**Modeling Credit Losses for Multiple Loan Portfolios**

Petr Gapko[[1]](#footnote-1), Martin Šmíd1

**Abstract**

We propose a dynamic macroeconomic model of credit risk for multiple portfolios with two factors for each portfolio. We follow the common approach that the credit risk on a loan portfolio can be decomposed into a probability of default and a loss given default and assume that both are driven by two underlying factors: one common for all borrowers in the portfolio and one individual for each single borrower. Our model additionally to the current research estimates the interconnectedness of the portfolios through the risk factors and enhances the current research by introducing dynamics and incorporating the external (macroeconomic) influence. We estimate the model on a set of two large real estate loan portfolios (one residential and one commercial) and show how the portfolios are interconnected and how the credit risk is influenced by macroeconomic environment.

**Keywords:** credit risk, mortgage, loan portfolio, dynamic model, estimation, interconnectedness, cointegration

**JEL Classification:** G32

1. **Introduction**

At the end of the last decade, when the financial crisis fully hit the US economy, losses from real estate loans in the US increased ten times, compared with the period of economic growth ending in 2007. Natural challenge for econometrician is to discover which factors drove this increase. The aim of this paper is to construct an easily manageable model which naturally links credit losses with exogenous variables.

The model we propose is structural factor one, which means that the losses of individual loans are driven by multiple factors. In particular, defaults depend on assets and liabilities of the debtors while the loss given default (LGD) is naturally dependent on collateral prices. Our model is multi-portfolio one as it allows simultaneous modeling of default rates (DRs) and LGDs of several portfolios. Finally, our model is dynamic in the sense that it naturally translates the dynamics of factors into the dynamics of losses.

Rather than a concrete model, we propose a way of connecting a credit risk model with a dynamic model for the factors. In this paper, VAR model is chosen to describe the dynamics of the factors; however, many different econometric models may be used to fit the dynamics of the factors. What we see as a contribution of our work is the way of translation of statistical results of the model for factors into the language of credit losses.

Clearly, Ideas behind our model are no way new. The first concept of a factor model is due to Merton (1974), the analytical expression of the loss distribution was derived by Vasicek (1987,1991), extended further by (Pykhtin, 2003) for the portfolios secured by collaterals. further Frye (Frye, 2000), Jimenez & Mencia (Jimenez & Mencia, 2009) or Witzany (Witzany, 2011) mainly in the idea of decomposition of the credit risk into underlying factors.

There are also many works dealing with interconnectedness of credit risk data with macroeconomy

Probably closest to our work is the paper by (Pesaran, Schuermann, Treutler, & Weiner, 2003) who also model losses of multiple (nationwide) portfolios in dependence on macroeconomic variables, following a VECM model. The main difference is that, while they use a credit rating sub-model for losses, we use a factor one. They are some additional differences, which we see as advantages of our mode: While they work only with losses, we separately treat both DR and LGD. While their model has many parameters in addition to those of the VECM model, there is only one such parameter per a prtfolio – the variance of the LGD individual factor – in our model.

There one rather surprising theoretical implication of our model: there is no difference in statistical distribution of DRs and LGDs whether the portfolios contain the same or different debtors. The theoretical reason for this is clear: the possible stochastic dependence of the individual factors across different portfolios diversifies out similarly as the individual factor themselves. Ih practice, however, it could simplify risk management substantially.

In the empirical part of our paper, we apply our model to two nationwide US real estate loan portfolios: the residential and the commercial one. As candidate exogenous variables, we took several US macroeconomic time series, which had been recently connected to credit risk in the literature, namely GDP, commercial and retail House Price Index, FED rate and unemployment, all of which showed to be significant except for the commercial HPI.

In explaining DRs and LGDs of both the portfolio, our model gives highly significant results, explaining PDs significantly better than the LGDs. which suggests that it was defaults which ignited the crisis rather than poor recovery.

Not surprisingly, the losses of both the portfolios are highly related TBD

Finally, we show that the amount of (hypothetical) economic capital recommended by our model is significantly less than that resulting from the standard formula. This is not surprising, as our model exploits more information that the standard one; however, it could also lead to substantial savings if oiur model was used a [ten model, co si mohou banky dělat samy].

The paper is organized as follows. In the following section we provide a description of the model methodology. In Section 3 we describe the data, the empirical analysis and our results. Finally, Section 4 concludes.

