Index

1. Introduction of the work

The scope of this study is to investigate the relationship between our dependent variable, the growth rate of the deflator of dwellings, (y), and our explanatory variables, the domestic inflation (x_1) and the growth rate of the housing price index (x_2). By using STATA and gretl, we denoted the variables as follows:

Variable	STATA	gretl
$\Delta\, log\ (diflab)$	Y	d_l_diflab
$\Delta\, log\ [dvcft * (1 + tiin)]$	X1	d_l_x1
$\Delta\, log\ (nhpi)$	X2	d_l_nhpi

All these variables refer to Italy, and are represented through quarterly data, starting from 1970 to 2023[1]. With regards to the long-term relationship, we expect the results to be in line with the economic literature, which states that if the growth rate of price index goes up, as well as domestic inflation, then we will encounter a positive effect on the growth rate of the deflator in the long-run. Thus, we expect both parameters to have positive signs.

We can justify the direction of causality by referring to economic theory. First of all, we consider the effect of inflationary pressures: an increase in domestic inflation can influence housing prices in several ways. For instance, a general rise in prices of goods and services can increase construction and maintenance costs for housing. Moreover, the increase in inflation can affect buyers' expectations regarding housing, leading to higher demand in the real estate market. These effects can contribute to the increase in the growth rate of the housing deflator. Furthermore, it is worth mentioning a possible minor short-run effect of monetary policy. Indeed, monetary policies aimed at containing domestic inflation can impact the real estate market. For instance, an increase in interest rates to combat inflation can immediately make mortgage loans more expensive, lowering the demand for housing and, consequently, housing prices. We know from economic literature and common sense that market prices are forcing on the newly built houses prices. Indeed, the prices of newly built houses are based on those of the dwellings already on the market (this reasoning also holds if we consider the growth rate of these prices). It is also reasonable to assume that a change in the prices of new houses does not influence much the prices of dwellings already on the market, since we know from economic theory that some prices are sticky, so it makes sense to consider x_2 as weakly exogenous. If market prices are rising, prices of newly built houses might rise, and conversely, if market prices are falling, prices of newly built houses might fall.

To begin with, we proceed with a single equation model in which we will consider as forcing our two explanatory variables. Despite the fact that this model let us achieve an approximate understanding of the real long-run dynamics, it does not account for feedbacks from the dependent variable to the independent ones. For this reason, in the last part of our work we will introduce the Vector Autoregressive Model as well as the Vector Error Correction Model (VECM), in order to account for this feature.

2. Main Findings

Our study focused on the following topics and related findings:

- *Stationarity*: through univariate analysis, we computed the time series plots, correlograms and Augmented Dickey-Fuller (ADF) Unit Roots test. The first findings showed that the log-levels of the variables were I(1), whereas the simultaneous differences were I(0). The ADF confirmed all these outcomes with the exception of the variable x_2, which turned out to be stationary even in log-levels.
- *ARDL and PSS testing*: due to our data periodicity (quarters), we decided to impose an ARDL (4,4,4). After having performed the main misspecification tests (Godfrey, inspection of correlogram, White,

[1] x_1 actually starts from 1960, but the first ten years are ignored since y has been collected since 1970.

ARCH and normality tests), we observed white noise residuals, although heteroskedastic and not normally distributed. This was sufficient to prove the goodness of the model and allowed us to proceed with the PSS testing which confirmed the presence of the long-run relationship. In this way, we detected a long-run positive elasticity for both explanatory variables, and the speed of adjustment parameter equal to -0.498.

- *Engle-Granger cointegration*: after running the OLS estimation of the static long-run relationship, we computed the Unit Roots test of the cointegration regression residuals. Hence, we found cointegration between the variables and, as a consequence, superconsistent estimators. After having computed the Engle-Granger cointegration test to identify the long-term relationships between non-stationary variables, we computed the ECM model to understand the short-run dynamics. The results were coherent with ARDL model: even though the SOA[2] was slightly lower, we still observed a positive change in y induced by a change in the explanatory variables.
- *Johansen cointegration test*: this method is the most informative since it enables us to check for more than one cointegrating relationships. As a matter of fact, through this test we found that the rank of the matrix is equal to 2, meaning that there are two cointegrating relationships. Consequently, by computing the VECM, we investigated deeper these relationships: the first one between y and x_1 and the second one between x_1 and x_2. This technique proved that x_2 is actually I(1) and x_1 is not weakly exogenous. Furthermore, the loading parameters matrix showed coefficients coherent with common sense and economic literature.
- *HP filter*: through the decomposition of our variables into permanent and transitory components, and by performing the Johansen cointegration test, we obtained further evidence that our three variables are I(1), since their trend components are non-stationary (not a full rank matrix).

