Inflation Forecast (p* concept)

Gabin BEUZEVAL Alexis JOHNSON Théo LINALE Noah MARTINEZ

May 2025

Abstract

This work presents a rigorous empirical analysis of the P* (P-star) model, an innovative theoretical framework introduced by Hallman, Porter, and Small (1991) to forecast inflation based on monetary aggregates. Based on the quantity theory of money, the model talks about the existence of an equilibrium price level (P*) determined by the money supply (M2), the long-term velocity of money (v*), and potential output (q*). The gap between the observed price level (p) and P* generates inflationary adjustments, which are formalized through an error correction model (ECM).

Contents

Introduction

In economics, the forecast of inflation remains a critical issue as it directly conditions monetary policy decisions as well as the expectations of various economic agents. The link between money supply and inflation has been the subject of intense debate for decades, with frequent empirical questioning. Historically, monetarism postulated a direct relationship between inflation and money supply growth under the assumption of short-term stability in money velocity. However, during economic crises or periods of significant financial system transformation, this stability assumption has proved to be particularly fragile. This fragility has inevitably called into question the predictive capacity of strictly monetarist models.

Hallman, Porter and Small (1991) introduce in this context the critical concept of P (P-star), extending quantitative theory by proposing a equilibrium price level based on money supply (M2), long-run money velocity as well as potential GDP. This thesis therefore aims precisely to examine and empirically verify the validity of this P concept as an advanced indicator of inflation through a comparative study on different countries.
The methodology combines graphical analysis with an error correction econometric model (ECM). This allows the observation of both short-term dynamics and long-term adjustments around the theoretical equilibrium price level (P). The results of the error correction model (ECM) will be systematically compared to those of a random walk model in order to rigorously evaluate its predictive relevance using the statistical test of Diebold and Mariano, known for for incorporating error differential variability.

This thesis is structured in two main parts. The first part examines the theoretical foundations of the P concept, its contributions to the economic literature, as well as its limits and extensions proposed in various recent academic works. The second part presents the detailed empirical application of the error correction model, the results obtained from historical data across different countries and time periods, and finally the comparative evaluation of its predictive performance using the Diebold and Mariano test.

1 Theoritical foundations of inflation forecast with p*

1.1 Traditional approaches of inflation forecast

Inflation forecasting has historically used various analytical and theoretical approaches that have shaped monetary policies and economic strategies. Among these traditional approaches we can find the monetarism, the Phillips curve in its various forms, and also VAR models (Vector Auto-Regressive).

Monetarism

Monetarism is one of the most important traditional approach for inflation prevision. This way of thinking is mainly based on the quantitative theory of money, for which the modern reformulation has been proposed by Milton Friedman in 1956 in his article called *The Quantity Theory of Money: A Restatement.* (**?**). For Friedman, the general level of price is determined by the money supply available in the economy and by the velocity of money. The monetarism consider then inflation as phenomenon mainly monetary, resulting from an excessive growth of money supply compared to that of the real output.

If we follow this approach, the economical mechanisms are based on a stable relation between money supply, the velocity of money, the general level of prices and the real revenue, often represented by Fisher equation: MV=PT. The presumed stability of velocity V, in the short term at least, imply that all the increase in money supply M beyond the growth of real income T lead mechanically speaking to a proportional increase of the general level of prices $P*$.

Friedman and Schwartz (1963) in their major article called *"Money and Business Cycles"* give a significant empirical illustration of this relation. By studying cyclical fluctuations of money supply in the United States for a century (1867-1960), they showed that major economical crisis like the Great Depression in the 30s were systematically linked to important contractions of money supply. Particularly, the 1929-1933 period has known an exceptional fall of more than 35% of money supply, showing the major role of this variable in the explanation of the deflation and the severe economical depression.

Even though the theoretical and empirical robustness of monetarism, several critics were addressed. Firstly, the hypothesis about the stability of monetary velocity has been questioned, especially the economic crisis period or the structural transformation of financial system. Friedman himself in *"The Role of Monetary Policy"* (Friedman, 1968), recognize that velocity can vary based on economic agent's anticipations concerning the future evolution of prices and interest rate, which put into perspective the effects of monetary policy.

In addition, if the quantitative theory put clearly in evidence a correlation between money supply and inflation, the causality is still complex to establish precisely. In fact, the money supply can also be endogenous, influenced by general economical conditions and the anticipation by market participants. Thus, the Keynesians critics show that in certain conditions, a rise of money supply do not necessarily lead to an immediate inflation but can be accompanied by variations in employment or in the short term production, particularly when the economy is far from full employment.

To conclude, if the monetarism bring a pertinent analytical framework and demonstrated its predictive capacity in numerous historical circumstances, its limits lead to a diversification of approaches when we talk about inflation prevision. It's precisely in this context that the Phillips curve and its extensions appeared, in response to the necessity of integrate explicitly the dynamic beta-ween inflation, unemployment and short term price adjustment, thus overtaking the restrictives hypothesis of classic monetarism.

Phillips curve and its evolutions

The Phillips curve is a great improvement in the analysis approaches of inflation forecast, by setting an inverse correlation between the unemployment rate and the inflation rate. Initially introduced by Phillips (1958), this curve was showing a negative empirical correlation between the variation in the nominal wage rate and the unemployment rate in United Kingdom between 1861 and 1957. According to Phillips, a weak workforce availability leaded to an increase in nominal wages given the competition increase between employer in order to acquire rare workers.

This relation was then extent by Samuelson and Solow in 1960 who transformed the original one between unemployment and wage variations into a relation between unemployment and price inflation. This rework offered to political deciders an apparently simple prevision tool: inflation could be anticipated given the observed level of unemployment.

Phelps (1967) and Lucas (1972) introduced an « augmented » version by adding inflationary anticipations. According to the authors, the initial negative relation between inflation and unemployment can last only if economic agents are victims of monetary illusions, which means that they can't distinguish their nominal wage and their purchasing power. Lucas insisted on the temporary nature of the negative relation, underlining that when these agents represent rational anticipations, this relation tends to disappear in the long run and the economy goes back to her natural unemployment rate.

The major role of anticipations was then explored deeply by Phelps (1967), by showing that the augmented Phillips curve is underlining the existence of a natural unemployment rate or NAIRU (Non-Accelerating Inflation Rate of Unemployment). At this rate, the expected inflation is equal to actual inflation, thus stabilizing the inflation rate. When unemployment rate drops below NAIRU, inflation increase; on the other hand, it decreases when unemployment is above NAIRU.

