The Hierarchical Structure of Price Changes and Core Inflation

Paulo Picchetti

paulo.picchetti@fgv.br

Instituto Brasileiro de Economia/Escola de Economia de São Paulo
Fundação Getulio Vargas - Brazil

March 5, 2015

Abstract

The approach proposed here explores the hierarchical nature of item-level data on price changes. On one hand, price data is naturally organized around a regional strucuture, with variations being observed on separate cities. Moreover, the itens that comprise the natural structure of CPIs are also normally interpreted in terms of groups that have economic interpretations, such as tradables and non-tradables, energy-related, raw foodstuff, monitored prices, etc. The hierarchical dynamic factor model allow the estimation of multiple factors that are naturally interpreted as relating to each of these regional and economic levels.

JEL: C32, C42, C53, E32
Keywords: Core Inflation, Dynamic Hierarchical Factor Models.

1 Introduction

Consumer Price Indexes are normally reported in terms of the time variation of a weighted average of individual prices changes. While this measure synthesizes a large amount of information about the economy, it masks some interesting information contained in the disaggregated price movements. First, there is regional heterogeneity, given that national-level indicators such as this are composed by information captured in different regional units. Second, individual prices reflect both aggregate shocks - steming from macroeconomic fundamentals, as well as idiosyncratic shocks - resulting from specific supply conditions. Attempts to decompose price variations into orthogonal latent variables designed to capture aggregate demand and individual supply shocks are the essence of the literature on Core Inflation.

Core inflation measures have become an essential tool for policy making and forecasting, but there is still no consensus on the best statistical procudure to evaluate it. Among the best practices, dynamic factor models have been widely adopted. One difficulty however remains regarding the interpretation of endogenously determined multiple factors, insofar as the statistical procedure has no direct theoretical structural counterparts. The approach proposed here explores the hierarchical nature of item-level data on price changes. On one hand, price data is naturally organized around a regional strucuture, with variations being observed on separate cities. Moreover, the itens that comprise the natural structure of CPIs are also normally interpreted in terms of groups that have economic interpretations, such as tradables and non-tradables, energy-related, raw foodstuff, monitored prices, etc. The hierarchical dynamic factor models allow the estimation of multiple factors that are naturally interpreted as relating to each of these regional and economic levels, so that overall price changes can be decomposed into a common factor for the whole economy (the measure of core inflation), and shocks which are specific to each of the proposed levels. This paper details the statisitcal procedure for estimating these dynamic factors, and provides an empirical application to a dataset used to produce a Brazilian CPI.

2 Data

The Brazilian Broad Consumer Price Index (henceforth IPCA) is calculated by Instituto Brasileiro de Geografia e Estatística (IBGE) on a monthly basis since 1980. In its current configuration, IPCA covers 10 metropolitan regions and 3 cities. Although the index has been calculated since 1980, the analysis here will start on January 1995, the reason being the stabilization of the Braziian economy achieved in mid-1994, after two decades of inflation characterized by extremely large rates of variation and volatility, implying that any meaningful statistical analysis cannot be performed for this period. Figure 1 shows the evolution of the monthly IPCA from January 1995 to December 2014.