3. Univariate Analysis (1/3)

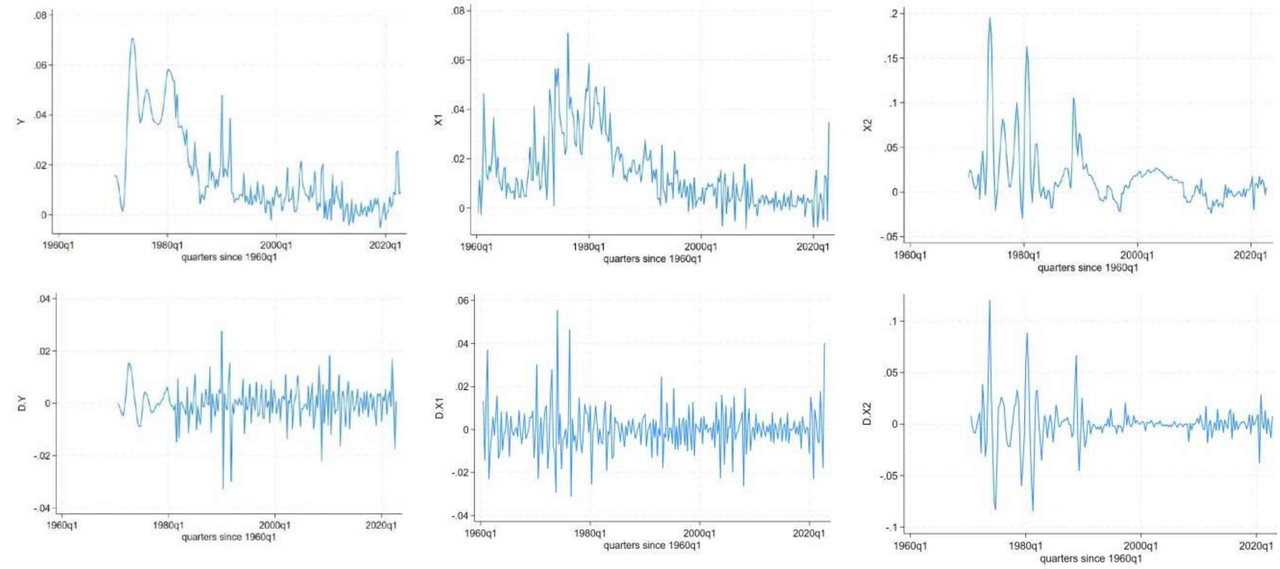


Figure 1: Time Series plots in log-levels (first line) and their first differences (second line)

Above this paragraph there are all the plots illustrating our variables in log-level (first line) and their first difference (second line). From these graphs, we can notice that the log-levels of y, x_1, x_2 show persistence and hence we think they will deliver a correlogram with high bars outside the confidence intervals, with correlation coefficients of notable magnitude. Hence, we also expect that the ADF test will fail to reject the null of the unit roots process. On the other hand, their simultaneous differences show stationarity and, as a consequence, their correlogram will have bars of decreasing length which become extremely short in order to indicate a decreasing

[2] Speed of adjustment.

memory. If ADF test was able to confirm this feature, then we would know that our available finite sample is enough informative to estimate the model's parameters.

From an economic perspective, these graphs provide highlights about remarkable economic phenomena that occurred in Italy during the analyzed period. More specifically, we observe several peaks in the plots of y, x_1 and x_2 which coincide with a combination of economic and political factors that affected the domestic inflation and the price of dwellings. For example, the rapid growth in the series during the '70s matches the inflationary shocks due to the oil crisis[3] and the increase in the price of building-land (Cannari, D'Alessio, & Vecchi, 2016).

4. Univariate Analysis (2/3)

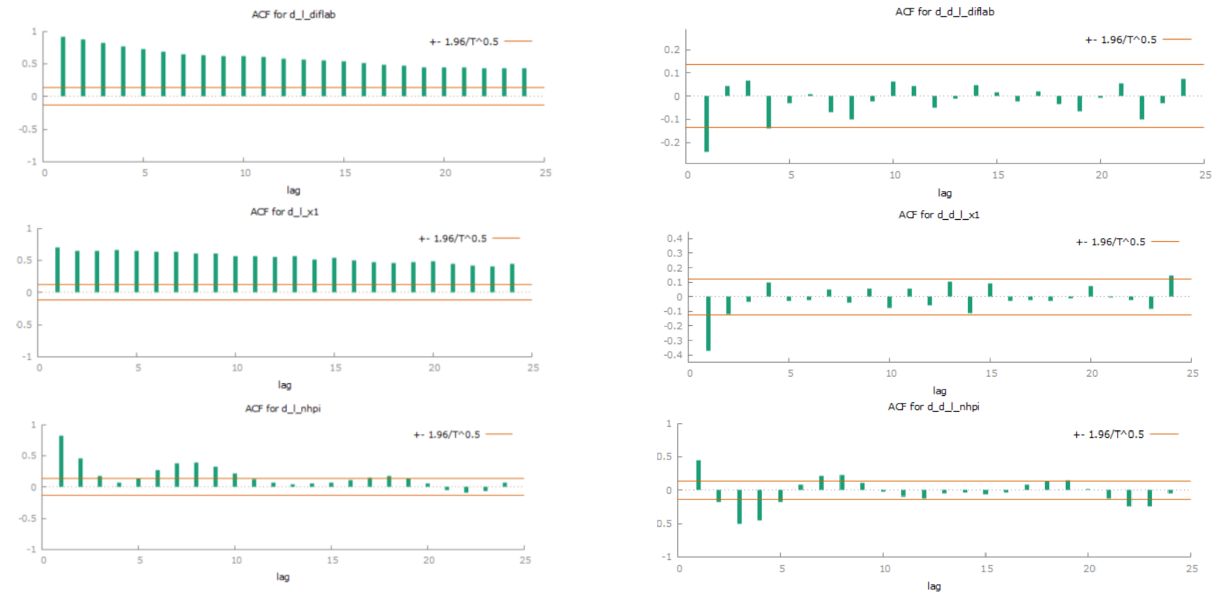


Figure 2: Correlograms of log-level variables (left) and their first differences (right)

To begin with, the correlograms of the variables and their first differences show what we expected and described before. Indeed, the variables in log-levels show high persistence, whereas their simultaneous differences show stationarity through the decreasing-in-length bars inside the interval. With regards to y, x_1 and x_2, it is worth noting that the correlation coefficients are very close to one, especially for the first few lags. Furthermore, for x_2 we observe a periodical oscillation, which could indicate seasonality in the growth rate of price index. This fact is particularly interesting since we were not able to spot it before.

Regarding Δy, Δx_1 and Δx_2, the autocorrelation coefficient decreases very rapidly as the order of lags increases, describing, as we already mentioned, a decreasing memory. This confirms our previous statement. It means that the covariance between our variable in period t and in period t-k will increase as the number of lags k increases as well.