Gordon (2013) goes beyond by adding the effects of supply shocks an inflationary inertia. According to him, inflation shows inertia due to late adjustments of prices and incomes, which signify that unemployment variations effects on inflation are delayed and spread out over time. Also, supply shocks such as important commodities price variations can temporary move the Phillips curve and perturb the stable relation between unemployment and inflation. Gordon then demonstrate that even though there exist some fluctuations, the Phillips curve is robust in its capacity to explain inflationary dynamics in the medium term.

Still, the empirical pertinence of the Phillips curve is subject to debates. Atkeson and Ohanian (2001) severely criticized its predictive efficiency, claiming that previsions based on the Phillips curve do not surpass a simple naive prevision on past inflation. This criticism is based on the empirical observation of a persisting instability of the relation between unemployment and inflation, particularly since 1970.
The major empirical criticism is then about the structural instability of the curve which makes its use incertain for short term inflationary forecast, especially in periods of significant macroeconomic changes.

To conclude, the evolution and the criticism about the Phillips curve have led to rethink the mechanisms of inflation predictions, putting under the light the important role of rational anticipations. Taking this into account about anticipations has deeply modified the traditional point of view about inflationary forecasts.

Rational Anticipations

The rational anticipations theory represents a break with the traditional Keynesian approaches when we talk about inflation forecast. Given this theory, economic agent's anticipations are made by exploiting all the information available and by following the forecasts of the relevant economic model (Muth, 1961). Muth champions the fact that these anticipations are rational in the sense that they are not systematically biased and they are linked with the objective forecasts of the economic model of reference.

The introduction of this analytical framework has changed traditional Keynesians models, particularly by questioning relations such as the augmented Phillips curve which is based on the idea that inflationary forecast are made from adaptation from past errors. Lucas (1976) criticized precisely these Keynesian econometrical models, showing that forecast based on such relations cannot be considered as reliable when they don't take into account the evolution of anticipations facing changes in economic policies. In fact, according to Lucas, when agents anticipate inflation rationally, any systematic attempt by the authorities to reduce unemployment sustainably by stimulating inflation are not working. Unemployment then returns to its natural rate, confirming empirically the hypothesis given by Friedman and Phelps.

The impact of rational anticipations on the modeling of inflation is important: it implies a deep revision of traditional implications of monetary policies. Sargent (1973) insist on this point by showing that, under rational anticipations, the real interest rate is independent from the systemic component of the monetary policy. Then, only the unexpected components of monetary policies have a real impact, confirming then the importance of rational anticipations to understand the limits of government economical interventions.

Still, even with its theoretical contribution, the approach by rational anticipations encounters several limits. First, it is based on really strong hypothesis about information and the cognitive capacity of economic agents. In fact, it suppose that this agents have access to exhaustive informations and their anticipations match perfectly with the forecasts made by the economic model of reference. In a second time, as stated by Mankiw, this assumption neglects real rigidities such as the adjustment costs or the cost to acquire information, which affects from a practical standpoint the process of formation of anticipations (Mankiw and Reis, 2002). These rigidities limit forecast efficiency of rational anticipations models by partially disconnecting some empirical observed behaviors.

These methodological and empirical limits of the approach based on rational anticipations have progressively led to the emergence of more agnostic approaches, less dependent of these strong theoretical assumptions. Vector autoregressive models (VAR) particularly, offer an interesting empirical alternative by modeling directly dynamical interactions of economical time series without restrictions based on a rigid theoretical structure. It's precisely to these VAR models that the transition is going, as a response to critics and practical limits given by the exclusive use of rational anticipations in modeling inflation.

VAR Models

Vector autoregressive models (VAR) represents a major breakthrough in the empirical macroeconomic analysis, especially when we are talking about inflation forecast. Introduced by Sims in 1980, these models are seen as an alternative for highly parameterized structural models, especially in a context where we questioned the great macroeconomic equations from traditional Keynesian theory and also with evolutions based on rational anticipations.

A VAR model is based on a linear equations system in which each endogenous variables are explained by their own past values and also the ones from the other variables in the system. This multivariate autoregressive approach enable to capture dynamic interdependencies between several time series, without imposing a rigid theoretical structure. Such as stated by Stock and Watson (2001), this flexibility makes VAR useful to describe data, to produce forecast, to infer underlying economical structures, and analyse economical policies effects.

One of the major VAR contribution is about its capacity to model inflationary dynamics from an empirical point of view. By estimating the response of inflation to exogenous shocks often due to monetary policies, VAR models allow to generate fonctions such as "impulse response functions" which gives information about the delay in the transmission of effects. The empirical works show that non anticipated increases in the key interest rate have a progressive and delayed impact on inflation, that univariate models or structure-rigid models does not always capture realistically.

VAR models, even though known for their adaptability can have issues. From the interpretation point of view, they do not allow to automatically distinguish the causality from the simple correlation. These identification of shocks difficulties often impose to use more restrictions based on economic theory. Sims himself, in his founding article (Sims, 1980), insist on the fact that every causal interpretations need to introduce explicit economical assumptions, exogenous to the estimated statistical system. Also, VAR models can suffer from their fragility in structural breaks periods, especially during economical crises or when we observe changes in monetary regimes. These estimated coefficients from past datas are then less reliable, and forecasts coming from these can be biased. Stock and Watson show also that forecast performances from VAR, even though they can be good in a stable framework, they can also not be satifying when historical regularities are disrupted. Then, the increase in the number of variables in a VAR model generate a boom in the number of parameters to estimate, which can be challenging methodologicaly speaking in term of over-parameterization and statistical robustness. In order to counter these difficulties, some bayesian versions of VAR models have been developed (Litterman, 1986).

Despite these limits, VAR models represent an essential contribution for modern econometrics. Their rising adoption in central banks and macroeconomic forecast institutions shows their capacity to produce empirical results. This approach underlines a turning point to a more inductive modeling, which opens the gate to more hybrid approaches adding theoretical and also empirical point of view at the same time, or using**Error correction models** .

These theoretical and empirical lack have lead to look for other approaches in order to add more efficiently long term relations and monetary dynamics. It's precisely in this context of research of robustness and theoretical coherence that the **(concept of p* or p-star)**,emerged and proposed an original response to the limits identified in classical approaches.