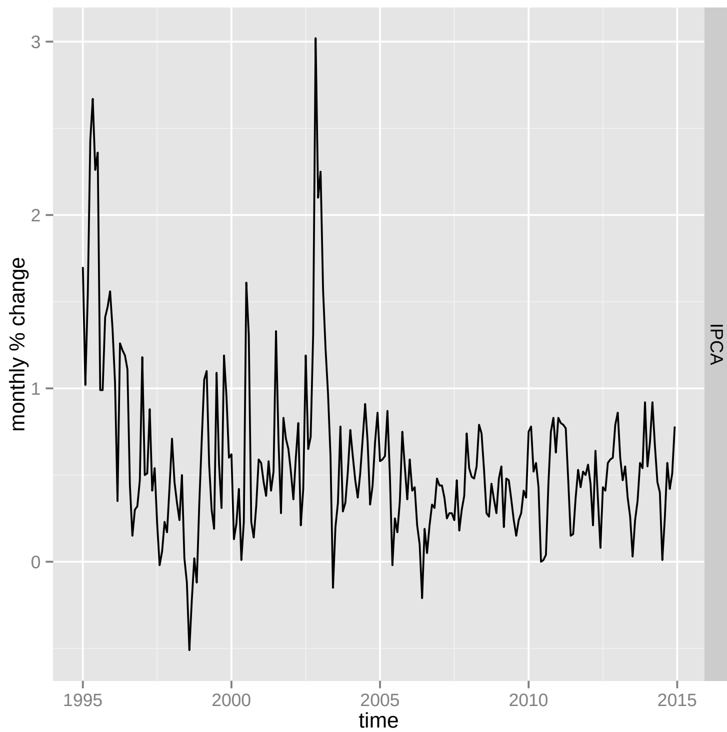

Figure 1: IPCA

In its current structure, IPCA is divided into 8 classes (Foodstuff, Transportation, Household expenses, Health, Education, Clothing, Leisure and Communications), which are then divided into 52 Items (see Table 1), and finally in 373 sub-items. This will provide the hierarchical structure which will be explored here.

3 Model

3.1 Three Levels

The statistical model applied here seeks to estimate the commonality in the evolution of item prices, at different levels of aggregation. These latent variables are structured as factors which are formally dependent on higher levels of the hierarchy. Each individual series (items of the IPCA) is denoted by X_{ibt}, i being the series subscript, b the block where the series belongs and

t the particular moment in time. G_b stands for the factor related to block b, while e_{Xibt} is the idiosyncratic variation of a particular series at a particular point in time. The block factors G_{bt} are related to the aggregate factors F_t, which are in turn dynamically formulated as depending on their own past values, and on their own unexplained variations (interpreted as error terms).

Using the notation where individual series and factors as grouped in vectors

$$X_{bt} = (X_{bt.1}, X_{bt.2}, \dots, X_{bt.N_b})$$
$$G_{bt} = (G_{bt.1}, G_{bt.2}, \dots, G_{bt.k_b})$$

the model can be concisely formulated including dynamics as

$$\begin{aligned}
X_{bt} &= \Lambda_{G.b0} G_{bt} + \dots + \Lambda_{G.bs_{Gb}} G_{b,t-s_{Gb}} + e_{Xbt} \\
G_{bt} &= \Lambda_{F.b0} F_t + \dots + \Lambda_{F.bs_F} F_{t-s_F} + e_{Gbt} \\
F_t &= \Psi_{F.1} F_{t-1} + \dots + \Psi_{F.q_F} F_{t-q_F} + \epsilon_{Ft} \\
e_{Gbt} &= \Psi_{G.b1} e_{Gb,t-1} + \dots + \Psi_{G.bq_{Gb}} e_{Gb,t-q_{Gb}} + \epsilon_{Gbt} \\
e_{Xbit} &= \Psi_{X.bi1} e_{Gb,t-1} + \dots + \Psi_{X.biq_{Xb}} e_{Xbit,t-q_{Xb}} + \epsilon_{Xbit}
\end{aligned}$$

The error terms in all equations follow the usual distributional assumptions (normality, zero mean and constant variance).

3.1.1 Parameter Estimation

The parameters in the matrices Λ and Ψ, along with the variances of the error terms in all equations are jointly estimated from data with the factors in all levels, in the context of a state-space model where a joint probability distribution is assumed for all these random quantities. For details of inference using the Gibbs sampling algorithm for this model, see Moench E. , Ng S. and Potter S. (2009).

3.1.2 Groups

The blocks are defined by IPCA items group in five categories, representing different economic dynamics for prices:

Foodstuff Food items consumed inside the house.

Non-regulated services Services provided by independent professionals, not subject to direct government intervention.

Regulated Prices Services either directly provided by the government, or whose prices are subject to direct intervention by the government.

Non and semi-durable consumer goods Goods not purchased on long-term credit.

Durable consumer goods Goods normally purchased on several installments or directly financed.

Table 1 contains every IPCA item along with its code grouped according to the five blocks listed above.

Group	Item
Foodstuff	1101.Cereals and oilseeds 1102.Flour, starches and pastas 1103.Tubercles, roots and greenstuff 1104.Sugar and derivatives 1105.Vegetables 1106.Fruits 1107.Meats 1108.Fish 1109.Processed meats and fish 1110.Poultry and eggs 1111.Milk and derivatives 1112.Bakery 1113.Oils and fats 1114.Drinks 1115.Canned and preserved 1116.Salt and spices
Non-regulated services	1201.Meals (non-home) 2101.Rents and fees 2103.Repairs 3301.Maintenances 6201.Medical and dental services 6202.Hospital and lab services 7101.Personal services 7201.Leasure 8101.Regular courses 8104.General courses