1. **The Model**

Similarly to (Vasicek 1987), we say that a loan *defaults* when

where is the value the debtor’s (hypothetical) assets and is the value of his debts, such that

where, are factors, common to all the loans, and are jointly normal individual factors.

The *relative recovery* is, in line with (Pythkin 2003), computed as

where is the outstanding principal of the loan, is the price of the collateral, is another common factor and is a normally distributed individual factor, independent of .

Now, consider an infinitely large portfolio of loans and define three important quantities: the *default rate*, usually imprecisely called probability of default (PD):

the *loss given default*, abbreviated as LGD:

and the *charge-off rate* (relative loss):

Here, denotes limit in probability. Not surprisingly,

(see [Kallenberg, Foundations of modern probability, Springer NY, 2001] Corollary 4.5.)

Assume further, that the portfolio is homogeneous in the sense that its loans have the same principal and their individual factors are Gaussian, independent between the loans. Then the default rate is given by

where is a standard normal c.d.f. and is the standard deviation of . Further, if is independent of , then

where is the standard deviation of (see Appendix X, where the formulas are proved given a the general setting, which is discussed below).Thanks to strict monotonicity of and (see Appendix of [GŠ FU 2012] for the latter), the correspondence between and is one-to-one.

Now consider homogeneous portfolios evolving in time. Assume that a loan from the -th portfolio may default at any time which happens when , where

and that the corresponding relative recovery is where is the outstanding principal and

Here, are general stochastic processes and, for each , vectors are Gaussian i.i.d. with independent of ,

Analogously to the static case, we have

where is the standard deviation of and is the standard deviation of – for a proof, see Appendix X.

Further, assume that the common factors follow a VAR model, i.e.

where is a deterministic matrix, is a Gaussian white noise and is a matrix of regressors possibly including trend, constants, lagged values of , their differences, and exogenous variables. Consequently, the dynamics of s and s is given by

where , , and is a Gaussian white noise. Here, , , is component-wise multiplication, is identity matrix, and , is the matrix consisting of the -th third of rows of ,

As are uniquely determined by the common factors and the parameters , their distribution depends only on these parameters and the parameters of the factors’ distribution; in particular, it does not depend on possible mutual inter-portfolio correlations of individual factors.

The econometrics of the model is straightforward. Once are observed and parameters and are known, the adjusted factors and may be retrieved by inverse relations

Consequently, standard techniques may be used to estimate and and the variance of the residuals.

Note that the estimation procedure does not depend on whether or not is known because the distribution of the adjusted factors depends on only through . Moreover, in case that , the values of need not be known; instead, , , may be estimated by becoming or being added to the trend of the VAR model.

Forecasting of and is easy in the model, because analytical formulas exist for conditional distributions of and given information up to . Namely, as

for some (-measurable) and in the VAR model, we have, according to TBD,

with the point forecast given by

Similarly, as

for some , we have

the mean loss of a loan given (see (iii) of Proposition [prop:h]).

Moreover, as we may equivalently express,

where , are assets, liabilities, respectively, of the -th debtor in the -th portfolio, we see that our formula generalizes the well known Vasicek’s formula for the distribution loss [Vasicek, Oldrich A. "Limiting loan loss probability distribution." Finance, Economics and Mathematics (1991): 147-148.].

Thanks to strict monotonicity of the functions transforming the factors to the rates, the confidence intervals for and G may be obtained by the same transformation. In particular, once , are confidence intervals for future values of , the intervals , may serve as confidence sets for respectively.

Contrary to and , the distribution of is not generally analytically tractable, because it is obtained by a nontrivial nonlinear transformation of dependent random variables. It may be, however, efficiently evaluated by Monte Carlo simulation.

1. **Data**

Loan corporate portfolio data are usually confidential, therefore hard to obtain. Thus we decided to apply our model to two US nationwide portfolios: the [X] and [Y], for which loss data are publically accessible. Namely, for both the portfolios, particular mortgage delinquency rates, which are proportions of loans more than 30 days past due (30+) on the total balance, and residential and commercial mortgage charge off rates, which are proportions of charged off loans (net of recoveries) on the average total balance, are available at the United States Federal Reserve System and thus includes the US nationwide statistics. The time period covered ranges from 1991 to 2016 in a quarterly granularity.