1.2 P* concept and its value

The P* concept, or "P star", emerged in the 1990s as a monetary analysis tool aimed at anticipating inflationary pressures. Although it falls within the tradition of the quantity theory of money, it was Hallman, Porter, and Small who formalized its empirical use in the context of U.S. monetary policy, in their article ***Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?*** (Hallman et al., 1991). Their objective was to provide a measure of the equilibrium price level based on monetary fundamentals, thereby allowing for the assessment of potential gaps between the observed price level and its "normal" or sustainable level.

1.2.1 Theoretical Framework of the Foundational Article

Clearly aligned with the quantitative theory of money tradition, the authors posit that, in the long run, there exists a stable relationship between the money supply, the velocity of money, and the aggregate price level. This relationship is formalized through the classical quantitative theory of money equation:

$$MV = PT$$

where M is the money supply, V is the velocity of money, P is the general price level, and T represents the volume of real transactions (here approximated by potential GDP). From this identity, the authors define the long-term equilibrium price level (P^*) as:

$$P^* = \frac{M2 \cdot V2^*}{Q^*} \tag{1}$$

Where:

- P^* is the Long-run equilibrium price level.
- $M2$ is observed M2 monetary aggregate.
- $V2^*$ is Long-run velocity of M2, estimated as the historical mean of V2 over the period, because authors
- Q^* is Potential GDP.

. When the level of observed prices P gets away from P^*, it generates inflationary adjustments, modeled through an Error correction model (ECM). This approach adds at the same time long-term stationary relations and short-term transitory dynamics.

1.2.2 Empirical application and mechanisms

In order to evaluate empirically this relation, Hallman and his colleagues have used quarterly US datas for the period between 1955 and 1988, under the assumption that velocity is stationary over that interval, using the GDP deflator as a measure of price. Their methodology is principally based on the econometric modeling of gaps between observed price and equilibrium price (from a theoretical point of view) (p - p*)[1] , that they link to gaps of velocity (v - v*) and production gaps (q* - q). Their empirical strategy is characterized by a dynamic estimation, using lags in order to capture the inflationary persistence observed from an empirical point of view. The model P^* is based on two important mechanisms :

[1]For each variable K, k is its natural logarithm : $k = ln(K)$

1. **Long-term monetary relation** : the relative stability of velocity of M2 on the long run implies that M2's variations are mechanically due to adjustments in the level of prices. This strengthen the idea that inflation is a monetary variable in the end. More precisely :

   - **Velocity gap** $(v-v^*)$: his term measures at what point the pace of circulation of M2 is getting away from its long term equilibrium value (v^*).A negative value of $v-v^*$ means that money circulate slower that what is predicted for the long run, which means that economic agents have too much money compared to their needs. It can generate a pressure to the increase of prices mid term because agents are progressively looking to reduce their excess cash balances by increasing their spendings.
   - **Production gap** (q^*-q) :This indicator corresponds to the output gap, which means that the difference between the potential GDP (q^*) and the real observed GDP (q). A positive q^*-q means that the economy is running below its full employment capacity, which puts a decrease pressure on inflation. On the other hand, if $q > q^*$, the economy is overused, which strengthen inflationary pressures.

   These two gaps combined are called *price gap* :

   $$p - p^* = (v - v^*) + (q^* - q) \tag{2}$$

   Thus, the model is based on the idea that inflation accelerates or decelerates depending on these imbalances :

   - If both gaps are negative (velocity is low **and** output is high), this indicates strong inflationary pressures.
   - If both are positive, then inflation tends to slow down.
   - The return to price level equilibrium therefore occurs gradually, through the adjustment of nominal demand and the utilization of productive capacity.

2. **Short-term adjustment dynamics** : When $P < P^*$, That exerts inflationary pressure. On the other hand, if $P > P^*$, this tends to slow down inflation. These adjustments are modeled using an equation of the type :

   $$\Delta\pi_t = \alpha(p_{t-1} - p^*_{t-1}) + \gamma\pi_{t-1} + \sum_{j=1}^{n} \beta_j \Delta\pi_{t-j} \tag{3}$$

   From the equation (11) of the article.

   Where:

   - $\Delta\pi_t$ is the acceleration of inflation between periods $t-1$ and t.
   - p_{t-1} is the log of the actual price level at time $t-1$.
   - p^*_{t-1} is the log of the long-run equilibrium price level at time $t-1$.
   - $(p_{t-1} - p^*_{t-1})$ is the *price gap* at t-1.
   - π_{t-1} is the observed inflation rate at time $t-1$.

- $\Delta\pi_{t-j}$ are the lagged inflation accelerations, capturing short-run dynamic adjustments.
- $\alpha < 0$ ensures convergence of actual prices toward their monetary equilibrium level.
- ε_t is a white noise disturbance term.

1.2.3 Hallman Porter and Small Results

First, it was important to test the stationarity of velocity. The authors use an ADF test, which proves to be conclusive : "*As is also shown in Table 1, a unit root for velocity can be rejected for the period after 1955, thereby justifying the use of the sample mean of velocity as a measure of long-run velocity. The unit root for velocity is not rejected for the earlier period*" (Hallman et al., 1991). This assumption was not verified for the period 1870–1954

The empirical results obtained in the article confirm a robust convergence of the general price level toward the theoretical equilibrium level P*, constructed from monetary fundamentals (M2, potential output, and long-term velocity). This convergence dynamic is highlighted by the estimation of a "price-gap" model (Equation 9, p. 847), in which the gap between the observed price level and P* exerts a significant and negative effect on inflation acceleration. Thus, the model accurately reproduces the way in which monetary imbalances gradually translate into inflationary pressures.

Moreover, the P* model outperforms competing approaches based solely on the output gap or the Phillips curve in terms of forecasting accuracy. The authors demonstrate, through a series of in-sample and out-of-sample comparisons (Section III, particularly Tables 2, 3, 4, 5, and 6), that the P* model provides more accurate and stable forecasts than classical models, including ARIMA models and forecasts from major private institutions such as Merrill Lynch, DRI, and the Conference Board.