5. Univariate Analysis (3/3)

To test for the stationarity of our variables, we perform the Augmented Dickey-Fuller (ADF) test, where the null hypothesis is that our variable follows a unit root process and so is I(1), meanwhile the alternative hypothesis is that our variable is generated by a stationary process. In order to run the tests we used the

[3] Which contributed to higher prices across various asset classes including real estate

commands "dfuller" and "adfmaxur", we obtained the results that are summarized below, where the augmentation is made in order to obtain white noise residuals and taking into account the data periodicity.

Variable	Test statistic	DF c.v. (5%)	Augmentation Order[4]	Conclusion	Integration order
y_t	-2.324	-2.883	1	We fail to reject H_0	I(1) Not stationary
Δy_t	-7.671	-2.883	0	We reject H_0	I(0) stationary
x_{1_t}	-2.076	-2.880	4	We fail to reject H_0	I(1) not stationary
Δx_{1_t}	-11.324	-2.880	4	We reject H_0	I(0) stationary
x_{2_t}	-3.675	-2.883	4	We reject H_0	I(0) stationary
Δx_{2_t}	-10.203	-2.883	4	We reject H_0	I(0) stationary

The table above confirms that a fifth order dynamics – AR(5) - is appropriate to explain the path of the explanatory variables under scrutiny and their related differences; whereas, to explain y_t, it is suitable to use an AR(2) model[5]. In conclusion, ADF tests proved all our previous expectations. Indeed, it is a well-known fact that inflation is a heavily persistent phenomenon and it takes time to return to its long-run value after a shock. The only outcome different from our expectations concerns the variable x_2, which, according to the Dickey-fuller test, results to be stationary even in log-levels. We will further investigate this topic with the help of other techniques such as the Johansen cointegration test.

6. ARDL Model

In order to obtain the final ARDL model, we focus on our first aim to have white noise residuals. As before, due to our data periodicity, we decide to use an ARDL(4,4,4) model which, through reparameterization, let us obtain the multivariate single equation model in the ECM form

$$\Delta y = c + \sum_{j=1}^{3} \gamma_j \Delta y_{t-j} + \sum_{j=0}^{3} \beta_j \Delta x_{1_{t-j}} + \sum_{j=0}^{3} \delta_j \Delta x_{2_{t-j}} + \pi_1 y_{t-1} + \pi_2 x_{1_{t-1}} + \pi_3 x_{2_{t-1}} + \epsilon_t$$

```
      Source |       SS           df       MS      Number of obs   =       207
-------------+----------------------------------   F(14, 192)      =      4.11
       Model |  .002504811        14  .000178915   Prob > F        =    0.0000
    Residual |  .008355456       192  .000043518   R-squared       =    0.2306
-------------+----------------------------------   Adj R-squared   =    0.1745
       Total |  .010860267       206   .00005272   Root MSE        =     .0066

------------------------------------------------------------------------------
         D.Y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
```

[4] Computed by using the testing down procedure.
[5] And for Δy_t we obtained an AR(1).

Y						
LD.	.1036832	.099064	1.05	0.297	-.0917104	.2990767
L2D.	.233207	.092556	2.52	0.013	.0506499	.4157641
L3D.	.2152167	.0803799	2.68	0.008	.0566757	.3737577
X1						
D1.	.2567684	.0602383	4.26	0.000	.1379546	.3755821
LD.	-.1744263	.0916766	-1.90	0.059	-.3552489	.0063963
L2D.	-.0813743	.0768371	-1.06	0.291	-.2329276	.0701789
L3D.	-.029262	.0585255	-0.50	0.618	-.1446974	.0861734
X2						
D1.	.0977067	.0338113	2.89	0.004	.0310174	.1643959
LD.	-.0461837	.0382463	-1.21	0.229	-.1216206	.0292532
L2D.	-.0088573	.0349996	-0.25	0.800	-.0778904	.0601758
L3D.	-.02071	.0332589	-0.62	0.534	-.0863097	.0448897
Y						
L1.	-.4982412	.0965079	-5.16	0.000	-.688593	-.3078895
X1						
L1.	.4690362	.108496	4.32	0.000	.255039	.6830333
X2						
L1.	.0672033	.0304146	2.21	0.028	.0072137	.1271929
_cons	.0003009	.0006856	0.44	0.661	-.0010514	.0016531

To verify the accuracy of our ARDL model, we performed the main misspecification tests. Firstly, we started with the Breusch-Godfrey LM test (see table below):

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.007	1	0.9354
4	3.388	4	0.4952
8	6.957	8	0.5413

H0: no serial correlation

We considered lags of 1st, 4th and 8th order due to the fact that they are multiples of our data periodicity. As we can notice from the table, we can state that we have white noise residuals[6]. We then proceed to perform the other misspecification tests: inspection of residuals correlogram, White's Heteroskedasticity test, the Engle ARCH (q) test and Jarque-Bera normality test.

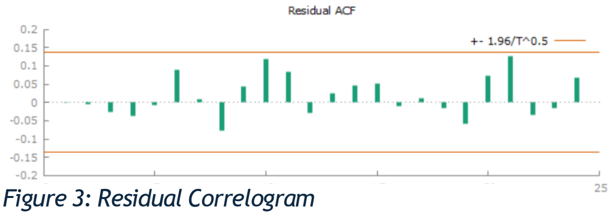


Figure 3: Residual Correlogram

- Inspection of the correlogram of the residuals allows us to see that we have white noise residuals, thanks to the fact that all the bars are inside the bar interval.

- White's Heteroskedasticity test: with this test we reject the null of no-heteroskedasticity.

```
White's general test statistic : 175.6129    Chi-sq(119)    P-value = 5.7e-04
```

- LM test for autoregressive conditional heteroskedasticity (ARCH): we reject the null hypothesis of homoskedasticity.

lags(p)	chi2	df	Prob > chi2
1	33.621	1	0.0000

H0: no ARCH effects vs. H1: ARCH(p) disturbance

- Skewness and kurtosis tests for normality: we reject the null hypothesis of normality of residuals.

