudyal (2012). Thus, variance heterogeneity can be modeled without applying the error-fixing strategy (such as the heteroskedasticity-robust estimators).

The respecification with the StDLR model produces a one-lag model with a linear time trend. Table 5.5 shows the M-S testing results of the best-effort model. There still exists assumption departures from the respecified StDLR(1) model. The most noticeable incorrect assumption is the (2) Linearity with a high F-value (38.55). Also, the (1) Student's t distribution and (4) second-order Markov and assumptions are not satisfied. These findings suggest the StDLR model is not adequate to model the behavior of the wage growth rate.

The modeling with both the Normal and StDLR models exposes their limitations that these models do not allow nonlinearity of the RHS variables (such as squares of wage change or unemployment). The comprehensive approach to statistical modeling with nonlinear parameters in the conditional regression function is beyond the scope of this dissertation.

Although statistical adequacy was not achieved, it is interesting to see how the value of estimated coefficients for the inflation rate when the lags of wage change present on the RHS. Table 5.6 shows the estimation results for the Normal DLR and StDLR models. In both results, the lag of the dependent variable (the wage change) plays a significant role; and the values of the sum of inflation coefficients shrink to less than 0.45. It suggests that in a correct model with adequate lags of wage change, the effects of price change will be considerably less than in the models using the theory-driven expectations-augmented Phillips equation.

Table 5.5: M-S Testing Results for the Respecified models: the Expectations-Augmented Phillips Curve

Normal DLR(1) model		
(1) Normal		$\chi(2)$)=4.53 [0.104]
(2) Linearity		**F(3,77)=232.96 [0.000]**
(3) Homoskedasticity		F(2,84)=0.14 [0.870]
(4) Markov(1) dependence	Regression function	**F(3,77)=3.58 [0.018]**
	Skedastic function	F(1,84)=0.080 [0.372]
(5) t-invariance	Regression function	F(1,77)=3.72 [0.058]
	Skedastic function	**F(1,84)=7.87 [0.006]**
StDLR(1) model		
(1) Student's t		**$\chi(2)$=9.14 [0.010]**
(2) Linearity		**F(3,76)=38.55 [0.000]**
(3) Heteroskedasticity		F(3,83)=0.48 [0.701]
(4) Markov(1) dependence	Regression function	F(3,76)=1.85 [0.144]
	Skedastic function	**F(1,83)=11.09 [0.001]**
(5) t-invariance	Regression function	F(1,76)=1.06 [0.306]
	Skedastic function	F(1,83)=0.002 [0.885]

Table 5.6: Estimation Results for the Respecified Models

Estimation results		
	Normal DLR(1) model	StDLR(1) model
π^w_{t-1}	0.611	0.516
	(0.081)**	(0.095)**
$\sum_{i=0}^{1} \pi^p_{t-i}$	0.351	0.425
	(0.078)**	(0.012)**
$\sum_{i=0}^{1} U_{t-i}$	-0.112	-0.107
	(-0.102)	(-0.060)

Note: The reported coefficients for π^p_t and U_t are sum of the contemporaneous and its one lag. The standard errors are in parentheses. (*) and (*) represent significance levels at 5% and 1% respectively.

5.3 Conclusions

This chapter provides a brief literature survey of the empirical expectations-augmented Phillips curve and selects two specifications from Gordon (1977a) as the representative works for statistical evaluations. As in the previous two chapters, the M-S testing results also identify the same departure of the Markov(p) assumption. In particular, this assumption departure stems from excluding the lags of the regressor (wage change) from the explanatory variable set. It is an understandable result of the theory-driven commitment to only use lags of the price inflation to approximate the expectation. However, there is no particular a priori statistical reason why one limits the wage change to be a static process, which is hardly true for any economic time series.

These misspecifications indicate that the empirical evidence obtained from Gordon's representative works and, generally, the empirical literature employing the expectations-augmented specification in the 1970s is untrustworthy. On the negative side, unlike the previous two periods, the respecifications with the Normal and Student's t DLR models fail to achieve statistical adequacy. It opens a line of further research in the future.

Chapter 6

The New Keynesian Phillips Curve

6.1 Introduction

Apart from the 'old' expectations-augmented Phillips curve, the New Keynesian paradigm also generated its own Phillips curve equation, which relates the current inflation rate to future inflation expectation and the market condition (real marginal cost). In the formulation of the New Keynesian Phillips Curve (NKPC), the aggregate market price level is assumed to be driven by the behavior of a representative firm. Each firm sets its prices (rationally) basing on its expectation of the overall price level and excess demand on the market. Firms' price expectation is assumed to be a weighted average of the infinite past periods and future contracts (Calvo 1983's model). Solving the first-order condition of the optimization each firm is facing and then aggregating for the whole market, it yields an equation that the current inflation rate is determined by the expectation of the future inflation rate and the economy-level real marginal cost as:

$$\pi_t^p = \gamma E_t(\pi_{t+1}^p) + \lambda mc_t, \tag{6.1}$$

where $E_t(\pi_{t+1}^p)$ is the next-period inflation expectation, and mc_t is the unobserved real marginal cost which can be proxied by the output gap or the labor income share. In addition to the 'New Keynesian' label, it is also called the 'forward-looking' curve to be distinguished with the old 'backward-looking', which uses the past inflation rates to approximate the expectation. The NKPC in Equation 6.1 was further developed into a hybrid version which also included the past inflation rate into the RHS to resemble the backward-looking Phillips curve as:

$$\pi_t^p = \gamma_f E_t(\pi_{t+1}^p) + \gamma_b \pi_{t-1}^p + \lambda mc_t \tag{6.2}$$

In the estimation of the NKPC as a stand-alone equation, it is argued that the OLS method is unreliable because there is a potential correlation between the real marginal cost x_t and the attached error terms u_t (assumed as supply shocks). Also, the future inflation expectation $E_t(\pi_{t+1}^p)$ is potentially endogenous and cannot be observed with precision. Consequently, much of the literature focuses on identifying the NKPC and the related problems; see Mavroeidis, Plagborg-Moller, and Stock (2014) and Abbas, Bhattacharya, and Sgro (2016) for comprehensive surveys.

In a ground-breaking paper, Galı and Gertler (1999) first propose the generalized method of moments (GMM) approach to identify the NKPC. Their method firstly rewrites Equation 6.2 as:

$$\pi_t = \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda x_t + \underbrace{u_t - \gamma_f[\pi_{t+1} - E_t(\pi_{t+1})]}_{\tilde{u_t}} \tag{6.3}$$

Then, the residual function $h_t(\vartheta)$ $(\vartheta = (\gamma_f, \gamma_b, \lambda))$ is defined as:

$$\tilde{u_t} \equiv h_t(\vartheta) = \pi_t - \gamma_f \pi_{t+1} - \gamma_b \pi_{t-1} - \lambda x_t \tag{6.4}$$

This function consists of observable variables. Now, the GMM estimators can be estimated by identifying valid instruments Z_t which satisfies:

$$E[Z_t h_t(\vartheta)] = 0 \tag{6.5}$$

By assuming on zero unconditional expectation of the error terms ($E_{t-1}(u_t) = 0$) and rational expectation, it can be shown *theoretically* that $E_{t-1}(\tilde{u}_t) = 0$. This implies that the orthogonality condition between $h_t(\vartheta) \equiv \tilde{u}_t$ and the instrument set Z_t can be met when Z_t is any one-lag or earlier predetermined variables ($Z_t = Y_{t-i}\,, i = 1, 2, 3...$). In other words, there is no limit on the choices of admissible instrumental variables as long as the instruments are lagged variables. It is the baseline strategy for identification in the literature.

Under this identification approach, the GMM estimations from Galı and Gertler (1999) and the follow-up Galı, Gertler, and Lopez-Salido (2001) first prove the empirical fit of the NPKC. The estimators for γ_f mostly fall into the range 0.7-0.8, and the estimators for γ_b are mostly from 0.2 to 0.3. These numbers suggest the forward-looking behavior plays a more prominent role in the formation of inflation than the past rate does. This finding contradicts the standard expectations-augmented Phillip curve in which only the past information matters.

However, the GMM identification approach and its empirical validity have been heatedly debating in the literature. The GMM identification of the NPKC was derived as the theoretical results; on equal terms, it can also be rejected by theoretical arguments. Mavroeidis (2005) assume the marginal cost x_t an autoregressive (AR) process of order 2, and shows that the NKPC cannot be identified when there is little dynamics in the marginal cost process. Nason and Smith (2008) put the NPKC into a triple-equation system consisting of the inflation rate, output gap, and central bank policy rate, and derived that the GMM estimation cannot be employed to identify the single-equation NKPC. However, theoretical proof of the weak identification is also subjected to counter-theoretical arguments. In particular, Krogh (2015) generalizes the theoretical model in Nason and Smith (2008) and demonstrates that the weak identification proved by Nason and Smith (2008) is not true in all applicable scenarios.

In addition, the single-equation NKPC was also demonstrated not to fit well with data. Rudd and Whelan (2005, 2006) employ the same GMM method to estimate various closed-form solutions of the NKPC and find poor empirical fits. In turn, Gali, Gertler, and Lopez-Salido (2005) replied that Rudd and Whelan (2005, 2006) do not establish the correct connection between the structural parameters in Equation 6.2 and their close-formed coefficients. Newly estimation results from Gali, Gertler, and Lopez-Salido (2005) also provide that their estimate of a slightly different close-formed solution resembles their valid results in Galı and Gertler (1999) and Galı, Gertler, and Lopez-Salido (2001).

Aside from weak identification, some other studies also detect a structural break in the NKPC at the end of Volcker's disinflation in 1983-1984. Zhang, Osborn, and Kim (2008) apply the standard coefficients stability test for unknown break points (Andrews 1993; Andrews and Ploberger 1994) and find evidence for parameter changes in 1975, 1981, and 2000. Kleibergen and Mavroeidis (2009) apply an identification-robust test for the structural break and detect a break point in 1984.

The identifying assumption $E_{t-1}(u_t) = 0$ is also challenged by Zhang and Clovis (2010). Performing a serial correlation test proposed by Godfrey (1994) for the GMM estimation, they find the presence of autocorrelation from the residuals of Galı and Gertler (1999) and suggest that inclusion of longer lags can help to solve the problem.

Apart from the GMM approach, the single-equation NKPC is also estimated via a reduced-form VAR system Sbordone (2002, 2005) and Kurmann (2007). Firstly, inflation and marginal cost are assumed to belong to a finite-order VAR system. Then, the structural NKPC in Equation 6.2 enters as restrictions on the reduced-form VAR coefficients. Estimates of the structural NKPC coefficients are obtained by solving the constrained VAR.

Because the NKPC is still developing literature, the landmark Galı and Gertler (1999), which put forward the popular GMM approach, is selected as the representative work to be evaluated. The GMM estimation of the NKPC is different from the OLS estimations of the backward-looking curve that were explored in the last three chapters. Hence, the next section will unveil the statistical model on which the GMM estimators are founded.

6.2 The Statistical Models Underlying the NKPC's GMM Estimation

6.2.1 Is There a Common Close-form Statistical Model behind the Structural NKPC?

Since the single-equation NKPC in Equation 6.2 is formulated as the linear rational expectations model, one needs to solve to the model to get rid of the expectation terms $E_t(\pi^p_{t+1})$ to derive an estimable equation. Mavroeidis (2004) provides an analysis of the closed-form solutions of the NKPC. He shows that there exist two types of solutions, contingent upon the number of explosive roots of the following equation:

$$-L^{-1} + \frac{1}{\gamma_f} - \frac{\gamma_b L}{\gamma_f} = 0 \tag{6.6}$$

If Equation 6.6 has exactly one explosive root, there is a unique 'forward' solution as:

$$\pi^p_t = \delta_1 \pi^p_{t-1} + \left(\frac{\lambda}{\gamma_f}\right) \sum_{k=0}^{\infty} \frac{1}{\delta_2} E_t x_{t+k}, \tag{6.7}$$

where δ_1 and δ_2 are respectively the stable and unstable of root of Equation 6.6. When no unstable roots are found, the solution is not unique under a backward form as:

$$\pi^p_t = \frac{1}{\gamma_f}\pi^p_{t-1} - \frac{\gamma_b}{\gamma_f}\pi^p_{t-2} - \frac{\lambda}{\gamma_f}x_{t-1} + \zeta_t, \tag{6.8}$$

where ζ_t is a Martingale Difference process.

Mavroeidis (2004) analyzes the two solutions and shows that the single-equation NPKC alone does not provide sufficient information to be identified. In other words, additional information about the process x_t or a bigger system of equations must be assumed to make the NKPC identifiable.

On the other hand, the GMM approach offers an advantage that researchers do not need to be explicit about a complete structural model. With a minimum number of assumptions, the orthogonality conditions holding for some certain instruments are sufficient to deliver an identification strategy. However, it does not mean that there is no implicit closed-form solution guiding the evaluations of the GMM approach. In fact, the big full-information DGSE systems, in which the NKPC is an included equation, are strongly subjected to the VAR representation (Fernández-Villaverde et al. 2007; Ravenna 2007; Giacomini 2013). A close look at the literature also confirms a presumption of the similar VAR statistical representations generating the NKPC.

Nason and Smith (2008) approach their solution from a system of triple equations as:

$$\begin{aligned} \pi_t^p &= \gamma_f E_t \pi_{t+1}^p + \gamma_b \pi_{t-1}^p + \lambda x_t + \varepsilon_{\pi t} \\ x_t &= \beta_f E_t x_{t+1} + \beta_b x_{t-1} - \beta_R (R_t - E_t \pi_{t+1}^p) + \varepsilon_{xt} \\ R_t &= \omega_\pi \pi_t^p + \omega_y x_t + \varepsilon_{Rt} \end{aligned}$$

where R_t is the interest rate.

Choosing $\mathbf{Z}_t = (\pi_t^p, x_t)'$, they show that the unique solution for the system (under certain restrictions) has a VAR presentation of $\mathbf{Z}_t$. The VAR model is then utilized to reach the paper's conclusion on the non-identifiability of the GMM approach.

Krogh (2015) employs a system consisting of four equations as:

$$\begin{aligned} \pi_t^p &= \gamma_f E_t \pi_{t+1}^p + \gamma_b \pi_{t-1}^p + \lambda x_t + \varepsilon_{\pi t} \\ x_t &= \gamma_w x_{t-1} + (1-\gamma_w)(1+\varphi) y_t^g + \varepsilon_{xt} \\ y_t^g &= \frac{1}{1+h}[E_t y_{t+1}^g + h y_{t-1}^g - (1-h)(r_t - E_t \pi_{t+1}^p)] + \varepsilon_{y^g t} \\ R_t &= \phi R_{t-1} + (1-\phi)(\phi_\pi \pi_t^p + \phi y_t^g) + \varepsilon_{Rt}, \end{aligned}$$

where $(\varepsilon_{\pi t}, \varepsilon_{xt}, \varepsilon_{y^g t}, \varepsilon_{Rt})$ is a AR(1) process and y_t^g is the output gap, x_t as the marginal cost.

The reduced-form solution of this system is a VAR(2) process. Krogh next shows that the NKPC can be identified by the GMM.

Kurmann (2007) considered a (arbitrary) general bivariate case as:

$$\begin{aligned} \pi_t^p &= \gamma_f E_t \pi_{t+1}^p + \gamma_b \pi_{t-1}^p + \lambda y_t + \varepsilon_{\pi t} \\ x_t &= b_1 \pi_t + b_2 x_{t-1} + b_3 \pi_{t-1}^p + \varepsilon_{xt} \end{aligned}$$

Under certain conditions, the unique non-explosive solution for this system also has a VAR(1) form.

Apart from the GMM method, the reduced-form VAR presentation is always assumed as the initial step to forward other estimation strategies. Sbordone (2002, 2005, 2006) assume a VAR specification and use Equation 6.2 to get parameter restrictions on a distance function consisting of structural parameters. The estimation for the structural parameters is obtained by finding the minimizer of the distance function. In Kurmann (2007), instead of minimizing a distance function, the author uses the Maximizing Likelihood (ML) method to solve the constrained VAR. Finally, the GMM orthogonality condition from a single equation NPKC can be extended to the general VAR form as an identification strategy - the VAR-GMM method in Mavroeidis, Plagborg-Moller, and Stock (2014).

In light of this discussion, it can be concluded that the VAR form is implicitly assumed as an umbrella process generating the NKPC. The GMM method is an extension of the instrument variable (IV) estimation method. Spanos (2007) provides the PR approach's analysis of the underlying statistical models behind the IV estimation. The next part will discuss the statistical model underpinning the IV-GMM estimators and relate it to the umbrella VAR form implicitly assumed by the literature.

6.2.2 The Statistical Model Underpinning the GMM Estimators

Rewrite the structural NPKC in Equation 6.3 in a generic form as:

$$y_t = \boldsymbol{\gamma}_1' \mathbf{X}_{endo,t} + \boldsymbol{\gamma}_2' \mathbf{X}_{exgo,t} + \tilde{u}_t, \tag{6.9}$$

where y_t is the LHS variable, $\mathbf{X}_{endo,t}$ is a $p_1 \times 1$ set of endogenous variables, $\mathbf{X}_{exgo,t}$ is a $p_2 \times 1$ set of exogenous variables and $\tilde{u}_t$ is the residual. The GMM method requires that:

(i) $E(\mathbf{X}_{endo,t}\tilde{u}_t) \neq \mathbf{0}$ and $E(\mathbf{X}_{exgo,t}\tilde{u}_t) = \mathbf{0}$

A $m \times 1$ set of instrument $\mathbf{Z}_t$ is chosen be satisfy the following conditions (see Stock, Wright, and Yogo 2002 for a complete discussion):

(ii) $E(\mathbf{Z}_t\tilde{u}_t) = \mathbf{0}$

(iii) $E(\mathbf{Z}_t'\mathbf{X}_{endo,t}) \neq \mathbf{0}$

(iv) $E(\mathbf{Z}_t'\mathbf{Z}_t) > \mathbf{0}$

There exist several tests to probe the validity of these two conditions. It is well-known in the literature that all of these tests, no matter how informative they are, cannot deliver definite outcomes. Economic theories are the main source of justification for validity of the condition (ii).