For both the portfolios, we used the delinquency and charge-off rates as proxies for default and charge-off rates. Consequently, the LGDs were computed by (lgt)

Table 4.1 and Figure 4.1 summarize descriptive statistics and show the time series of the DRs and LGDs of the residential and commercial portfolios From the Figure 4.1 it is obvious that the time series are correlated. Also, the recent economic crisis, which started in the US in late 2007 and impacted the US mortgage and real estate markets excessively is visible, as all the time series rocketed up to multiples of their preceding values between 2007 and 2010

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistic** | **30+ delinquency rate residential** | **Charge-off rate residential** | **30+ delinquency rate commercial** | **Charge-off rate commercial** |
| Mean value | 0.041188 | 0.004693 | 0.038408 | 0.009065 |
| Median | 0.023051 | 0.001584 | 0.022500 | 0.002884 |
| Minimum | 0.013358 | 0.000673 | 0.008500 | 0.000100 |
| Maximum | 0.110150 | 0.027057 | 0.120600 | 0.036297 |
| Standard deviation | 0.030813 | 0.006504 | 0.031300 | 0.010536 |
| Variance | 0.74810 | 1.3858 | 0.81495 | 1.1623 |
| Skewness | 1.1154 | 1.8399 | 1.1312 | 1.1850 |
| Excess kurtosis | -0.332530 | 2.1346 | 0.091479 | -0.082452 |
| 5% percentile | 0.010670 | 0.001345 | 0.015785 | 0.0008 |
| 95% percentile | 0.11055 | 0.031084 | 0.10596 | 0.02177 |

Table 3.1: Descriptive statistics of input data

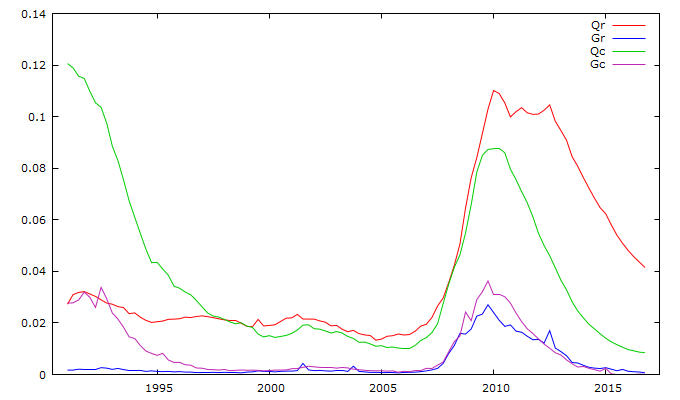


Figure 3.1: Development of the 30+ delinquency rates (Qr = residential, Qc = commercial) and charge-off rates (Gr = residential, Gc = commercial)

Having the DRs and LGDs, the factors were extracted by (extract) where the values of were obtained from the series of the residential and commercial house price indices the way described in Appendix A1. The resulting time series of the extracted common factors Y (default rate) and I (loss given default) for both commercial (Yc, Ic) and residential (Yr, Ir) mortgage portfolios are illustrated in Figure 3.2. As the factors were obtained by a monotone transformation of the loss rates, there is again a strong visual correlation, especially between Yr and Yc, and Ir and Ic.

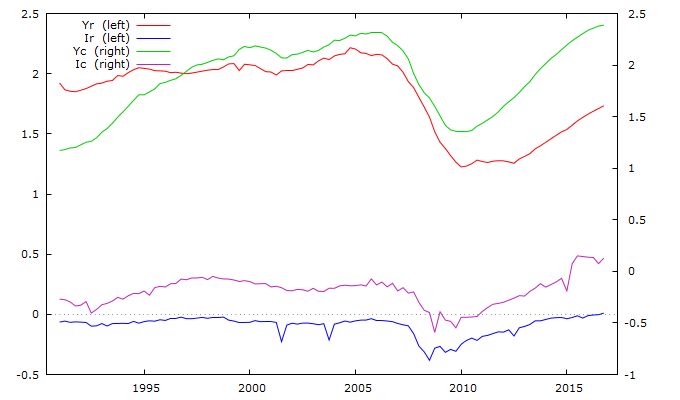


Figure 3.2: The development of the extracted common factors Yr and Ir (left axis), and Yc and Ic (right axis)

As candidates for explanatory macroeconomic variables, we choose FED base interest rate, unemployment, GDP, personal income (PI), industrial production (IP) and Case-Shiller HPI index, which we anticipated to influence the loss rates. In particular, the default rates were suspected to be driven by the wealth, represented by the GDP, unemployment, personal income and industrial production. The HPI index and the interest rate on the other hand were expected to be a driver of property prices, i.e. determine the charge-off rate. In the choice of the macroeconomic factors we were inspired by Pesaran (Pesaran, Schuermann, Treutler, & Weiner, 2003). All time series of macroeconomic variables were obtained from the FED. Except for FEDR, logarithms of all the variables were used in the actual estimation.