The implications for monetary policy are significant. As the authors emphasize, the P* model serves as an operational tool for assessing the gap between the observed price level and its benchmark level based on monetary aggregates. It thus provides a clear and quantifiable signal of potential medium-term inflationary pressures. This role as a leading indicator for monetary policy based on a rigorous connection between theory and data enhances its relevance, especially in light of the limitations of traditional cyclical indicators.

It is precisely this threefold strength of the model — its conceptual clarity, empirical robustness, and practical usefulness for policymakers — that makes this article the true foundational text of the P* concept. It offers a rigorous reformulation of the quantity theory of money, adapted to contemporary conditions and suitable for integration into a long-term price forecasting and policy framework.

1.3 Contributions of P* model to literature

Following the initial work of Hallman, Porter, and Small (1991), the P* model has generated considerable interest in monetary macroeconomics, particularly in the empirical analysis of the relationship between money and inflation. Several authors have sought to apply and refine the model in various contexts, notably by adjusting the assumptions

regarding the velocity of money. Tödter and Reimers (1994), for example, introduced a significant innovation by incorporating a long-term money demand function to capture the empirically observed downward trend in the velocity of the M3 aggregate in Germany. They thus demonstrate that the P* model remains relevant for explaining German inflation, thereby providing ex post validation of the Bundesbank's monetary targeting policy implemented since 1975. Gerlach and Svensson (2001) they extend the use of the model to the euro area, explicitly introducing the concept of *real money gap*, which they consider as an indactor superior to the *money-growth indicator* traditionally used by the European Central Bank.
Finally, Orphanides and Porter (2000) propose an innovative approach for the United States by using non-parametric methods (regression trees) to dynamically estimate a time-varying equilibrium velocity, thereby addressing empirical criticisms regarding the constancy of velocity initially assumed by Hallman et al. (1991).

1.3.1 Empirical validity of the P* model at an international scale

The validity of the P* model has been tested beyond the United States and Germany. Hoeller and Poret (1991) have examined the relevance of the model in several OECD countries, showing mixed performance depending on national contexts and the stability of money demand functions. In particular, the application of the P* model to Japan by Kole and Leahy (1991) reveals encouraging results, although monetary velocity also shows a marked trend that requires specific adjustments similar to those proposed by Tödter and Reimers (1994). It does, however, confirm the model's relevance in Germany, using the M3 aggregate.

More recently, the validity of the model in the euro area, as studied by Gerlach and Svensson (2001), indicates an empirical robustness. These authors show that *real money gap*, Directly derived from the P* model, offers a superior predictive power compared to other traditional indicators such as the output gap or simple monetary growth, in the European context of the period 1980s–2000s.

1.3.2 Theoretical and empirical limits of the model

Despite these encouraging empirical results, the P* model presents several fundamental limitations. First, the initial assumption of a constant or predictable long-term velocity central for Hallman et al. (1991) has been strongly challenged by certain empirical studies.(Orphanides and Porter, 2000). Structural instability related to financial and regulatory innovations alters money demand and, consequently, monetary velocity, making the model's predictions potentially fragile if this instability is not properly accounted for (Pecchenino and Rasche, 1990).
Moreover, the P* model implicitly assumes stable long-term relationships between the money supply, potential GDP, and velocity. However, episodes of economic crises, structural breaks, or monetary integration (for example German unification, analyzed by Tödter and Reimers (1994)) questions the stability of these relations.

The P* model faces several major criticisms in the recent literature. Woodford (2003) has questioned the very relevance of a direct monetary anchor, arguing that contemporary monetary policies focused on direct inflation targeting and the use of interest rate rules (such as the Taylor rule) provide more robust theoretical and empirical frameworks. A rigid monetary targeting approach, particularly in economies where financial innovations

continuously alter money demand, cannot be justified by the model P, who, such as stated by Svensson (2000), lacks explicit macroeconomics foundations.
Moreover, Rudebusch and Svensson (1999) highlights the lack of any significant predictive advantage of the *money-growth indicator* derived from the P* model compared to models based on interest rates or on expected inflation itself.

1.3.3 Extensions and upgrade proposed

In response to these criticisms, several extensions of the P* model have been proposed in the literature.Orphanides and Porter (2000) propose to replace the assumption of constant velocity with dynamic estimates using non-parametric methods, allowing for better consideration of structural changes.
Similarly, Gerlach and Svensson explicitly recommend prioritizing the *real money gap* rather than the simple monetary growth gap, thereby directly incorporating the velocity gap into the analysis. This approach is also supported by the ECB itself in its recent studies (ECB, 2001).

Finally, alternative approaches based on by targeting inflation directly and the use of interest rate based rules (for example, modified Taylor rules or structural VAR models) are increasingly emphasized in the literature as credible alternatives to strictly monetary frameworks represented by the P* model(Woodford, 2003; Clarida et al., 1999)).

In conclusion, although the P* model retains a predictive value in certain contexts, it requires significant empirical and theoretical adaptations to remain relevant in contemporary monetary policy..

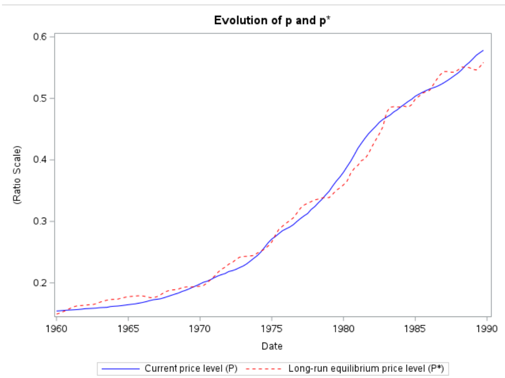
2 Empirical application of the p* concept

This section aims at empirically test the validity of the concept of P^* (P-star) developed by Hallman, Porter, and Small (1991), applied to U.S. economic data from 1960 to 1990, and then from 1990 to the end of 2018, as well as to Japanese data from 1994 to 2016, and British data from 1983 to 2016. This temporal choice avoids major disruptions related to the COVID-19 pandemic, thereby ensuring the homogeneity of the time series studied, and is also determined by data availability.