For the NKPC's GMM estimation, as discussed previously, the condition (i) is obtained as a theoretical result by Galı and Gertler (1999); and it is not questioned in the literature. Many criticisms of the identification strategy focus on the condition (iii). When the endogenous variables $\mathbf{X}_{endo,t}$ is weakly correlated with the instruments $\mathbf{Z}_t$, the GMM estimations will suffer the weak identification problem; see Mavroeidis, Plagborg-Moller, and Stock (2014) for a detailed discussion.

Looking from the pure statistical ground, Spanos (2007) provides an analysis that these conditions (i)-(iv) can be evaluated directly with observable sample data. From a multivariate joint distribution, he showed that the IV estimator (and the GMM as a derivative) is embedded into a statistical Implicit Reduced Form (IRF) (directly related to the reduced form in the Simultaneous Equations) as:

$$y_t = \boldsymbol{\beta}_1' \mathbf{X}_{exgo,t} + \boldsymbol{\beta}_2' \mathbf{Z}_t + u_{1t} \tag{6.10a}$$

$$\mathbf{X}_{endo,t} = \mathbf{B}_1' \mathbf{X}_{exgo,t} + \mathbf{B}_2' \mathbf{Z}_t + \mathbf{u}_{2t} \tag{6.10b}$$

$$\text{with } \begin{pmatrix} u_{1t} \\ \mathbf{u}_{2t} \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \omega_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix} \right) \tag{6.10c}$$

To create the link with the structural model, the IRF in Equations 6.10a and 6.10b can be reparameterized as:

$$y_t = \boldsymbol{\gamma}_0'\mathbf{X}_t + \boldsymbol{\alpha}_0'\mathbf{Z}_t + \varepsilon_{0t} \tag{6.11a}$$

$$\mathbf{X}_{endo,t} = \mathbf{B}_1'\mathbf{X}_{exgo,t} + \mathbf{B}_2'\mathbf{Z}_t + \mathbf{u}_{2t} \tag{6.11b}$$

$$\text{with } \begin{pmatrix} \varepsilon_{0t} \\ \mathbf{u}_{2t} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{22} \end{pmatrix}\right), \tag{6.11c}$$

where the exogenous and endogenous variables are stacked into $\mathbf{X}_t$ as $\mathbf{X}_t = (\mathbf{X}'_{endo,t}, \mathbf{X}'_{exog,t})'$

To make the IV structural equation identifiable, Spanos shows that one must impose (non-testable) $\boldsymbol{\gamma}_0 = \mathbf{0}$ from Equation 6.11a, in conjunction with $\mathbf{B}_1, \mathbf{B}_2 \neq \mathbf{0}$ in Equations 6.10b and 6.11b plus $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \neq \mathbf{0}$ in Equation 6.10a. In other words, the *structural* equation in Equation 6.9 is a reparameterization of the *statistical* IRF in Equations 6.10a and 6.10b under certain restrictions.[1] These restrictions are not ordinary parameter constraints on regression coefficients; however, they can shed light on various problems associated with the IV estimators. In particular, conditions (i) and (ii) relate to the substantive parameterizations of interest and (iii) and (iv) are testable; see Spanos (2007). Moreover, it provides the basis for testing the overidentifying restrictions $(m - p)$ relating to:

$$H_0\text{: } \mathbf{G}(\theta,\varphi){=}\mathbf{0},\ H_1\textbf{:}\ \mathbf{G}(\theta,\varphi){\neq}\mathbf{0},\ \theta{\in}\mathbf{\Theta}{\subset}\mathbb{R}^m,\ \varphi{\in}\mathbf{\Phi}{\subset}\mathbb{R}^p,\ m \geq p. \tag{6.13}$$

However, before testing the validity of conditions [iii]-[iv] and Equation 6.13, one needs to establish the statistical adequacy of the IRF in Equations 6.10a and 6.10b. Since the GMM estimation is an IV-based specification with a weighting matrix, with the IRF consisting of assumptions analogous to (1)-(5) of Table 2.6 which are statistical assumptions of the standard multivariate linear regression model. When any of these assumptions are invalid for the particular data, the invoked sampling distributions of the estimated structural parameters will be erroneous and result in inferences unreliable. That is, if all the assumptions are valid for the particular data, statistical adequacy is secured, and further tests for weak instruments, exogeneity and overidentifying restrictions can follow. Otherwise, any findings of assumption violation will expose the GMM estimation to be built on a misspecified model and, put differently, to be unreliable.

To evaluate the statistical adequacy of the GMM estimation of the NKPC, the first step is to rewrite the generic form with the explicit variables. Note that the instrument $\mathbf{Z}_t$ in the NKPC's GMM estimation includes the lags of the LHS variable π_t, the endogenous forcing variable x_t and a column vector of some external instruments $\mathbf{Z}_t$; thus, the IRF in Equations 6.10a

1. The full statistical condition for identification from the IFR is:

$$\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_1 - \mathbf{B}_2\boldsymbol{\gamma}_0 = 0, \text{ subject to } \mathbf{B}_{12} \neq 0,\ \mathbf{B}_{22} \neq 0,\ \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \neq 0 \tag{6.12}$$

and 6.10b can be written as:

$$\pi_t^p = \sum_{i=1}^{l} \mathbf{b}'_{1i}\mathbf{Z}_{t-i} + u_{1t} \tag{6.14a}$$

$$x_t = \sum_{i=1}^{l} \mathbf{b}'_{2i}\mathbf{Z}_{t-i} + u_{2t} \tag{6.14b}$$

$$\pi_{t+1}^p = \sum_{i=1}^{l} \mathbf{b}'_{3i}\mathbf{Z}_{t-i} + u_{2t} \tag{6.14c}$$

The Equation 6.14a corresponds for Equation 6.10a and Equations 6.14b and 6.14c for Equation 6.10b. Among the three, the first two - Equations 6.14a and 6.14b - are identical to two components of the umbrella VAR models that are implicitly assumed behind the structural NKPC solutions (as discussed in the previous section). The statistical adequacy of these two equations is key to ensure the reliability of the GMM estimators. It is decided to put aside the Equation 6.14c because the LHS variable π_{t+1} in Equation 6.14c is one-period lead of the LHS variable π_t in Equation 6.14a. The statistical adequacy of Equation 6.14a mostly ensures a similar condition for Equation 6.14c. From this point, the combination of Equations 6.14a and 6.14b is called the implicit statistical model (ISM).

The next section will evaluate the statistical adequacy of this ISM, beginning with an description of instrument sets, data series, and sample periods.

6.3 Statistical Adequacy Evaluation for the GMM Estimation of the NKPC

6.3.1 Evaluation Description

Since the literature does not use a single set of instrument set for the GMM estimation, this section imitates the exhaustive study on the GMM estimation of the NKPC by Mavroeidis, Plagborg-Moller, and Stock (2014) to pick the three most defining instrument sets. These three consist of the small set (with only the lags of inflation and forcing variables), the set from Galı, Gertler, and Lopez-Salido (2001) and the set from Galı and Gertler (1999). Afterward, they will be called the small, the GGLS2001, and the GG1999 sets, respectively. Table 6.1 show all instrumental variables of each set. It is important to note that the exogenous variable π_{t-1} is already included in each instrument set.

The literature also found evidence of structural breaks in the 1980s. To consider this aspect, three sample periods will be selected based on the two major economic events in the last 50 years to ensure this aspect to be explored. The most extended - the Full sample - covers all available data from 1960 to 2019. Another sample is chosen to leave out the 1970s stagflation and the subsequent recession happening under the Volcker disinflation. Because the last quarter of 1982 is considered as the end of the economic recession induced by the Volcker disinflation, it is reasonable to start this sample at the first quarter of 1983 and end it at the last available observation. This one is named the Post-1983 sample.

After the Financial Crisis, the Federal Reserve (FED) conducted an unusual expansionary monetary policy (Quantitative Easing) to prevent further economic collapse and support the economy to recover. Then, the subsequent recovery during most of the 2010s suggests the breakdown of the 'traditional' Phillips curve relationship that literature has also struggled to explain. It is not unreasonable to consider the Financial Crisis as a structural change of the overall economy. Therefore, the post-Financial Crisis years are kept out of the Post-1983 sample to create the Great Moderation sample: from 1983 to the third quarter of 2007.

The two principal measurements of the forcing variable (the real marginal cost) x_t as the output gap and the labor income share are also chosen to make the evaluation comprehensively. Table 6.1 shows the description of all variable choices. All data are obtained from FRED Economic Data. In total, there are 18 ISMs from the combinations between three instrument sets, two forcing variables, and three sample periods. The evaluation of the statistical adequacy of these ISMs (via the M-S testings) will be discussed next.

Table 6.1: Description of Instrument Sets, Forcing Variables and Sample Periods

Variables in the structural model $(y_t, \mathbf{X}_t)$	
$y_t = \{\pi_t^p\}$: inflation rate which is the percent change in implicit GDP deflator index. $\mathbf{X}_{endo,t} = \{\pi_{t+1}^p, x_t\}$: the one-period forward inflation rate and the marginal cost. $\mathbf{X}_{exgo,t} = \{\pi_{t-1}^p\}$: one-lag inflation rate. The proxy for the marginal cost x_t is either log of labor non-farm business income share ($share_t$) or CBO output gap ($ygap_t$).	
Three set of instruments $(\mathbf{Z}_t)$	
Set	Instruments
1. Small set	$\mathbf{Z}_t$ = 4 lags of inflation, and 3 lags of share.
2. Gali et al. (2001)	$\mathbf{Z}_t$ = 4 lags of inflation, 2 lags of income share, output gap, and wage inflation.
3. Gali and Gertler (1999)	$\mathbf{Z}_t$ = 4 lags of inflation, labor income share, output gap, long-short interest rate spread, wage inflation, and commodity price inflation.
Three sample periods	
1. The Full sample: from 1960Q1 to 2019Q3.	
2. The Post-1983 sample: from 1983Q1 to 2019Q3.	
3. The Great Moderation sample: from 1983Q1 to 2007Q3.	

6.3.2 M-S Testings Results

The M-S testing results for the 18 ISMs are reported in Tables 6.2 to 6.4. Each table covers one of the three sample periods, starting from the shortest Great Moderation and ending at the most extended Full sample. In each table, the M-S testing results are grouped by the instrument sets. There are two results (based on the choice of the forcing variable x_t) for each instrument set. The testing results of each ISM, which have two equations (Equations 6.14a and 6.14b), have two columns. The relevant information from the M-S testing results can be summarized as:

1. Except for the ISMs of the small set in the Great Moderation period, the remaining 17 ISMs are found statistically adequate. Of inadequate models, the other five ISMs under the Great Moderation sample appear nearly statistically adequate. Overall, these results cast down on the statistical reliability of the IV/GMM estimations.

2. The estimated equation from the three regressors (inflation and the two forcing variables) do not produce identical MS test results. Particularly, $ygap_t$ has the least assumption departures; π_t and $share_t$ have a similar degree of assumption invalidity. However, compared to π_t, $share_t$ distinguishes itself by the invalidity of the (1) Normality assumption in all combinations of instrument sets and sample periods.

3. The M-S testing results greatly vary between the Full sample and the other shorter two. In the Great Moderation sample, the (3) first-order Markov dependence is the most common assumption departure. Evidence for the departures of the second order of the assumptions (4) and (5) are not detected. This finding is consistent with the general observation of low volatility during the Great Moderation. For the Post-1983 sample, its M-S testing results are different from the Great Moderation's. The departure of the second-order of the (5) t-invariance assumption is discovered twice. Also, the (1) Normality assumption is not met in all ISMs with both forcing variables. Finally, considering the Full sample, which covers the turbulent 1970s, it is not surprising that the large differences in volatility are transformed into the detection of variance changing and heteroskedasticity (departures of the (3) assumption and the second order of the (4) assumption). Besides, the (1) Normality assumption does not hold for all M-S testings for this sample period.

4. Comparing the M-S testing results among the three instrument sets, the larger instruments set tend to have fewer invalid assumptions. This testing result pattern is expected because the inclusion of more instruments and their lags offers more explanatory power to the regression outcomes.

In summary, seventeen of the eighteen ISMs in the evaluation are found to be statistically inadequate with various degrees of assumption departures. Hence, Section 6.4 will proceed to respecify the misspecified ISMs. Subsequently, the new GMM estimators under the statistical adequately ISMs will be estimated and compared with the original ones in Section 6.5.

Table 6.2: M-S Testing Results for the ISMs behind the NKPC, the Great Moderation Sample (1983Q1-2007Q3)

		Small set			
		Labor share		Output gap	
		π_t^p	x_t	π_t^p	x_t
(1) Normality		χ(2)=1.93 [0.380]	**χ(2)=16.34 [0.000]**	χ(2)=2.73 [0.255]	χ(2)=0.23 [0.893]
(2) Linearity		F(2,83)=0.30 [0.744]	F(2,83)=0.08 [0.925]	F(2,83)=0.10 [0.904]	F(2,83)=2.84 [0.064]
(3) Homoskedasticity		F(2,91)=0.43 [0.653]	F(2,91)=0.98 [0.377]	F(2,91)=0.54 [0.585]	F(2,91)=1.40 [0.253]
(4) Markov(4) dependence	R	**F(2,83)=4.00 [0.022]**	F(2,83)=0.99 [0.377]	F(2,83)=1.21 [0.304]	F(2,83)=0.64 [0.529]
	S	F(1,91)=0.91 [0.342]	F(1,91)=0.02 [0.897]	F(1,91)=1.41 [0.239]	F(3,91)=1.14 [0.339]
(5) T-invariance	R	F(3,83)=1.99 [0.122]	F(3,83)=2.24 [0.090]	F(3,83)=0.66 [0.577]	F(3,83)=0.71 [0.546]
		GGLS2001			
		Labor share		Output gap	
		π_t^p	Labor share	π_t^p	Output gap
(1) Normality		χ(2)=1.1 [0.578]	**χ(2)=7.23 [0.027]**	χ(2)=1.1 [0.578]	χ(2)=1.26 [0.533]
(2) Linearity		F(4,73)=0.50 [0.734]	F(4,73)=0.97 [0.428]	F(4,73)=0.50 [0.734]	**F(4,73)=3.01 [0.024]**
(3) Homoskedasticity		F(4,89)=0.30 [0.874]	F(4,89)=1.33 [0.267]	F(4,89)=0.30 [0.874]	F(4,89)=1.12 [0.354]
(4) Markov(4) dependence	R	F(4,73)=1.00 [0.413]	F(4,73)=0.64 [0.634]	F(4,73)=1.00 [0.413]	F(4,73)=1.76 [0.147]
	S	F(1,89)=0.41 [0.521]	F(1,89)=0.02 [0.894]	F(1,89)=0.41 [0.521]	F(1,89)=0.00 [0.974]
(5) T-invariance	R	F(3,73)=1.06 [0.371]	**F(3,73)=5.97 [0.001]**	F(3,73)=1.06 [0.371]	F(3,73)=0.79 [0.501]
	S	F(3,89)=1.00 [0.399]	F(3,89)=1.48 [0.225]	F(3,89)=1.00 [0.399]	F(3,89)=0.25 [0.862]
		GG1999			
		Labor share		Output gap	
		π_t^p	x_t	π_t^p	x_t
(1) Normality		χ(2)=1.71 [0.425]	**χ(2)=7.13 [0.028]**	χ(2)=1.71 [0.425]	χ(2)=2.9 [0.234]
(2) Linearity		F(6,58)=0.92 [0.490]	F(6,58)=1.70 [0.137]	F(6,58)=0.92 [0.490]	F(6,58)=1.58 [0.170]
(3) Homoskedasticity		F(6,87)=0.18 [0.980]	F(6,87)=0.68 [0.667]	F(6,87)=0.18 [0.980]	F(6,87)=0.59 [0.737]
(4) Markov(4) dependence	R	**F(3,58)=3.04 [0.036]**	F(6,58)=1.49 [0.196]	**F(3,58)=3.04 [0.036]**	F(6,58)=0.79 [0.583]
	S	F(1,87)=0.08 [0.779]	F(1,87)=0.16 [0.692]	F(1,87)=0.08 [0.779]	F(1,87)=2.44 [0.122]
(5) T-invariance	R	F(6,58)=0.65 [0.691]	**F(3,58)=8.52 [0.000]**	F(6,58)=0.65 [0.691]	F(3,58)=0.38 [0.771]
	S	F(3,87)=0.51 [0.674]	F(3,87)=0.21 [0.890]	F(3,87)=0.51 [0.674]	F(3,87)=0.36 [0.785]

Note: R and S denote the auxiliary regression and skedastic functions, respectively.