1. **Results**

First, we tested all the factors and the exogenous variables for unit roots. Using the ADF test we could not reject the unit root only for FEDR. Thus, we further regarded all the remaining exogenous variables , as well as all the factors as integrated with order one. The detailed results of the ADF test can be found in the Appendix in the table A.2.1.

In line with the usual procedure, we further tested for cointegration between the integrated variables. First we examined the exogenous variables and alone, then the factors alone and, finally, all the variables together. The results for exogenous variables were mixed as the Engle-Granger test [macro\_EG] did not confirm the cointegration while the Johansen test [macro\_JOH] suggested conintegration rank one. As for the factors, cointegration was found between [EG\_I] ]but not between (EG\_Y) and cointegration rank 2 was suggested by the Johansen test [factors\_joh]. As for all the variables, cointegration was found between and exogenous variables, between and exogenous variables, but not between and exogenous variables [Yr\_EG, Yc\_EG, Yr\_EG, Yc\_EG]. Finally, after choosing lag order 2 by a lag selection test, the Johansen test applied to all the variables suggested rank 3 (by Lmax test) or more (by trace test) [joh]. Thus, we decided to futher work with cointegration rank 3.

Consequently, we constructed the final model. First, we estimated a nine-equation VECM model for with rank 3 and 2 lags, and with as exogenousl variables (these variables were previously found suitable candidates for explaining factors by preliminary statistical analysis). Consequently, we restricted the cointegration matrix so as to reflect the relations, tested by the EG tests (namely to have the only factor in the first equation, only factor in the second equation and only in the third one). Finally, we removed insignificant variables and re-estimated the equations with factors on their right hand sides. The results are summarized in the Table 4.1. represents the error correction term of the -th cointegration equation.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | d\_Ic | d\_Ir | d\_Yc | d\_Yr |
| Constant | -1.944 \*\*\* | 1.117 \*\* | -0.719 \*\* | -2.437 \*\*\* |
| d\_Ic (lag1) | -0.369 \*\*\* | - | - | 0.061 \* |
| d\_Ir (lag1) | - | - | - | -0.138 \*\*\* |
| d\_Yc (lag1) | 0.658 \*\*\* | 0.779 \*\*\* | 0.348 \*\*\* | - |
| d\_HPI (lag1) | -1.355 \*\*\* | - | 0.719 \*\*\* | - |
| d\_U (lag1) | - | - | -0.184 \*\*\* | -0.222 \*\*\* |
| FEDR (lag1) | -0.006 \*\*\* | 0.009 \*\*\* | -0.005 \*\*\* | -0.009 \*\*\* |
| d\_IP (lag1) | - | - | -0.365 \*\* | - |
| EC1 | -0.103 \*\*\* | 0.073 \*\*\* | 0.072 \*\*\* | -0.039 \*\*\* |
| EC2 | - | -0.657 \*\*\* | 0.087 \* | 0.199 \*\*\* |
| EC3 | 0.335 \*\*\* | 0.229 \*\*\* | -0.087 \*\* | 0.161 \*\*\* |
| Adjusted R-square | 31 % | 36 % | 82 % | 73 % |

Table 4.1: Results of the VECM estimation (significance: \* - 90%, \*\* - 95%, \*\*\* - 99%)

[asi nějaká ta obecna věta o tom, jak je ten model dobrej, jistě ve smazaných něco najdeš:-]

A clear distinction between DR and LGD factors emerges from our results. The LGD factors are cointegrated with the macroeconomic environment, but poorly predicted, while the DR factors are predicted well, but not cointegrated. This suggests that [no a teď přemejšlejme].