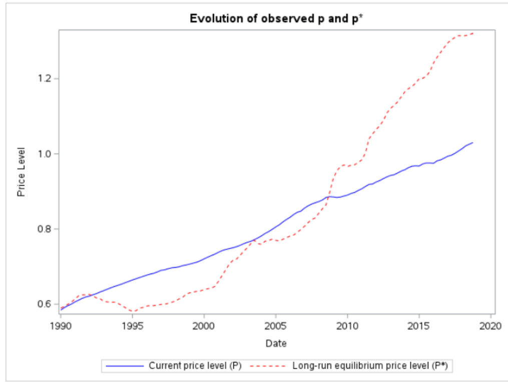
2.1 Data Preparation and Initial Processing

The data used include the M2 money supply, monetary velocity, nominal GDP, and real GDP for the United States, Japan, and the United Kingdom. These series come from official sources and cover the specified periods.
After import, these datasets were merged by date of observation. The variable P, corresponding to our price index, is obtained by dividing nominal GDP by real GDP, thus representing the implicit GDP deflator. In order to estimate the level of potential output, denoted Q^*, we applied a 4-period moving average to the quarterly real GDP (Q). This operation was done by using the SAS procedure `PROC EXPAND`. This transformation aims to smooth short-term fluctuations in observed GDP by capturing its long-term trend. In practice, the 4-period moving average (equivalent to one year for quarterly data) helps to neutralize seasonal and cyclical effects, in order to better isolating the structural component of GDP, in line with the approach followed by Hallman, Porter, and Small in their modeling of the concept P^*, while the long-term velocity V^* is estimated by the average of the observed velocity over the entire study period.
Based on these elements, the P^* series, representing the long-term equilibrium price level, is constructed according to the formula in the equation. (1).

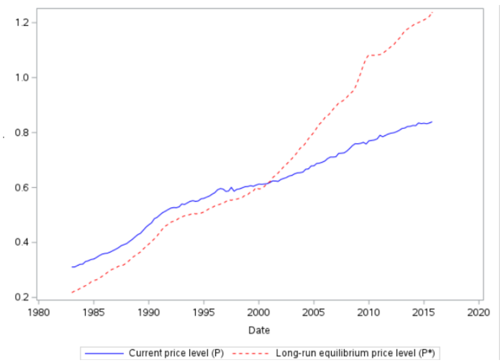
The evolution of the variables P and P^* then obtained were visualized for a preliminary graphical analysis of long-term trends :



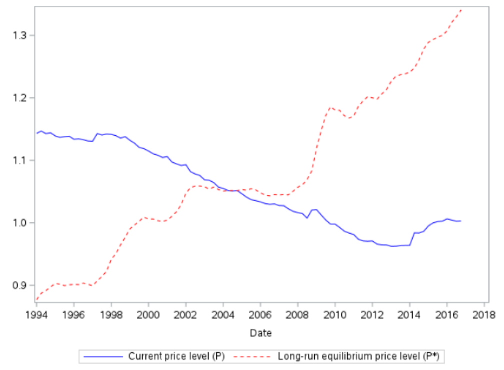
(a) USA 1960–1990



(b) USA 1990–2018



(c) UK 1983–2016



(d) Japan 1994–2016

Figure 1: Observed Price Level Evolution (p) and equilibrium price evolution (p^*) in different countries and periods

The presented graphs compare the evolution of the observed price level P (blue line) at the theoretical equilibrium level P*(dotted red line) for different countries and periods. We observe for the United States a strong co-movement for the period 1960 - 1990, which becomes much weaker in the following period. n fact, the resulting graph is almost identical to Figure (1) in Hallman et al. (1991) on the first period. For the United Kingdom, the relationship seems to be relatively strong between 1983 and 2002, before deteriorating, possibly due to the euro adoption by its European neighbors. Finally, for Japan, there is a complete divergence between P and P*, which suggests that the P* model may not be relevant for explaining price fluctuations in this country.

In a second step, all variables were transformed into natural logarithms in order to stabilize variance and facilitate econometric testing. First differences were also calculated to determine the order of integration of the series.
Augmented Dickey-Fuller (ADF) tests (with 5 lags) were performed on the variables in levels and in first differences to assess their stationarity. These tests allow for a precise identification of the order of integration of each time series. It is necessary to obtain series integrated of order 1 : $I(1)$, That is, the variables are non-stationary in levels but stationary in first differences. This condition ensures that the variables may potentially share a stable long-run relationship, which is a prerequisite for testing cointegration and properly modeling economic dynamics through an error correction model. According to

the ADF tests, velocity is indeed stationary for the three countries studied, which supports the choice of using the average including for the United States over the 1990–2018 period despite this approach being questioned in several articles discussed in section 1.3.

Tests de racine unitaire de Dickey-Fuller augmentés							
Type	Retards	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Moyenne zéro	0	-42.0478	<.0001	-5.06	<.0001		
	1	-44.4551	<.0001	-4.67	<.0001		
	2	-37.4225	<.0001	-3.97	0.0001		
	3	-30.2899	<.0001	-3.40	0.0008		
	4	-34.6699	<.0001	-3.35	0.0010		
	5	-31.9946	<.0001	-3.09	0.0023		
Moyenne simple	0	-43.6616	0.0010	-5.17	<.0001	13.37	0.0010
	1	-46.8894	0.0010	-4.78	0.0002	11.43	0.0010
	2	-40.0075	0.0010	-4.05	0.0017	8.21	0.0010
	3	-32.8398	0.0010	-3.48	0.0104	6.06	0.0114
	4	-39.1165	0.0010	-3.47	0.0108	6.01	0.0125
	5	-37.5193	0.0010	-3.24	0.0207	5.24	0.0315
Tendance	0	-48.7147	0.0004	-5.48	<.0001	15.03	0.0010
	1	-56.3777	0.0004	-5.20	0.0002	13.52	0.0010
	2	-53.4816	0.0004	-4.56	0.0019	10.38	0.0010
	3	-48.8042	0.0004	-4.00	0.0114	7.99	0.0099
	4	-69.1640	0.0004	-4.00	0.0113	8.02	0.0094
	5	-77.7166	0.0004	-3.72	0.0249	6.98	0.0325

Figure 2: Results of the stationarity test on first differences of velocity, USA 1990–2018

The other series are also $I(1)$. We can test the cointegration.

2.2 Cointegration Test and Error Correction Model (ECM)

2.2.1 Cointegration Tests on Prices

To verify the existence of a long-term relationship between the general price level (p) and its theoretical equilibrium level (p^*), we conducted a cointegration test using the two-step method proposed by Engle and Granger (1987). First, a regression of p on p^* was performed, and the residuals from this regression were extracted. An Augmented Dickey-Fuller (ADF) test was then applied to these residuals to test for stationarity. If the residuals are stationary (reject $H0$), this shows that p and p^* are cointegrated, hat is, they follow a stable long-term equilibrium relationship despite being individually non-stationary. This result also serves as a condition for estimating an error correction model, which will structure short-term dynamics around this long-term relationship.