Table 6.3: M-S Testing Results for the ISMs behind the NKPC, the Post-1983 Sample (1983Q1-2019Q3)

		Small set			
		Labor share		Output gap	
		π_t^p	x_t	π_t^p	x_t
(1) Normality		χ(2)=0.16 [0.923]	**χ(2)=9.28 [0.010]**	χ(2)=0.09 [0.954]	**χ(2)=19.45 [0.000]**
(2) Linearity		F(2,131)=0.33 [0.718]	F(2,131)=0.72 [0.490]	F(2,131)=0.87 [0.421]	F(2,131)=2.05 [0.133]
(3) Homoskedasticity		F(2,139)=1.31 [0.274]	F(2,139)=1.84 [0.163]	F(2,139)=0.76 [0.471]	F(2,139)=1.53 [0.220]
(4) Markov(4) dependence	R	F(2,131)=2.23 [0.111]	F(2,131)=0.13 [0.882]	F(2,131)=0.36 [0.700]	F(2,131)=0.23 [0.798]
	S	**F(1,139)=3.86 [0.051]**	F(1,139)=0.73 [0.394]	F(1,139)=1.67 [0.199]	F(1,139)=1.12 [0.291]
(5) T-invariance	R	F(3,131)=0.41 [0.747]	**F(3,131)=3.64 [0.015]**	F(3,131)=0.76 [0.519]	F(3,131)=0.98 [0.405]
	S	F(3,139)=1.30 [0.276]	**F(3,139)=3.81 [0.012]**	**F(3,139)=2.98 [0.034]**	F(3,139)=1.70 [0.171]
		GGLS2001			
		Labor share		Output gap	
		π_t^p	Labor share	π_t^p	Output gap
(1) Normality		χ(2)=0.95 [0.622]	**χ(2)=8.62 [0.013]**	χ(2)=0.95 [0.622]	**χ(2)=11.79 [0.003]**
(2) Linearity		F(4,121)=0.73 [0.576]	**F(4,121)=2.56 [0.042]**	F(4,121)=0.73 [0.576]	F(4,121)=2.55 [0.043]
(3) Homoskedasticity		F(4,137)=0.84 [0.502]	F(4,137)=0.67 [0.616]	F(4,137)=0.84 [0.502]	F(4,137)=1.01 [0.405]
(4) Markov(4) dependence	R	**F(4,121)=2.42 [0.052]**	F(4,121)=0.85 [0.496]	**F(4,121)=2.42 [0.052]**	F(4,121)=1.39 [0.241]
	S	F(1,137)=1.96 [0.164]	F(1,137)=0.01 [0.933]	F(1,137)=1.96 [0.164]	F(1,137)=1.54 [0.217]
(5) T-invariance	R	F(3,121)=1.08 [0.359]	**F(3,121)=7.26 [0.000]**	F(3,121)=1.08 [0.359]	F(3,121)=1.49 [0.221]
	S	F(3,137)=1.45 [0.230]	F(3,137)=2.37 [0.073]	F(3,137)=1.45 [0.230]	F(3,137)=1.51 [0.216]
		GG1999			
		Labor share		Output gap	
		π_t^p	x_t	π_t^p	x_t
(1) Normality		χ(2)=0.20 [0.906]	**χ(2)=13.02 [0.002]**	χ(2)=0.20 [0.906]	**χ(2)=5.95 [0.051]**
(2) Linearity		F(6,106)=1.93 [0.083]	F(6,106)=1.12 [0.355]	F(6,106)=1.93 [0.083]	F(6,106)=1.68 [0.133]
(3) Homoskedasticity		F(6,135)=0.29 [0.942]	F(6,135)=0.49 [0.817]	F(6,135)=0.29 [0.942]	F(6,135)=0.90 [0.498]
(4) Markov(4) dependence	R	F(6,106)=0.88 [0.513]	F(6,106)=0.65 [0.692]	F(6,106)=0.88 [0.513]	F(6,106)=0.10 [0.997]
	S	F(1,135)=0.02 [0.897]	F(1,135)=0.15 [0.701]	F(1,135)=0.02 [0.897]	F(1,135)=0.02 [0.885]
(5) T-invariance	R	F(3,106)=1.12 [0.345]	**F(3,106)=6.98 [0.000]**	F(3,106)=1.12 [0.345]	F(3,106)=1.80 [0.152]
	S	F(3,135)=0.68 [0.568]	F(3,135)=1.06 [0.367]	F(3,135)=0.68 [0.568]	F(3,135)=1.86 [0.139]

Note: R and S denote the auxiliary regression and skedastic functions, respectively.

Table 6.4: M-S Testing Results for the ISMs behind the NKPC, the Full sample (1960Q1-2019Q3)

		Small set			
		Labor share		Output gap	
		π_t^p	x_t	π_t^p	x_t
(1) Normality		**χ(2)=10.2 [0.006]**	**χ(2)=7.25 [0.027]**	**χ(2)=7.35 [0.025]**	**χ(2)=14.97 [0.001]**
(2) Linearity		F(2,223)=1.96 [0.143]	F(2,223)=0.45 [0.641]	F(2,223)=2.95 [0.055]	F(2,223)=0.55 [0.576]
(3) Homoskedasticity		**F(2,231)=4.17 [0.017]**	F(2,231)=2.85 [0.060]	**F(2,231)=8.16 [0.000]**	**F(2,231)=4.86 [0.009]**
(4) Markov(4) dependence	R	**F(2,223)=3.21 [0.042]**	F(2,223)=0.23 [0.791]	F(2,223)=1.09 [0.340]	F(2,223)=0.23 [0.796]
	S	F(1,231)=0.05 [0.825]	F(1,231)=0.00 [0.968]	F(1,231)=0.19 [0.663]	F(1,231)=0.01 [0.943]
(5) T-invariance	R	**F(3,223)=3.53 [0.016]**	**F(3,223)=3.11 [0.027]**	F(3,223)=2.48 [0.062]	**F(3,223)=2.74 [0.044]**
	S	**F(3,231)=2.81 [0.040]**	**F(3,231)=2.98 [0.032]**	**F(3,231)=3.10 [0.028]**	F(3,231)=1.74 [0.159]
		GGLS2001			
		Labor share		Output gap	
		π_t^p	Labor share	π_t^p	Output gap
(1) Normality		**χ(2)=9.98 [0.007]**	**χ(2)=9.98 [0.007]**	**χ(2)=9.98 [0.007]**	**χ(2)=10.6 [0.005]**
(2) Linearity		**F(4,216)=2.60 [0.037]**	**F(4,216)=5.37 [0.000]**	**F(4,216)=2.60 [0.037]**	F(4,216)=1.75 [0.140]
(3) Homoskedasticity		**F(4,229)=4.82 [0.001]**	F(4,229)=1.18 [0.319]	**F(4,229)=4.82 [0.001]**	F(4,229)=2.07 [0.085]
(4) Markov(4) dependence	R	F(4,216)=0.69 [0.602]	F(4,216)=0.99 [0.417]	F(4,216)=0.69 [0.602]	F(4,216)=0.81 [0.518]
	S	F(1,229)=0.20 [0.658]	F(1,229)=0.11 [0.740]	F(1,229)=0.20 [0.658]	F(1,229)=0.20 [0.652]
(5) T-invariance	R	F(3,216)=2.57 [0.055]	**F(3,216)=8.56 [0.000]**	F(3,216)=2.57 [0.055]	**F(3,216)=3.63 [0.014]**
	S	F(3,229)=2.54 [0.058]	**F(3,229)=3.15 [0.026]**	F(3,229)=2.54 [0.058]	F(3,229)=1.07 [0.364]
		GG1999			
		Labor share		Output gap	
		π_t^p	x_t	π_t^p	x_t
(1) Normality		**χ(2)=11.6 [0.003]**	**χ(2)=15.17 [0.001]**	**χ(2)=11.6 [0.003]**	**χ(2)=11.12 [0.004]**
(2) Linearity		**F(6,197)=2.76 [0.013]**	**F(6,197)=3.03 [0.007]**	**F(6,197)=2.76 [0.013]**	F(6,197)=2.14 [0.051]
(3) Homoskedasticity		**F(6,227)=2.57 [0.020]**	F(6,227)=1.19 [0.315]	**F(6,227)=2.57 [0.020]**	F(6,227)=1.00 [0.425]
(4) Markov(4) dependence	R	**F(3,197)=4.44 [0.005]**	**F(3,197)=6.14 [0.001]**	**F(3,197)=4.44 [0.005]**	**F(3,197)=3.44 [0.018]**
	S	F(1,227)=0.01 [0.913]	F(1,227)=0.02 [0.895]	F(1,227)=0.01 [0.913]	F(1,227)=0.61 [0.436]
(5) T-invariance	R	F(6,197)=0.69 [0.656]	F(6,197)=0.62 [0.717]	F(6,197)=0.69 [0.656]	F(6,197)=0.35 [0.911]
	S	**F(3,227)=2.74 [0.044]**	F(3,227)=1.85 [0.140]	**F(3,227)=2.74 [0.044]**	F(3,227)=0.92 [0.433]

Note: R and S denote the auxiliary regression and skedastic functions, respectively.

6.4 Respecification

6.4.1 The General Approach

The respecification for the 18 ISMs is divided into two groups. The first group includes three ISMs with the output gap from the small instrument set of the Great Moderation sample. The second group consists of the remaining 15 cases.

The M-S testing results from the first group identify only one statistically adequate ISM - the one with the small instrument set. The remaining two are nearly adequate with one departure of the (2) linearity assumption for the GGLS2001 instrument set and one of the first-order (4) Markov dependence for the GG1999 set. The invalidity of the (4) assumption can be fixed within the Normal VAR/DLR models by choosing the appropriate lag lengths.

In the second group, more serious assumptions departures are needed to be taken into account. Firstly, evidence for heteroskedasticity and variance-heterogeneity are widespread detected, especially in the Full sample. Secondly, the Normality assumption is found invalid for all models with the labor share and the output gap as the regressors. These findings suggest that a Normality-based homogeneous-only model, such as the Normal VAR/DLR models, cannot be used for respecification. Thirdly, the (5) first-order t-invariance is not often met across the fifteen cases. It suggests a presence of mean-changing trends.

This information leads to the usage of the Student's *t* VAR (StVAR) model, as developed in Spanos (1994) and Poudyal (2012), for respecification. The StVAR model is an extension of the classic multivariate Normal VAR in a way that heteroskedasticity and variance-heterogeneity are modeled by as a set of internally consistent and coherent assumptions. Then, the quadratic trends are also included to account for departures from the mean t-invariance assumption. Table 6.5 provides a full description of the quadratic-trend StVAR(3) that is the main model in this part's respecification[2]. The parameter ν of the StVAR is selected to be 3. In the respecification process, one outlier for the output gap at the fourth quarter of 2008 and one for the labor share at the first quarter of 2001 were detected; the dummy variables for these time points are included in the appropriate samples.

The M-S testing for the StVAR model is described in Appendix E. The next three subsections will present the M-S testing results from each of the three sample periods.

6.4.2 Great Moderation Sample

Table 6.6 provided the M-S testing results for the Great Moderation sample. The number of observations is not sufficient for the R package to perform the StVAR estimation with GG1999 instrument set[3]. Respecification with the GG1999 instrument set is then not performed. Therefore, there are only four M-S testing results for the four respecified ISMs. From the M-S testing results, all respecified models are found to be statistically adequate.

2. As in other chapters, the two trends are transformed by the Chebyshev polynomials to prevent the problem of multicollinearity.

3. All estimations for the StVAR models use the "StVAR" package in R.

Table 6.5: Quadratic-trend Student's t VAR(3;ν) Model

Statistical GM: $\mathbf{Z}_t = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t + \boldsymbol{\delta}_2 t^2 + \mathbf{A}_1'\mathbf{Z}_{t-1} + \mathbf{A}_2'\mathbf{Z}_{t-2} + \mathbf{A}_3'\mathbf{Z}_{t-3} + \mathbf{u}_t,\ t \in \mathbb{N}$	
(1) Student's t:	$D(\mathbf{Z}_t, \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, \mathbf{Z}_{t-3}; \nu; \boldsymbol{\Theta})$ is Student's t with v d.f.
(2) Linearity:	$E(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^0) = \boldsymbol{\delta}_0 + \boldsymbol{\delta}_1 t + \boldsymbol{\delta}_2 t^2 + \mathbf{A}_1'\mathbf{Z}_{t-1} + \mathbf{A}_2'\mathbf{Z}_{t-2} + \mathbf{A}_3'\mathbf{Z}_{t-3} + \mathbf{u}_t$
(3) Heteroskedasticity:	$\mathrm{Var}(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^0)) = \left(\frac{\nu}{\nu+3k-2}\right) \boldsymbol{\Omega} q(\mathbf{Z}_{t-1}^0)$
	$q(\mathbf{Z}_{t-1}^0) = \left\{ 1 + \frac{1}{\nu} \begin{bmatrix} \mathbf{Z}_{t-1} - \boldsymbol{\mu}_z(t) \\ \mathbf{Z}_{t-2} - \boldsymbol{\mu}_z(t) \\ \mathbf{Z}_{t-3} - \boldsymbol{\mu}_z(t) \end{bmatrix}' \boldsymbol{\Sigma}_{22}^{-1} \begin{bmatrix} \mathbf{Z}_{t-1} - \boldsymbol{\mu}_z(t) \\ \mathbf{Z}_{t-2} - \boldsymbol{\mu}_z(t) \\ \mathbf{Z}_{t-3} - \boldsymbol{\mu}_z(t) \end{bmatrix} \right\}$
(4) Markov:	$\mathbf{Z}_t$ is a Markov(3) process
(5) t-invariance:	$\boldsymbol{\Theta} := (\boldsymbol{\delta}_0, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\mu}_z, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \boldsymbol{\Omega}, \boldsymbol{\Sigma}_{22})$ are not changing with t, $t \in \mathbb{N}$

Note: see Poudyal (2012), page 89 for a complete explanation of all symbols.

6.4.3 Post-1983 Sample

Table 6.7 shows the M-S testing results for the post-1983 sample. As in the Great Moderation, two ISMs with the GG1999 instrument set are not provided due to an insufficient number of observations. From the M-S testing results, two respecified models under the Small set are statistically adequate. The other two respecified models under the GGLS2001 set are nearly statistically adequate. There are two assumption departures: one is the (2) linear assumption for the labor share's ISM, and the other is the (3) Markov(3) dependence for the output gap's ISM. Although statistical adequacy can only be achieved when all assumptions are valid, one concession is made in this case. The two departures are now considered spurious; the four respecified ISMs are statistically adequate. This choice provides two more ISMs to forward the GMM estimation later.

6.4.4 Full Sample

Under the Full sample, the M-S testing results in Table G.1 (Appendix G) show that none of the respecified StVAR ISMs are statistically adequate. The least assumption departures come under the small instrument set, and the most are from the largest GG1999 set. Among the five assumptions, the (2) linearity is found to be the most invalid. The other departures spread nearly evenly among the (1) Student's t distribution, the (3) heteroskedasticity, and the (5) second-order t-invariance. These results indicate that StVAR is not adequate to model inflation for the Full sample. Misspecifications of the respecified StVAR ISMs under the Full sample can be explained by the inclusion of the highly volatile 1970s. Due to major institutional changes in the US economy during the early 1980s, the behaviors of macroeconomic indicators before and after the first years of the 1980s can be seen as to be completely different. It may cause the StVAR model's failure as a single statistical generation mechanism to accommodate all complex mean and variances heterogeneity of the inflation series.

Table 6.6: M-S Testing Results for the Respecified Normal VAR and St-VAR ISMs: the Great Moderation Sample (1983Q1-2007Q3)

		Small instrument set			
		x_t=labor share, StVAR(3)		x_t=output gap, VAR(3)	
		π_t^p	x_t	π_t^p	x_t
(1) Normal/Student's t		$\chi(2)$=1.44 [0.487]	$\chi(2)$=4.02 [0.133]	$\chi(2)$=1.90 [0.386]	$\chi(2)$=0.15 [0.927]
(2) Linearity		F(2,80)=0.61 [0.548]	F(2,80)=2.34 [0.103]	F(2,81)=0.08 [0.926]	F(2,81)=1.06 [0.350]
(3) Homosk./Heterosk.		F(2,89)=0.76 [0.471]	F(2,89)=0.59 [0.556]	F(2,88)=0.66 [0.517]	F(2,88)=1.16 [0.320]
(4) Markov(p) dependence	R	F(2,80)=2.80 [0.067]	F(2,80)=1.21 [0.328]	F(2,81)=0.24 [0.787]	F(2,81)=0.49 [0.613]
	S	F(1,89)=3.26 [0.074]	F(1,89)=0.21 [0.649]	F(1,88)=1.96 [0.166]	F(1,88)=0.01 [0.929]
(5) t-invariance	R	F(3,80)=0.67 [0.575]	F(3,80)=1.17 [0.328]	F(3,81)=0.50 [0.686]	F(3,81)=0.42 [0.740]
	S	F(1,89)=1.42 [0.236]	F(1,89)=0.40 [0.527]	F(3,88)=2.19 [0.095]	F(3,88)=1.36 [0.261]
		GGLS2001 instrument set			
		x_t=labor share, StVAR(3)		x_t=output gap, VAR(3)	
		π_t^p	x_t	π_t^p	x_t
(1) Normal/Student's t		$\chi(2)$=0.94 [0.625]	$\chi(2)$=3.10 [0.212]	$\chi(2)$=0.73 [0.693]	$\chi(2)$=0.61 [0.738]
(2) Linearity		F(4,70)=0.79 [0.537]	F(4,70)=2.25 [0.072]	F(4,71)=0.53 [0.713]	F(4,71)=1.69 [0.163]
(3) Homosk./Heterosk.		F(4,87)=0.19 [0.942]	F(4,87)=0.26 [0.904]	F(4,86)=0.18 [0.948]	F(4,86)=0.31 [0.871]
(4) Markov(p) dependence	R	F(4,70)=0.89 [0.473]	F(4,70)=2.15 [0.083]	F(4,71)=0.92 [0.458]	F(4,71)=1.15 [0.340]
	S	F(1,87)=2.63 [0.108]	F(1,87)=0.54 [0.467]	F(1,86)=0.51 [0.476]	F(1,86)=0.26 [0.613]
(5) t-invariance	R	F(1,70)=0.85 [0.360]	F(1,70)=1.54 [0.219]	F(3,71)=1.27 [0.291]	F(3,71)=0.62 [0.605]
	S	F(1,87)=1.02 [0.316]	F(1,87)=0.03 [0.856]	F(3,86)=0.62 [0.607]	F(3,86)=0.68 [0.567]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.

Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

Table 6.7: M-S Testing Results for the Respecified St-VAR ISMs: the Post-1983 Sample (1983Q1-2019Q3)

		Small instrument set			
		Labor share, StVAR(3)		Output gap, StVAR(3)	
		π_t^p	x_t	π_t^p	x_t
(1) Student's t		$\chi(2)$=2.03 [0.327]	$\chi(2)$=3.863 [0.145]	$\chi(2)$=5.37 [0.057]	$\chi(2)$=2.03 [0.362]
(2) Linearity		F(2,128)=0.34 [0.711]	F(2,128)=1.22 [0.300]	F(2,128)=0.87 [0.421]	F(2,128)=0.19 [0.827]
(3) Heteroskedasticity		F(2,137)=0.15 [0.849]	F(2,137)=1.03 [0.361]	F(2,137)=0.89 [0.411]	F(2,137)=0.04 [0.960]
(4) Markov(p) dependence	R	F(2,128)=2.01 [0.138]	F(2,128)=0.04 [0.598]	F(2,128)=0.37 [0.694]	F(2,128)=0.13 [0.880]
	S	F(1,137)=2.70 [0.103]	F(1,137)=0.28 [0.000]	F(1,137)=2.26 [0.135]	F(1,137)=0.34 [0.562]
(5) t-invariance	R	F(3,128)=0.23 [0.875]	F(1,128)=0.22 [0.640]	F(3,128)=0.66 [0.577]	F(1,128)=2.21 [0.140]
	S	F(3,137)=0.88 [0.103]	F(1,137)=2.53 [0.114]	F(1,137)=0.67 [0.414]	F(1,137)=0.13 [0.717]
		GGLS2001 instrument set			
		Labor share, StVAR(4)		Output gap, StVAR(3)	
		π_t^p	x_t	π_t^p	x_t
(1) Student's t		$\chi(2)$=0.244 [0.885]	$\chi(2)$=1.356 [0.508]	$\chi(2)$=1.990 [0.370]	$\chi(2)$=2.703 [0.257]
(2) Linearity		F(4,113)=0.83 [0.510]	**F(4,113)=2.78 [0.030]**	F(4,119)=0.53 [0.710]	F(4,119)=2.09 [0.086]
(3) Heteroskedasticity		F(4,134)=0.64 [0.638]	F(4,135)=2.27 [0.065]	F(4,135)=0.79 [0.535]	F(4,135)=0.62 [0.647]
(4) Markov(p) dependence	R	F(4,113)=0.82 [0.513]	F(4,113)=0.76 [0.555]	F(4,119)=0.59 [0.673]	F(4,119)=0.61 [0.654]
	S	F(1,134)=1.87 [0.174]	F(1,135)=1.04 [0.311]	F(1,135)=2.71 [0.102]	**F(1,135)=10.90 [0.000]**
(5) t-invariance	R	F(1,113)=1.88 [0.173]	F(1,113)=0.60 [0.441]	F(1,119)=1.52 [0.220]	F(1,119)=0.75 [0.390]
	S	F(1,134)=0.00 [0.987]	F(1,135)=0.28 [0.599]	F(1,135)=1.77 [0.185]	F(1,135)=0.11 [0.739]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.

Note 2: In Markov(p) dependence, p indicates the highest lags in each model.

6.5 GMM Estimations under Variance-heterogeneity Implicit Statistical Models (ISMs)

6.5.1 GMM Estimation Results

The statistically adequate ISMs provided in the previous section have provided a reliable foundation to explore the GMM estimation of the NKPC. This section will employ these respecified ISMs to obtain reliable GMM estimations and perform the related tests.

As discussed in Chapter 2, the heteroskedastic and autocorrelation consistent (HAC) estimators do not ensure reliable inferences. On the other hand, a statistically adequate ISM can offer a reliable estimator of the variance and covariance matrix $\mathbf{\Omega}$ of the residual u_t. It opens a way to obtain reliable GMM estimators without provoking the HAC estimators. The key point is to identify the weighting matrix $\mathbf{W}$ of the GMM procedure to ensure the estimators' consistency. Appendix F explains in detail the choice of the weighting matrix $\mathbf{W}$ as well as the respectively appropriate GMM estimators.

The new GMM estimation results with the eight statistically adequate respecified ISMs are provided on Table 6.8. For comparison, the original robust GMM estimators, which used the unreliable weighting matrices from the statistically inadequate ISMs, are also included.

In general, the GMM estimations using two types of models generate very similar results. The significances of all estimators are nearly identical; whenever it is significant for an estimator of the original models, it is also significant for the respecified models. All estimated values fall into the ranges of the general findings in the literature (0.7-0.9 for γ_f, 0.1-0.3 for γ_b and very small for λ). In all original and respecified ISMs, the estimated values of the inflation parameters (the forward expected inflation γ_f and the past inflation γ_b) do not differ much. There is a small difference that the estimated values of γ_f in the respecified model become smaller in most cases, while the estimated values of γ_b in the respecified model, in turn, gets bigger. However, the crucial result of γ_f always getting dominant values over γ_b does not change.

The overall result for forcing values under the original and respecified I suggests that both labor share and output gap do not play a role in the formation of the current inflation rate. There are some cases of changing signs (positive turns to negative and reversely) of the estimators for the forcing variable x_t. Nevertheless, their very small values and insignificances make the sign differences unimportant. In short, the estimation results for x_t are analogous to the general results in the literature.

Table 6.8: GMM Estimators of the Original Estimations and the Respecified ISMs: the Post-1983 and Great Moderation Samples

	The Post-1983 sample (1983Q1-2019Q3)							
	x_t= labor share				x_t= output gap			
	Small set		GGLS2001 set		Small set		GGLS2001 set	
	Original	Respecified	Original	Respecified	Original	Respecified	Original	Respecified
λ	-0.001	0.001	-0.0005	0.0001	0.005	0.004	0.003	0.005
	(0.005)	(0.005)	(0.002)	(0.005)	(0.003)	(0.013)	(0.004)	(0.010)
γ_f	0.806**	0.698**	0.798**	0.774**	0.916**	0.608**	0.758**	0.717**
	(0.197)	(0.205)	(0.135)	(0.142)	(0.212)	(0.228)	(0.114)	(0.134)
γ_b	0.279*	0.275*	0.257**	0.243*	0.167	0.320	0.276**	0.273**
	(0.131)	(0.133)	(0.090)	(0.098)	(0.126)	(0.161)	(0.080)	(0.101)

	The Great Moderation sample (1983Q1-2007Q3)							
	x_t= labor share				x_t= output gap			
	Small set		GGLS2001 set		Small set		GGLS2001 set	
	Original	Respecified	Original	Respecified	Original	Respecified	Original	Respecified
λ	0.001	0.001	-0.0003	-0.002	-0.001	0.005	0.001	0.006
	(0.005)	(0.009)	(0.006)	(0.009)	(0.005)	(0.011)	(0.006)	(0.011)
γ_f	0.735**	0.757**	0.929**	0.863**	0.806**	0.879**	0.931**	0.848**
	(0.179)	(0.212)	(0.151)	(0.188)	(0.222)	(0.239)	(0.150)	(0.199)
γ_b	0.304**	0.228	0.163	0.166	0.246	0.156	0.162	0.176
	(0.115)	(0.150)	(0.103)	(0.139)	(0.154)	(0.163)	(0.102)	(0.141)

Note: The original GMM estimators use the two-step procedure with the 4-lag Newey-West estimate of the covariance matrix. The standard errors are reported in parentheses. (*) and (**) represent significance levels at 5% and 1% respectively.

6.5.2 GMM Testing Results

One of the most heated debates on the GMM estimation of the NKPC is the problem of weak identification. To shed light on this issue, the first-stage F-statistics are computed to evaluate the strength of the instruments. Although the relevance of the instruments can be assessed directly from the ISMs, this approach is not chosen here. The reason is that one of the endogenous variables - the forward inflation rate π^p_{t+1} - does not include in the ISMs. However, the usage of F test under the GMM equation still offers a distinct advantage. All of the GMM estimations under the respecified ISMs are performed under the condition of homoskedasticity (in the two VAR ISMs for the Great Moderation sample and in the adjusted GMM estimation method for variance-heterogeneity of the remaining StVAR VAR ISMs); see Appendix E. Therefore, the normal F-test can be used to evaluate the hypothesis of the joint significance of the instruments. There is no need to utilize the thresholds suggested by Stock, Wright, and Yogo (2002) to control for the potential presence of heteroskedasticity and serial autocorrelation. Aside from this F-test, the tests for overidentifying restrictions and exogeneity are also performed further to assess the statistical reliability of the GMM estimations.

Table 6.9 shows the testing results for the three types of tests. The weak instrument test results are reported by two F-statistics for the two endogenous variables (π^p_{t+1} and x_t). For comparison, the testing results for the original estimations are also provided. In all original and respecified ISMs, the F-statistics are consistently small for π^p_{t+1} and identically very large for x_t. These big F-statistics values for x_t provide p-values well below the 1% significance level and much exceed the rule-sof-thumb 10. Thus, it clearly shows that all instrument sets for x_t are truly relevant.

However, the case for π^p_{t+1} is less straightforward. In the original GMM estimations, all of the F-values for π^p_{t+1} are lower than the rule-of-thumb threshold suggested by Stock, Wright, and Yogo (2002) at 10. The literature interprets this result as a concern for the weak identification; see Mavroeidis, Plagborg-Moller, and Stock (2014). From the statistically adequate respecified ISMs, the F-statistics are similarly small across the two samples and the two choices of x_t. On the other hand, when the normal F-test critical values are used, the relatively small F-statistics values result in all p-values (of the hypothesis that all instruments jointly are different from zero) are *smaller* than 1%. In other words, there is no evidence that the instruments are not weak. This result suggests the concern of the weak instruments owes mostly to the usage of the rule-of-thumb threshold rather than to tiny correlations between the instruments and the endogenous π^p_{t+1}. In short, the weak instrument test results suggest the problem of weak instruments does not exist for the GMM estimations under the respecified models.

For the overidentifying restrictions tests, the Sargan score's test results under the respecified models provide no evidence against the validity of the overidentifying restrictions (the Sargan score is selected because of homoskedasticity). It is slightly different from the original estimations that evidence of overidentifying restrictions (from Hansen's J statistics) is only found in a few cases across the eight ISMs.

The test results for exogeneity between the original and the respecified estimations are more contrasting. For the estimations under the respecified ISMs, the high values of the Durbin score present evidence for the problem of exogeneity (the Durbin test is used due to homoskedasiticity). On the contrary, the GMM-general C statistics for all original estimations do not detect the presence of endogeneity; see Hayashi (2000) for the explanation of C statistics.

In conclusion, the GMM estimations from the statistically adequate ISMs produce similar results to the misspecified ISMs' ones. It suggests that various choices of the weighting matrix do not alter much the relative estimated values of γ_f and γ_b. The following section will analyze the computation aspect of the GMM estimations to understand why it is the case.

Table 6.9: GMM Testing Results for the Original Estimations and the Respecified ISMs: the Post-1983 and Great Moderation Samples

	The Post-1983 sample (1983Q1-2019Q3)								
		x_t= labor share				x_t= output gap			
		Small set		GGLS2001 set		Small set		GGLS2001 set	
		Original	Respecified	Original	Respecified	Original	Respecified	Original	Respecified
Weak instrument	π^p_{t+1}	2.814	2.834	3.841	2.601	3.321	3.258	3.841	3.834
	x_t	505.708	256.716	428.014	143.890	658.943	182.365	380.961	102.252
Overidentifying restriction test		11.286	7.216	6.933	11.212	3.039	4.548	6.555	11.829
		[0.024]	[0.301]	[0.436]	[0.796]	[0.551]	[0.603]	[0.477]	[0.377]
Exogeneity test		1.754	4.815	4.421	10.363	4.371	2.601	3.167	7.867
		[0.416]	[0.090]	[0.110]	[0.006]	[0.112]	[0.272]	[0.205]	[0.020]
	The Great Moderation sample (1983Q1-2007Q3)								
		x_t= labor share				x_t= output gap			
		Small set		GGLS2001 set		Small set		GGLS2001 set	
		Original	Respecified	Original	Respecified	Original	Respecified	Original	Respecified
Weak instrument	π^p_{t+1}	2.741	3.052	5.264	2.283	4.470	3.226	5.264	2.052
	x_t	148.047	90.854	88.631	63.761	392.555	191.422	221.116	92.066
Overidentifying restriction test		6.464	4.906	5.951	9.521	5.158	4.049	6.031	8.292
		[0.167]	[0.427]	[0.545]	[0.574]	[0.272]	[0.256]	[0.536]	[0.505]
Exogeneity test		0.929	5.226	2.485	10.178	1.150	6.100	2.018	9.006
		[0.629]	[0.073]	[0.289]	[0.006]	[0.563]	[0.047]	[0.365]	[0.011]

Note: The weak instrument tests report the F-statistics from the first stage. The overidentifying restrictions tests report the Sargan score for the respecified models and the Hansen J statistics for the original models. The exogeneity tests report the Durbin test for the respecified models and C statistics for the original models. The numbers in square brackets represent the p-values.

6.6 Understanding Computation of the GMM Estimators of the NKPC

Mavroeidis, Plagborg-Moller, and Stock (2014) run a 2-million point-estimate sensitivity check on the GMM estimations for NKPC across various GMM variants, variable measurements, sample periods, instrument sets, and other choices. The paper finds that the estimators of the forward inflation π^p_{t+1} (γ_f) always dominates the estimator of the past inflation π^p_{t-1} (γ_b). The median of the estimators for γ_f is approximately at 0.75 and λ at nearly zero. However, they only found approximately 600,000 of these estimates belonging to the confidence intervals of the estimated values from the ground-breaking paper Galı and Gertler (1999). They interpret their findings as a sign of weak identification.

In the previous section, the GMM estimations using the statistically adequate ISMs provide similar results as previously found in the literature. The section will decipher why it is the case from the computation point of view. Differed from Mavroeidis, Plagborg-Moller, and Stock (2014), it is argued that the GMM estimation of NPKC reflects the general autoregressive structure of the underlying stochastic process rather than offering any real insights into the behavior of the inflation rate.

From the computation point of view, the linear GMM method, as in the case of the NKPC, is an extension of the two-stage least-square (2SLS) with a weighting matrix. In the 2SLS, the first step is to regress the endogenous variables against the instruments and obtain the fitted values as the new RHS variables:

$$\widehat{\mathbf{X}}_t = \mathbf{Z}_t(\mathbf{Z}_t'\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{X}_t \tag{6.15}$$

Then, the new fitted values replace the original RHS variables; under the assumption of homoskedasticity, a new OLS estimator is computed as the 2SLS's:

$$\hat{\boldsymbol{\beta}}_{2\text{SLS}} = (\widehat{\mathbf{X}}_t'\widehat{\mathbf{X}}_t)^{-1}\widehat{\mathbf{X}}_t'\mathbf{y}_t = (\mathbf{X}_t'\mathbf{Z}_t(\mathbf{Z}_t'\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Z}_t(\mathbf{Z}_t'\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{y}_t \tag{6.16}$$

The GMM estimator is only different from the 2SLS estimator by adding a weighting matrix $\mathbf{W}$:

$$\hat{\boldsymbol{\beta}}_{\text{GMM}} = (\mathbf{X}_t'\mathbf{Z}_t\mathbf{W}\mathbf{Z}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Z}_t\mathbf{W}\mathbf{Z}_t'\mathbf{y}_t, \tag{6.17}$$

with $\mathbf{W} = (\frac{1}{n}E(\mathbf{Z}_t'u_tu_t'\mathbf{Z}_t))^{-1} = (\frac{1}{n}E(\mathbf{Z}_t'\boldsymbol{\Omega}\mathbf{Z}_t))^{-1}$ as the inverse of a variance-covariance matrix of the moment conditions.

To obtain the weighing matrix $\mathbf{W}$, there are two computation approaches: the two-step (2STEP) and the continuously update (CEU) GMMs. The 2STEP assumes that the obtained residual $\hat{u}_t$ is a consistent estimate of the error term u_t. Then, $\hat{u}_t$ is used to obtain a 'hat' matrix:

$$\hat{\Omega} = \begin{pmatrix} \hat{u}_1^2 & & & & 0 \\ & \ddots & & & \\ & & \hat{u}_t^2 & & \\ & & & \ddots & \\ 0 & & & & \hat{u}_n^2 \end{pmatrix} \tag{6.18}$$

The $\hat{\Omega}$ is then plugged into Equation 6.16 of the 2SLS - as the weighting matrix $\widehat{\mathbf{W}} = (\frac{1}{n}(\mathbf{Z}_t'\hat{\Omega}^{-1}\mathbf{Z}_t))^{-1}$ - to get the 2STEP GMM estimators:

$$\hat{\boldsymbol{\beta}}_{\text{GMM-2STEP}} = (\mathbf{X}_t'\mathbf{Z}_t(\mathbf{Z}_t'\hat{\Omega}^{-1}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Z}_t(\mathbf{Z}_t'\hat{\Omega}^{-1}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{y}_t \tag{6.19}$$

Further adjusting of the weighting matrix $\hat{\Omega}$ can be performed to provide the HAC GMM estimators. Under the CUE-GMM, the second step of the 2STEP is repeatedly performed until some convergence criteria for $\widehat{\mathbf{W}}$ are met.

If homoskedasticity is assumed, the $\hat{\Omega}$ equals the multiplicity between an identity matrix and a constant ($\hat{\Omega} = \sigma\mathbf{I}_n$). Then, the 2SLS estimators are equal to the GMM estimators $\hat{\boldsymbol{\beta}}_{\text{2SLS}} = \hat{\boldsymbol{\beta}}_{\text{GMM-2STEP}}$.