[předpovědi, konfidentní intervaly]

We constructed a set of predictions to compare the model with the currently commonly used Vasicek’s distribution suggested losses on the 95% probability level. The Vasicek’s distribution is implemented in practice in the IRB formula for unexpected losses. For our comparison, we used the IRB formula and replaced the 99.9% probability level with the 95% (as 99.9% is an unrealistic quantile). The comparison was performed on one year horizon, i.e. we constructed forecasts for 4 periods/quarters by our model. The comparison shows that our model suggests higher loss on the 95% probability level. Therefore, the IRB formula might underestimate the far quantiles loss rate and suggest lower capital requirement. This shortcoming might be a reason why the CRR regulation requires banks to hold capital to cover the 99.9% quantile loss. The comparison of the two models is summarized in the table 4.2.

|  |  |  |
| --- | --- | --- |
| Segment/Model | IRB | Our |
| Retail | 0.198% | 0.374% |
| Commercial | 0.034% | 0.054% |

Table 4.2: Comparison of the predictions of IRB vs. our model – 12 month loss on the 95% probability level

[The economic interpretation of the non-significant cointegration here might be that the retail mortgage default rate was one of the key triggers of particularly the 2007-2009 economic crisis in the United States, in other words, the Yr enters the system as an exogenous variable.]

1. **Conclusion**

We constructed a multi-period multi-portfolio dynamic macroeconomic model of credit losses and applied it on two U.S. national portfolios. The empirical analysis showed that there exists a clear and estimable relationship between the credit risk and the macroeconomic environment. Additionally, we proved that the default rate on the portfolio and the loss given default are not independent, as well as there exists interconnectedness between portfolios. Thus, a reasonable model of credit risk has to incorporate the interconnectedness between defaults (represented e.g. by a probability of default) and losses (or, in other words, loss given default) and among risk factors of different portfolios. Finally, we demonstrated the possibility of prediction of the credit risk.

The empirical comparison of the model with the Vasicek’s distribution, the key element of the IRB approach in capital requirement calculation, shows that the Vasicek’s distribution based IRB formula tends to underestimate the quantile loss and therefore suggests to hold lower capital amount. As a consequence, the credit risk might be underestimated.

Generally, our results confirm strong interconnectedness of the factors and the macroeconomic environment and our final model proves itself to be potentially useful in estimation the future credit losses given the macroeconomic factors.

# **References**

Frontczak, R., & Rostek, S. (2015). Modeling loss given default with stochastic collateral. *Economic Modelling*, 44 (2015), pp.162-170.

Frye, J. (2000). Collateral Damage. *Risk*.

Gapko, P., & Šmíd, M. (2012). Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors. *Czech Journal of Economics and Finance*, 62(2): 125-140.

Gapko, P., & Šmíd, M. (2012a). Modeling a Distribution of Mortgage Credit Losses. *Ekonomicky casopis*, 1005-1023.

Hamerle, A., Dartsch, A., Jobst, R., & Plank, K. (2011). Integrating macroeconomic risk factors into. *The Journal of Risk Model Validation, vol.5/2*, pp. 3-24.

Hochguertel, S., & Ohlsson, H. (2011). Wealth Mobility and Dynamics Over Entire Individual Working Life Cycles. *ECB Working Paper Series no. 1301*.

Jimenez, G., & Mencia, J. (2009). Modelling the distribution of credit losses with observable and latent factors. *Journal of Empirical Finance*, 235-253.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journla of Finance 29*, Chapter 12.

Pesaran, M. H., Schuermann, T., Treutler, B.-J., & Weiner, S. M. (2003). Macroeconomic Dynamics and Credit Risk: A Global Perspective. *CESifo Working Paper No. 995*.

Pykhtin, M. V. (2003). Unexpected Recovery Risk. *Risk*, 74-78.

Sommar, P. A., & Shahnazarian, H. (2009, September). Interdependencies between Expected Default Frequency and the Macro Economy. *International Journal of Central Banking*, pp. 83-110.

Šmíd, M. (2015). Model of risk and losses of a multigeneration mortgage portfolio. *10th International Scientific Conference Financial management of firms and financial institutions.* Available at http://ssrn.com.

Šmíd, M., & Dufek, J. (2016). Multi-Period Factor Model of a Loan Portfolio. *Available at SSRN: http://dx.doi.org/10.2139/ssrn.2703884*.

Vasicek, O. A. (1987). *Probability of Loss on Loan Portfolio.* KMV.

Virolainen, K. (2004). Macro Stress Testing with a Macroeconomic Credit Risk Model for Finland. *Bank of Finland Discussion Paper No. 18/2004*.