Group	Item
Regulated Prices	2201.Fuels (home use) 2202.Electrical Energy (residential) 5101.Public transportation 5104.Fuels (automobiles) 6101.Pharmaceuticals 6203.Health plans 7202.Smokes 9101.Communications
Non and semi-durable consumer goods	2104.Cleansing products 3102.Utensils and decorations 3103.Bed, table and bath cloths 4101.Menswear 4102.Womenswear 4103.Childrenswear 4201.Shoes and accessories 4301.Jewelery 4401.Cloths 6102.Optical products 6301.Personal hygiene 7203.Photography 8102.Books and magazines 8103.Stationery
Durable consumer goods	3101.Furniture 3201.Appliances 3202.TVs, stereo systems and computers 5102.Personal vehicles

Table 1: Groups

The three-level model is estimated using data structured along these groups. The results can be visualized through the factors estimated along the hierarchy. Figure 2 shows the headline IPCA along with $\hat{F}_t$.

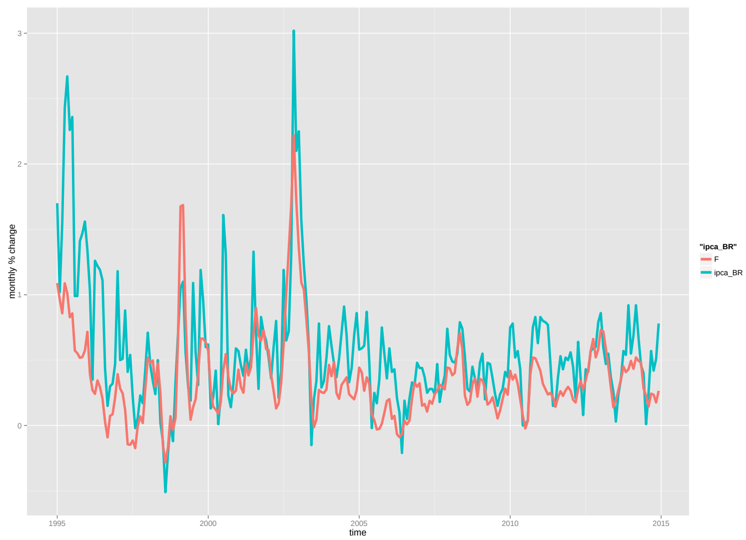

Figure 2: IPCA and aggregate factor

$\hat{F}_t$ can be interpreted as the commonality of behavior of price changes across groups, and as so provides an alternative measure of core inflation. One striking feature of Figure 2 is that $\hat{F}_t$ is much smoother than the headline IPCA, and also leads its behavior on cyclical trends, both desirable features of this kind of measurement.

Figure 3 shows the estimated group factors $\hat{G}_1, \ldots, \hat{G}_5$:

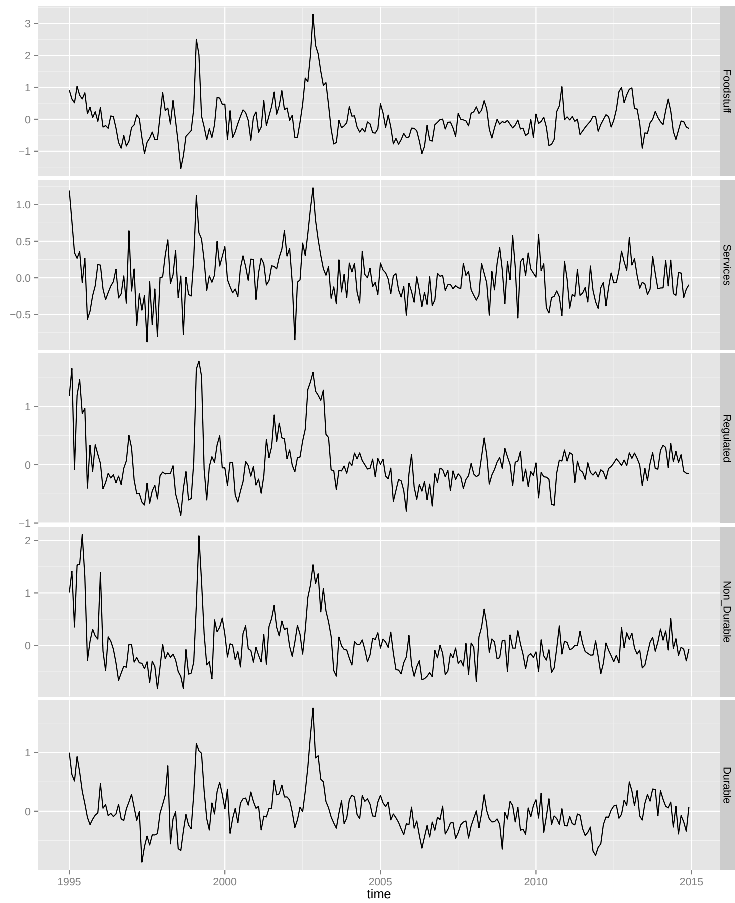

Figure 3: Block Factors (3 levels)