As explored in the GMM estimations using the weighting matrices from the statistically adequate ISMs, the choices of the weighting matrix do not significantly change the estimators of the NPKC. It is reasonable to understand the nature of the computation of the GMM estimators from the simplest case of the 2SLS.

The forcing variable x_t is next ignored to simplify the analysis. As shown by Mavroeidis, Plagborg-Moller, and Stock (2014), the estimated values of the forcing variables x_t (both the output gap and the labor share) are nearly zero and often insignificant at 5%. Also, Atkeson and Ohanian (2001) and Stock and Watson (2007) find that the US's inflation has become more difficult to forecast; this result suggests that the forcing variable is an irrelevant factor. There, putting x_t aside will not dramatically change the analysis's overall conclusion due to its negligible role. The simplified estimated equation only with the inflation variables is now:

$$\pi_t^p = \gamma_0 + \gamma_b\pi_{t-1}^p + \gamma_f\pi_{t+1}^p + u_t, \tag{6.20}$$

where π_{t+1}^p is the only endogenous variable and four lags of π_t^p are used as the instruments.

Under the normal OLS, the estimated γ_b and γ_f are computed as:

$$\hat{\gamma}_b = \frac{\text{Cov}(\pi_t^p, \pi_{t-1}^p)\text{Var}(\pi_{t+1}^p) - \text{Cov}(\pi_t^p, \pi_{t+1}^p)\text{Cov}(\pi_{t-1}^p, \pi_{t+1}^p)}{\text{Var}(\pi_{t-1}^p)\text{Var}(\pi_{t+1}^p) - (\text{Cov}(\pi_{t-1}^p, \pi_{t+1}^p))^2} \tag{6.21}$$

$$\hat{\gamma}_f = \frac{\text{Cov}(\pi_t^p, \pi_{t+1}^p)\text{Var}(\pi_{t-1}^p) - \text{Cov}(\pi_t^p, \pi_{t-1}^p)\text{Cov}(\pi_{t-1}^p, \pi_{t+1}^p)}{\text{Var}(\pi_{t-1}^p)\text{Var}(\pi_{t+1}^p) - (\text{Cov}(\pi_{t-1}^p, \pi_{t+1}^p))^2} \tag{6.22}$$

To simplify the notations, the variance of π_{t-i}^p is called Var_i, and the covariance between π_{t-i}^p and π_{t-j}^p ($i, j \in \{-1, 0, 1\}$) as $\text{Cov}_{i,j}$. Then, Equations 6.21 and 6.22 can be written as:

$$\hat{\gamma}_b = \frac{\text{Cov}_{0,-1}\text{Var}_{+1} - \text{Cov}_{0,+1}\text{Cov}_{-1,+1}}{\text{Var}_{-1}\text{Var}_{+1} - \text{Cov}_{-1,+1}^2} \tag{6.23}$$

$$\hat{\gamma}_f = \frac{\text{Cov}_{0,+1}\text{Var}_{-1} - \text{Cov}_{0,-1}\text{Cov}_{-1,+1}}{\text{Var}_{-1}\text{Var}_{+1} - \text{Cov}_{-1,+1}^2} \tag{6.24}$$

When the OLS is used to estimate Equation 6.20, the values of $\hat{\gamma}_b$ and $\hat{\gamma}_f$ are simply equal. It is because the covariance between inflation and its one-period lag are nearly identical ($\text{Cov}_{-1,0} = \text{Cov}_{0,+1}$), and its variance is a constant ($\text{Var}_{-1} = \text{Var}_{+1}$) within the same sample. Their sum $\hat{\gamma}_b + \hat{\gamma}_f \approx \frac{2\text{Cov}_{0,-1}}{\text{Var}_{-1}+\text{Cov}_{0,-1}}$ is a little lower than 1 (because $\text{Cov}_{0,-1}$ is a little smaller than Var_{-1}).

Now, to understand the relative dominance of $\hat{\gamma}_f$ over $\hat{\gamma}_b$, it is needed to look at how the fitted values $\hat{\pi}^p_{t+1}$ are transformed after the first step of the 2SLS procedure. Considering the exemplary case that four lags of π^p_t are used as the instruments, the first step of the 2SLS procedure is to regress the one-period forward rate π^p_{t+1} on its four lags (the current is rate is not included). This regression is equivalent to regressing the current rate π^p_t on the second to the fifth lags.

Next, to understand the first step of the 2SLS, we need to take a look at the autoregressive structure of π^p_t. The autoregressive structure can be revealed by regressing π^p_t on its four lags. The estimated equation with the full sample is obtained as[4]:

$$\pi^p_t = 0.051 + 0.622\pi^p_{t-1} + 0.096\pi^p_{t-2} + 0.136\pi^p_{t-3} + .084\pi^p_{t-4}, \tag{6.26}$$

with the sum of estimated coefficients of all lags being at 0.938.

This estimation shows that the autocorrelation function is positive and uniformly decreasing (Figure 6.1 also exhibits the autocorrelation graph). In other words, the further backward, the less correlation it is between the current rate π^p_t and its lags. As a weighted sum of the many lags, the correlations between the fitted values $\hat{\pi}^p_{t+1}$ and π^p_t or π^p_{t-1} are smaller than from the original π^p_{t+1}. For example, $\text{Cov}(\pi^p_t, \widehat{\pi}^p_{t+1}) < \text{Cov}(\pi^p_t, \pi^p_{t+1})$ with $\widehat{\pi}^p_{t+1}$ being the fitted value from the first step of the 2SLS.

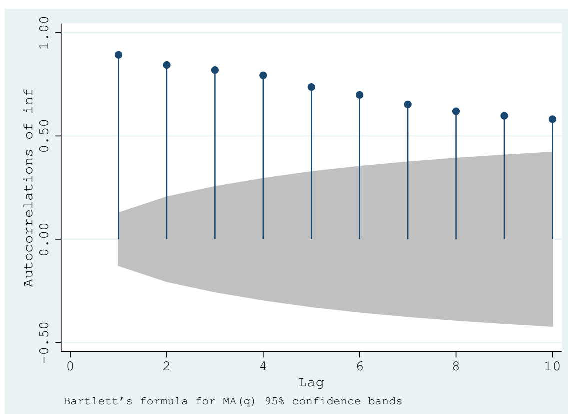


Figure 6.1: The Autocorrelation (AC) Function for π^p_t, the Full Sample

When the fitted value of π^p_{t+1} replaces the RHS variables in the second step of the 2SLS procedure, the 2SLS estimator is obtained as:

$$\hat{\gamma}_b^{2SLS} = \frac{\text{Cov}_{0,-1}\widehat{\text{Var}}_{+1} - \widehat{\text{Cov}}_{0,+1}\widehat{\text{Cov}}_{-1,+1}}{\text{Var}_{-1}\widehat{\text{Var}}_{+1} - \widehat{\text{Cov}}^2_{-1,+1}} \tag{6.27}$$

$$\hat{\gamma}_f^{2SLS} = \frac{\widehat{\text{Cov}}_{0,+1}\text{Var}_{-1} - \text{Cov}_{0,-1}\widehat{\text{Cov}}_{-1,+1}}{\text{Var}_{-1}\widehat{\text{Var}}_{+1} - \widehat{\text{Cov}}^2_{-1,+1}} \tag{6.28}$$

4. The estimated equation of π^p_t over its four lags is:

$$\pi^p_{t+1} = 0.091 + 0.508\pi^p_{t-1} + 0.203\pi^p_{t-2} + 0.249\pi^p_{t-3} - .070\pi^p_{t-4} \tag{6.25}$$

where $\widehat{\text{Var}}_{+1}$, $\widehat{\text{Cov}}_{0,+1}$, and $\widehat{\text{Cov}}_{-1,+1}$ represent the variance and covariances (with π_t^p and π_{t-1}^p) of the fitted value $\hat{\pi}_{t+1}^p$.

The smaller correlations of $\hat{\pi}_{t+1}^p$ provide:

- $\widehat{\text{Var}}_{+1} < \widehat{\text{Cov}}_{0,+1}$ that makes $\hat{\gamma}_f^{2SLS}$ bigger than $\hat{\gamma}_b^{2SLS}$ in Equations 6.27 and 6.28.

- $\widehat{\text{Cov}}_{0,+1} < \text{Cov}_{0,-1}$ that makes $\hat{\gamma}_b^{2SLS}$ bigger than $\hat{\gamma}_f^{2SLS}$ in Equations 6.27 and 6.28.

Because the denominator is the same, these two differences can cancel out each other and cause $\hat{\gamma}_b^{2SLS} = \hat{\gamma}_f^{2SLS}$. However, because the multiplier factor $\text{Var}_{-1} > \text{Cov}_{0,-1}$, it provides that $\hat{\gamma}_f^{2SLS} > \hat{\gamma}_b^{2SLS}$. This is reason why the estimators of the forward inflation rate dominate over the estimators of the past rate.

Although it is not logically exhausted, this analysis suggests that the dominant role of γ_f is caused by regressing the forward inflation rate on its lags in the first stage. Any time series, which share the same decreasingly positive autocorrelation structure, will provide the exact 2SLS estimation result. In other words, the GMM estimation of the NKPC does not reveal any insight into the behavior of the inflation rate.

Next, a simple simulation is run to see how the relative values of $\hat{\gamma}_f^{2SLS}$ and $\hat{\gamma}_b^{2SLS}$ change from time series with different autoregressive structures. Different values of the 2SLS estimators for γ_f and γ_b obtaining from the simulated series can help to illustrate the main argument of this section. Table 6.10 provides the 2SLS estimators under six different autoregressive models Y_t (Model 1 to Model 6). Four lags of Y_t are used as the instruments under the assumption of homoskedasticity. Model 1 is designed to have a similar autoregressive structure as the estimation obtained from the inflation rate π_t^p under the Full sample. The 2SLS estimation result produces the same dominant $\hat{\gamma}_f^{2SLS}$ at 0.877 over $\hat{\gamma}_b^{2SLS}$ at 0.147. In Model 2, in which the coefficients for the three lags are nearly the same, the 2SLS estimation produces the same superiority of $\hat{\gamma}_f^{2SLS}$ and a negative $\hat{\gamma}_b^{2SLS}$. For Model 3 to Model 5, one of the lags' coefficients is designed to be negative. They result in three different outcomes: both $\hat{\gamma}_f$ and $\hat{\gamma}_b$ are negative in Model 3, nearly equal in Model 4, and $\hat{\gamma}_b^{2SLS}$ dominates over $\hat{\gamma}_f^{2SLS}$ in Model 5. In Model 6, the similar dominance of $\hat{\gamma}_f^{2SLS}$ is preserved although the relative values of the lags reverse (this model is also used to illustrate for the problem of the weak instrument below). In short, the 2SLS estimations from the simulated series clearly show the relative values of $\hat{\gamma}_f^{2SLS}$ and $\hat{\gamma}_b^{2SLS}$ primarily depend on the autoregressive structures of Y_t.

Finally, how can the wide dispersion of the two million estimators getting by Mavroeidis, Plagborg-Moller, and Stock (2014) be explained? The big spread of the point estimators can be seen as a departure from the simplest 2SLS, single-variable case discussed in this section. The variations in the GMM approaches, sample periods, variable choices, and instruments modify the simplest case and lead to a high level of sensitivity. More works can be done with the more complicated cases provided by Mavroeidis, Plagborg-Moller, and Stock (2014) to achieve a full analysis of their results.

The analysis on the computation of the NKPC also sheds light on the weak instrument problem. As discussed in Section 5.5.2, the literature has a big concern about the weak correlation between π_{t+1}^p and the instruments. The weak correlation can be explained from the partial autocorrelation (PAC) graphs.

The first stage of the GMM regresses the forward rate π_{t+1}^p on its four lags. It is equivalent

Table 6.10: GMM/2SLS Estimators under Six Data Simulations

	Data generation mechanism	γ_f^{2SLS}	γ_b^{2SLS}	F-stats
Model 1	$Y_t = 0.031 + 0.664Y_{t-1} + 0.059Y_{t-2} + 0.214Y_{t-3} + u_t$	0.877	0.147	3.40
Model 2	$Y_t = 0.136 + 0.247Y_{t-1} + 0.231Y_{t-2} + 0.251Y_{t-3} + u_t$	1.157	-0.119	8.28
Model 3	$Y_t = 0.676 - 0.653Y_{t-1} + 0.110Y_{t-2} + 0.191Y_{t-3} + u_t$	-0.425	-0.446	7.81
Model 4	$Y_t = 0.275 + 0.377Y_{t-1} - 0.181Y_{t-2} + 0.254Y_{t-3} + u_t$	0.366	0.359	4.03
Model 5	$Y_t = 0.224 + 0.534Y_{t-1} + 0.126Y_{t-2} - 0.107Y_{t-3} + u_t$	0.151	0.563	1.42
Model 6	$Y_t = 0.135 + 0.028Y_{t-1} + 0.131Y_{t-2} + 0.570Y_{t-3} + u_t$	0.689	0.084	30.76

Note: $\hat{\gamma}_f^{2SLS}$ and $\hat{\gamma}_b^{2SLS}$ are estimated by the GMM procedure under the homoskedastic assumption. The instrument includes four lags of y_t. F-stats are the F-statistics values from the first stage.

to regress the current rate on the second to the fifth lags. Figure 6.2 draw the PAC graphs for the three sample periods. The three PAC graphs show the biggest conditional correlations at the first lag. From the second lag, the conditional correlations drop considerably and approach zero at the fourth lag. Moving t one period forward generates the same, little correlation between the forward inflation rate and its lags. Because the RHS of the forcing variable x_t in the first stage has all the closest lags, their F-statistics in the first stage is much higher than ones of π_{t+1}^p. That is the reason why there are huge discrepancies between the F-statistics from the two endogenous variables. The simple simulation above also helps to shed light on the values of the first-stage F-statistics. In the first five models in Table 6.10, the five first-stage F-statistics are lower than the rule-of-thumb threshold at 10 because the coefficient of x_{t-1} is relatively large compared to other lags. However, when it is adjusted to be much smaller (at 0.028) in Model 6, the F-statistic increases to 30.76 that is well above the 10 threshold.

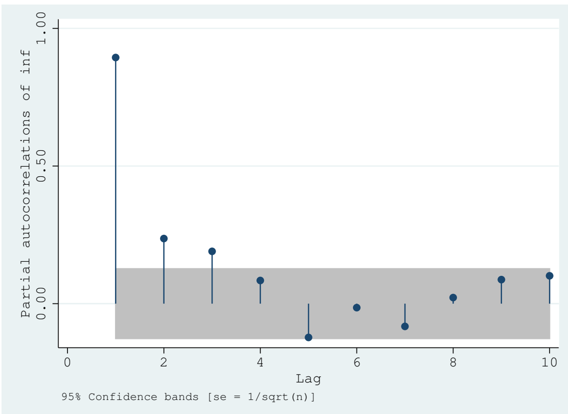
In summary, this section argues that the estimation results from the GMM estimation of the NKPC only reflect the general autoregressive structure of the inflation rate. Any time series with the same autoregressive behavior will generate the same estimation result, and no real understanding for the behavior of inflation is actually revealed. The same autoregressive structure also helps to decipher the problem of weak instruments proposed by the literature.

6.7 Conclusions

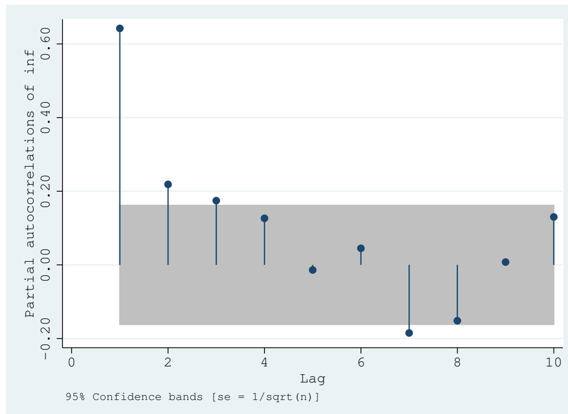
In this chapter, the implicit statistical models behind the GMM estimations of the NKPC are delineated and evaluated. Except for one, all of the implicit statistical models across the three sample periods, three instrument sets, and two forcing variables are found to be statistically inadequate. Respecification with Normal VAR and the StVAR models for the total 18 achieves statistical adequacy for the 8 ISMs under the Post-1983 and the Great Moderation samples. The respecified StVAR models under the Full sample, which contains the high-volatile 1970s remain problematic and exhibit many assumptions departures; this leaves much space for further works on statistical modeling with the samples including the 1970s.

The new GMM estimations with the eight statistically adequate ISMs yield similar results as in the misspecified ISMs. This finding leads to a short analysis of the computation of the

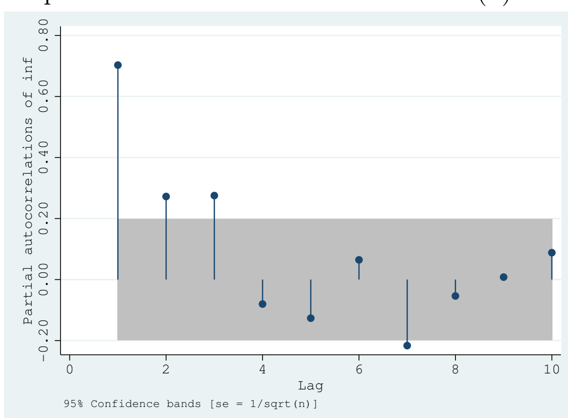
Figure 6.2: The Partial Autocorrelation (PAC) Function for π_t^p



(a) Full sample



(b) Post-1983 sample



(c) Great Moderation sample

GMM estimation for the NKPC. The analysis suggests the general autoregressive structure is the main factor behind the strong dominance of the forward inflation expectation over the past rate. The GMM estimation, indeed, does not provide a profound understanding of the behavior of inflation. However, the computation analysis is just sketchy and leaves much room for improvement.