Witzany, J. (2011). A Two-Factor Model for PD and LGD Correlation. *Bulletin of the Czech Econometric Society*.

**Appendix**

**A.1: Determination of**

Assume that, at time , the portfolio contains multiple “generations” of loans namely the loans originated at (the loans older than k are no longer present in the portfolio). Assume further that the inflow of fresh loans into the portfolio is constant in time. Finally, assume that all the collaterals securing loans from the generation which started at have been bought for the same price and that the price of each of them at is where is a normal random walk, specific to the loan, with variance ,

Denote the age of a loan randomly chosen at . Clearly, after periods, the ratio of the generations within the portfolio is: , which uniquely determines .

Let be the price of a randomly chosen collateral. By the Law of Iterated Variance, we then get

,

Even though the is a mixture of normal distributions rather than a normal distribution, it is thin tailed so it will not make a big harm to approximate it by .

**A.2: Results of the ADF tests and the Engle-Granger cointegration tests for individual factors**

|  |  |
| --- | --- |
| **Variable** | **P-value of the ADF unit root test** |
| Ic | 0.87 |
| Ir | 0.52 |
| Yc | 0.19 |
| Yr | 0.11 |
| U (unemployment) | 0.97 |
| PI (personal income) | 0.79 |
| IP (industrial production) | 0.83 |
| GDP | 0.93 |
| HPI | 0.97 |
| FEDR (FED interest rate) | 0.005 |

Table A.2.1: Results of the ADF unit root tests

**Testing for a unit root in Ic**

Dickey-Fuller test for Ic

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.0497293

test statistic: tau\_ct(1) = -1.33714

p-value 0.8729

1st-order autocorrelation coeff. for e: -0.275

**Testing for a unit root in Ir**

Dickey-Fuller test for Ir

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.100647

test statistic: tau\_ct(1) = -2.13345

p-value 0.5209

1st-order autocorrelation coeff. for e: -0.170

**Testing for a unit root in Yc**

Augmented Dickey-Fuller test for Yc

including 2 lags of (1-L)Yc

sample size 101

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + ... + e

estimated value of (a - 1): -0.0184176

test statistic: tau\_ct(1) = -2.82117

asymptotic p-value 0.1895

1st-order autocorrelation coeff. for e: -0.032

lagged differences: F(2, 96) = 134.873 [0.0000]

**Testing for a unit root in Yr**

Augmented Dickey-Fuller test for Yr

including 3 lags of (1-L)Yr

sample size 100

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + ... + e

estimated value of (a - 1): -0.0293095

test statistic: tau\_ct(1) = -3.07102

asymptotic p-value 0.1133

1st-order autocorrelation coeff. for e: -0.021

lagged differences: F(3, 94) = 46.857 [0.0000]