The heterogeneity of factors across different groups seems to justify the hierarchical structure assumed by the model. Again, these could be potentially estimated as different factors along a single level, but their interpretation would not be nearly as straightforward. Moreover, the quality of the statiscal estimates would also not be guaranteed.

A number of interpretations can be constructed based on these estimated factors, along different economic scencarios during the sample period in question. The first evident comovement of the factors occurs during the generalized deceleration in prices that consolidated throughout 1995 and 1996 the stabilization of the Brazilian currency started in mid-1994. Around this common trend, the factors for each group show distinctive behaviors regarding their volatility, with Services prices as the most volatile, and Foodstuff prices as the least. There is common spike around 1999 characterizing price responses from the Brazilian economy to internal (energy shortages) and external (dot.com bubble) shocks. Even so, the response functions of prices from these shocks vary considerably across sectors, both in terms of amplitude and timing. Another common spike occurs during 2002. This period is mainly characterized by a sharp rise in risk perception following change in political power, the most evident economic effect being a depreciation of the Brazilian Real. Expectations improved rapidly during 2003 when a healthy economic policy became credible. During this transition, the spike in prices for Durable Consumer Goods was rapidly reversed, while the one on the Regulated Prices sector decreased more slowly. Even after currency stabilization, contracts (among private sector entities and also on entreprises subject to regulation) typically include an indexation clause for annual price revisions, which naturally creates a lagging effect of shocks that affect any of the items included in the indexes used in these contracts. The period following the global crisis initiated in 2008 witnesses a somewhat smoother tracectory of prices. However, some episodes of economic interest are clearly depicted by the estimated factors. In the case of Durable Consumer Goods, there is a pronunciated dip after 2011. This is the period where attempts to boost weakened consumption included tax exemptions for nearly every Durable Consumer Good. The fall in commodity prices after 2008 had a direct impact on the most commonly used index for indexating contracts in the Brazilian economy. Accordingly, regulated prices oscilate along a clear downward trend between 2008 and 2011.

3.2 Four Levels

The structure of the IPCA panel also allows for further groupings of economic interest. The regional dimension can be explored, allowing specific factors to be estimated for the metropolitan areas where the index is calculated. In terms of the statistical model employed, this can be naturally represented by a four-level hierarchical structure, where we now have the items at level 1, grouped by the economic aggregates already employed in the 3-level structure, but now under a third level comprised by the nine metropolitan areas considered here. The headline IPCA sits above these three levels, representing the fourth level of the hierarchy. The inclusion of this new level is justified by the variability of the IPCA across regions, and the economic interest behind analysis such as the regional effects of monetary policy. Figure 4 presents the behavior of the healine IPCA across the nine considered metropolitan regions.

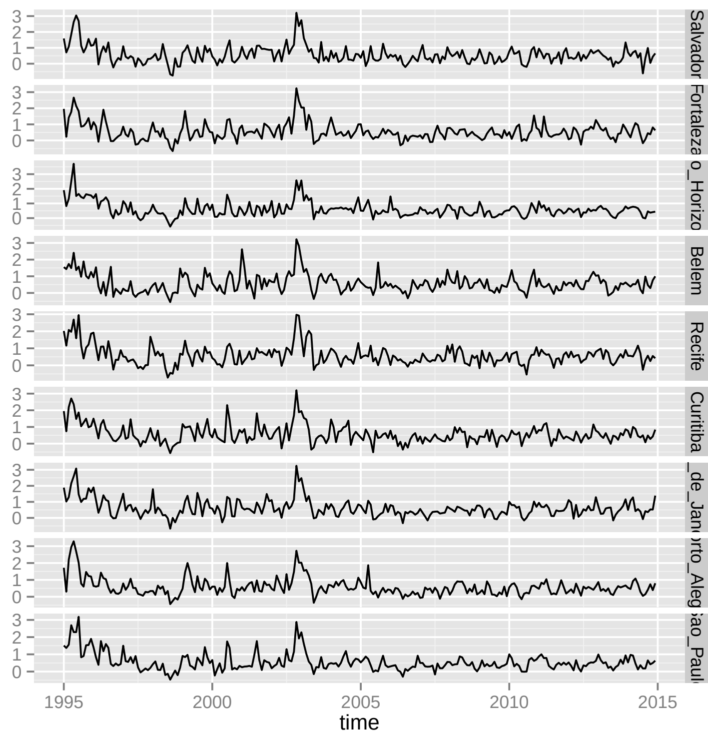

Figure 4: Block factors (4 levels)