Chapter 7

Final Conclusions

During the last 60 years, the Phillips curve as a price-expectation-unemployment relationship has survived various significant changes (Kuhnian paradigm shifts) in macroeconomic theory. The curve's survival can be explained by the endless empirical studies produced in the applied macroeconomic literature that provides evidence for some forms of the Phillips curve. This dissertation revisits several important, representative papers in four historical periods to offer a retrospective examination of the empirical evidence presented throughout the four periods, starting from the landmark paper by Phillips (1958) and ending with the latest New Keynesian Phillips Curve (NKPC).

The rereading of the selected papers along the curve's long history highlights the preeminence of the theoretical considerations over the choices of statistical specifications and the validity of the estimated models vis-à-vis the probabilistic structure of the data used. This is typical of many other empirical relationships that are estimated using the theory-driven methodology of textbook econometrics. Due to the commitment to a priori theory, the textbook approach does not sufficiently secure the trustworthiness of the evidence by establishing the validity of the probabilistic assumptions imposed (implicitly or explicitly) on one's data (the implicit statistical model). Departures from any of the probabilistic assumptions (statistical inadequacy) often lead to inconsistent estimators and unreliable inferences due to the actual type I and type II error probabilities being different from their nominal (assumed) levels.

Therefore, this dissertation chooses to apply the Probabilistic Reduction approach (Spanos 2006, 2010), in which the statistical models are viewed as particular parameterizations of the observable stochastic process $\{\mathbf{Z}_t, t \in \mathbb{N}\}$ underlying the data $\mathbf{z}_0$, to forward the statistical evaluations. The evaluations of the key estimated models from several influential and highly cited papers reveal that statistical inadequacy is almost ubiquitous. In other words, the empirical evidence derived in these papers is untrustworthy.

Apart from the almost universal overall misspecification of these key papers, each period also has its specific variants of untrustworthiness. In the first two periods - the curve of Phillips and the wage-determination literature in the 1960s, the Mis-Specification (M-S) testing of the replicated models using the vintage data show that the researchers in that period tended to ignore the temporal dependence in the explanatory variables. The respecification for these early wage Phillips equations succeeds in achieving statistical adequacy. The new estimated results suggest that apart from the unemployment and the price change, the past behavior of wage growth plays a crucial role in explaining the dynamics of the dependent variable.

The focus on the third period during the 1970s is on the empirical validity of the second variant of the Phillips curve - the expectations-augmented Phillips curve. Due to the lack of exploration for the history of empirical Phillips curve during the 1970s, a short survey on the eight most cited papers brings out a disproportionate contribution from Robert Gordon. In light of that, two of his empirical models are chosen to be evaluated. However, given that the original data are lost, a serious attempt was made to reconstruct the vintage data to preserve the historical context as in the previous two periods. The M-S testing results discover similar neglect of the temporal dependence of the explanatory variables as in the previous periods. Despite persistent attempts, the respecification for the original Gordon's models fails to achieve statistical adequacy, leaving several questions of interest unanswered.

The evaluation of the NKPC for the fourth period focuses on its GMM (Generalized Methods of Moments) estimation established by Galı and Gertler (1999), viewed as an extension of the IV (Instrumental Variables) approach. For the GMM and IV estimation methods, the underlying statistical model needs to be explicitly derived before testing its probabilistic assumptions. It turns out that the implicit statistical models underlying the NKPC models are often some variant of the Vector Autoregressive (VAR(p)) and Dynamic Linear Regression (DLR(p)) models whose adequacy needs to be established before their overidentifying restrictions can be reliably tested and the reliability of any inferences based on the structural model can be secured. The extensive M-S testing results show that most of the implicit statistical models underlying the GMM estimation of the NKPC are statistically misspecified. Attempts to respecify them lead to the usage of the Student's t multivariate to replace the universally employed Normal distribution. The crucial difference of the Student's t distribution for the VAR(p) and DLR(p) models is its ability to account for the heteroskedasticity existing from the involving variables. The revised GMM estimations with the respecified models produce similar estimated coefficients as in the original ones. These results led to a short analysis of the nature of the GMM computation of the NKPC that sheds new light on several issues raised by the literature. The dominance of the estimators for the forward inflation over the past rate's and the source of the weak instrument problem can be attributed to the general autoregressive lag structure of the stochastic process. In other words, the GMM estimation does not reveal much insight into the behavior of the inflation rate. The respecification using the St-VAR model for the most extended sample, covering the turbulent 1970s, does not succeed in containing several major structural breaks; it is a topic for further research.

Overall, the results from a small but representative set of evidence based on key papers for each of the four periods can be used to infer the untrustworthiness of the broader literature. The untrustworthiness of the empirical evidence can be viewed as supplementary to a novice historical work, say Forder (2014a), to render the Phillips curve neither as a concrete concept nor as a well-founded empirical relationship in macroeconomics. One might argue that focusing on widely quoted papers could result in missing less noticed studies which might have done a better job in establishing the statistical adequacy of their estimated structural models. The choice of papers to view retrospectively is based on their role in establishing the traditional narrative of the Phillips curve, and the examination of the evidence was broader than all models reported in the dissertation.

Bibliography

Abbas, Syed K., Prasad Sankar Bhattacharya, and Pasquale Sgro. 2016. "The new Keynesian Phillips curve: An update on recent empirical advances." *International Review of Economics & Finance* 43:378–403.

Andrews, Donald W. K. 1993. "Tests for parameter instability and structural change with unknown change point." *Econometrica: Journal of the Econometric Society,* 821–856.

Andrews, Donald W. K., and Werner Ploberger. 1994. "Optimal tests when a nuisance parameter is present only under the alternative." *Econometrica: Journal of the Econometric Society,* 1383–1414.

Atkeson, Andrew, and Lee E. Ohanian. 2001. "Are Phillips curves useful for forecasting inflation?" *Federal Reserve bank of Minneapolis quarterly review* 25 (1): 2–11.

Ball, Laurence M., and N. Gregory Mankiw. 2002. "The NAIRU in theory and practice." *Journal of economic Perspectives* 16 (4): 115–136.

Ball, Laurence M., N. Gregory Mankiw, and David Romer. 1991. "The New Keynesian economics and the output-inflation trade-off." *New Keynesian Economics* 1:147–211.

Ball, Laurence M., and Sandeep Mazumder. 2011. "Inflation dynamics and the great recession." *Brookings Papers on Economic Activity* 42 (1): 337–405.

———. 2019. *The nonpuzzling behavior of median inflation.* Working Paper, 25512. National Bureau of Economic Research.

Blanchard, Olivier J. 2016. "The Phillips Curve: Back to the'60s?" *American Economic Review* 106 (5): 31–34.

———. 2018. "Should we reject the natural rate hypothesis?" *Journal of Economic Perspectives* 32 (1): 97–120.

Bowden, Roger J., and Darrell A. Turkington. 1990. *Instrumental variables.* Cambridge University Press.

Calvo, Guillermo A. 1983. "Staggered prices in a utility-maximizing framework." *Journal of monetary Economics* 12 (3): 383–398.

Canova, Fabio. 2007. *Methods for applied macroeconomic research.* Princeton University Press.

De Vroey, Michel. 2016. *A history of macroeconomics from Keynes to Lucas and beyond.* Cambridge University Press.

Desai, Meghnad. 1975. "The Phillips curve: A revisionist interpretation." *Economica* 42 (165): 1–19.

Dicks-Mireaux, Leslie A., and J. Christopher R. Dow. 1959. "The determinants of wage inflation: United Kingdom, 1946-56." *Journal of the Royal Statistical Society. Series A (General)* 122 (2): 145–184.

Dow, J. Christopher R., and Leslie A. Dicks-Mireaux. 1958. "The excess demand for labour. A study of conditions in Great Britain, 1946-56." *Oxford Economic Papers* 10 (1): 1–33.

Eckstein, Otto, and Roger E. Brinner. 1972. "The inflation process in the United States; a study prepared for the use of the Joint Economic Committee, Congress of the United States."

Fernández-Villaverde, Jesús, Juan F Rubio-Ramırez, Thomas J Sargent, and Mark W Watson. 2007. "ABCs (and Ds) of understanding VARs." *American economic review* 97 (3): 1021–1026.

Forder, James. 2014a. *Macroeconomics and the Phillips curve myth.* Oxford University Press.

———. 2014b. "Nine views of the Phillips curve: Eight authentic and one inauthentic," Oxford University Department of Economics Discussion Paper Series.

Friedman, Milton. 1968. "The role of monetary policy." *American Economic Review* 58 (1): 1–17.

———. 1977. "Nobel lecture: inflation and unemployment." *Journal of political economy* 85 (3): 451–472.

Fuhrer, Jeffrey C., Yolanda K. Kodrzycki, Jane S. Little, Giovanni P. Olivei, and Paul A. Samuelson. 2009. *Understanding inflation and the implications for monetary policy: A Phillips curve retrospective.* MIT Press Cambridge.

Gali, Jordi, Mark Gertler, and J. David Lopez-Salido. 2005. "Robustness of the estimates of the hybrid New Keynesian Phillips curve." *Journal of Monetary Economics* 52 (6): 1107–1118.

Galı, Jordi, and Mark Gertler. 1999. "Inflation dynamics: A structural econometric analysis." *Journal of monetary Economics* 44 (2): 195–222.

Galı, Jordi, Mark Gertler, and J. David Lopez-Salido. 2001. "European inflation dynamics." *European economic review* 45 (7): 1237–1270.

Giacomini, Raffaella. 2013. "The relationship between DSGE and VAR models." *VAR models in macroeconomics–new developments and applications: Essays in honor of Christopher A. Sims (Advances in Econometrics, Vol. 32).*

Gilbert, Christopher L. 1976. "The original Phillips curve estimates." *Economica* 43 (169): 51–57.

Godfrey, Leslie G. 1994. "Testing for serial correlation by variable addition in dynamic models estimated by instrumental variables." *The Review of Economics and Statistics,* 550–559.

Goldstein, Morris. 1972. "The trade-off between inflation and unemployment: a survey of the econometric evidence for selected countries." *International Monetary Fund Staff Papers* 19 (3): 647–698.

Gordon, Robert J. 1970. "The recent acceleration of inflation and its lessons for the future." *Brookings Papers on Economic Activity* 1970 (1): 8–47.

———. 1971. "Inflation in recession and recovery." *Brookings Papers on Economic Activity* 1971 (1): 105–166.

———. 1972. "Wage-price controls and the shifting Phillips curve." *Brookings Papers on Economic Activity* 1972 (2): 385–421.

———. 1977a. "Can the inflation of the 1970s be explained?" *Brookings Papers on Economic Activity* 1977 (1): 253–279.

———. 1977b. "The theory of domestic inflation." *The American Economic Review* 67 (1): 128–134.

———. 1981. *Inflation, flexible exchange rates, and the natural rate of unemployment.* Working Paper, 708. National Bureau of Economic Research.

———. 1997. "The time-varying NAIRU and its implications for economic policy." *Journal of economic Perspectives* 11 (1): 11–32.

———. 2011. "The history of the Phillips curve: Consensus and bifurcation." *Economica* 78 (309): 10–50.

———. 2013. *The Phillips curve is alive and well: Inflation and the NAIRU during the slow recovery.* National Bureau of Economic Research.

———. 2018. "Friedman and Phelps on the Phillips curve viewed from a half century's perspective." *Review of Keynesian Economics* 6 (4): 425–436.

Goutsmedt, Aurélien, and Goulven Rubin. 2018. *Robert J. Gordon and the introduction of the natural rate hypothesis in the Keynesian framework.* Documents de travail du Centre d'Economie de la Sorbonne, 18013. Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.

Granger, Clive W. J., and Yongil Jeon. 2011. "The evolution of the Phillips curve: a modern time series viewpoint." *Economica* 78 (309): 51–66.

Greene, William H. 2012. *Econometric Analysis.* Prentice Hall.

Hayashi, Fumio. 2000. *Econometrics.* Princeton University Press.

Heckman, James J. 2000. "Causal parameters and policy analysis in economics: A twentieth century retrospective." *The Quarterly Journal of Economics* 115 (1): 45–97.

Hendry, David F. 2003. "J. Denis Sargan and the origins of LSE econometric methodology." *Econometric Theory,* 457–480.

Hooper, Peter, Frederic S. Mishkin, and Amir Sufi. 2020. "Prospects for inflation in a high pressure economy: Is the Phillips curve dead or is it just hibernating?" *Research in Economics* 74 (1): 26–62.

Hoover, Kevin D. 2006. "The Methodology of Econometrics." In *New Palgrave Handbook of Econometrics,* edited by T. C. Mills and K. Patterson, 1:61–87. London: Macmillan.

Kleibergen, Frank, and Sophocles Mavroeidis. 2009. "Weak instrument robust tests in GMM and the new Keynesian Phillips curve." *Journal of Business & Economic Statistics* 27 (3): 293–311.

Klein, Lawrence R., and James R. Ball. 1959. "Some econometrics of the determination of absolute prices and wages." *Economic Journal* 69:465–482.

Krogh, Tord S. 2015. "Macro frictions and theoretical identification of the New Keynesian Phillips curve." *Journal of Macroeconomics* 43:191–204.

Kurmann, André. 2007. "VAR-based estimation of Euler equations with an application to New Keynesian pricing." *Journal of Economic Dynamics and Control* 31 (3): 767–796.

Kydland, Finn E., and Edward C. Prescott. 1977. "Rules rather than discretion: The inconsistency of optimal plans." *Journal of political economy* 85 (3): 473–491.

———. 1991. "Hours and employment variation in business-cycle theory." In *Business Cycles,* 107–134.

Lipsey, Richard G. 1960. "The relationship between unemployment and the rate of change of money wage rates in the United Kingdom, 1862-1957." *Economica* 27:1–31.

Mavroeidis, Sophocles. 2004. "Weak identification of forward-looking models in monetary economics." *Oxford Bulletin of Economics and statistics* 66:609–635.

———. 2005. "Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve." *Journal of Money, Credit and Banking,* 421–448.

Mavroeidis, Sophocles, Mikkel Plagborg-Moller, and James H. Stock. 2014. "Empirical evidence on inflation expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52 (1): 124–88.

McGuirk, Anya, and Aris Spanos. 2009. "Revisiting Error-Autocorrelation Correction: Common Factor Restrictions and Granger Non-Causality." *Oxford Bulletin of Economics and Statistics* 71 (2): 273–294.

Mizon, Grayham E. 1995. "Progressive Modeling of Macroeconomic Time Series The LSE Methodology." In *Macroeconometrics,* edited by Hoover Kevin D., 107–180. Springer.

Nalewaik, Jeremy. 2016. "Non-linear Phillips curves with inflation regime-switching," FEDS Working Paper.

Nason, James M., and Gregor W. Smith. 2008. "Identifying the new Keynesian Phillips curve." *Journal of Applied Econometrics* 23 (5): 525–551.

Perry, George L. 1964. "The determinants of wage rate changes and the inflation-unemployment trade-off for the United States." *The Review of Economic Studies* 31 (4): 287–308.

———. 1970. "Changing labor markets and inflation." *Brookings Papers on Economic Activity* 1970 (3): 411–448.

Phelps, Edmund S. 1967. "Phillips curves, expectations of inflation and optimal unemployment over time." *Economica,* 254–281.

———. 1969. "The new microeconomics in inflation and employment theory." *The American Economic Review* 59 (2): 147–160.

Phillips, Alban W. 1958. "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957 1." *economica* 25 (100): 283–299.

Poudyal, Niraj. 2012. "Confronting Theory with Data: the Case of DSGE Modeling." PhD diss., Virginia Tech.

Qin, Duo. 2011. "The Phillips curve from the perspective of the history of econometrics." *History of Political Economy* 43 (supplement): 283–308.

Ravenna, Federico. 2007. "Vector autoregressions and reduced form representations of DSGE models." *Journal of monetary economics* 54 (7): 2048–2064.

Rudd, Jeremy, and Karl Whelan. 2005. "New tests of the new-Keynesian Phillips curve." *Journal of Monetary Economics* 52 (6): 1167–1181.

———. 2006. "Can rational expectations sticky-price models explain inflation dynamics?" *American Economic Review* 96 (1): 303–320.

Samuelson, Paul A, and Robert M Solow. 1960. "Analytical aspects of anti-inflation policy." *The American Economic Review* 50 (2): 177–194.

Santomero, Anthony M., and John J. Seater. 1978. "The inflation-unemployment trade-off: A critique of the literature." *Journal of economic literature* 16 (2): 499–544.

Sargan, John D. 1964. "Wages and Prices in the UK: A Study in Econometric Methodology." In *Econometric Analysis for National Economic Planning,* edited by P. E. Heat, G. Mills, and J. K. Whittaker. London: Butterworth.

Sbordone, Argia M. 2002. "Prices and unit labor costs: a new test of price stickiness." *Journal of Monetary economics* 49 (2): 265–292.

———. 2005. "Do expected future marginal costs drive inflation dynamics?" *Journal of Monetary Economics* 52 (6): 1183–1197.

———. 2006. "U.S. Wage and Price Dynamics: A Limited-Information Approach." *International Journal of Central Banking* 2 (3): 155–191.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (1): 1–48.

Snowdon, Brian, and Howard R. Vane. 2005. *Modern macroeconomics: its origins, development and current state.* Edward Elgar Publishing.

Solow, Robert M. 1969. *Price expectations and the behavior of the price level: lectures given in the University of Manchester.* Manchester University Press.