**Testing for a unit root in U**

Dickey-Fuller test for U

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.0128755

test statistic: tau\_ct(1) = -0.690306

p-value 0.9708

1st-order autocorrelation coeff. for e: 0.675

**Testing for a unit root in PI**

Dickey-Fuller test for PI

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.0428271

test statistic: tau\_ct(1) = -1.5951

p-value 0.7884

1st-order autocorrelation coeff. for e: -0.148

**Testing for a unit root in IP**

Dickey-Fuller test for IP

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.0262377

test statistic: tau\_ct(1) = -1.47207

p-value 0.833

1st-order autocorrelation coeff. for e: 0.503

**Testing for a unit root in GDP**

Dickey-Fuller test for GDP

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.0141484

test statistic: tau\_ct(1) = -1.05409

p-value 0.9309

1st-order autocorrelation coeff. for e: 0.322

**Testing for a unit root in HPIr**

Dickey-Fuller test for HPIr

sample size 103

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + e

estimated value of (a - 1): -0.00810067

test statistic: tau\_ct(1) = -0.646854

p-value 0.9738

1st-order autocorrelation coeff. for e: 0.750

**Testing for a unit root in FEDR**

Augmented Dickey-Fuller test for FEDR

including 3 lags of (1-L)FEDR

sample size 100

unit-root null hypothesis: a = 1

with constant and trend

model: (1-L)y = b0 + b1\*t + (a-1)\*y(-1) + ... + e

estimated value of (a - 1): -0.0898696

test statistic: tau\_ct(1) = -4.17634

asymptotic p-value 0.004789

1st-order autocorrelation coeff. for e: -0.005

**A.3: Results of the cointegration testing of the macroeconomic variables**

**Engle-Granger Step 1: cointegrating regression**

Dependent variable: GDP

coefficient std. error t-ratio p-value

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

const 0.128411 0.621682 0.2066 0.8368

IP 0.431848 0.0484118 8.920 1.98e-14 \*\*\*

PI 0.691862 0.0881088 7.852 4.26e-12 \*\*\*

U 0.0918087 0.00926205 9.912 1.27e-16 \*\*\*

HPIr 0.0973953 0.0123797 7.867 3.96e-12 \*\*\*

time 0.00227997 0.000506206 4.504 1.78e-05 \*\*\*

**Johansen’s rank test:**

Rank Eigenvalue Trace test [p-value] Lmax test [p-value]

0 0.27947 80.345 [0.0398] 34.744 [0.0902]

1 0.18937 45.601 [0.2688] 22.254 [0.3908]

2 0.11800 23.347 [0.4927] 13.309 [0.6609]

3 0.080769 10.038 [0.4845] 8.9271 [0.5106]

4 0.010422 1.1105 [0.2920] 1.1105 [0.2920]

Corrected for degrees of freedom (df = 94)

Rank Trace test [p-value]

0 80.345 [0.0503]

1 45.601 [0.2873]

2 23.347 [0.5015]

3 10.038 [0.4856]

4 1.1105 [0.2980]