				----- Joint test -----	
Variable	Obs	Pr(skewness)	Pr(kurtosis)	Adj chi2(2)	Prob>chi2
residuals	207	0.8673	0.0010	9.73	0.0077

To summarise, the table below illustrates the conclusion of our tests with regards to the residuals of our ARDL(4,4,4) model. Even though our residuals are just white noise and do not show homoskedasticity nor normality, it is sufficient to prove the goodness of the model and allows us to proceed with our study.

Under the null	ARDL(4,4,4) residuals
No serial autocorrelation	Not rejected
No heteroskedasticity	Rejected
Normality	Rejected

Finally, to study the long-run relationship between our variables, we perform the PSS test (Pesaran, Shin, & Smith, 2001), which is illustrated below. We follow the third case (the most common) of the restrictions on the intercept and the trend term, since our variables are rates. This test is particularly important, since, when looking at the estimates of the variable for the ARDL model, we should not be fooled by the reported p-values and confidence intervals into believing that they are statistically significant. Hence, we rely on the critical values reported by STATA.

[6] Since we do not reject the null hypothesis of no serial correlation.

```
PESARAN, SHIN AND SMITH (2001) COINTEGRATION TEST

Obs: 207

No. Regressors (k): 2

Case: 3

-------------------------------------------------------------

                                  F-test

-------------------------------------------------------------

                   <----- I(0) ---------- I(1) ----->

10% critical value       3.170           4.140

5% critical value        3.790           4.850

1% critical value        5.150           6.360


F-stat. =     9.378

-------------------------------------------------------------

                                  t-test

-------------------------------------------------------------

                   <----- I(0) ---------- I(1) ----->

10% critical value      -2.570          -3.210

5% critical value       -2.860          -3.530

1% critical value       -3.430          -4.100


t-stat. =    -5.163

-------------------------------------------------------------

F-statistic note: Asymptotic critical values used.


t-statistic note: Asymptotic critical values used.
```

From these tables, looking at t-statistic and F-statistic, we reject both the two null hypotheses of:

- $H_0: \pi_1 = 0$
- $H_0: \pi_1 = \pi_2 = \pi_3 = 0$

Thanks to these tests, we can affirm that we have a long-run relationship and calculate the long run parameters.

```
   b_x1: _b[l.X1]/-_b[l.Y]

        b_x2: _b[l.X2]/-_b[l.Y]


------------------------------------------------------------------------------

         D.Y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]

-------------+----------------------------------------------------------------

        b_x1 |   .9413837   .1021319     9.22   0.000     .7412088    1.141559

        b_x2 |   .1348811   .0532694     2.53   0.011     .0304751    .2392871

------------------------------------------------------------------------------
```

The estimates of the long-run parameters are:

$$\beta_{x1} = \frac{\pi_2}{-\pi_1} = \frac{0.469}{0.498} = 0.941$$

$$\beta_{x2} = \frac{\pi_3}{-\pi_1} = \frac{0.067}{0.498} = 0.135$$

Through this computation, we observe almost unit elasticity between the growth rate of the deflator and the domestic inflation. This means that an increase of 1% in domestic inflation will lead to a 0.941% increase in the growth rate of the deflator.

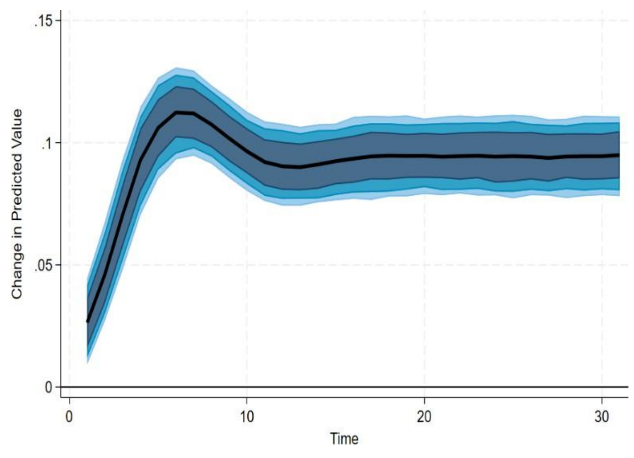
On the other hand, the growth rate of the deflator follows the growth rate of price index with a steady-state long-run relationship with a parameter equal to 0.135. We can interpret this estimate as the elasticity of the growth rate of the deflator to the growth rate of price index: a 1% increase in the growth rate of price index implies a 0.135% increase in the growth rate of the deflator.

Moreover, the estimate of the loading parameter (speed of adjustment) is $\hat{\alpha} = -0.498$, indicating a stable process of adjustments[7] : more specifically, each quarter, the growth rate of the deflator adjusts by about 50% towards the target (equilibrium) level given by:

$$y^* = 0.941x_1 + 0.135x_2$$

As a further investigation of the long-run parameters, we interpret the ARDL dynamics through shock simulations. Indeed, since the ARDL model has a fairly complex lag structure, with lags, contemporaneous values, first differences and lagged first differences of the independent variable appearing in the model specification, interpreting short and long-run effects may be difficult. Thus, to better interpret the substantive significance of our results, we dynamically simulated our model through the use of the command "`dynardl`", which enables us to see the impact of a counterfactual change in one of the independent variables at a single point, holding everything else equal, using stochastic simulation techniques (Jordan & Philips, 2018).

This method first runs a regression using ordinary least squares and then, using a self-contained procedure, it takes 1000 draws of the vector parameter from a multivariate normal distribution. With regards to the possible options, we ask STATA to show an area plot, with the predicted means and 75%, 90%, and 95% confidence intervals. The graph will illustrate predicted changes (from the sample mean) across time, starting with the time at which the shock occurs, similar to an impulse-response reaction (Jordan & Philips, 2018).