Spanos, Aris. 1994. "On Modeling Heteroskedasticity: The Student's t and Elliptical Linear Regression Models." *Econometric Theory* 10 (2): 286–315.

———. 1999. *Probability theory and statistical inference: econometric modeling with observational data.* Cambridge University Press.

———. 2006. "Econometrics in Retrospect and Prospect." In *New Palgrave Handbook of Econometrics,* edited by T. C. Mills and K. Patterson, vol. 1. London: Macmillan.

———. 2007. "The instrumental variables method revisited: On the nature and choice of optimal instruments," edited by Gary D. A. Phillips and Elias Tzavalis, 34–59. Cambridge University Press, Cambridge.

———. 2010. "Statistical adequacy and the trustworthiness of empirical evidence: Statistical vs. substantive information." *Economic Modelling* 27 (6): 1436–1452.

———. 2014. "Reflections on the LSE Tradition in Econometrics: a Student's Perspective." *Œconomia. History, Methodology, Philosophy,* nos. 4-3, 343–380.

———. 2018. "Mis-specification testing in retrospect." *Journal of Economic Surveys* 32 (2): 541–577.

———. 2019. *Probability Theory and Statistical Inference: Empirical Modeling with Observational Data.* Cambridge University Press.

Spanos, Aris, and Anya McGuirk. 2001. "The model specification problem from a probabilistic reduction perspective." *American Journal of Agricultural Economics* 83 (5): 1168–1176.

Stock, James H., and Mark W. Watson. 2007. "Why has US inflation become harder to forecast?" *Journal of Money, Credit and banking* 39:3–33.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A survey of weak instruments and weak identification in generalized method of moments." *Journal of Business & Economic Statistics* 20 (4): 518–529.

Summa, Ricardo, and Julia Braga. 2019. "The (conflict augmented) Phillips curve is 'alive and well'." *TD IE-UFRJ* 3.

Turnovsky, Stephen J., and Michael L. Wachter. 1972. "A test of the "expectations hypothesis" using directly observed wage and price expectations." *The Review of Economics and Statistics,* 47–54.

Zhang, Chengsi, and Joel Clovis. 2010. "The New Keynesian Phillips Curve of rational expectations: A serial correlation extension." *Journal of Applied Economics* 13 (1): 159–179.

Zhang, Chengsi, Denise R. Osborn, and Dong Heon Kim. 2008. "The new Keynesian Phillips curve: from sticky inflation to sticky prices." *Journal of Money, Credit and Banking* 40 (4): 667–699.

Appendices

Appendix A

Vector Autoregressive (VAR) and Dynamic Linear Regression (DLR) Models

In the PR approach, the VAR model is formulated by imposing the three Markov chain (M), stationary (S), and Normal distribution (N) assumptions on the joint distribution $D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \Phi)$, with $\mathbf{Z}_t = (y_t, \mathbf{X}_t)$ as a $(l \times 1)$ vector of l variables. The first two assumptions produce a transformation as:

$$
\begin{aligned}
D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \mathbf{\Phi}) =& D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_p; \boldsymbol{\varphi_1}) \prod_{t=p}^{T} D_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}^0; \boldsymbol{\varphi_t}) \\
\overset{M}{=}& D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_p; \boldsymbol{\varphi_1}) \prod_{t=p}^{T} D_t(\mathbf{Z}_t | \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, ..., \mathbf{Z}_{t-p}; \boldsymbol{\varphi_t}) \\
\overset{M\&S}{=}& D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_p; \boldsymbol{\varphi_1}) \prod_{t=p}^{T} D(\mathbf{Z}_t | \mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, ..., \mathbf{Z}_{t-p}; \boldsymbol{\varphi})
\end{aligned}
\tag{A.1}
$$

When the Normality distribution is next assumed, the specification of the VAR(p) model (p as the length of lags) with its testable assumptions is obtained in Table A.1; see step-by-step derivations on Spanos (1994) and Poudyal (2012).

Table A.1: Normal Vector Autoregressive (VAR) Model

Statistical GM: $\mathbf{Z}_t = \Psi_0 + \sum_{i=1}^{p} \Psi_i' \mathbf{Z}_{t-i} + \mathbf{u}_t,\ t \in \mathbb{N}$	
(1) Normality:	$D(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^0); \boldsymbol{\theta}) \sim \mathrm{N}(.,.)$, for $\mathbf{Z}_{t-1}^0 \equiv (\mathbf{Z}_{t-1}, ..., \mathbf{Z}_t),\ t \in \mathbb{N}$
(2) Linearity:	$E(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^0)) = \Psi_0 + \sum_{i=1}^{p} \Psi_i' \mathbf{Z}_{t-i},\ t \in \mathbb{N}$
(3) Homoskedasticity:	$Var(\mathbf{Z}_t \vert \sigma(\mathbf{Z}_{t-1}^0)) = \mathbf{\Omega_u}$ is free of $\mathbf{Z}_{t-1}^0,\ t \in \mathbb{N}$
(4) Markov:	$\{\mathbf{Z}_t,\ t \in \mathbb{N}\}$ is a Markov process.
(5) t-invariance:	$\Phi \equiv (\Psi_0, \Psi_1, ..., \Psi_p, \mathbf{\Omega_u})$ are not changing with t, $t \in \mathbb{N}$

The dynamic multivariate VAR model is not suitable for evaluating estimation models for the single-equation Phillips curve. It is needed to be reparameterized into a single-variable (univariate) autoregressive model. Under the same assumptions imposing on the joint distribution, the VAR model can be reparameterized into the single-variable Dynamic Linear Regression (DLR) model as in Table A.2; see Spanos (1999, 2019) for a full discussion. The key difference between the VAR model and the DLR model is that the RHS of the DLR model also includes contemporaneous dependent variables $\mathbf{X}_t$ while the VAR model's only have the lagged variables.

Table A.2: Dynamic Linear Regression (DLR) Model

Statistical GM:	$y_t = \boldsymbol{\sigma}_0(t) + \sum_{i=1}^{p} \alpha_i y_{t-i} + \mathbf{B}_0'\mathbf{X}_t + \sum_{j=1}^{p} \mathbf{B}_j'\mathbf{X}_{t-j} + u_t,\ t \in \mathbb{N}$
(1) Normality:	$D(y_t\|\sigma(\mathbf{Z}_{t-1}^0);\boldsymbol{\theta}) \sim \mathrm{N}(.,.)$, for $\mathbf{Z}_{t-1}^0 \equiv (y_{t-1},...,y_p,\mathbf{X}_{t-1},...,\mathbf{X}_p)$, $t \in \mathbb{N}$
(2) Linearity:	$E(y_t\|\sigma(\mathbf{Z}_{t-1}^0)) = \boldsymbol{\sigma}_0(t) + \sum_{i=1}^{p} \alpha_i y_{t-i} + \mathbf{B}_0'\mathbf{X}_t + \sum_{j=1}^{p} \mathbf{B}_j'\mathbf{X}_{t-j},\ t \in \mathbb{N}$
(3) Homoskedasticity:	$Var(y_t\|\sigma(\mathbf{Z}_{t-1}^0)) = \mathbf{\Omega}_\mathbf{u}$ is free of $\mathbf{Z}_{t-1}^0$, $t \in \mathbb{N}$
(4) Markov(p):	$\{\mathbf{Z}_t = (y_t, \mathbf{X}_t)\}$ is a Markov process, $t \in \mathbb{N}$
(5) t-invariance:	$\Phi \equiv (\boldsymbol{\sigma}_0, \alpha_1, ..., \alpha_p, \mathbf{B}_1, ..., \mathbf{B}_p, \mathbf{\Omega}_\mathbf{u})$ are not changing with t, $t \in \mathbb{N}$

Note: $\boldsymbol{\sigma}_0(t)$ can be a constant term or a function of t (trend terms).

Appendix B

Simulation Design

Because the VAR and DLR models are two parameterizations of the same joint distribution, the corresponding VAR reparameterization of the DLR model is used to simulate the data. First, the joint distribution in Experiment 1 can be rewritten in the general form as:

$$\begin{pmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1) \\ \boldsymbol{\Sigma}(1)' & \boldsymbol{\Sigma}(0) \end{bmatrix} \right), \tag{B.1}$$

where $\mathbf{Z}_t = (Y_t, X_t)'$, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ are the (2×1) mean vectors, $\boldsymbol{\Sigma}(0)$ and $\boldsymbol{\Sigma}(1)$ are the partitioned (2×2) vectors of the variance-covariance matrix. Note that $\mathbf{Z}_t$ can be expanded to any finite number of variables.

The data is then simulated recursively from the following marginal distribution:

$$D(\mathbf{Z}_t | \mathbf{Z}_{t-1}) \sim \mathrm{N}(\mathbf{A}_0 + \mathbf{A}_1 \mathbf{Z}_{t-1}, \boldsymbol{\Omega}), \tag{B.2}$$

where the parameterization is as:

$$\mathbf{A}_0 = \boldsymbol{\mu}_1 - \mathbf{A}_1 \boldsymbol{\mu}_2, \ \mathbf{A}_1 = \boldsymbol{\Sigma}(1)\boldsymbol{\Sigma}(0)^{-1}, \ \boldsymbol{\Omega} = \boldsymbol{\Sigma}(0) - \boldsymbol{\Sigma}(1)\boldsymbol{\Sigma}(0)^{-1}\boldsymbol{\Sigma}(1)'$$

The first observations (Y_1, X_1) are drawn from the normal distributions with the corresponding mean $\boldsymbol{\mu}_1$ and variance $\boldsymbol{\Sigma}(0)$. Experiment 2 uses the similar parameterization of the VAR to simulate the data except that $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are not constant, but functions of time t:

$$\begin{pmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu} + \boldsymbol{\gamma} t \\ \boldsymbol{\mu} + \boldsymbol{\gamma}(t-1) \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1) \\ \boldsymbol{\Sigma}(1)' & \boldsymbol{\Sigma}(0) \end{bmatrix} \right), \tag{B.3}$$

where $\boldsymbol{\gamma}$ is a constant (2×1) vector adding to the mean vector $\boldsymbol{\mu}$ at each time t.

Appendix C

M-S Testing under Small Sample Sizes

When the sample size is small (around 50 observations), the full-form auxiliary regressions have too many RHS variables; and there will remain a low number of degrees of freedom. It is especially pronounced when the two or three lags of two or three lags of the LHS variables are included. In this case, the lags and squares of the residuals can replace the lags of the LHR and RHS variables; and the squares of the fitted values can be used to probe the (2) linearity assumption. In particular, the indicative auxiliary regressions in Equations 2.11 and 2.12 are adapted for the small sample sizes as:

$$\hat{u}_t = \gamma_0 + \gamma_1 x_t + \overbrace{\gamma_2 \widehat{u_t}^2}^{\text{(2) Linearity}} + \overbrace{\gamma_3 t}^{\text{(5) t-invariance}} + \overbrace{\gamma_4 \widehat{u}_{t-1}}^{\text{(4) Independence}} + v_{1t} \tag{C.1}$$

$$\hat{u}_t^2 = \delta_0 + \overbrace{\delta_1 \widehat{y_t}^2}^{\text{(3) Homoskedasticity}} + \overbrace{\delta_2 t}^{\text{(5) t-invariance}} + \overbrace{\delta_3 \widehat{u}_{t-1}^2}^{\text{(4) Independence}} + v_{2t} \tag{C.2}$$

In ensuring the precision of testing for the Markov(p) dependence, the one-lag residuals $\hat{u}_{t-1}$ are also replaced by the explicit variables. For example, when the model under the M-S testing is assumed to be a Markov(2) process, two further one-period variables $\mathbf{x}_{t-3}$ and Y_{t-3} are used. This check-up step is always performed but is not reported in the summary of the M-S testing results.

Appendix D

Vintage Data Reconstruction and Modern Data Choices

Vintage data reconstruction

The full form of the first reproduced specification (G1977a) - column (1) in Table 3 of Gordon (1977a) - is:

$$\pi^w_{a,t}{=}\beta_0 + \boldsymbol{\beta}_1'\mathbf{U}_t + \gamma E_t(\pi^{PCE}_t) + \sum\nolimits_{j=1}^{8} \gamma^d_j \pi^{\text{diff}}_{t-j} + \alpha Tax_t + \varepsilon_t, \tag{D.1}$$

where $\pi^w_{a,t}$ is an adjusted wage inflation, $\mathbf{U}_t$ is a vector of three "adjusted" unemployment measurements (unemployment dispersion, disguised unemployment rate and unemployment rate of hours), $E_t(\pi^p_t)$ is expected inflation which is approximated as a single-variable weighted average of the 12 lags of a consumer price inflation (equivalent to the modern Personal Consumption Expenditure (PCE)), π^{diff}_t is the difference between the PCE and a producer price inflation rates, and Tax_t is a tax-rate measurement. Both wage and price inflation were computed as two-quarter growth rates.

The full form of the second specification (G1977b), which is the column (7) in Table 3 in Gordon 1977a, is:

$$\pi^w_{a,t}{=}\beta_0{+}\boldsymbol{\beta}_1'\mathbf{U}_t{+}\sum\nolimits_{i=1}^{8} \gamma_i \pi^p_{core,t-i}{+}\sum\nolimits_{j=1}^{8} \gamma^d_j \pi^{\text{diff}}_{core,t-j}{+}\alpha Tax_t + \varepsilon_t \tag{D.2}$$

In this specification, the expected inflation was approximated by the eight lags of the inflation rate from Gordon's newly developed "nonfood product net of energy" (deflator) price index (a precursor of the modern core price index). The sample period for this specification is 1954-1976.

As discussed, the three essential wages, unemployment, and price variables were reconstructed. It is noted that the coefficients of individual lags for π^{diff}_t, in Gordon's original estimations, were constrained by a third-degree Almon polynomial to produce a decreasing order with a zero limit. Because the reproduced models do not keep π^{diff}, the normal OLS estimation (rather than the constrained OLS) is the appropriate choice. Table D.1 provides a summary of the data sources of the reconstructed variables. The reconstruction process is next described for the three variables.

The unemployment variable

A single vintage official rate of unemployment is employed to represent the three "adjusted" unemployment measurements.

The wage inflation variable

In the first specification (G1977a) with the 1954-1970 sample period, the wage inflation rate $\pi^w_{a,t}$ was computed from a wage index with a full name as "nonfarm private hourly earnings adjusted for fringe benefits, including employers' social security tax payments". To obtain this wage index, Gordon began with an intermediate "private average hourly earnings" which was only available from 1964. To extend the sample to earlier years, Gordon approximated the pre-1964 "private average hourly earnings" as a weighted average of the seven disaggregated industries. The weights were fixed as the employment share of each industry in 1963 (the middle point of the sample period) to adjust for any changes in the industry mix. To get the final wage series, the aggregated hourly earning was further amended for fringe benefits (payments added to the official salary) by multiplying with the corresponding ratio between "total compensation of employees" and "wage plus salary incomes".

There were two small adjustments in the data reconstruction. Firstly, the finance sector (with a small 6% share) is omitted in the final weighted average because it lacks sufficient details to perform an interpolation as Gordon did. Secondly, for the period before 1964, the retail sector was provided as "retail exclude data at eating and drinking places"; and this old series (for the pre-1964 sample) was combined with a newer standard retail series (for the post-1964 sample) to get the final series for the disaggregated retail trade sector.

In Gordon (1977a), Gordon discussed a better choice for the wage series. A new index of hourly earnings, which had already been devised for industry mix and fringe benefits from the Bureau of Labor Statistics, was picked because it requires no other adjustment. This new hourly earnings index could not be found on a monthly or quarterly basis before 1972 from public sources. It was provided in Gordon (1972) that Gordon's final wage series in Gordon (1977a) was a combination of Gordon (1971) adjusted hourly earnings for the 1954-1963 period and the new hourly earnings index for the later part. It was not clearly described how Gordon could access the new hourly earnings index from 1964 to 1972. Thus, it is decided to best match his method by merging Gordon (1971)'s adjusted hourly earnings for the 1954-1972 period and the new earnings index for the 1972-1976 period.

The expected inflation variable

The compilation of the implicit deflator index π^{PCE}_t in the G1977a utilized two editions - the 1929-1976 and the 1928-82 - of The National Income and Product Accounts of the United States because neither of both editions provides sufficient data available for the 1954-1976 period. Then, the same weights as Gordon used were kept to compute the expected inflation rate in the reproduced models for G1977a.

To construct the "nonfood product net of energy" index in G1977b, Gordon first subtracted spending on foods and direct/indirect energy expenditures from the aggregate output at the constant and the current prices. The final non-food non-energy deflator was then computed as the ratio between the outputs at the two prices. As discussed in the main body of the chapter, it is impossible to reconstruct this core price index because some of the data sources were not published or completely lost, and it lacks of sufficient technical instructions from Gordon. The modern core PCE is used as the replacement.

Table D.1: Data Sources of the Reconstructed Series

Series	Data sources
Official unemployment rate	Employment and Earnings, January and July issues of each year from 1954 to 1977 from the U.S. Bureau of Labor Statistics.
Private average hourly earnings	Employment and Earnings, United States, 1909-84, from the U.S. Bureau of Labor Statistics.
Total compensation of employees	The National Income and Product Accounts of the United States, 1929-82 from the U.S. Bureau of Economic Analysis.
Wage plus salary incomes	The National Income and Product Accounts of the United States, 1929-82 from the U.S. Bureau of Economic Analysis.
Index of hourly earnings	Survey of Current Business, January and July issues of each year from 1973 to 1977 from the U.S. Bureau of Economic Analysis.
Personal consumption expenditure - implicit price deflator	The National Income and Product Accounts of the United States with the 1929-76 edition for the 1955-1958 sample period, and the 1928-1982 edition for the 1959-1976 sample period from the U.S. Bureau of Economic Analysis.
Personal consumption expenditure price excluding food and energy	Personal Consumption Expenditures: Chain-type Price Index Less Food and Energy from FRED Economic Data.