**A.4: Results of the cointegration testing of all variables**

**Cointegrating regression -**

**OLS, using observations 1991:1-2016:4 (T = 104)**

**Dependent variable: Ic**

coefficient std. error t-ratio p-value

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

const 30.9510 3.15466 9.811 3.43e-016 \*\*\*

U −0.369376 0.0651256 −5.672 1.46e-07 \*\*\*

PI −1.81483 0.573778 −3.163 0.0021 \*\*\*

IP 2.11787 0.324598 6.525 3.10e-09 \*\*\*

GDP −2.79685 0.539503 −5.184 1.18e-06 \*\*\*

HPIr 0.299283 0.0806026 3.713 0.0003 \*\*\*

time 0.0298391 0.00283753 10.52 1.04e-017 \*\*\*

Mean dependent var −0.190888 S.D. dependent var 0.144941

Sum squared resid 0.332984 S.E. of regression 0.058590

R-squared 0.846113 Adjusted R-squared 0.836594

Log-likelihood 151.1211 Akaike criterion −288.2422

Schwarz criterion −269.7314 Hannan-Quinn −280.7429

rho 0.510871 Durbin-Watson 0.977351

Testing for a unit root in uhat

Dickey-Fuller test for uhat

sample size 103

unit-root null hypothesis: a = 1

model: (1-L)y = (a-1)\*y(-1) + e

estimated value of (a - 1): -0.489129

test statistic: tau\_ct(6) = -5.70672

p-value 0.01346

1st-order autocorrelation coeff. for e: -0.047

**Cointegrating regression -**

**OLS, using observations 1991:1-2016:4 (T = 104)**

**Dependent variable: Ir**

coefficient std. error t-ratio p-value

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

const 24.0517 2.08559 11.53 6.89e-020 \*\*\*

U 0.0223668 0.0430554 0.5195 0.6046

PI −1.09478 0.379333 −2.886 0.0048 \*\*\*

IP 2.17336 0.214596 10.13 7.11e-017 \*\*\*

GDP −2.89900 0.356673 −8.128 1.43e-012 \*\*\*

HPIr 0.399876 0.0532875 7.504 2.99e-011 \*\*\*

time 0.0243888 0.00187593 13.00 5.67e-023 \*\*\*

Mean dependent var −0.092010 S.D. dependent var 0.079836

Sum squared resid 0.145538 S.E. of regression 0.038735

R-squared 0.778312 Adjusted R-squared 0.764600

Log-likelihood 194.1593 Akaike criterion −374.3186

Schwarz criterion −355.8079 Hannan-Quinn −366.8193

rho 0.511374 Durbin-Watson 0.971527

Testing for a unit root in uhat

Dickey-Fuller test for uhat

sample size 103

unit-root null hypothesis: a = 1

model: (1-L)y = (a-1)\*y(-1) + e

estimated value of (a - 1): -0.488626

test statistic: tau\_ct(6) = -5.73702

p-value 0.0124

1st-order autocorrelation coeff. for e: -0.056

**Cointegrating regression -**

**OLS, using observations 1991:1-2016:4 (T = 104)**

**Dependent variable: Yc**

coefficient std. error t-ratio p-value

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

const 22.8334 5.70629 4.001 0.0001 \*\*\*

U −0.562029 0.0933475 −6.021 3.02e-08 \*\*\*

PI −4.68746 0.800614 −5.855 6.37e-08 \*\*\*

IP 4.04688 0.452070 8.952 2.27e-014 \*\*\*

HPIr 0.659988 0.112513 5.866 6.06e-08 \*\*\*

time 0.0334555 0.00460100 7.271 8.78e-011 \*\*\*

Mean dependent var 1.883028 S.D. dependent var 0.353531

Sum squared resid 1.108573 S.E. of regression 0.106358

R-squared 0.913886 Adjusted R-squared 0.909493

Log-likelihood 88.57890 Akaike criterion −165.1578

Schwarz criterion −149.2915 Hannan-Quinn −158.7299

rho 0.859357 Durbin-Watson 0.256658

Step 7: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat

including 2 lags of (1-L)uhat

sample size 101

unit-root null hypothesis: a = 1

model: (1-L)y = (a-1)\*y(-1) + ... + e

estimated value of (a - 1): -0.233003

test statistic: tau\_ct(5) = -4.53415

asymptotic p-value 0.07904

1st-order autocorrelation coeff. for e: -0.018

lagged differences: F(2, 98) = 8.805 [0.0003]

**Cointegrating regression -**

**OLS, using observations 1991:1-2016:4 (T = 104)**

**Dependent variable: Yr**

coefficient std. error t-ratio p-value

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

const −0.190922 0.307739 −0.6204 0.5364

U −0.882586 0.0421590 −20.93 3.85e-038 \*\*\*

HPIr 0.858367 0.0553383 15.51 3.00e-028 \*\*\*

FEDR −0.0552038 0.00661670 −8.343 4.39e-013 \*\*\*

time −0.0116112 0.000453633 −25.60 1.92e-045 \*\*\*

Mean dependent var 1.825992 S.D. dependent var 0.310982

Sum squared resid 0.436184 S.E. of regression 0.066377

R-squared 0.956211 Adjusted R-squared 0.954442

Log-likelihood 137.0827 Akaike criterion −264.1654

Schwarz criterion −250.9435 Hannan-Quinn −258.8088

rho 0.868036 Durbin-Watson 0.254628

Testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat

including 3 lags of (1-L)uhat

sample size 100

unit-root null hypothesis: a = 1

model: (1-L)y = (a-1)\*y(-1) + ... + e

estimated value of (a - 1): -0.195126

test statistic: tau\_ct(4) = -3.64373

asymptotic p-value 0.2673

1st-order autocorrelation coeff. for e: 0.050

lagged differences: F(3, 96) = 2.247 [0.0877]

**Johansen’s rank test:**

**Variables Yr, Yc, Ic, Ir, FEDR, U, IP, HPI**

Number of equations = 8

Lag order = 4

Estimation period: 1992:2 - 2016:4 (T = 102)

Case 5: Unrestricted trend and constant

Log-likelihood = 2217.97 (including constant term: 2217.97)

Rank Eigenvalue Trace test p-value Lmax test p-value

0 0.47193 231.70 [0.0000] 63.214 [0.0014]

1 0.38684 168.49 [0.0000] 48.423 [0.0242]

2 0.32386 120.07 [0.0003] 38.744 [0.0670]

3 0.30333 81.323 [0.0038] 35.783 [0.0253]

4 0.23654 45.540 [0.0800] 26.720 [0.0616]

5 0.099636 18.820 [0.5166] 10.391 [0.7124]

6 0.056243 8.4296 [0.4279] 5.7308 [0.6527]

7 0.026892 2.6988 [0.1004] 2.6988 [0.1004]

Corrected for sample size (df = 66)

Rank Trace test p-value

0 231.70 [0.0000]

1 168.49 [0.0001]

2 120.07 [0.0013]

3 81.323 [0.0088]

4 45.540 [0.1088]

5 18.820 [0.5455]

6 8.4296 [0.4421]

7 2.6988 [0.1072]

1. Econometric Department, Institute of Information Theory and Automation, Czech Academy of Sciences [↑](#footnote-ref-1)