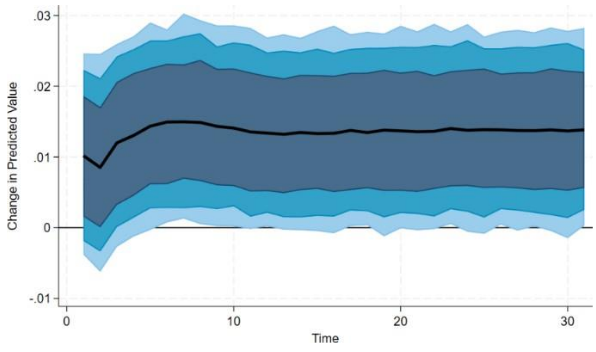


Graph 1: shock of 10% to the variable x_1

In the first graph, we introduced a shock in x_1 with a 10% increase (Graph 1): in the first period, we have a change in y by about 0.03. As time passes, it increases rapidly before reaching a steady state level of 0.1. This is in line with our parameter β_{x1}= 0.941.

By following the same procedure for x_2, we can see through Graph 2 that a 10% increase in period one of x_2, would cause around 1% change in the predicted value. Again, this is in line with the previously estimated parameter β_{x2}.

[7] $-1 < \hat{\alpha} < 0$



Graph 2: shock of 10% to the variable x_2

7. Engle-Granger and ECM

We now proceed to investigate if our variables have short-run and long-run relationship with each others, in the sense that while short-run perturbation may move the series apart, as time passes the disequilibrium is corrected because the series moves back towards a stable long-run relationship. This will show that the variables are cointegrated. To verify a potential cointegration, we computed the Engle-Granger test through the command "`egranger`" in STATA.

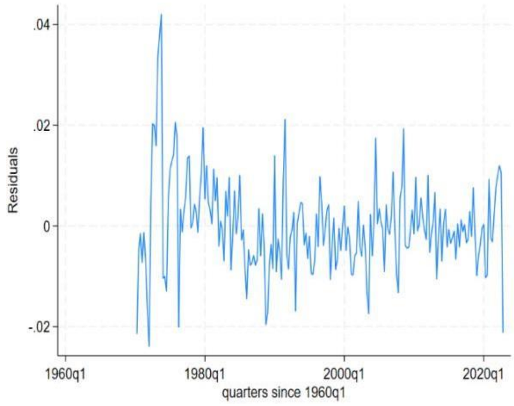
```
Augmented Engle-Granger test for cointegration          N (1st step)  =       211
Number of lags    =   4                                  N (test)      =       206
------------------------------------------------------------------------------
                   Test           1% Critical       5% Critical     10% Critical
                Statistic            Value             Value            Value
------------------------------------------------------------------------------
 Z(t)              -5.868           -4.363            -3.782           -3.482

Critical values from MacKinnon (1990, 2010)
------------------------------------------------------------------------------
Engle-Granger 1st-step regression
------------------------------------------------------------------------------
           Y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          X1 |   .7803344   .0480948    16.22   0.000     .6855186    .8751502
          X2 |   .1547156   .0208394     7.42   0.000     .1136321     .195799
       _cons |   .0024754   .0009079     2.73   0.007     .0006854    .0042653
------------------------------------------------------------------------------
Engle-Granger test regression
------------------------------------------------------------------------------
  D._egresid | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
```

```
------------+----------------------------------------------------------------
    _egresid |
         L1. |  -.6185862   .1054136    -5.87   0.000    -.8264446   -.4107278
         LD. |   .0380807   .1017886     0.37   0.709    -.1626299    .2387912
        L2D. |   .0444453   .0929462     0.48   0.633    -.1388294      .22772
        L3D. |    .014656   .0832862     0.18   0.860    -.1495708    .1788827
        L4D. |   .0290066   .0714486     0.41   0.685    -.1118784    .1698915
```

Following the two-steps procedure, we first run a static regression for the long-run model and then we test for the stationarity of the residuals. We can see that the test statistic is -5.868 meaning that we can reject the null hypothesis that residuals are I(1).

By looking at the plot of the residuals for the static regression, we have further evidence that they are I(0). Since we have obtained stationary residuals, we can affirm that our variables are cointegrated. In light of this, the OLS method yields a "superconsistent" estimator of the cointegrating parameters since the effect of the common trend dominates the effect of the stationary component (Stock, 1987).



Graph 3: Stationary residuals of the static model

Now, to estimate the ECM model we rely on the second step of Engle-Granger procedure, where the first difference of y is regressed on the lagged level of first-step residuals and the lagged first difference of x_1 and x_2 using OLS. Conversely to the regression of the static model (first step), in order to inspect the long-run relationship, we can now have a safe look at the p-values and confidence intervals since these variables are stationary by definition (they are all in differences). This was not possible for the static regression since the t-stat of those coefficients does not follow the standard distribution.

The coefficient on the lagged residuals is an estimate of the ECM "speed of adjustment" parameter towards the long-run equilibrium. After having estimated a first model, given that residuals are white noise, we test for dropping some lags: the null hypothesis of the F-test for all lags is not rejected with a p-value of 0.3525. Furthermore, residuals checks do not show relevant problems [8]. As a result, we obtained the final restricted ECM model:

```
Engle-Granger 2-step ECM estimation                 N (1st step)  =      211
                                                    N (2nd step)  =      210
```

[8] Checked through the main misspecification tests.