Test de cointégration de Engle et Granger			
Type	Retards	Tau	Pr < Tau
Moyenne simple	5	-3.8558	0.0044
Tendance	5	-3.7159	0.0614

Figure 3: Engle and Granger cointegration test - USA 1960 - 1990

The results of the Engle and Granger cointegration tests show that only the 1960–1990 period in the United States allows for the rejection of the null hypothesis of no cointegration between p and $p*$. The p-values obtained for this sample are significantly below the significance thresholds of 5 and 10 %, indicating the existence of a long-term relationship between the two variables. In contrast, for the other periods and countries analyzed (Japan, the United Kingdom, and the United States post-1990), the p-values from the tests are well above conventional thresholds (reaching up to 0.97), which prevents us from concluding that cointegration exists. In these cases, the use of an error correction model (ECM) is not econometrically justified, as such modeling explicitly requires the existence of a long-term relationship between the variables. Without cointegration, short-term dynamics cannot be properly specified through a mechanism of adjustment toward equilibrium, thus undermining the validity of the model. Consequently, the ECM analysis will be conducted only on U.S. data from 1960 to 1990. Although the cointegration test does not formally confirm the existence of a long-term relationship between p and $p*$ on the period after 1990 (p-value $>$ 10 %), n ECM will be estimated on an exploratory basis. The objective here is to examine whether the short-term dynamics remain consistent with the theoretical intuition of the model $P*$, even in the absence of a statistically established long-term relationship.

2.2.2 Error-Correction Model (ECM) Specification and Estimation

To capture both the short-term dynamics and the long-run equilibrium adjustment, we estimate an Error-Correction Model (ECM) based on the gap between the observed price level p and the theoretical equilibrium price level p^* derived from the P* model.

Construction of the error-correction term

First, we compute the cointegration residual:

$$ECM_t = p_t - p_t^*$$

This variable represents the deviation from the long-run price equilibrium. Its lagged value, denoted as `lag_ECM`, captures the degree of disequilibrium at time $t-1$, which the inflation rate is expected to correct progressively over time if cointegration is present.

Model specification

The estimated ECM is specified as follows:

$$\Delta p_t = \gamma \cdot \texttt{lag\_ECM} + \sum_{i=1}^{3} \delta_i \cdot \Delta p_{t-i} + \sum_{i=1}^{3} \theta_i \cdot \Delta m_{t-i} + \sum_{i=1}^{3} \phi_i \cdot \Delta q_{t-i} + \sum_{i=1}^{3} \psi_i \cdot \Delta v_{t-i} + \varepsilon_t \tag{4}$$

where:

- Δp_t is the inflation rate, measured as the first difference of the logarithm of the observed price level.
- Δm_t, Δq_t, and Δv_t denote the first differences of the logarithms of the observed money supply, real output, and velocity, respectively.
- γ measures the speed at which inflation adjusts to the long-run disequilibrium. A significantly negative value of γ supports the presence of a stable error-correction mechanism;
- The coefficients δ_i, θ_i, ϕ_i, and ψ_i capture short-run effects through the inclusion of three lags of inflation, money growth, real output growth, and velocity changes, respectively.

Results

Valeurs estimées OLS Paramètre non linéaires				
Paramètre	Estimation	Err type approx.	Valeur du test t	Approx Pr > |t|
gamma	-0.02418	0.0109	-2.21	0.0290
delta1	1.076237	1.5817	0.68	0.4978
delta2	0.215823	1.8920	0.11	0.9094
delta3	0.745432	1.5778	0.47	0.6376
theta1	-0.47994	1.5814	-0.30	0.7621
theta2	-0.04534	1.8890	-0.02	0.9809
theta3	-0.52514	1.5729	-0.33	0.7392
phi1	0.408123	1.5806	0.26	0.7968
phi2	0.089332	1.8834	0.05	0.9623
phi3	0.532307	1.5686	0.34	0.7350
psi1	-0.43189	1.5801	-0.27	0.7851
psi2	-0.08434	1.8849	-0.04	0.9644
psi3	-0.49934	1.5703	-0.32	0.7511

Figure 4: Significance test on the coefficients of the ECM model

The error correction coefficient γ is significantly negative at 5 %, which confirms the existence of a partial adjustment of inflation toward its long-term equilibrium level determined by 2,4 % of the gap corrected each quarter. The other coefficients associated with short-term effects (lags of inflation, money supply, real GDP, and velocity) are not statistically significant. This suggests that, during this period, inflation dynamics are primarily driven by the error correction mechanism, while transitory short-term shocks do not appear to exert any systematically detectable effect.

In order to test the robustness of the empirical framework, two variants were estimated. On one hand, the GDP deflator was replaced by the Consumer Price Index (CPI) as the measure of the price level. On the other hand, the monetary aggregate$M2$ has been substitute by $M3$, a larger aggregate, potentially more representative of the overall liquidity. In both cases, the results do not show any significant improvement in the model's performance P, either in terms of cointegration or predictive quality. These results are consistent with those of Kuttner (1990), who also test the robustness of P faced with different price measures (other than the GDP deflator), and concludes that the model's predictive performance remains limited when alternative indices such as the CPI are used.

2.3 Analysis of the Performance of p*

2.3.1 In-sample performance

A preliminary in-sample evaluation of the model's forecasts was also used in order to assess its internal predictive capacity an essential preliminary step before any subsequent out-of-sample validation.
This methodology enables a complete evaluation consistent with the approach originally

proposed by Hallman et al., while providing a solid foundation for further analyses, including robustness checks and out-of-sample forecasting.