Modern data choices

Because the modern all-section wage series is only available from 1960, the disaggregate manufacturing series is selected as the replacement to have a sufficient sample length. The price series is the Personal Consumption Expenditure index as the modern version of Gordon's price series. The unemployment variable is the standard all-civilian rate. All of the modern data were obtained from FRED Economic Data.

Appendix E

M-S Testing for the Student's t VAR(p) Model

Consider the following a quadratic-trend St-VAR(p) (order of p) model:

$$\mathbf{Z}_t = \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 t + \boldsymbol{\psi}_2 t^2 + \boldsymbol{\psi}_3 D + \sum_{i=1}^{p} \mathbf{A}_i' \mathbf{Z}_{t-i} + \mathbf{u}_t, \tag{E.1}$$

where $\mathbf{Z}_t$ is a vector of variables and D is a dummy variable to capture outliers.

The M-S testing for the Student's t VAR model is performed through the two auxiliary regressions as:

$$\hat{u}_t = \delta_{10} + \underbrace{\boldsymbol{\delta}_{11}' \mathbf{Z}_{t-1}^0}_{\text{All RHS vars.}} + \underbrace{\boldsymbol{\delta}_{12}' (\mathbf{Z}_{t-1})^2}_{\text{(2) Linearity}} + \underbrace{\boldsymbol{\delta}_{13}' \boldsymbol{\sigma}_1(t)}_{\text{(5) No further t-invariance}} + \underbrace{\boldsymbol{\delta}_{14}' \mathbf{Z}_{t-p-1}}_{\text{(4) Dependence up to order-}p} + \varepsilon_{10} \tag{E.2}$$

$$\hat{u}_t^2 = \delta_{20} + \underbrace{\delta_{21} \hat{\sigma}_{zt}^2}_{\text{Conditional variance}} + \underbrace{\boldsymbol{\delta}_{22}' (\mathbf{Z}_{t-1})^2}_{\text{(3) No further heteroskedasticity}} + \underbrace{\delta_{23} t^5}_{\text{(5) No further 2nd t-variance}} + \underbrace{\delta_{24} \hat{u}_{t-1}^2}_{\text{(4) No 2nd-order dependence}} + \varepsilon_{20} \tag{E.3}$$

where:

- $\hat{u}_t = \mathbf{L}_t'(\mathbf{Z}_t - \hat{\mathbf{Z}}_t)$ is the standardized residuals with $\hat{\mathbf{Z}}$ as the fitted value and $\mathbf{L}_t \mathbf{L}_t' = \widehat{\text{Cov}}(\mathbf{Z}_t | \sigma(\mathbf{Z}_{t-1}^0))$.
- $\mathbf{Z}_{t-1}^0$ are all RHS variables in Equation E.1 which include $(\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, ..., \mathbf{Z}_{t-p})$ plus t, t^2 and dummy variables.
- $(\mathbf{Z}_{t-1})^2$ is squared of one-period lag of the variables $(\mathbf{Z}_t)$ in the St-VAR model, excluding polynomial trend and the dummy variable.
- $\boldsymbol{\sigma}_1(t)$ is the one-order higher trend polynomial than the highest order in the Student's t VAR model (t^2).
- $\hat{\sigma}_{zt}^2 = \widehat{\text{Var}}(\mathbf{Z}_t | \sigma(\mathbf{Z}_{t-1}^0))$ is the estimated values of the conditional variance function; see Poudyal (2012).

The Skewness Kurtosis test for the Student's t distribution is used test the (1) Normality assumption. Considering a Student's t VAR(3) model which has two variables π_t and x_t plus

t and t^2 as an example, the fully explicit forms of the auxiliary regressions for π_t^p are:

$$\hat{u}_t = \delta_{10} + \underbrace{\delta_{t1} t + \delta_{t2} t^2 + \sum_{i=1}^{3} \delta_{\pi i} \pi_{t-i} + \sum_{j=1}^{3} \delta_{xi} x_{t-j}}_{\text{All RHS variables}} + \underbrace{\delta_{11} \pi_{t-1}^2 + \delta_{12} x_{t-1}^2}_{(2)} + \underbrace{\delta_{13} t^3}_{(5)} + \underbrace{\delta_{14} \pi_{t-4} + \delta_{15} x_{t-4}}_{(4)} + \varepsilon_{10} \tag{E.4}$$

$$\hat{u}_t^2 = \delta_{20} + \underbrace{\delta_{21} \hat{\sigma}_{yt}^2}_{\text{Conditional variance}} + \underbrace{\delta_{22} \pi_{t-1}^2 + \delta_{23} x_{t-1}^2}_{(3)} + \underbrace{\delta_{24} t^5}_{(5)} + \underbrace{\delta_{25} \hat{u}_{t-1}^2}_{(4)} + \varepsilon_{20} \tag{E.5}$$

Appendix F

Computation of the GMM Estimators Under the Implicit Statistical Models as the Student's t VAR(p) Models

GMM estimators under non-homoskedasticity

Consider a n-observation linear single-equation model as:

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \qquad E(u_t'u_t) = \boldsymbol{\Omega} \tag{F.1}$$

The linear GMM estimator is usually as:

$$\hat{\beta}_{GMM} = (\mathbf{X}_t'\mathbf{Z}_t\mathbf{W}\mathbf{Z}_t'\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{Z}_t\mathbf{W}\mathbf{Z}_t'y_t \tag{F.2}$$

where $\mathbf{W} = \mathbf{S}^{-1}$ is the weighting matrix. $\mathbf{S}$ is a variance-covariance matrix of the moment conditions:

$$\mathbf{S} = \frac{1}{n}E(\mathbf{Z}_t'u_tu_t'\mathbf{Z}_t) = \frac{1}{n}E(\mathbf{Z}_t'\boldsymbol{\Omega}\mathbf{Z}_t) \tag{F.3}$$

When the error term in Equation F.1 is assumed to be homoskedastic, the matrix Ω in Equation F.3 can be written down as:

$$\boldsymbol{\Omega} = \sigma^2\mathbf{I}_n \tag{F.4}$$

where $\mathbf{I}_n$ is an identity matrix of size n and σ is the constant variance of u_t. The GMM estimator in this case is identical with the estimators of the 2SLS method.

When the assumption of homoskedasticity is supposed not to hold, the heteroskedastic and autocorrelation consistent (HAC) estimators, such as the Newey-West, are used. As explored in Chapter 2, the error-fixing approach does not guarantee reliability of statistical inferences. On the contrary, the statistically adequate ISMs provide a direct estimator of the variance-covariance matrix $\boldsymbol{\Omega}$ of the residual u_t. Then, the strategies proposed by Bowden and Turkington (1990) are utilized to compute the new GMM estimators when the heteroskedastic $\boldsymbol{\Omega}$ can be reliably estimated.

Bowden and Turkington (1990) proposes two methods to compute the consistent GMM estimators under the case of the non-homoskedasticity. The first method is to plug the estimator of $\boldsymbol{\Omega}$ into the moment condition:

$$\hat{\boldsymbol{\beta}}_{GMM} = [\mathbf{X}_t'\mathbf{Z}_t(\mathbf{Z}_t'\hat{\Omega}\mathbf{Z}_t)^{-1}\mathbf{Z}_t'\mathbf{X}_t]^{-1}\mathbf{X}_t'\mathbf{Z}_t(\mathbf{Z}_t'\hat{\Omega}\mathbf{Z}_t)^{-1}y_t \tag{F.5}$$

where $\hat{\boldsymbol{\Omega}}$ is the estimator of $\boldsymbol{\Omega}$.

The second method is similar to the Generalized Least Square (GLS) method. Firstly, the $\boldsymbol{\Omega}$ is factorized as $\boldsymbol{\Omega}^{-1} = \mathbf{D}'\mathbf{D}$. Then, $\mathbf{D}$ is multiplied to the original equation as:

$$\mathbf{D}y_t = \mathbf{D}\mathbf{X}_t\boldsymbol{\beta} + \mathbf{D}u_t \tag{F.6}$$

The transformed $\mathbf{Z}_{D,t} = \mathbf{D}\mathbf{Z}_t$ is now used as the new instrument set. The GMM objective function is now $(\mathbf{D}u_t)'\mathbf{Z}_{D,t}\mathbf{W}\mathbf{Z}'_{D,t}(\mathbf{D}u_t)$ with the optimal choice of the weight matrix as $\mathbf{W} = (E(\mathbf{Z}_{D,t}\mathbf{Z}'_{D,t}))^{-1} = (\mathbf{Z}'_t\boldsymbol{\Omega}^{-1}\mathbf{Z}_t)$; see more details on Bowden and Turkington (1990), page 70-71. By definition, we have $E(u_tu'_t) = \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}^{-1} = \mathbf{D}'\mathbf{D}$. It can be transformed as $E(\mathbf{D}u_tu'_t\mathbf{D}') = \sigma^2\mathbf{I}_n$ with σ^2 as a constant. The transformed residual $\mathbf{D}u_t$ is now homoskedastic. Therefore, $\boldsymbol{\beta}$ can be estimated using the 2SLS (the simplest GMM approach).

Link-up with the Implicit Statistical Models (ISMs)

The next step is to compute $\boldsymbol{\Omega}$. As shown in the Section 5.2, the general ISM can be written:

$$y_t = \boldsymbol{\beta}'_1\mathbf{X}_{exog,t} + \boldsymbol{\beta}'_2\mathbf{Z}_t + u_{1t} \tag{F.7a}$$

$$\mathbf{X}_{endo,t} = \mathbf{B}'_1\mathbf{X}_{exog,t} + \mathbf{B}'_2\mathbf{Z}_t + \mathbf{u}_{2t} \tag{F.7b}$$

$$\text{with } \begin{pmatrix} u_{1t} \\ \mathbf{u}_{2t} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \omega_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}\right) \tag{F.7c}$$

as well as the reparameterized link-up model is:

$$y_t = \boldsymbol{\gamma}'_0\mathbf{X}_t + \boldsymbol{\alpha}'_0\mathbf{Z}_t + \varepsilon_{0t} \tag{F.8a}$$

$$\mathbf{X}_{endo,t} = \mathbf{B}'_1\mathbf{X}_{exog,t} + \mathbf{B}'_2\mathbf{Z}_t + \mathbf{u}_{2t} \tag{F.8b}$$

$$\text{with } \begin{pmatrix} \varepsilon_{0t} \\ \mathbf{u}_{2t} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{pmatrix}\right) \tag{F.8c}$$

The estimators of all parameters of the ISM are first obtained from the estimation of the Student's t VAR model. Now, it is needed to compute the estimator σ_0^2 from Equation F.8c which is analogous to the variance-covariance $\boldsymbol{\Omega}$ of u_t in the previous step. It is shown in Spanos (2007) that the link of parameters between the two models is as follow:

$$\boldsymbol{\Omega} \equiv \sigma_0^2 = \omega_{11} - \boldsymbol{\omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\omega}_{21} \tag{F.9}$$

Note that the Normal distribution is assumed for both the ISM and the reparameterized link-up model. As provided in Spanos (1994), under the elliptical family, to which both Normal and Student's t distributions belong, the parameterization is the same. Under the StVAR model, all parameters in Equation F.9 is now a function of time t as:

$$\boldsymbol{\Omega} \equiv \sigma_0^2(t) = \omega_{11}(t) - \boldsymbol{\omega}_{12}(t)\boldsymbol{\Omega}_{22}^{-1}(t)\boldsymbol{\omega}_{21}(t) \tag{F.10}$$

As shown in Poudyal (2012), it can be further simplified by taking out the common time-varying quadratic factor as:

$$\boldsymbol{\Omega} \equiv \sigma_0^2(t) = \omega_{11}(t) - \boldsymbol{\omega}_{12}(t)\boldsymbol{\Omega}_{22}^{-1}(t)\boldsymbol{\omega}_{21}(t) = (\omega_{11} - \boldsymbol{\omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\omega}_{21})q(t) \tag{F.11}$$

Due to the statistical adequacy of the ISM (with the correct lag lengths), the problem of autocorrelation does not exist. Using the estimators of parameters in Equation F.11, the ultimate estimator of $\hat{\mathbf{\Omega}}$ is obtained as:

$$\hat{\mathbf{\Omega}} = \begin{pmatrix} \hat{\sigma}_0^2(1) & & & & 0 \\ & \ddots & & & \\ & & \hat{\sigma}_0^2(t) & & \\ & & & \ddots & \\ 0 & & & & \hat{\sigma}_0^2(n) \end{pmatrix} \tag{F.12}$$

Between the methods proposed by Bowden and Turkington (1990), a simulation is run to evaluate which method is better. The simulation results favor the second approach (the GLS-like). Thus, this approach is used to compute the GMM estimators under the StVAR ISMs.

Appendix G

M-S Testing Results for the Respecified Student's t VAR Models: the Full sample

The M-S testing results for the respecified St-VAR models under the full sample is shown on Table G.1 on the next page.

Table G.1: M-S Testing Results for the Respecified St-VAR Models: the Full sample (1960Q1-2019Q3)

		Small instrument set			
		Labor share, StVAR(3)		Output gap, StVAR(3)	
		π_t	x_t	π_t	x_t
(1) Student's t		$\chi(2)$=3.35 [0.187]	$\chi(2)$=5.72 [0.057]	$\chi(2)$=5.57 [0.062]	$\chi(2)$=2.74 [0.254]
(2) Linearity		F(2,220)=0.85 [0.429]	F(2,220)=0.10 [0.905]	F(2,220)=2.60 [0.077]	F(2,220)=0.25 [0.781]
(3) Heteroskedasticity		F(2,225)=0.57 [0.566]	F(2,225)=0.41 [0.664]	**F(2,229)=3.30 [0.039]**	F(2,229)=2.59 [0.077]
(4) Markov(p) dependence	R	F(2,220)=0.66 [0.517]	F(2,220)=0.01 [0.992]	F(2,220)=2.08 [0.127]	F(2,220)=0.93 [0.395]
	S	F(1,229)=0.63 [0.427]	F(1,229)=0.17 [0.682]	F(1,229)=0.60 [0.492]	F(1,229)=0.82 [0.367]
(5) t-invariance	R	F(1,220)=2.89 [0.091]	F(1,220)=0.08 [0.777]	F(1,220)=1.12 [0.291]	F(1,220)=1.38 [0.242]
	S	**F(1,229)=5.07 [0.025]**	F(1,229)=0.02 [0.880]	**F(1,229)=4.02 [0.046]**	F(1,229)=1.38 [0.241]

		GGLS2001 instrument set			
		Labor share, StVAR(4)		Output gap, StVAR(3)	
		π_t	x_t	π_t	x_t
(1) Student's t		**$\chi(2)$=7.05 [0.030]**	$\chi(2)$=2.48 [0.290]	$\chi(2)$=4.87 [0.088]	$\chi(2)$=1.67 [0.433]
(2) Linearity		**F(4,205)=2.87 [0.024]**	**F(4,205)=4.94 [0.001]**	**F(4,210)=2.88 [0.024]**	F(4,210)=2.11 [0.081]
(3) Heteroskedasticity		**F(4,226)=5.81 [0.000]**	F(4,227)=1.80 [0.129]	**F(4,227)=6.03 [0.000]**	**F(4,227)=5.57 [0.000]**
(4) Markov(p)	R	F(4,205)=1.72 [0.148]	F(4,205)=1.57 [0.183]	F(1,210)=2.03 [0.155]	F(4,210)=1.88 [0.116]
	S	F(1,226)=2.88 [0.091]	F(1,227)=3.76 [0.054]	F(1,227)=1.54 [0.216]	F(1,227)=0.00 [0.991]
(5) t-invariance	R	F(1,205)=2.55 [0.112]	F(1,205)=0.05 [0.831]	F(1,210)=2.03 [0.099]	F(1,210)=1.49 [0.224]
	S	F(1,226)=1.49 [0.223]	F(1,227)=0.01 [0.930]	F(1,227)=1.59 [0.209]	F(1,227)=0.75 [0.388]

		GG1999 instrument set			
		Labor share, StVAR(4)		Output gap, StVAR(3)	
		π_t	x_t	π_t	x_t
(1) Student's t		$\chi(2)$=2.39 [0.303]	$\chi(2)$=0.18 [0.398]	**$\chi(2)$=6.60 [0.037]**	$\chi(2)$=5.86 [0.053]
(2) Linearity		F(6,194)=2.02 [0.065]	**F(6,193)=2.54 [0.022]**	F(6,201)=1.97 [0.071]	**F(6,201)=2.38 [0.030]**
(3) Heteroskedasticity		**F(6,224)=2.52 [0.022]**	F(6,221)=1.08 [0.374]	F(6,225)=1.88 [0.086]	F(6,221)=0.98 [0.437]
(4) Markov(p)	R	F(6,194)=0.47 [0.830]	F(1,193)=0.07 [0.790]	F(6,201)=1.99 [0.068]	F(6,201)=1.82 [0.097]
	S	F(1,224)=0.92 [0.338]	F(1,221)=1.36 [0.244]	F(1,225)=0.82 [0.367]	F(1,221)=0.72 [0.398]
(5) t-invariance	R	**F(1,194)=5.43 [0.021]**	F(6,193)=0.54 [0.776]	**F(1,201)=7.11 [0.008]**	F(1,201)=1.66 [0.199]
	S	F(1,224)=2.73 [0.100]	F(1,221)=0.03 [0.870]	F(1,225)=0.03 [0.856]	F(1,221)=0.37 [0.544]

Note 1: R and S denote the auxiliary regression and skedastic functions, respectively.

Note 2: In Markov(p) dependence, p indicates the highest lags in each model.