Engle-Granger 2-step ECM

D.Y	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
X1						
D1.	.2307485	.0469928	4.91	0.000	.1380999	.323397
X2						
D1.	.0934981	.023694	3.95	0.000	.0467843	.1402118
_egresid						
L1.	-.3144955	.0572954	-5.49	0.000	-.427456	-.201535
_cons	.0000129	.0004574	0.03	0.978	-.000889	.0009148

From the table, we elaborate the retained uniequational model:

$$\Delta y_t = 0.000129 + \ 0.230\Delta x_{1_t} + 0.093\Delta x_{2_t} - 0.314\ ECM_{t-1} + \hat{\epsilon}_t$$

The table above illustrates the changes of y induced by the fluctuations of the explanatory variables x_1 and x_2 (which are the short-run effects or impulses) respectively highlighted in blue and green. As we previously saw in Graph 1, the response of y following a change in x_1 is positive: as a matter of fact, in the first quarters the change in the predicted value is positive and increasing. This is in line with the economic theory, which suggests that the domestic inflation has a positive impact on the growth rate of the deflator.

The same reasoning can be done for x_2, despite the fact that the simulation of a shock in Graph 2 showed us a negative short-run effect. By looking at the error correction term, we can notice that the discrepancies from the long-run equilibrium are corrected gradually through a series of partial short-run adjustments. More specifically, we see that around 31% of the discrepancy is adjusted within a quarter.

Finally, through a comparison between ARDL and ECM model, we can see that the speed of adjustment is slightly different. This is due to the fact that while ECM explicitly models the adjustment speed through the error correction term, in ARDL models, SOA is reflected in the lag order and the dynamic effect of variables over time.

8. Johansen Cointegration Test

So far, we considered the Pesaran and the 2-step Engle-Granger model by single equation dynamic modelling through the concepts of level relationships and cointegration. However, these methods may not be accurate if the explanatory variables are not weakly exogenous and there are potentially more than one cointegrating relationships. To test their validity, we perform the Johansen cointegration test, which is related to the framework of the Vector Autoregressive (VAR) model, useful to describe the variation of the data, and then formulating questions concerning structural economic relations. These hypotheses are tested using likelihood ratio statistics, and allow us to check interesting economic hypotheses against the data (Johansen, 1991).

To begin with, we compute the VAR lag selection through gretl: by referring to the Akaike criterion (AIC), the optimal number of lags turns out to be 5. We then proceed with the Johansen cointegration test by selecting five lags and, as far as the deterministic component of the VAR is concerned, we decide to select "unrestricted constant". The reason behind this choice lies on the fact that our variables are rates and therefore we choose not to include a trend in our analysis.

```
Johansen test:
Number of equations = 3
Lag order = 5
Estimation period: 1971:3 - 2022:4 (T = 206)
Case 3: Unrestricted constant
Log-likelihood = 2644.6 (including constant term: 2059.99)


Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0    0.17193     72.154 [0.0000]     38.863 [0.0000]
   1    0.13568     33.291 [0.0000]     30.038 [0.0000]
   2   0.015667     3.2529 [0.0713]     3.2529 [0.0713]


Corrected for sample size (df = 190)
Rank Trace test p-value
   0      72.154 [0.0000]
   1      33.291 [0.0000]
   2      3.2529 [0.0730]
```

From the table above, we reject the null hypothesis that the rank is equal to 0 (case of no cointegration) and rank equal to 1, but we fail to reject the hypothesis that the rank is equal to 2. Thus, we have evidence that there are two long-run relationships and one common stochastic trend (common source of non-stationarity). Furthermore, we can state that, given the rank is greater than one, PSS and Engle-Granger cointegration tests are not valid.

To identify the long-run relationships, we can estimate the VECM representation with cointegrating rank equal to two (with unrestricted constant).

$$\Delta X_t = c_0 + \Pi \mathrm{X}_{\mathrm{t-1}} + \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{\mathrm{p-1}} \Delta X_{t-p+1} + u_t$$

As we have already mentioned, the VECM model is the multivariate extension of the univariate ADF test. Instead of testing $H_0\colon \alpha = 0$ (no unit roots), it tests the rank of the matrix Π.

```
VECM system, lag order 5
Maximum likelihood estimates, observations 1971:3-2022:4 (T = 206)
Cointegration rank = 2
Case 3: Unrestricted constant


beta (cointegrating vectors, standard errors in parentheses)
```

```
d_l_diflab       1.0000       0.00000
               (0.00000)    (0.00000)
d_l_X1          0.00000       1.0000
               (0.00000)    (0.00000)
d_l_nhpi       -0.91191      -0.75653
              (0.099150)   (0.095674)
```

alpha (adjustment vectors)

```
d_l_diflab      -0.37294      0.39169
d_l_X1           0.38998     -0.50855
d_l_nhpi         0.36113     0.024091
```

Log-likelihood = 2058.367

Determinant of covariance matrix = 4.2031291e-013

AIC = -19.5181

BIC = -18.7427

HQC = -19.2045

Equation 1: d_d_l_diflab

	coefficient	std. error	t-ratio	p-value	
const	−0.000252481	0.000493149	−0.5120	0.6093	
d_d_l_diflab_1	0.0505131	0.109583	0.4610	0.6454	
d_d_l_diflab_2	0.263701	0.106308	2.481	0.0140	**
d_d_l_diflab_3	0.283802	0.0996742	2.847	0.0049	***
d_d_l_diflab_4	0.0335613	0.0869664	0.3859	0.7000	
d_d_l_X1_1	−0.345953	0.108281	−3.195	0.0016	***
d_d_l_X1_2	−0.227613	0.0956170	−2.380	0.0183	**
d_d_l_X1_3	−0.153827	0.0803320	−1.915	0.0570	*
d_d_l_X1_4	−0.0779343	0.0622087	−1.253	0.2118	
d_d_l_nhpi_1	0.00185656	0.0372559	0.04983	0.9603	
d_d_l_nhpi_2	−0.0246019	0.0406840	−0.6047	0.5461	
d_d_l_nhpi_3	−0.0277517	0.0370890	−0.7482	0.4552	
d_d_l_nhpi_4	0.00461299	0.0351589	0.1312	0.8958	
EC1	−0.372944	0.108909	−3.424	0.0008	***
EC2	0.391690	0.115418	3.394	0.0008	***