The heteregeneity across the nine metropolitan regions can be captured by the inclusion of a fourth level in the hierarchy. Now we have the same individual series (IPCA items) at level 4, which are modeled as dependent on the five economic groups - the sub-blocks at level 3, in turn dependent on the nine metropolitan regions - the blocks at level 2, which finally depend on the aggregate factor at level one.

In terms of the above established notation, the individual series are now represented by Z_{bsit}, where the indices b, i and t have the same meanings as in the three levels model, but now augmented by s, the index for the sub-blocks. Denoting H_{bst} as the vector of factors at the sub-block level, the model becomes

$$
\begin{aligned}
Z_{bsit} &= \lambda_{H.bsi}(L)H_{bst} + e_{Xbsit}\\
H_{bst} &= \Lambda_{G.bs}(L)G_{bt} + e_{Hbst}\\
G_{bt} &= \Lambda_{F.b}(L)F_t + e_{Gbt}\\
F_t &= \Psi_F(L)F_{t-1} + \epsilon_{Ft}\\
e_{Gbt} &= \Psi_{G.b1}e_{Gb,t-1} + \ldots + \Psi_{G.bq_{Gb}}e_{Gb,t-q_{Gb}} + \epsilon_{Gbt}\\
e_{Hbst} &= \Psi_{H.bs1}e_{Hbs,t-1} + \ldots + \Psi_{H.bsq_{Hbs}}e_{Hbs,t-q_{Hbs}} + \epsilon_{Hbst}\\
e_{Xbit} &= \Psi_{X.bi1}e_{Gb,t-1} + \ldots + \Psi_{X.biq_{Xb}}e_{Xbit,t-q_{Xb}} + \epsilon_{Xbit}
\end{aligned}
$$

Again, all error terms follow the usual distributional assumptions. The estimated factors for the regional level are very close, indicating that regional heterogeneity results mainly from idiosyncratic shocks to the individual series. The next section quantifies the relative importance of each level in the hierarchy for explaining the individual series' behavior.

3.3 Variance Decomposition

Each series in the four-level model can have its sample variance decomposed into four different sources of variances:

$share_F$ Aggregate level.

$share_G$ Block-level (regional).

$share_G$ Sub-block level (economic groups).

$share_Z$ Individual items.

Table 2 shows the estimated variance decompostions. Technical details of the statistical procedures behind this decomposition can be found in Moench E., Ng S.and Potter S. (2009).

The first point to note is that the main source of variation for each individual series comes from their own idiosyncratic shocks: more than 80% in Foodstuff and Consumer Goods, and around only 71% for Regulated prices. Regulated prices, on the other hand, display the biggest participation for the group level, at almost 30%, which intuitively results from the institutional aspect of Regulated prices by indexation in Brazil, which typically evolve as specified by contracts using rules conditioned to past behavior of general price indices. The regional dimension does not play a significant role in explaining the variances of the individuals series, accounting for at most

Item	$share_F$	$share_G$	$share_H$	$share_Z$
Foodstuff	0.130	0.027	0.027	0.815
Services	0.012	0.003	0.194	0.792
Regulated	0.001	0.000	0.291	0.708
Consumer Non-durable	0.057	0.012	0.043	0.888
Consumer Durable	0.052	0.011	0.132	0.805

Table 2: Variance Decomposition

around 3% for the Foodstuff group, which can probably be explained by local supply conditions of perishable items. Consumer goods have the greater share of explanation through the aggregate level variation, a little more than 5%, reflecting national macroeconomic fundamentals such as disposable income and credit conditions. The importance of the factor at the aggregate level is inversely proportional to its share in explaining the variance of the individual series: the headline index would be very close to alternative measures of core inflation if most of the variation for each series resulted from a single common factor.

4 Conclusions

There appears to be a lot of useful information contained in the estimated factors of a multi-level model for inflation and relative price changes. The two results presented here can be extended on a number of dimensions:

Alternative groupings There are other types of aggregation of the index sub-items. For instance two groups which would overlap with the ones considered here are tradables vs. non-tradables, which factors could be correlated with macroeconomic fundamentals such as the exchange rate and salaries.