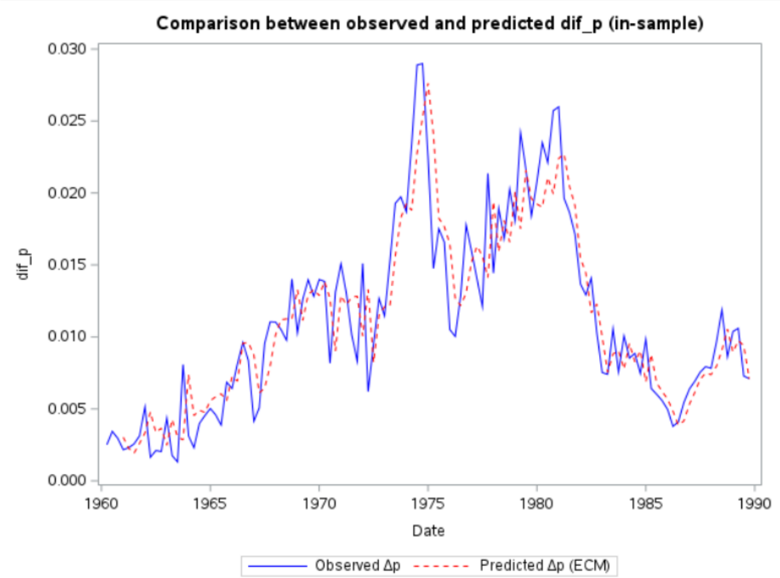


Figure 5: Comparison Between observed and predicted dif p - in sample- USA 1960-1990

The in-sample evaluation of the ECM model over the 1960–1990 period shows strong explanatory power, with a coefficient of determination of $R^2 = 0.82$. The comparative graph between observed and predicted inflation values confirms this performance, with the model accurately capturing the main inflationary dynamics. These results support the validity of the P* framework for modeling inflation in the United States during this period, highlighting both the importance of long-term disequilibrium and the relevance of short-term adjustments.

In the absence of formally established cointegration, the post-1990 ECM model lacks solid econometric foundations. Nevertheless, its estimation remains informative: the explanatory power is more limited than in the earlier period $R^2 = 0.45$, and the fit between observed and predicted inflation remains partial. This reinforces the idea that the relationship between p and p* has likely weakened or transformed since the 1990s.

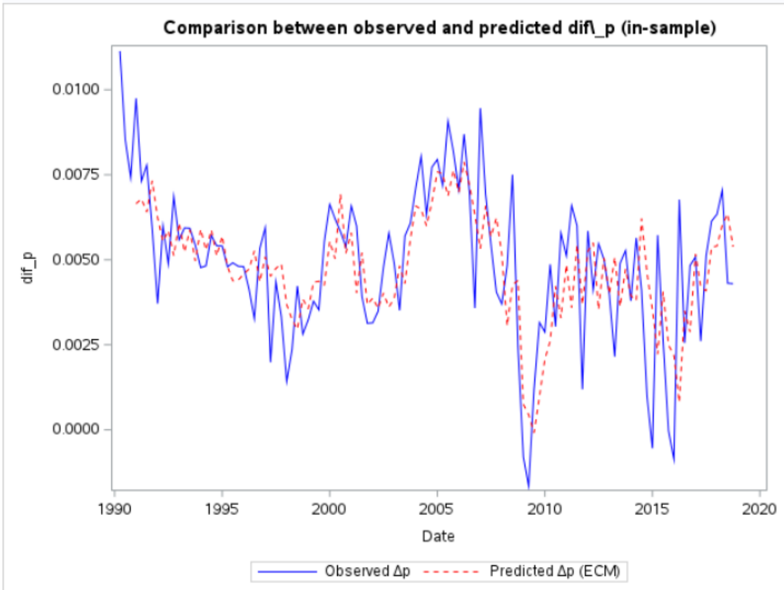


Figure 6: Comparison Between observed and predicted dif p - in sample- USA 1990 - 2018

2.3.2 Out-sample Performance

The evaluation of in-sample performance, while informative, is not sufficient to assess the predictive robustness of a model. A high capacity of adjustments on the estimation sample can overshadowed overfitting or also a structural mismatch when the model is applied to non observed data.

The sample was split into two sub-periods: a first period up to 1985, used for estimating the ECM model, and a second period from 1985 to 1990, used for prediction *out-of-sample*. This split allows testing the model's ability estimated solely on historical data to reproduce inflation dynamics over a period not included in the estimation.

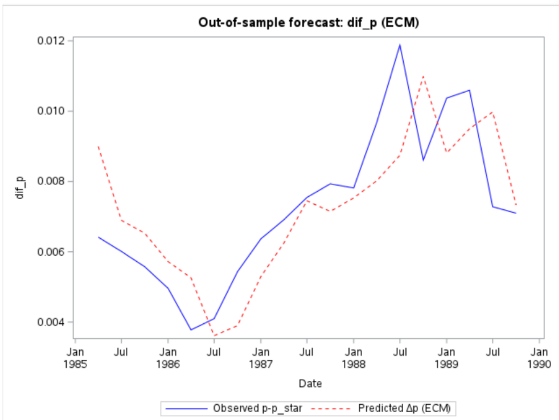


Figure 7: Out-of-sample forecast of quarterly inflation (Δp) using the ECM model (1985–1990)

The model seems to capture the overall dynamics well, although some occasional deviations remain visible, particularly when the inflation peaks. The quantitative evaluation of these forecasts is summarized by the RMSFE (Root Mean Square Forecast Error), which amounts to 0.0015. This low root mean square forecast error indicates a reasonable performance of the model when we talk about out-sample forecasting. However, the RMSFE only becomes meaningful when compared to a benchmark model. At this stage, it does not allow us to conclude that the performance is "good" in absolute terms. This forecast will be compared to a random walk in the following subsection using a Diebold-Mariano test.

2.4 Comparison with a Random Walk: Diebold-Mariano Test

To complete our analysis, we used the Diebold and Mariano test in order to statistically compare the forecasts from the error-correction model (ECM) to those of a random walk, namely a naive forecast of constant inflation. This test allows us to go further in our analysis and move beyond a simple graphical or numerical comparison of average errors by incorporating the variability of the differences in errors between the two models.

The procedure of the test begins with the calculation of squared errors. For each observation date, the forecast errors of both models are squared in order to highlight extreme deviations. The second step of the test consists in calculating the differences between the errors of the ECM model and those of the naive model. We thus construct the series d_t as follows:

$$d_t = \text{error}^2_{\text{ECM}} - \text{error}^2_{\text{naive}} \tag{5}$$

Finally, the Diebold and Mariano statistic is calculated under the null hypothesis H_0 of equal forecasting accuracy. The standardized, asymptotically normal statistic is given by:

$$DM = \frac{\text{mean}_d}{\sqrt{\text{var}_d / n}} \tag{6}$$

With mean_d the mean of the differences, var_d their variance, and n the number of observations. The associated p-value allows us to conclude on significance.