Mean dependent var	7.16e-06	S.D. dependent var	0.007273
Sum squared resid	0.009309	S.E. of regression	0.006981
R-squared	0.141471	Adjusted R-squared	0.078542
rho	-0.007575	Durbin-Watson	2.006206

Equation 2: d_d_l_X1

	coefficient	std. error	t-ratio	p-value	
const	0.000334082	0.000562791	0.5936	0.5535	
d_d_l_diflab_1	-0.216060	0.125058	-1.728	0.0857	*
d_d_l_diflab_2	0.0310913	0.121321	0.2563	0.7980	
d_d_l_diflab_3	0.109770	0.113750	0.9650	0.3358	
d_d_l_diflab_4	0.0274426	0.0992477	0.2765	0.7825	
d_d_l_X1_1	-0.429314	0.123572	-3.474	0.0006	***
d_d_l_X1_2	-0.423425	0.109120	-3.880	0.0001	***
d_d_l_X1_3	-0.425233	0.0916764	-4.638	6.50e-06	***
d_d_l_X1_4	-0.160950	0.0709937	-2.267	0.0245	**
d_d_l_nhpi_1	-0.0362620	0.0425171	-0.8529	0.3948	
d_d_l_nhpi_2	0.000668698	0.0464293	0.01440	0.9885	
d_d_l_nhpi_3	-0.0120674	0.0423266	-0.2851	0.7759	
d_d_l_nhpi_4	0.0674067	0.0401240	1.680	0.0946	*
EC1	0.389978	0.124289	3.138	0.0020	***
EC2	-0.508546	0.131718	-3.861	0.0002	***

Mean dependent var	0.000133	S.D. dependent var	0.010710
Sum squared resid	0.012123	S.E. of regression	0.007967
R-squared	0.484468	Adjusted R-squared	0.446680
rho	-0.041894	Durbin-Watson	2.012106

Equation 3: d_d_l_nhpi

	coefficient	std. error	t-ratio	p-value	
const	-0.000274137	0.000992822	-0.2761	0.7828	
d_d_l_diflab_1	-0.155088	0.220615	-0.7030	0.4829	
d_d_l_diflab_2	0.0185105	0.214023	0.08649	0.9312	
d_d_l_diflab_3	0.263179	0.200667	1.312	0.1913	
d_d_l_diflab_4	0.295541	0.175083	1.688	0.0930	*
d_d_l_X1_1	-0.0141313	0.217995	-0.06482	0.9484	
d_d_l_X1_2	0.215935	0.192499	1.122	0.2634	

d_d_l_X1_3	0.453420	0.161727	2.804	0.0056	***
d_d_l_X1_4	0.255469	0.125240	2.040	0.0427	**
d_d_l_nhpi_1	0.585564	0.0750046	7.807	3.76e-013	***
d_d_l_nhpi_2	−0.0839748	0.0819062	−1.025	0.3065	
d_d_l_nhpi_3	−0.0488445	0.0746686	−0.6541	0.5138	
d_d_l_nhpi_4	−0.121736	0.0707830	−1.720	0.0871	*
EC1	0.361125	0.219258	1.647	0.1012	
EC2	0.0240912	0.232364	0.1037	0.9175	

Mean dependent var	−1.00e-06	S.D. dependent var	0.021296
Sum squared resid	0.037729	S.E. of regression	0.014055
R-squared	0.594205	Adjusted R-squared	0.564461
rho	−0.022856	Durbin-Watson	2.043264

Thanks to this model, we can detect the case of "two-in-two" cointegration.

$$\Pi = \alpha\beta' = \begin{pmatrix} -0.373 & 0.392 \\ 0.390 & -0.509 \\ 0.361 & 0.024 \end{pmatrix} \begin{pmatrix} 1 & 0 & -0.912 \\ 0 & 1 & -0.757 \end{pmatrix}$$

Since the model has a good performance that is in line with the economic theory, we can say that our variable x_2 is actually I(1). This result is in contrast with our previous findings (ADF test), but, as we have already noticed, the correlogram of x_2 showed for the first lag an autocorrelation coefficient of 0.8, which could be a sign of integrating variable. By looking at the cointegrating vectors (beta matrix), we can conclude that we have a long-run relationship between y and x_2 and another one between x_1 and x_2.

In particular, the long-run elasticity between y and x_2 is positive: a rise in x_2 by 1% will increase y by 0.911%. With regards to the second cointegrating relationship, we observe that a rise in x_2 by 1% will lead to a positive effect of 0.757% in x_1. The signs of these coefficients (positive) are coherent with the results obtained with Engle-Granger and PSS representations. Nonetheless, we must emphasize that, in our context, these two latter tools are not enough accurate since they do not take into account the possibility of more than one cointegrating relationship, as well as the violation of the weak exogeneity assumption.

The VECM permits us to analyze the three equations related to our variables. In particular, it is observable that some lags have a significant impact on the changes in the variables in all regressions, as well as the error correction terms in the first two equations. As discussed, the error correction term indicates the speed of adjustment to restore the equilibrium in the dynamic model. As a matter of fact, in the alpha matrix (adjustment vectors), the first column indicates the speed of adjustment of the first cointegrating relationship, meanwhile, the second column indicates the SOA for the second one. Hence, these coefficients are the ones of the Error Correction Terms for the three equations. For the first one, EC1 and EC2 are both significant (highlighted in red in the table above), meaning that y adjusts to both the discrepancies in the first and the second cointegrating relationships (respectively by 37% and 39% for each quarter). The negative sign of EC1 suggests that y tends to decrease when its value is higher than what the long-run relationship indicates as target value. On the other hand, the positive sign of EC2 implies that y increases when the value of x_1 is higher than the long-run target value. This makes sense, as it implies that the growth rate of the dwellings deflator surges when the domestic inflation exceeds its long-run value.