Dynamics The model is completely general in its possible specifications for lags between observed series and factor, between factors of different levels, and between the random variables behind the error terms. In parallel with results from time-series models that do not include latent variables, the results can be very sensitive to alternative specifications in the dynamics.

Number of Factors Like in one-level factor models, it is possible to estimate more than one factor for each level in the model adopted here.

As already mentioned, the reason for adopting the hierarchical structure is that it provides a direct interpretation for each distinct factor. However, the information contained in the data may not be enough to exhaust the possibilities of meaningfully estimated latent variables. In this sense, more than one factor can be potentially estimated for each of the model's levels.

Data from different sources In principle, there is no restriction for including data on price changes calculated by different institutions in differents regions. This opens the possibility of defining new types of groupings and levels.

The reason why these variations have not yet been tried is mainly computational. The MCMC algorithm used to provide samples values from which factors and parameters of the model can be estimated is very demanding in CPU time. Parallelization of computation and efficiency of the algorithms is something to be seriously considered.

Also, whichever specificication of the model along the above dimensions performs best by some criterium, the ultimate performance test of the core inflation measure proposed here would be its dynamic correlations with macroeconomic fundamentals such as the interest and exchange rates, unemployment level, etc., when compared to alternative measures. The IPCA has a number of core inflation measures routinely calculated, such as variations of trimmed means measures. If any significant differences appear between these alternatives, one should be able to evaluate the relative merits of each in terms of interpretability, computation costs and forecasting efficiency.

5 References

Altissimo F, Cristadoro R. Forni M., Lippi M., Veronese G. (2010): "New Eurocoin: Tracking economic growth in real time." The Review of Economics and Statistics 92(4), 1024-1034.

Bakhshi, H. and Yates, T. (1999): "To Trim or not to trim ? An application of a trimmed mean in‡ation estimator to the United Kingdom", Bank of England working paper series, ISSN number 1368-5562.

Ball, L. and Mankiw, G. (1992): "Relative-price changes as aggregate supply shocks", Quarterly Journal of Economics, 110, pages 161-93.

Blanchard, O., Quah, D., (1989): "The dynamic effects of aggregate supply and demand disturbances." American Economic Review 79 (September), 655–673.

Bryan, M. and Cecchetti, S. G. (1994): "Measuring Core In‡ation" , in Mankiw, N. G. (ed.) Monetary Policy, Chicago: The University of Chicago Press, 195-215.

Cristadoro R., Forni, M, Reichlin L. Veronese G. (2005): "A Core Inflation indicator of the Euro Area." Journal of Money Credit and Banking 37(3), 539-560.

Giannone D. Matheson T.D. (2007): "A new core inflation indicator for New Zealand." International Journal of Central Banking 3(4), 145-180.

Moench E. , Ng S. and Potter S. (2009): "Dynamic Hierarchical Factor Models", Federal Reserve Bank of New York - Staff Report no. 412.

Picchetti P., Toledo C. (2000): "How much to trim ? A methodology for calculating Core In‡ation, with an application for Brazil", Economia Aplicada, Vol 4, No. 4

Picchetti, P., Kanczuk, F. (2001): "An Application of Quah and Vahey's SVAR Methodology for Estimating Core In‡ation in Brazil", proceedings of XXIV Encontro da Associação Nacional de Pós-Graduação em Economia.

Quah, D., Vahey, S.P., 1995. "Measuring core inflation.", Economic Journal 105 (September), 1130–1144.

Reis R., Watson M. (2010): "Relative Goods' Prices, Pure Inflation, and the Phillips Correlation." American Economic Journal: Macroeconomics, 2(3): 128-57.

Selvanathan, E.A. and Prasada Rao, D. S. (1994): "Index Numbers: A Stochastic Approach", Michigan: University of Michigan Press.

Stock J., Watson M. (1991): "A ProbabilityModel of the Coincident Economic Indicators", in Kajal Lahiri and Geoffrey H. Moore, eds., Leading Economic Indicators: New Approaches and Forecasting Records.

Cambridge: Cambridge University Press, pp. 63-89.

Stock J., Watson M. (2002). "Forecasting Using Principal Components from a Large Number of Predictors." Journal of the American Statistical Association, vol. 97, 1167-1179.

Wynne, M.A., 1999. "Core inflation: A review of some conceptual issues", European Central Bank working paper no. 5 (May).