The results of the test in the context of the United States between 1960 and 1990 indicate a DM statistic of -3.236233 with a p-value of 0.0012112. If the p-value is less than 0.05, H_0 is significantly rejected at the 5% level. This means that the errors of the two models are significantly different and that one model is better than the other. Then, in order to know which model is significantly better, we look at the DM statistic. If $DM < 0$, the ECM model has significantly better forecasting accuracy than the naive model. If $DM > 0$, the naive model is better. If the p-value is greater than or equal to 0.05, we cannot conclude a significant difference between the two models, and thus neither model statistically outperforms the other.

In our case, the p-value of 0.0012112 and the DM statistic of -3.236233 allow us to reject H_0 at the 5% level and thus confirm the statistical superiority of the ECM model over the naive model.

We can also address the limitations and perspectives of the Diebold and Mariano test. First, concerning autocorrelation of errors. In fact, the standard DM test assumes uncorrelated differences. For multi-period forecasts, a correction of the variance would be necessary. The use of alternative benchmarks could also be interesting. Indeed, although the naive model is a classical reference, comparisons with other models (ARIMA, Machine

Learning) could strengthen the robustness of the conclusions.

This analysis using the Diebold and Mariano test thus validates the contribution of the ECM model in the context of the United States between 1960 and 1990, while underlining the importance of statistical testing in verifying comparisons between forecasting models.

Conclusion

This study rigorously examined the predictive validity of the P model in multiple economic contexts, and led to three key conclusions. First, the model demonstrates strong empirical performance for U.S. data during the 1960-1990 period, with cointegration tests confirming a stable long-term relationship between observed prices and their monetary equilibrium level. The error correction model is statistically significant, indicating that quarterly inflation adjusts toward the implicit equilibrium P. This is consistent with the original results of Hallman, Porter and Small (1991).

Secondly, the predictive superiority of the model is confirmed by Diebold and Mariano tests, showing significantly lower forecast errors compared to a random walk benchmark. However, this robustness decreases after 1990 as the cointegration tests fail and the explanatory power decreases sharply. Similar limitations appear in applications to other countries. Indeed, while the UK data (1983–2002) shows a temporary alignment with the dynamics of P, Japanese inflation (1994–2016) is completely out of line with the model forecast.

Regarding theoretical and policy implications, the results highlight two key points. First, monetary anchors hold conditional importance. Indeed, P retains its value when demand for money is stable (before 1990 in the United States), but financial innovations and velocity changes undermine its reliability in contemporary contexts. Furthermore, the model serves better used as a diagnostic tool to identify price deviations than as a real-time forecasting device, especially given the 1 to 2 year lag in monetary transmission. These results support the argument of Gerlach and Svensson (2001) for using P derived gaps as additional indicators rather than primary policy targets.

Then, regarding limitations and research extensions, we identify three main constraints. Structural ruptures are a first limit. Indeed, the constant velocity model hypothesis is particularly fragile during periods of financial deregulation. In addition, comparisons were limited to random walks. Future work should be tested against modern benchmark models. Finally, data-related constraints emerge. Indeed, estimates of potential output through moving averages may exceed the effects of the business cycle.

Promising extensions could incorporate time-varying velocity using state-space methods, develop hybrid specifications combining P deviations with Phillips curve elements or even expanding the sample of countries to emerging markets with different monetary regimes.

While the P model provides a theoretically effective synthesis of quantity theory and error correction mechanisms, its operational relevance has declined in an era of unstable monetary demand. Nevertheless, as a diagnostic framework for the detection of monetary imbalances, particularly when complemented by dynamic velocity adjustments, it retains analytical value for central banks monitoring long-term inflation risks. This study ultimately reaffirms Friedman's (1963) sustainable vision: inflation remains a long-term monetary phenomenon, but the "long-term" itself evolves with the complexity of the financial system.

References

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Quarterly Review*, 25(Win):2–11.

Clarida, R., Gali, J., and Gertler, M. (1999). The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature*, 37(4):1661–1707.

Engle, R. F. and Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276.

Friedman, M. (1968). The Role of Monetary Policy. *The American Economic Review*, 58(1):1–17.

Friedman, M. and Schwartz, A. J. (1963). Money and Business Cycles. *The Review of Economics and Statistics*, 45(1):32–64.

Gerlach, S. and Svensson, L. E. O. (2001). Money and inflation in the euro area: A case for monetary indicators? *Journal of Monetary Economics*, 50(8):1649–1672.

Gordon, R. J. (2013). The Phillips Curve is Alive and Well: Inflation and the NAIRU During the Slow Recovery.

Hallman, J. J., Porter, R. D., and Small, D. H. (1991). Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run? *The American Economic Review*, 81(4):841–858.

Hoeller, P. and Poret, P. (1991). P-Star as an Indicator of Inflationary Pressure. *OECD Economics Department Working Papers*, (101).

Kole, L. S. and Leahy, M. P. (1991). The usefulness of P* measures for Japan and Germany. *International Finance Discussion Papers*, (414).

Kuttner, K. N. (1990). Inflation and the growth rate of money. *Economic Perspectives*, 14(Jan):2–11.

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, 4(2):103–124.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

Mankiw, N. G. and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, 117(4):1295–1328.

Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3):315–335.

Orphanides, A. and Porter, R. D. (2000). $P*$ revisited: Money-based inflation forecasts with a changing equilibrium velocity. *Journal of Economics and Business*, 52(1):87–100.

Phelps, E. S. (1967). Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time. *Economica*, 34(135):254.

Rudebusch, G. and Svensson, L. E. O. (1999). Policy Rules for Inflation Targeting. *NBER Chapters*, pages 203–262.

Sargent, T. (1973). Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment. *Brookings Papers on Economic Activity*, 4(2):429–480.

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

Stock, J. H. and Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15(4):101–115.

Svensson, L. E. O. (2000). Does the P* Model Provide Any Rationale for Monetary Targeting? *German Economic Review*, 1(1):69–81.

Tödter, K.-H. and Reimers, H.-E. (1994). P-Star as a Link between Money and Prices in Germany. *Weltwirtschaftliches Archiv*, 130(2):273–289.

Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press.