Regarding the second equation, both EC1 and EC2 are again significant. This implies that x_1 adjusts to discrepancies in both cointegrating relationships. In particular, the positive sign of EC1 shows that x_1 increases when the value of y is higher than its long-run level. This entails that when the growth rate of the deflator of dwellings is higher than its equilibrium value, the domestic inflation must go up. Conversely, the negative sign of EC2 implies that the value of x_1 tends to diminish when it is higher than its long-run equilibrium. Hence, the assumption of weak exogeneity is violated, since x_1 is related to both y and x_2. This is coherent with reality, since presumably the bundle used to estimate domestic inflation contains dwellings (both new and old) and hence is influenced by its prices.

Finally, the third equation shows us that there is evidence that the two Error Correction Terms are not significant, meaning that in this case the weak exogeneity assumption holds.

Overall, the results so far obtained are in line with the main points of economic literature regarding this topic. As a matter of fact, we already knew that domestic inflation has a positive impact on the growth rate of dwellings deflator. This could be due to the many reasons we have already discussed in the first point, where we explicitly talked about agents' expectations and the effect of inflationary pressures.

9. HP Filter

Now we proceed to analyse our dependent variable through the filtering/decomposing of the actual data y_t in two unobservable components: $y_t = y_t^* + c_t$, where y_t^* is the trend that explains the long-run path of y_t and the c_t is the cycle that drives the short-run fluctuations.

In the two graphs below, we decompose our dependent variable y: on the left we observe the actual data and the trend, whereas on the right we plot the cycle (gap), which represents the discrepancies between the trend and actual data. It is possible to see that the trend is smoother than the actual series. We assume that the trend (permanent) component is I(1), meanwhile, by definition, we know that the cycle (transitory) component is I(0).

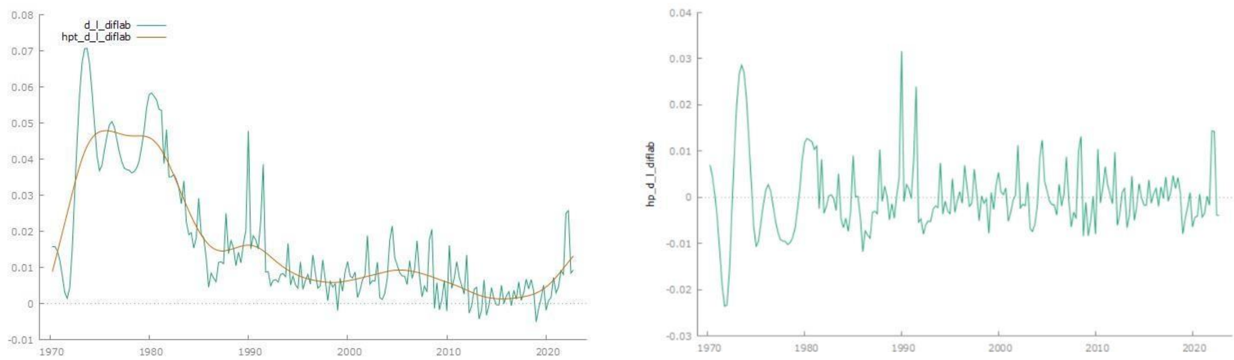


Figure 4: Decomposition of the dependent variable into the trend (on the left) and the cycle (right)

We apply the same filter to our explanatory variables x_1 and x_2 in order to run the two rank cointegration tests for the permanent components and the transitory ones, with the aim of proving that the trends are I(1) while the cycles are stationary I(0).

The first table reports the result of the Johansen cointegration test for the permanent components: after selecting the optimal number of lags we choose the option "unrestricted trend" (case 5).

```
Johansen test:

Number of equations = 3

Lag order = 6

Estimation period: 1971:4 - 2022:4 (T = 205)
```

```
Case 5: Unrestricted trend and constant

Log-likelihood = 7372.14 (including constant term: 6790.38)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0    0.12009     54.924 [0.0001]     26.226 [0.0246]
   1    0.11608     28.698 [0.0010]     25.294 [0.0019]
   2   0.016467     3.4038 [0.0650]     3.4038 [0.0651]
```

Given the decomposition of the HP filter, as we expected, the Johansen cointegration test between the three permanent components reports a not full rank matrix (rank = 2), because the variables are not stationary. On the other hand, since we know a priori that the cyclical components are I(0) by definition, we expect that the test will deliver a full rank matrix.

```
Johansen test:
Number of equations = 3
Lag order = 4
Estimation period: 1971:2 - 2022:4 (T = 207)
Case 3: Unrestricted constant

Log-likelihood = 2718 (including constant term: 2130.56)

Rank Eigenvalue Trace test  p-value  Lmax test  p-value
   0    0.37838     224.15 [0.0000]     98.412 [0.0000]
   1    0.33643     125.74 [0.0000]     84.894 [0.0000]
   2    0.17908     40.848 [0.0000]     40.848 [0.0000]
```

As further evidence, if we run a VECM model of our transitory components, we would obtain a beta matrix equal to an identity matrix.

References

Cannari, D'Alessio, & Vecchi. (2016). I prezzi delle abitazioni in Italia, 19927-2012. *Questioni di Economia e FInanza (Occasional Paper Banca d'Italia)*.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 1551-1580.

Jordan, S., & Philips, A. Q. (2018). Cointegration Testing and Dynamic Simulations of Autoregressive Distributed Lag Models. *The Stata Journal*,, 902-923.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *Applied Econometrics*, 289–326.

Stock, J. H. (1987). Asymptotic Properties of Least Squares Estimates of Cointegration Vectors. *Econometrica*, ,1035-1036.